

**TEST GÜVENLİĞİ AÇISINDAN BİREYLER ARASINDAKİ  
OLASI İŞ BİRLİĞİNİN İNCELENMESİ**

**DOKTORA TEZİ**

**SEHA YORMAZ**

**MERSİN ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI  
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**MERSİN  
ARALIK-2019**

**TEST GÜVENLİĞİ AÇISINDAN BİREYLER ARASINDAKİ  
OLASI İŞ BİRLİĞİNİN İNCELENMESİ**

**DOKTORA TEZİ**

**SEHA YORMAZ**

**MERSİN ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI  
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**






**Danışman  
Doç. Dr. ÖNDER SÜNBÜL**

**MERSİN**

**ARALIK-2019**

## ONAY

Seha YORMAZ tarafından Doç. Dr. Önder SÜNBÜL danışmanlığında hazırlanan "Test Güvenliği Açısından Bireyler Arasındaki Olası İş Birliğinin İncelenmesi" başlıklı bu çalışma aşağıda imzaları bulunan jüri üyeleri tarafından oy birliği ile Doktora tezi olarak kabul edilmiştir.

Görevi	Unvanı, Adı ve Soyadı	İmza
Başkan	Prof. Dr. Adnan ERKUŞ	
Üye	Prof. Dr. Cem Oktay GÜZELLER	
Üye	Doç. Dr. Dilara BAKAN KALAYCIOĞLU	
Üye	Doç. Dr. Önder SÜNBÜL (Danışman)	
Üye	Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL	

Yukarıdaki Jüri kararı Eğitim Bilimleri Enstitüsü Yönetim Kurulu'nun 15.01.20... tarih ve 03. / 19... sayılı kararıyla onaylanmıştır.

Prof. Dr. Gülşen AÇICI  
Eğitim Bilimleri Enstitüsü Müdürü



Bu tezde kullanılan özgün bilgiler, şekil, tablo ve fotoğraflardan kaynak göstermeden alıntı yapmak 5846 sayılı Fikir ve Sanat Eserleri Kanunu hükümlerine tabidir.

## ETİK BEYAN

Mersin Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliğinde belirtilen kurallara uygun olarak hazırladığım bu tez çalışmada,

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
  - Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlâk kurallarına uygun olarak sunduğumu,
  - Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
  - Atıfta bulunduğum eserlerin tümünü kaynak olarak kullandığımı,
  - Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
  - Bu tezin herhangi bir bölümünü Mersin Üniversitesi veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı,
  - Tezin tüm telif haklarını Mersin Üniversitesi'ne devrettiğimi
- beyan ederim.

## ETHIC DECLARATION

This thesis is prepared in accordance with the rules specified in Mersin University Graduate Education Regulation and I declare to comply with the following conditions:

- I have obtained all the information and the documents of the thesis in accordance with the academic rules.
- I presented all the visual, auditory and written informations and results in accordance with scientific ethics.
- I refer in accordance with the norms of scientific works about the case of exploitation of others' works.
- I used all of the referred works as the references.
- I did not do any tampering in the used data.
- I did not present any part of this thesis as an another thesis at Mersin University or another university.
- I transfer all copyrights of this thesis to the Mersin University.

27 Aralık 2019/27 December 2019

  
İmza / Signature

Seha YORMAZ

## ÖZET

### TEST GÜVENLİĞİ AÇISINDAN BİREYLER ARASINDAKİ OLASI İŞ BİRLİĞİNİN İNCELENMESİ

Bireyler test öncesinde ya da test sürecinde hile yoluyla iş birliği yaparak cevapları paylaşabilmektedirler. Yapılan bu hileler test puanlarının geçerliği için büyük bir tehdit oluşturmaktadır. Bu araştırmanın amacı çoktan seçmeli maddelerden oluşan testlerde hile yapan bireylerin olası iş birliğini belirlemede M4 benzerlik istatistiği kullanılarak makine öğrenme yöntemlerinin performansını ortaya koymaktır. Bu çalışmada, M4 istatistiği kullanılarak denetimsiz makine öğrenme türü olan hiyerarşik kümeleme ve yarı-denetimli makine öğrenme türü olan kendi kendine öğrenme ve SETRED (Self Training with Editing) yöntemlerinin hile yapan bireyleri ve grupları belirleme performansını ortaya koymak hedeflenmektedir. Araştırmada simülasyon verileriyle çalışılarak çeşitli koşullar altında yöntemlerin performans ölçütleri ortaya konmaktadır. Bu koşullar iş birliği yapan grup sayısı (1 ve 2), grupta yer alan bireylerin sayısı (5, 10, 20) ve kopya çekilmiş madde sayısının oranıdır (%10, %15, %20, %40). Bu koşulların çaprazlanmasıyla elde edilen verinin üretimi ve analizleri R-3.6.1 bilgisayar programı yardımıyla yapılmıştır. Çalışma sonucunda Ward kümeleme yönteminin, testteki maddelerin %40 ve üzeri kopya oranında hileyle iş birliği yapan bireyleri belirlemede oldukça güçlü bir yöntem olduğu; %20 ve altında ise yöntemin performansının yetersiz kaldığı ortaya çıkmıştır. Gruplardaki iş birlikçi birey sayısı arttıkça yöntemin kümeleme performansı da artmaktadır. Kendi kendine öğrenme ve SETRED yöntemlerinin sınıflandırma performanslarının her koşulda aynı olduğu ortaya çıkmıştır. Bu yöntemlerin iş birliği yapmayan bireyleri sınıflandırma performansının her koşulda yüksek olduğu; ancak iş birlikçi bir gruplu örneklerde iş birliği yapan bireylerin sınıfını tahmin etmede performansının yeterli olmadığı; gruptaki birey sayısının yöntemlerin performansına etkisinin ise düşük olduğu ortaya çıkmıştır. İki farklı iş birlikçi gruplu örneklerde bu yöntemlerin sınıflandırma performanslarının bir gruplu örneklemden daha iyi olduğu ortaya çıkmıştır. Genel olarak %40 ve üzeri kopya oranında bu yarı-denetimsiz yöntemlerin iş birlikçi bireyleri ve grupları oldukça iyi ayırt ettiği; gruplardaki iş birliği yapan birey sayısı arttıkça bu performansın da arttığı sonucuna ulaşılmıştır.

**Anahtar kelimeler:** Hile, bireyler arası iş birliği, M4, makine öğrenmesi

**Danışman:** Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi, Eğitim Bilimleri/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Mersin.

## ABSTRACT

### INVESTIGATION OF POSSIBLE COLLUSION BETWEEN EXAMINEES IN TERMS OF TEST SECURITY

Examinees will share their answers with collusion via cheating in or during test. Cheating is a big threat for validity of test scores. Purpose of this research is to determine the performance of machine learning methods by using M4 similarity statistic in investigating the possible collusion of individuals in the tests consisting of multiple-choice items. In this study, it is aimed to determine the performance of hierarchical clustering, self-learning and SETRED (Self Training with Editing) methods to identify collusion individuals and groups by using M4 statistics. In the research, the performance criteria of the methods are presented by working with simulation data under various conditions. These conditions are the number of collusion groups (1,2), the number of examinees in groups (5, 10, 20), and the amount of copying (%10, %15, %20, %40). Data generation with crossed of these conditions and analysis were done with R-3.6.1 software. At the end of the study, it was determined that Ward clustering method was a powerful method for identifying collusion examinees who collaborated at a rate of 40% or more of the items in the test and also the performance of the method was insufficient at a rate of 20% or less of the items. Clustering performance of the method increased with increasing the number of collusion examinees in the groups. The classification performance of self-learning and SETRED methods was found to be the same in all conditions. The performance of these methods to classify examinees who do not cooperate was high in all conditions; however, it was found that the performances in predicting the class of one collusion group was not sufficient and the effect of the number of examinees in the collusion group on the performance of the methods was low. The classification performance of these methods in samples with two different collusion groups was found to be better than samples with one collusion group. Generally, these methods quite well distinguished the collusion examinees and groups at 40% or more number of copied items and this performance increased as the number of collusion examinees in the groups increased.

**Keywords:** Cheating, collusion, M4, machine learning

## TEŞEKKÜR

Yüksek lisans ve doktora öğrenimim boyunca üzerimde büyük emeği olan, desteğini ve bilgisini hiç esirgemeyen, akademik anlamda beni geliştiren ve başarılı olmam konusunda beni güdüleyen değerli hocam ve danışmanım Doç Dr. Önder SÜNBÜL'e sonsuz teşekkürlerimi sunarım.

Lisansüstü öğrenimimin ilk yıllarından bu yana desteğini üzerimde hep hissettiğin, ne zaman yüzüne baksam o gülen gözleriyle beni güçlendiren ve motive eden, akademik yaşantımda her zaman örnek aldığım değerli bilim hocam Prof. Dr. Adnan ERKUŞ'a teşekkürlerimi sunarım.

Yüksek lisans öğrenimimden itibaren yardımları ve önerileriyle çalışmalarına katkıda bulunan ve çalışmalarımda ne zaman başım sıkışsa tereddütsüz tüm olanaklarını bana sunan Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBÜL'e çok teşekkür ederim.

Değerli fikirleriyle lisansüstü öğrenimim boyunca çalışmalarına katkı sağlayan değerli hocam Prof. Dr. Adnan KAN'a teşekkürlerimi sunarım.

Görüşleri ve yapıcı eleştirileriyle çalışmama katkıda bulunan diğer değerli jüri üyeleri Prof. Dr. Cem Oktay GÜZELLER ve Doç. Dr. Dilara BAKAN KALAYCIOĞLU'na teşekkürlerimi sunarım.

Lisansüstü öğrenimimiz boyunca her konuda desteğini benden esirgemeyen ve hep yanımda olan meslektaşım ve arkadaşım Semih AŞİRET'e çok teşekkür ederim.

Tüm eğitim yaşantım boyunca bana güvenen, inanan, her koşulda yanımda olan canım anne ve babama sonsuz sevgi ve desteklerinden dolayı teşekkür ederim. Lisansüstü öğrenimimdeki en zor ve yoğun anlarımda bile her zaman yanımda olduğunu hissettirerek bana güven ve cesaret veren, sevgi ve sabrıyla beni hiçbir zaman yalnız bırakmayan canım eşime çok teşekkür ederim.

*Sevgili çocuklarım Zeynep ve Ozan'a ithafen...*

## İÇİNDEKİLER

	Sayfa
İÇ KAPAK	
ONAY	
ETİK BEYAN	
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
TABLOLAR DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
GRAFİKLER DİZİNİ	viii
<b>1.GİRİŞ</b>	<b>1</b>
1.1.Kopya Belirleme Yöntemleri	2
1.1.1. Cevap Kopyalama İndeksleri	2
1.1.1.1. $\omega$ İndeks	2
1.1.1.2. Angoff'un A Değişimi	4
1.1.1.3. Kappa	5
1.1.1.4. $\tau$ İstatistikleri	6
1.1.1.5. VM (Variable Match) İndeks	6
1.1.1.6. İki Aşamalı Yöntem (KL- K)	7
1.1.2. Benzerlik İstatistikleri	7
1.1.2.1. GBT (Generalized Binomial Test)	7
1.1.2.2. M4 İstatistiği	9
1.2. Madde Ön bilgisi Belirleme Yöntemleri	11
1.2.1. FLOR (Final Log-Odds Ratio)	11
1.2.2. DGIRTM (Deterministic, Gated IRT Model)	12
1.2.3. CTAD	13
1.2.4. CUSUM <sub>LR</sub> (Cumulative Sum Statistic-Log Likelihood Ratio)	13
1.2.5. Faktör Analizi	14
1.2.6. Birey-Uyum İndeksleri	14
1.2.6.1. Gruba Dayalı (Parametrik Olmayan) Birey-Uyum İndeksleri	16
1.2.6.2. Madde Tepki Kuramına Dayalı Birey-Uyum İndeksleri	20
1.2.6.2.1. Artıklara Dayalı İndeksler	20
1.2.6.2.2. Olabilirliğe Dayalı İndeksler	22
1.2.6.3. Birey Tepki Fonksiyonu (Person Response Function, PRF)	24
1.2.6.4. Optimal Birey-Uyum İndeksleri	25
1.2.6.5. Uyarı (Caution) İndeksine Dayalı İstatistikler	26
1.2.6.6. Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT)	28
İçin Birey-Uyum İndeksleri	28
1.3. Silme ve Anormal Puan Artışı Belirleme Yöntemleri	29
1.3.1. Sıkça Görülen Yanlıştan Doğruya Cevap Değişimini Belirlemeye Yönelik Yaklaşımlar	29
1.3.2. Test Uygulamaları Arasında Oluşan Radikal Puan Farklaşmasını Belirlemeye Yönelik Yaklaşımlar	30
1.4. İş birliği Yapan Grupları Belirlemede Kullanılabilecek Yöntemler	30
1.4.1. Hiyerarşik Lineer Modelleme	30
1.4.2. Faktör Analizi	31
1.4.3. Makine Öğrenmesi	31
1.4.3.1. Denetimli Öğrenme	32
1.4.3.2. Yarı-Denetimli Öğrenme	32

1.4.3.2.1. Kendi Kendine Öğrenme (Self-Learning)	33
1.4.3.3. Denetimsiz Öğrenme	35
1.4.3.3.1. Kümeleme	35
1.4.3.3.1.1. Aşamalı (Hiyerarşik) Kümeleme Yöntemleri	36
1.4.3.3.1.2. Aşamalı Olmayan Kümeleme Yöntemleri	38
1.5. Yöntemlerin Performansını Belirlemede Kullanılabilecek Ölçütler	39
1.6. Araştırmanın Amacı	40
1.7. Problem Cümlesi	41
1.7.1. Alt Problemler	41
1.8. Araştırmanın Kapsam ve Sınırlılıkları	41
<b>2.KAYNAK ARAŞTIRMALARI</b>	42
2.1.Araştırmanın Önemi	43
<b>3.YÖNTEM</b>	44
3.1.Araştırmanın Türü	44
3.1.Veri Üretimi ve Analizi	44
<b>4. BULGULAR</b>	47
4.1. Ward Yönteminin Kümeleme Performansı	47
4.2. Kendi Kendine Öğrenme ve SETRED Yöntemlerinin Sınıflandırma Performansları	50
<b>5. TARTIŞMA, SONUÇ VE ÖNERİLER</b>	54
<b>6. KAYNAKLAR</b>	57
<b>7. ÖZGEÇMİŞ</b>	62

---

## TABLolar DİZİNİ

---

	<u>Sayfa</u>
Tablo 1.1. $\omega$ İndeksinin Farklı Değerleri İçin Yorumlamalar	4
Tablo 1.2.6.1. Birey-Uyum İndeksleri	15
Tablo 1.2.6.2. Guttman Modeline Dayalı İndekslere İlişkin Örnek Bir Uygulama	17
Tablo 1.2.6.3. <i>A</i> , <i>D</i> ve <i>E</i> İndeksleri İçin Örnek Bir Uygulama	18
Tablo 1.2.6.4. Korelasyon ya da Kovaryansa Dayalı İndeksler İçin Örnek Bir Uygulama	19
Tablo 1.5.1. Kontenjans Tablosu	39
Tablo 4.1.1. Ward Yönteminin İş Birliği Yapan Bir Grup Bireyi Kümeleme Performansına Ait Değerler	47
Tablo 4.1.2. Ward Yönteminin İş Birliği Yapan İki Farklı Grubu Kümeleme Performansına Ait Değerler	49
Tablo 4.2.1. Kendi Kendine Öğrenme ve SETRED Yöntemlerinin İş Birliği Yapan Bir Grup Bireyi Sınıflandırma Performansına Ait Değerler	50
Tablo 4.2.2. Kendi Kendine Öğrenme ve SETRED Yöntemlerinin İş Birliği Yapan İki Farklı Gruptaki Bireyleri Sınıflandırma Performansına Ait Değerler	52

---

## ŞEKİLLER DİZİNİ

---

	<u>Sayfa</u>
Şekil 1.4.1. Hiyerarşik Kümeleme Yöntemlerinin Yapılanması	36
Şekil 1.4.2. Hiyerarşik Kümeleme Yöntemlerinin Şematik Gösterimi	38
Şekil 1.4.3. Aşamalı Olmayan Kümeleme Yöntemlerinin Yapılanması	38

---



## GRAFİKLER DİZİNİ

---

	<u>Sayfa</u>
Grafik 3.1. Ward Kümeleme Analizi İçin Ağaç Diyagramı Örneği	45
Grafik 4.1.1. Ward Yönteminin İş Birliği Yapan Bir Grup Bireyi Kümeleme Performansı	48
Grafik 4.1.2. Ward Yönteminin İş Birliği Yapan İki Farklı Gruptaki Bireyleri Kümeleme Performansı	49
Grafik 4.2.1. Kendi Kendine Öğrenme ve SETRED Yönteminin İş Birlikçi Bir Grup Bireyi Sınıflandırma Performansı	51
Grafik 4.2.2. Kendi Kendine Öğrenme ve SETRED Yönteminin İş Birlikçi İki Farklı Gruptaki Bireyleri Sınıflandırma Performansı	53

---



## 1. GİRİŞ

Günümüzde ölçme ve değerlendirmede ve buna paralel olarak bilgisayar programlarındaki hızla artan gelişmelerle bireylerin ilgilenilen özellik ya da yetenek düzeyleri hakkında kolaylıkla bilgi elde edilebilmektedir. Birçok alanda geniş ölçekli sınavlardan elde edilen bu bilgiler kullanılarak bireylerin o alanla ilgili yeterlikleri kestirilmeye çalışılır. Özellikle iş alanlarına birey seçiminde ve eğitim kurumlarına öğrenci yerleştirmede kullanılan bu bilgiler oldukça büyük öneme sahiptir. Bu nedenle bireyler hem sosyal statü hem de sosyo-ekonomik yönden tatmin olma isteklerini karşılamak adına ilgi gören iş alanları ve üniversitelerde yer alabilmek için büyük çaba sarf etmektedirler. Bireyler arası rekabetin gün geçtikçe artması ve eşzamanlı olarak teknolojinin gelişmesi ve sosyal medyanın kullanımının artması ile hile yoluyla başarılı olma çabaları da artmaktadır. Ancak bu şekilde hilelerle elde edilen başarılar, bireylerin gerçekteki yetenek düzeylerini yansıtmamaktadır. Bu da bireyler hakkında yanlış kararlar verilmesine, ilgili özellik yönünden yeterli olmadığı halde bireye yeterliymiş gibi ya da ilgili özellik açısından yeterliken yetersizmiş gibi davranılmasına neden olmaktadır. Bu nedenle test güvenliğini bozan hileleri belirlemek testin geçerliği açısından büyük önem taşımaktadır.

Test içerisinde yapılan hilelerin, kopya çekme, testi uygulayan öğretmenin maddeler hakkında bireylere ipucu vermesi (educator cheating), maddeler yayınlanmamış olsa da maddeler hakkında ön bilgilerin (preknowledge) olması veya maddelerin teşhir (exposure) olması gibi birçok çeşidi vardır. Yapılan bu hileler bireysel ya da grupça da olabilmektedir. Eğer test uygulamasından önce ya da testin uygulanması esnasında bir iş birliğinden söz ediliyorsa buna, iş birliğiyle yapılan hile (collusion) adı verilmektedir. Bireyler arası iş birliği olarak da adlandırılabilen bu hile türü, en az bir bireye bireysel kazanç sağlamak için iki ya da daha fazla bireyin birlikte çalışması olarak tanımlanan test senaryosudur (Maynes, 2017, s.48). Bu iş birliği, bir bireyin teste başka bir bireyin yerine girmesi, testten önce bilgilerin paylaşımı, test esnasında birbirleriyle sözel ya da sözel olmayan bir şekilde (teknolojik araçlarla, yazılı olarak hazırlanan kopya gibi) iletişime geçmesi; test yöneticisi (uygulayıcısı) veya öğretmenin gizli olarak test hakkında test öncesi ya da esnasında bilgi paylaşması şeklinde olabilmektedir.

Bu hileleri belirlemede ise birçok yöntem kullanılmaktadır. Bunlar arasında kopya belirleme yöntemleri, madde önbilgisini (testten önce cevapları bilinen maddeler) ya da şaibeli maddeleri belirleme yöntemleri, silme ve anormal puan artışını belirleme yöntemleri sayılabilmektedir. Bu yöntemler ise kullanılan istatistiklerin özelliklerine göre kendi içinde farklı adlar almaktadır. Bu çalışmada alanyazında bu yöntemlerden sıklıkla tercih edilenlere yer verilmiştir.

## 1.1. Kopya Belirleme Yöntemleri

Bireyler daha iyi bir derece almak, sosyal yaşamda kabul görmek, ebeveynlerini memnun etmek için veya kişilik özelliklerinden dolayı kopya çekme eğiliminde olabilirler (Cizek, 1999). Test esnasında kopya çeken bireyler ya biri kaynak diğeri kopyacı olup kopyacı pasif olan kaynaktan maddenin cevabını hile yoluyla almakta ya da bireyler birbiriyle cevapları paylaşmaktadır.

İstatistiksel bulgulara dayalı kopya belirleme yöntemlerini Saupe (1960) deneysel ve şans olasılığına dayalı olmak üzere ikiye ayırmaktadır. Deneysel (empirical) yöntemlerde, kopya çektiği şüphelenilen bireylere ait dağılım ile kopya çekilmediği bilinen bir gruba ait cevapların dağılımları karşılaştırılmaktayken; şans olasılığına dayalı yöntemlerde ise kopya çektiği şüphelenilen bireylere ait bilinen bir dağılım ile binom, poisson, standart normal dağılım gibi teorik bir dağılım karşılaştırılmaktadır. Zamanla kopya belirlemede kullanılan yöntemlerin birçoğunun üretiminde araştırmacılar deneysel yöntemlerden uzaklaşarak, şans yöntemlerine yönelmiştir (Cizek, 1999, s.139).

Günümüzde kopya belirleme çalışmalarında çekilen kopyaya göre *cevap kopyalama (answer copying) indeksleri, benzerlik istatistikleri ve birey-uyum istatistikleri* kullanılmaktadır.

Cevap kopyalamada isteyerek ya da istemeyerek cevapları kopyacı tarafından kopya çekilen bir kaynak vardır. Bu tür kopyayı belirlemede kullanılan indeksler kopyacının kaynaktan çektiği cevaplara yani bu iki birey çifti arasındaki benzer cevaplara odaklanmaktadır. Benzerlik istatistiklerinde ise durum biraz farklıdır. Çünkü bireyler kopyacı ya da kaynak olarak adlandırılmamaktadır ve bireylerin birbiriyle anlaşmalı olarak çektiği kopyayı belirlemeye çalışmaktadır. Cevap kopyalamanın aksine bireylerin benzer cevaplarında simetriklik mevcuttur. A bireyi B'den, B bireyi de A'dan isteyerek ya da anlaşmalı olarak kopya çekmiştir. Birey-uyum istatistikleri ise tek başına kopya belirlemede zayıf kalmaktadır ve daha çok madde ön bilgisini belirlemede kullanılmaktadır. Bu bölümde cevap kopyalama ve benzerlik istatistiklerinden güncel ve tercih edilen istatistiklerin açıklamalarına yer verilmektedir:

### 1.1.1. Cevap Kopyalama İndeksleri

#### 1.1.1.1. $\omega$ İndeks

$\omega$  indeks, iki bireyin bir maddeye verdikleri cevap arasındaki gözlenen benzerliğin miktarı ile teorik olarak beklenen benzerliğin miktarını karşılaştırmaktadır (Wollack, 1997). Maddelere verilen cevapların olasılığı ise çoklu puanlanan madde tepki kuramı modellerinden

nominal tepki modeli ile kestirilmektedir. Bununla beraber ikili puanlama (1-0) durumları için de lojistik modeller çerçevesinde uyarlanmış versiyonları da bulunmaktadır.

Şans yöntemleri arasında yer alan  $\omega$  indeksiyle olası kopyacı ve kaynağın gözlenen aynı cevapları ile şansa bağlı benzerlik karşılaştırılarak kopya olasılığı belirlenmektedir. Bunun için olası kopyacı (C) ve kaynak (S) çifti için aynı cevaplanan maddelerin sayısı şu eşitlikle bulunur:

$$h_{CS} = \sum_{i=1}^n I[u_{iC} = u_{iS}] \quad (1)$$

Eğer olası çiftin cevapları eşleşiyorsa  $I=1$ , eşleşmiyorsa  $I=0$  olmaktadır.

Sırasıyla kopyacının yetenek düzeyi, kaynağın cevapları ve madde parametreleri koşulu altında kopyacının kaynaktan aldığı cevapların olasılığı şu şekildedir:

$$E(h_{CS}|\theta_C, U_S, \xi) = E \left[ \sum_{i=1}^n I(u_{iC} = u_{iS}|\theta_C, U_S, \xi) \right] = \sum_{i=1}^n [P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)] \quad (2)$$

MTK'ye göre bireylerin maddelere verdikleri cevapların yerel bağımsız olduğu kabul edilirse, kaynağın yeteneği ve doğru bir şekilde kestirilen kopyacının yetenek düzeyi ile madde parametreleri koşulları altında  $h_{CS}$  bağımsız Benoulli değişkenlerinin toplamı olup (Wollack, 1997), kopyacının kaynaktan aldığı cevapların olasılığı ve varyans aşağıda yer alan eşitlikler yardımıyla hesaplanmaktadır;

$$P(u_{iC} = u_{iS}|\theta_C, U_S, \xi) \quad (3)$$

ve

$$[P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)][1 - P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)] \quad (4)$$

Madde sayısı sonsuza yaklaştığında  $h_{CS}$  de normale yaklaşır ve  $\omega$  herhangi bir C ve S çifti için standart normal bir dağılım sergiler (Wollack, 1997). Böylece  $\omega$  şu eşitlikle hesaplanır;

$$\omega = \frac{h_{CS} - E(h_{CS}|\theta_C, U_S, \xi)}{\sigma_{h_{CS}} - E(h_{CS}|\theta_C, U_S, \xi)} = \frac{h_{CS} - \sum_{i=1}^n [P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]}{\sqrt{\sum_{i=1}^n P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)[1 - P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]}} \quad (5)$$

Özetle;  $\omega$  indeks testteki tüm maddeler için kopyacı ile kaynağın eşleşen cevaplarının olasılıklarının toplamına dayanmaktadır. Bu toplam ise beklenen eşleşen cevaplardır. Gerçekteki eşlenen cevapların sayısından bu beklenen cevapların sayısı çıkarılıp standart hataya bölündüğünde  $\omega$  indeks hesaplanmış olur.

$$\omega = \frac{\text{Eşleşen cevaplar} - \text{Beklenen eşleşen cevaplar}}{\text{Standart Hata}}$$

$\omega$  indeksinin farklı değerleri için verilen kararlar Tablo 1.1'de verilmektedir (Wollack, 2004, s.39).

**Tablo 1.1.**  
 $\omega$  İndeksinin Farklı Değerleri İçin Yorumlamalar

Düzyey	Yanlış Pozitif Oranı	$\omega$	Yorum
1	> 50	<0	Kopya olmadığına ilişkin ipucu
2	%5-%50	0-1.64	Zayıf ipucu
3	%1-%5	1.65-2.32	Az ipucu
4	%.1-%1	2.33-3.08	İyi ipucu
5	%.01-%.1	3.09-3.72	Güçlü ipucu
6	<%.01	> 3.72	Çok güçlü ipucu

#### 1.1.1.2. Angoff'un A Değişimi

Angoff (1974) kopya çekenleri belirlemede A'dan H'ye kadar sekiz indeks önermiştir. Bu indekslerden biri olan A indeksinin performansını belirlemek için McManus, Lissauer ve Williams (2005) Acinonyx adı verilen bilgisayar programı kullanmışlardır. Bu program yardımıyla doğrusal regresyon teknikleriyle birey çiftlerinin benzer cevaplarının artıklarını hesaplayarak Angoff'un A indeksini buna uyarlamışlardır. Çalışmaları sonucunda Acinonyx yönteminin I. Tip hatayı azalttığını; ancak sadece kopyayı belirlemede sınırlı olduğunu, bilgi paylaşımını belirleyemediğini görmüşlerdir. Bu yöntemin geniş ölçekli testlerde kopya belirlemede kullanışlı olduğunu vurgulamışlardır.

### 1.1.1.3. Kappa

Sotaridona, van der Linden ve Meijer (2006) Cohen'in kappa istatistiğini kullanarak bireylerin cevaplarının eşleşme oranına dayalı kopyanın belirlenebileceğini öne sürmüşlerdir. Bu istatistiğin diğer indekslerden ayıran özelliklerinden biri, kaynak ve kopyacının testteki maddelere verdikleri tepkiler için belli bir sayıltı getirmemiş olmasıdır. Tek öne sürülen sayıltı bireyler tarafından verilen tüm tepkilerin birçok olasılığının olmasıdır (Sotaridona ve diğ., 2006).

Bu istatistikte, maddelere verilen yanıtların seçenekleri " $v = 1, \dots, m$ " ve madde sayısının belirtildiği " $i = 1, \dots, N$ " sembolleri kullanılarak oluşturulan bir  $m \times m$  tablosu ile Kappa istatistiği Eşitlik 6 kullanılarak hesaplanmaktadır.

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \quad (6)$$

Eşitlikte yer alan  $\pi_0$ , iki birey arasındaki eşleşen cevapların olasılığını;  $\pi_e$ , kaynak ve kopyacının birbirinden bağımsız olduğu durumlarda iki birey arasındaki eşleşen cevapların şans olasılığını vermektedir.

Kappa istatistiği için kurulan hipotezler test edilerek, seçilen bireylerin kopya çekip çekmediği hakkında karar verilmektedir (Sotaridona ve diğ., 2006):

$$H_0 : \kappa = 0$$

$$H_1 : \kappa > 0$$

Cevaplayanlar evreninin bağımsız olduğu yokluk hipotezi reddedildiğinde, kopyanın olduğu söylenebilmektedir.

Kopyacı ve kaynağın yetenek düzeylerine göre aynı cevabı verme olasılıklarının I. Tip hata oranlarını belirlemeye çalışan Sotaridona ve diğerleri (2006) yaptıkları çalışmada yetenek düzeylerinin zıt yönde olduğu durumların dışında I. Tip hata oranlarının arttığını ortaya çıkarmıştır. Araştırmacılar bu oranı azaltmak için cevapları yeniden kodlamayı önermişlerdir. Hesaplamalara kodlama değerlerinin dahil edildiği çalışma sonrasında yetenek düzeyi aynı yönde olduğu yüksek puanlarda I. Tip hata oranının arttığı, zıt yönde olduğunda ise azaldığı görülmüştür. Ancak yazarlar bu durumun kopya belirleme çalışmalarında beklendiğini belirtmişlerdir. Kopya belirlemede kappanın %30-40 kopya oranında ve kaynağın yetenek düzeyinin orta olduğu durumlarda uç yetenek düzeylerine göre oldukça güçlü olduğunu belirtmişlerdir.

Kappa istatistiğinin hesaplanmasındaki kolaylığı yanında bazı sınırlılıkları da vardır. Kappa kopya belirleme dışında maddeler hakkındaki önbilgilerin ya da hatırlamaların belirlenmesi için kullanılamamaktadır.

#### 1.1.1.4. $\tau$ İstatistikleri

Küçük örneklerde çekilen kopyayı belirlemek amacıyla üretilen  $\tau_1$  ve  $\tau_2$  istatistikleri (van der Ark, Emons ve Sijtsma, 2008), sırasıyla bireylerin şüpheli cevaplarının belirlenmesi ve ardından şüpheli cevaplardan kopyacı ve kaynağın tespit edilmesi için kullanılır. Araştırmacılar tarafından madde güçlüğü de kopya belirlemeye dahil edilip istatistiklerin gücü artırılarak yeni bir çalışma yapılmıştır. Bu çalışma sonucunda, yine küçük örneklerde etkili olan, kestirimi ise  $\tau_1$  ve  $\tau_2$  istatistiklerine göre daha güç olan  $\tau_1^*$  ve  $\tau_2^*$  istatistikleri ortaya konmuştur (van der Ark, Emons ve Sijtsma, 2008).

van der Ark, Emons ve Sijtsma (2008), bu iki istatistiğin simülasyon çalışmalarında özellikle kaynak ve kopyacının rolleri değiştiğinde tutarsız sonuçlar verdiğini ve çok yüksek puanlarda kopyayı belirlemede daha az duyarlı olduğunu belirtmişlerdir.

#### 1.1.1.5. VM (Variable Match) İndeks

VM - indeks (Belov, 2011), bireylerin testteki farklı bölümlerde çektiği körlemesine kopyayı (blind-copy) veya bireylerin yanlışlıkla bir önceki veya bir sonraki maddelerde (kaydırma) çektiği kopyayı (shift-copy) belirlemede kullanılmaktadır. Belov (2011) testi iki bölüme ayırmıştır: testi alan her bireyin aynı olan cevaplardan oluşan *operasyonel (operational) bölüm* ve yakın oturan bireylerin farklı olan cevaplarından oluşan *değişken (variable) bölüm*dür. VM - indeksin hesaplanmasında bu bölümlerden elde edilen bilgiler kullanılmaktadır. Bu bilgilere dayalı eşitlik ise şöyledir:

$$\xi \equiv \xi(w_c, w_s, l, u) = \sum_{j=l}^u \xi_j(w_c, w_s) \quad (7)$$

$w_c$ , operasyonel bölümde kopyacının yanlış cevaplarının sayısı,

$w_s$ , operasyonel bölümde kaynağın yanlış cevaplarının sayısı,

$l$  ve  $u$  parametreleri çekilen kopyanın yöntemi.

1.  $l = u = 0$  (körlemesine kopya)
2.  $l \leq u < 0$  (negatif kaydırma)
3.  $0 < l \leq u$  (pozitif kaydırma)
4.  $l < 0 < u$  (yukarıdaki tüm kopya yöntemleri) (Belov, 2011, s:498).

Belov (2011) tarafından bu indeksin uzantısı olan VM - indeks\*, yalnızca kopyacının değişken bölümdeki yanlış cevapları ile aynı maddelere kaynağın verdiği doğru cevapları dikkate almaktadır. Bu şekilde çekilen kopyaya örnek olarak şüpheli kaynak ve kopyacıya ait cevaplar aşağıdaki şekilde verilebilir:

Şüpheli kaynak: C D A A B E D C B E D C

Şüpheli kopyacı: D B E A A B E D C E A B

Kopyacı kaynağın 3 ile 8. madde arasındaki doğru cevaplarını alıp, kendi cevap kağıdına kaydırma yaparak bu cevapları 4 ile 9 arasındaki maddeler için işaretlemiştir. Böylece birey kopya çekmiş; ancak maddeler yanlış yanıtlandığı için test puanında bir avantaj elde edememiştir.

Belov (2011) VM indeksleri ile K indeksini karşılaştırdığı çalışmasında K indeksinin bu indekslerden I. Tip hata oranının daha fazla olduğunu; ancak kopya belirlemede daha güçlü olduğunu ve VM indekslerinin ise birbirine yakın; fakat çok düşük güce sahip olduklarını ortaya koymuştur.

#### 1.1.1.6. İki Aşamalı Yöntem (KL - K)

Belov ve Armstrong (2010) kopya belirlemede iki aşamalı yöntem önermiştir. Bu yöntemin birinci aşamasında ilk kez Belov, Pashley ve Lewis (2007) tarafından ortaya konan birey-uyum indeksi Kullback Leibler (KL) istatistiği ile bir bireyin bir testin örtüşmeyen bölümleri (örneğin operasyonel bölümü ya da değişken bölümü) üzerindeki performansındaki farklılıklar belirlenmektedir. İkinci aşamada ise Holland (1996) tarafından önerilen K indeksi birinci aşamada belirlenen bireylere odaklanılarak birey çiftlerinin olağandışı eşleşen yanlış cevapları belirlenmektedir.

Araştırma sonucunda yalnız K indeksinin kullanıldığında olası kopyacılarının belirlenmesinin iki aşamalıdan 2.4 kat daha fazla olduğu ortaya çıkmıştır (Belov ve Armstrong, 2010).

#### 1.1.2. Benzerlik İstatistikleri

##### 1.1.2.1. GBT (Generalized Binomial Test)

van der Linden ve Sotaridona (2006) tarafından kopyayı belirlemek için önerilen GBT,  $\omega$  indeksi gibi MTK'ye dayalı bir indeks olup kopyacı ile kaynak arasındaki eşleşen cevapların

sayısının olasılık dağılımı yardımıyla hesaplanmasına dayanmaktadır.  $i$  maddesi için eşleşen cevapların olasılığı  $P_i$  olup olasılık için gerekli olan parametreler ikili puanlanan veya çoklu puanlanan MTK modelleri kullanılarak kestirilebilmektedir.

Kopyacının kaynaktan kopya çektiği maddelerin (bilinmeyen) sayısı  $\gamma_{cs}$  olarak gösterilmektedir. Kopyacı ve kaynağın eşleşen cevaplarının sayısı Eşitlik 8 yardımıyla hesaplanmaktadır;

$$M_{js} = \sum_{i=1}^n I_{jsi} \quad (8)$$

Ancak eşleşen cevapların sayısı her biri seçkisiz eşleşen farklı bir olasılıkla bir dizi bağımsız Bernoulli denemelerine dayanmaktadır ve böylelikle bu cevaplar genelleştirilmiş binom dağılımı sergilemektedir (van der Linden ve Sotaridona, 2006). Bu dağılımlara ait olasılıklar, olasılık fonksiyonu yerine üretilen fonksiyonlarla hesaplanmaktadır.

$$\prod_{i=1}^n [Q_i + zP_i] \quad (9)$$

$n$  tane denemeden  $i$ . denemede başarı olasılığı  $P_i$  olup  $Q_i = 1 - P_i$ 'dir.  $n$  denemede  $m$  tane başarı olasılığı  $f_n(m)$  ile gösterilmekte ve bu olasılık  $z^m$  katsayısı ile verilmektedir.

Bu olasılıklar döngü yardımıyla hesaplanabilmektedir. İlk deneme ile başarılı olma olasılığı hesaplanarak döngü süreci başlar ve bir sonraki döngü  $t = n$  olana kadar sonraki denemeye  $t$  eklenerek devam eder (van der Linden ve Sotaridona, 2006). Bu süreç Eşitlik 9 ile gösterilmektedir.

$$f_t(m) = Q_t f_{t-1}(m) + P_t f_{t-1}(m-1) \quad (10)$$

Eşitlik 10'da  $m < 0$  ya da  $m > t$  ise  $F_t(m) = 0$ 'dır. Eğer kopyacı ile kaynak  $i$  maddesini aynı cevaplamışsa  $t$ , 1'e eşittir; değilse 0'dır.

van der Linden genelleştirilmiş binomun birikimli dağılım fonksiyonuna ait şu ilişkiyi ortaya koymuştur (van der Linden ve Sotaridona, 2006):

$$F_n(m) = (1 - \pi_i) f_{n-1}(m) + F_{n-1}(m-1) \quad (11)$$

$\pi_i$ , keyfi bir  $i$  denemesi için başarı olasılığı,  $F_n(m)$  tüm  $n$  denemeleri için dağılım fonksiyonu ve  $f_{n-1}(m-1)$  diziden  $i$  denemesi çıkarılırsa kalan  $n-1$  deneme için dağılım fonksiyonudur (van der Linden ve Sotaridona, 2006).

van der Linden ve Sotaridona (2006) çalışmalarında GBT yönteminin özellikle düşük yetenek düzeyine sahip kopyacı ile yüksek yetenek düzeyine sahip kaynak çiftlerinde oldukça güçlü olduğunu göstermişlerdir. Zopluoğlu ve Davenport (2012) tarafından yapılan çalışmada kopyacı ve kaynağın yetenek düzeyine ve belirlenen kopya oranlarına göre GBT indeksinin  $\omega$  indeksine göre az farkla daha güçlü olduğu ortaya çıkmıştır.

### 1.1.2.2. $M4$ İstatistiği

Maynes (2014) tarafından ortaya konan  $M4$  istatistiği bireylerin maddelere verdiği cevapların benzerliğine dayalı bir benzerlik istatistiğidir. İki değişkenli bu istatistik, bireyler arasındaki hem benzer doğru cevapları hem de benzer yanlış cevapları dikkate almaktadır. Bireyler arasındaki benzer yanlış cevapların olma olasılığı benzer doğru cevaplardan daha düşüktür. Bu durum  $M4$  istatistiğini, sadece benzer doğru cevapların olasılığına dayalı istatistiklerden daha üstün kılmaktadır.

$M4$  benzerlik istatistiği için eşleşen cevapların istatistiksel ve yerel bağımsız olduğu sayılıtısının karşılanması gerekmektedir (Maynes, 2017). Bu sayılıtlar altında istatistiğin olasılık yoğunluk fonksiyonu, GBT istatistiğindeki iki değişkenli dağılım yerine genelleştirilmiş üç değişkenli dağılıma (generalized trinomial distribution) sahiptir. Bu fonksiyon (Feller, 1968) şu şekildedir:

$$G(x, y) = \prod_i (r_i + p_i x + q_i y) \quad (12)$$

Fonksiyonda yer alan  $r_i$ ,  $i$  maddesine verilen eşleşmeyen cevapların olasılığı;  $p_i$  eşleşen doğru cevapların olasılığı,  $q_i$  ise eşleşen yanlış cevapların olasılığıdır ve bu olasılıkların toplamı 1'dir. Bu olasılıklar birey çifti için sırasıyla şu üçlü kombinle görülebilir: (0,0,1), (0,1,0), (1,0,0). Bunu daha da özetlemek istersek, eşleşmeyen cevapların olasılığı, diğer iki olasılığın toplamının 1'den çıkarılmasıyla elde edilebilir. Bu durumda elimizde (0,1), (1,0), (0,0) kalır. Böylece iki değişkenli bu cevap olasılıkları çiftlerinin yardımıyla üç değişkenli bir dağılım ortaya koyulabilir. Fonksiyonda yer alan  $x$  ise gözlenmiş olan eşleşen doğru cevapların sayısı,  $y$  eşleşen yanlış cevapların sayısıdır.

Fonksiyon yeniden şekillendirilecek olursa  $M4$  istatistiğinin katışık (joint) olasılık dağılımı şöyle olmaktadır (Maynes, 2017);

$$T_{k+1}(x, y) = p_{k+1}(1,0)T_k(x-1, y) + q_{k+1}(0,1)T_k(x, y-1) + (1 - p_{k+1}(1,0) - q_{k+1}(0,1))T_k(x, y) \quad (13)$$

Fonksiyona ait sınır koşulu şu şekildedir:

$$T_0(0,0) = 1 \text{ ve } T_0(x, y) = 0 \forall (x, y) \neq (0,0) \quad (14)$$

Her iki birey çiftinin cevapladığı madde sayısı  $n$  olmak üzere,  $k+1 = 1, \dots, n$ 'dir. Yinelemeli bu eşitlikte yer alan  $1 - p_{k+1}(1,0) - q_{k+1}(0,1)$  bölümü eşleşmeyen cevapların olasılığı olan  $r_{k+1}$ 'e denk gelmektedir. Eşitlikte  $p_i$  ve  $q_i$  sabit olduğunda *genelleştirilmiş* üç değişkenli dağılım, üç değişkenli dağılım halini almaktadır (Maynes, 2017). Eşitlikle yapılan hesaplamalar yinelemeli bir süreç izlemektedir. Oldukça güç olan bu hesaplamalar, bilgisayar programları yardımıyla gerçekleştirilmektedir.

Maynes'in dikkat çektiği bir diğer konu ise bir grup bireyin ikili ikili ele alınıp değerlendirilmesinin oluşturacağı I. Tip hata miktarı ve bu durumu düzeltmek için kullanılacak olan kritik değerin seçimidir. Kritik değer belirlenmesi için geliştirmiş olduğu Bonferroni düzeltmesi önerisi bulunmaktadır (Maynes, 2009).

Maynes (2017) kendi üretmiş olduğu  $M4$  istatistiğinin benzerlik araştırmaları için en iyi sonuçları ürettiğini iddia etmektedir. Bu konuda dikkat edilmesi gereken başka bir nokta ise benzer olarak nitelendirilen bireylerin gruplamasına yönelik ek çalışmalar yapılmasının gerekliliğidir.

Maynes (2017)  $M4$  istatistiğinin performansını belirlemek için çalışmasında kopya oranını %5 arttırarak %0'dan %55'e kadar değişen on iki veri seti simüle etmiştir. Bunun için 101.608 birey çifti üretmiştir. Bununla birlikte iki cevap vektörü, belirlenen iki yetenek düzeyine göre üretilip bireylerin bunlardan kopya çekilmesi sağlanmıştır. Bu nedenle kopya çekilen maddeler iş birliği yapılan maddeler olarak adlandırılmıştır. Çalışma sonucunda istatistiğin 0.05, 0.01, 0.005, 0.001, 0.0005 ve 0.0001 alfa düzeylerinde I. Tip hatayı iyi kontrol ettiği ortaya çıkmıştır. Güç çalışması sonucunda ise kopya oranı arttıkça gücün de arttığı, özellikle %50 ve üzeri kopya oranında oldukça güçlü olduğu görülmüştür.

## 1.2. Madde Önbilgisi Belirleme Yöntemleri

Özellikle teknolojik araçların ve sosyal medya kullanımının toplumda yaygınlaşmasıyla beraber bireyler arasında yapılacak testlerin maddeleriyle ilgili olası bilgiler kolaylıkla paylaşılabilir. Sınav sorularının gizliliği için oluşturulan güvenlik önlemleri aşarak sınav öncesi sınav sorularına ulaşılabilir. Buna ek olarak sınava giren adaylar sınavda yer alan soruları ezberleyebilir. Sınav sorularının tekrarlı kullanımının söz konusu olduğu uygulamalarda, ezberlenmiş ve paylaşılmış olan sorular yine sınav güvenliğini tehdit edebilir. Sınav öncesi sınavda yer alma olasılığı olan maddelerin paylaşımlarının belirlenmesi için kullanılan istatistiksel yöntemler şöyledir:

### 1.2.1. FLOR (Final Log-Odds Ratio)

Hui (2010) çalışmasında üç parametrelili lojistik MTK modeline dayalı Bayes posterior log-odds oranı olan FLOR modelini madde önbilgisini belirlemek için kullanmıştır. Anormal tepki örüntülerinin olasılığını belirlemeye çalışan Hui (2010), sabit maddelerdeki ve test uzunluğundaki hileleri belirlemede FLOR'un etkili olduğunu ortaya koymuştur; ancak bu yöntemin özellikle BOBUT (Bilgisayar Ortamında Bireye Uyarlanmış Testler) uygulamalarında kullanışlı olduğunu vurgulamıştır.

Hui (2010), FLOR'un performansını belirlemede madde parametrelerinin ve model özelliklerinin sabitliği; madde önbilgisini belirlemede duyarlılığı ve BOBUT uygulamalarında sabit ve değişken test uzunluğunu test etmiştir.

Madde parametrelerinin ve model özelliklerinin sabitliğinin etkisini belirleme çalışmasında, yarısı teşhir olmuş madde olarak düzenlenen 10 maddelik veriyi araştırmak için birçok yöntem kullanmıştır.

Madde önbilgisi için duyarlılık çalışmasında yarısı teşhir olmuş -3 ile +3 arasında değişen güçlükte 20 madde ve -3 ile +3 arasında yetenek düzeyine sahip 20,000 kişilik bir veri simüle edilmiştir. Bireylerin yarısı normal, diğer yarısı madde önbilgisi olanlar olarak iki grubu ayrılıp FLOR'un madde önbilgisi için duyarlılığını birey uyum indekslerinden  $I_z$  (Drasgow, Levine ve Williams, 1985) ve CUSUM (Armstrong ve Shi, 2009) ile karşılaştırarak ortaya koymuştur.

Hui (2010) çalışmasında BOBUT uygulamalarında FLOR'un etkisini incelemek için güvenli maddelerin sayısı, teşhir olmuş maddelerin sayısı ve test uzunluğunu değiştirmiştir. Aynı zamanda bunu yaparken bahsedilen birey-uyum indeksleri ve yazarlar tarafından önerilen Düzeltilmiş FLOR ile karşılaştırılmıştır.

Çalışma sonucunda FLOR'un bireylerin tüm teşhir olan maddeleri doğru, tüm güvenli olan maddeleri yanlış cevapladıklarında en etkili teknik olduğu ortaya koyulmuştur, ancak yüksek yetenek düzeyinde duyarlılığı azalırken hile yapan bireylerin sayısından büyük ölçüde

etkilenmediği belirtilmiştir. Aynı zamanda madde parametrelerinin FLOR üzerinde çok az etkisinin olduğu görülmüştür. Diğer indekslerle karşılaştırıldığında sabit uzunluk ve sabit maddeli testlerde FLOR'un daha duyarlı olduğu ortaya konmuştur. Teşhir olan madde sayısı arttıkça FLOR'un değerlerinin de maksimuma yükseldiği görülmüştür. BOBUT uygulamalarında bireyin yeteneğinin FLOR ya da Düzeltilmiş FLOR değerlerine çok küçük etkisinin olduğu vurgulanmıştır. Böylece FLOR ve Düzeltilmiş FLOR madde ve yetenek parametrelerinden büyük ölçüde etkilenmediğinden karşılaştırılan diğer indekslere göre daha duyarlı olduğu ve önbilgiden gelen uyumsuzluğu diğer uyumsuzluklardan etkili bir şekilde ayırt ettiği ve madde sırası ve yerinden etkilenmediği bulunmuştur.

### 1.2.2. DGIRTM (Deterministic, Gated IRT Model)

Shu (2011) tarafından testlerde hileyi belirlemek için önerilen DGIRTM, puan kazancı (score gain) elde etmiş bireylerle ilgilenmektedir. Bunun için her birey için çok boyutlu Rasch yaklaşımına göre kestirilen gerçek yetenek ve hile yeteneği olmak üzere iki yetenek düzeyi belirlenmektedir. Bu iki yetenek arasındaki fark ise birey için hile yoluyla elde edilen puan kazancını vermektedir. Model kestiriminin standart hatası içinde kalan bireyler hile yapmış olsa da hile yapmadı olarak sınıflandırılmaktadır. Çünkü bu bireyler puan kazancı sağlamamıştır. Bu yaklaşımda diğer kopya belirleme yöntemlerinden ve birey-uyum indekslerinden farklı olarak önemli olan test puanlarını ve geçerliği etkileyen puan kazancı sağlamış hilelerdir.

Shu (2011) çalışmasında teşhir oluş ve olmamış maddelerden oluşan 40 ve 80 test uzunluğunda 2000 cevaptan oluşan bir veri simüle etmiştir. Bununla birlikte Shu (2011) çalışmasında, teşhir olan maddelerin oranı, güçlüğü ve madde güçlüklerinin standart hataları, ayrıca hile yapanların sayısı, hile derecesi (teşhir olan madde başına düşen hile yapanların sayısı) ve hilenin etkililiği (hileden elde edilen puan kazançları) değişimlemiştir. Yapılan çalışmada birey-uyum indekslerinden  $Iz$  indeksi ile karşılaştırılarak DGIRT modelinin duyarlılığı (sensitivity) ve belirleyiciliği (specificity) araştırılmıştır. Duyarlılık doğru olarak belirlenen hile yapan bireylerin oranı, belirleyicilik ise doğru olarak belirlenen hile yapmayan bireylerin oranıdır.

Shu (2011) çalışması sonucunda DGIRTM'in hile yapmayan tüm bireyleri doğru olarak belirlediğini dolayısıyla belirleyiciliğinin yüksek olduğunu bulmuştur. Duyarlılığın ise test uzunluğunun, hilenin etkililiğinin ve hilenin derecesinin artışıyla arttığını; ancak hile yapanların sayısı arttığında ise duyarlılığın azaldığını ortaya koymuştur. DGIRTM,  $Iz$  ile karşılaştırıldığında daha duyarlı olduğu; belirleyicilikte ise birbirine yakın düzeyde oldukları ancak DGIRT modelinin daha kararlı olduğu ortaya çıkmıştır.

DGIRT modelinin önemli bir katkısı puan kazancı ile geçerliğin kestirilebilmesidir. Puan kazancı azsa geçerlik muhtemelen yüksek, puan kazancı çoksa geçerlik muhtemelen düşüktür. Çalışma sonucunda modelin duyarlık ve belirleyiciliğinin yüksek olduğu ortaya çıkmıştır; ancak hileyi belirleme oranında hile yapan bireylerin artışıyla azalması, bu modelin grupça ya da organize hilede kullanılmasının uygun olmadığını göstermektedir.

### 1.2.3. CTAD

Deng (2008) tarafından hem madde önbilgisi hem de kopyayı belirleme adına önerilen CTAD, birey uyum indekslerinden  $Iz$  indeksinin hileyi belirlemede fazla olan I.Tip hata oranını azaltmak ve düşük olan gücünü arttırmak için geliştirilmiştir. Deng (2008),  $Iz$  indeksine ait bu sınırlılıkların sayılısının gerektirdiği koşullu yokluk dağılımından kaynaklandığını belirtmektedir. Bunun önüne geçebilmek için kestirimde elde edilen EAP (Expected a Posteriori; beklenen sonsal) ile kestirimlerin kesinliği geliştirilmeye çalışılmıştır. Bunun için  $Iz$  indeksinin koşullu dağılımının hesaplanmasında Bayes kestirim yöntemi kullanılmıştır. Bu düzeltmeler sayesinde bahsedilen sınırlılıkların üstesinden gelindiği öne sürülmüştür.

Deng (2008) simülasyon çalışmasında hileyi belirlemedeki CTAD performansını test uzunluğunu değişimleyerek ortaya koymaya çalışmıştır. Çalışmasının sonucunda diğer birey-uyum indekslerinden daha az I. Tip hata oranına sahip olduğunu, indekslerin %10 hile oranında kabul edilebilir güce sahip olmadığını, %30 hile oranında ise diğer indekslerle aynı ya da daha fazla güce sahip olduğunu ortaya koymuştur. Sonuç olarak CTAD, simülasyondaki kestirimler doğru yapılırsa  $Iz$  indeksinin hileyi belirlemede I. Tip hata oranını düşürmekte ve gücünü arttırmaktadır.

### 1.2.4. CUSUM<sub>LR</sub> (Cumulative Sum Statistic-Log Likelihood Ratio)

Armstrong ve Shi (2009), çoktan seçmeli testlerde yapılan kaydırmaları belirlemede kullanılan birikimli toplam istatistiğini (CUSUM, Cumulative Sum Statistic) log olabilirlik oranına dayalı olarak geliştirerek hem madde önbilgisinin hem de kopyanın belirlenebileceğini öne sürmektedirler. Bunu ortaya koymak için çalışmalarında gerçek ve simülasyon verisi kullanarak CUSUM<sub>LR</sub>, CUSUM<sub>VM</sub> (Meijer, 2001; van Krimpen-Stoop ve Meijer, 2001) ve birey-uyum indekslerinden  $Iz$  (Drasgow ve diğ., 1985),  $U$  (Wright ve Stone, 1979),  $W$  (Wright, 1980) ve  $UB$  (Smith, 1985) istatistikleriyle karşılaştırmışlardır. Özellikle simülasyon verilerinde anormal kaydırma tepkileri oluşturarak senaryo üretilip tepki sayısı değişimlenmiştir. Simülasyon çalışması sonucunda CUSUM istatistiklerinin diğer indekslere göre daha düşük I. Tip hata barındırdığı, CUSUM<sub>LR</sub> istatistiğinin CUSUM<sub>VM</sub> yaklaşımına göre daha güçlü ancak az miktarda daha büyük I. Tip hata oranına sahip olduğu ortaya koyulmuştur. Gerçek veri ile çalışıldığında

CUSUM istatistiklerinin diğer birey-uyum indekslerinden %5-10 daha fazla tepkileri anormal olarak adlandırdığı görülmüştür. Bunun nedeni olarak yazarlar, diğer birey-uyum indekslerine göre CUSUM istatistiklerinin oluşturulan senaryoları daha çok anormal tepki örüntüleri olarak değerlendirdiğini ve bunun yerel bağımsızlık sayılısının çığnenmesinden kaynaklandığını belirtmişlerdir. Bu nedenle  $CUSUM_{LR}$  istatistiğinin model-veri uyumsuzluğunun farklı çeşitlerini belirlemede daha duyarlı olabileceği belirtilmiştir.

### 1.2.5. Faktör Analizi

Clark (2010) çok boyutlu bir yaklaşımla hileyi belirlemede faktör analizi yöntemini önermiştir. Araştırmacı tarafından eğer hile teşhir olan maddelerin tek boyutluluğunu ihlal etmişse, hilenin sistematik bir hal alacağı ve tek-faktör ve iki-faktör modelleri ile hesaplanan birey-uyumlarının karşılaştırılmasıyla hilenin ortaya çıkarılabileceği düşünülmektedir. Bu doğrultuda yapılan simülasyon çalışmasında tek-faktör model olan *lco* ve iki-faktör model olan *M-lco* (Ferrando, 2007, 2009) istatistiklerinden elde edilen birey-uyum farkı (*lco* farkı) kullanılmıştır.

Çalışması sonucunda *lcz* istatistiğiyle karşılaştırılan *lco* farkı hile yapanların sayısının az olduğu durumlarda yakın, fazla olduğunda ise daha büyük hile belirleme oranına sahip olduğu; ancak bu sayı çok düşük ya da çok yüksek olduğunda çok daha düşük güce sahip olduğu ortaya konmuştur. Faktör yükleri ile ortaya konan teşhir olan maddelerin, hile yapanların sayısının en az %5 ve madde sayısının yarısından azının teşhir olduğu durumlarda belirlenebileceği ortaya konmuştur. Sonuç olarak yazar bu yaklaşımın hem teşhir olan maddelerin hem de hile yapanların belirlenmesinde kullanışlı olduğunu vurgulamıştır.

### 1.2.6. Birey-Uyum İndeksleri

Testte yapılan hile türlerinden özellikle madde ön bilgisini belirlemede sıklıkla kullanılan birey-uyum indeksleri, bireyin madde tepki örüntülerinin model ile tutarlılığını belirlemeye çalışan indekslerdir (Embretson ve Reise, 2000). Model ile tepkiler arasındaki uyumsuzluğu ortaya koyabilen bu indeksler böylelikle bireylerin tepkilerindeki anormalliklerin belirlenmesi sağlanabilmektedir.

Birey-uyum indeksleri oldukça fazla sayıdadır ve bu indeksleri Meijer ve Sijtsma (2001) ve Karabatsos (2003) açık bir şekilde özetlemiştir. Meijer ve Sijtsma (2001) birey-uyum indekslerini incelerken KTK'ye ya da gruba dayalı, Rasch modele, iki ve üç parametrelili lojistik modele dayalı ve BOBUT uygulamalarında görülen anormal tepki örüntülerini belirlemeye dayanan indeksler olarak ayırmıştır. Bu çalışmada alanyazın tarafından kabul görmüş bu

sınıflandırma ve buna ek olarak Karabatsos'un (2003) sınıflandırması göz önünde bulundurularak içlerinde sıklıkla kullanılan birey-uyum indeksleri ele alınacaktır.

**Tablo 1.2.1.**

Birey-Uyum İndeksleri (Meijer ve Sijtsma, 2001; Karabatsos, 2003)

<b>Birey-Uyum İndeksleri</b>
<b>KTK'ye veya Gruba Dayalı (Parametrik Olmayan) Birey-Uyum İndeksleri</b>
<i>G</i> (Guttman, 1944, 1950)
<i>G*</i> (van der Flier, 1977)
<i>rpbis, rbis</i> (Donlon ve Fischer, 1968)
<i>C</i> (Sato, 1975)
<i>U</i> (van der Flier, 1980; Meijer, 1994)
<i>A<sub>i</sub>, D<sub>i</sub>, E<sub>i</sub></i> (Kane ve Brennan, 1980)
<i>C*</i> (Harnisch ve Linn, 1981)
<i>ZU3</i> (van der Flier, 1982)
<i>NCl, ICI</i> (Tatsuoka ve Tatsuoka, 1983)
<i>H<sub>i</sub><sup>T</sup></i> (Sijtsma, 1986; Sijtsma ve Meijer, 1992)
<b>Rasch Model İçin Birey-Uyum İndeksleri</b>
<i>U</i> (Wright ve Stone, 1979)
<i>W</i> (Wright ve Masters, 1982)
<i>UB, UW</i> (Smith, 1985)
<i>M</i> (Molenaar ve Hoijtink, 1990)
<i>χ<sub>sc</sub><sup>2</sup></i> (Klauer ve Rettig, 1990)
<i>T(X)</i> (Klauer, 1991, 1995)
<b>2PL ve 3PL Model Uyum İndeksleri</b>
<i>I<sub>0</sub></i> (Levine ve Rubin, 1979)
<i>D(θ)</i> (Weiss, 1973; Trabin ve Weiss, 1983)
<i>ECI</i> (Tatsuoka, 1984)
<i>Iz</i> (Drasgow, Levine, ve Williams, 1985)
<i>JK, O/E</i> (Drasgow, Levine ve McLaughlin, 1987)
<i>Lzm</i> (Drasgow, Levine ve McLaughlin, 1991)
<i>c</i> (Levine ve Drasgow, 1988)
<b>CAT İçin Birey-Uyum İndeksleri</b>
<i>K</i> (Bradlow, Weiss ve Cho, 1998)
<i>T</i> (van Krimpen-Stoop ve Meijer, 2000)
<i>Zc</i> (McLeod ve Lewis, 1999)

### 1.2.6.1. Gruba Dayalı (Parametrik Olmayan) Birey-Uyum İndeksleri

Birey-uyum istatistiği bir bireyin cevaplarının bir dizi test maddesine olan uyum derecesini belirler (Karabatsos, 2003). Gruba dayalı birey-uyum indeksleri ya da parametrik olmayan birey-uyum indekslerinde (Guttman'a dayalı, uyuşmaya (agreement) dayalı indeksler) genellikle doğru-oran ( $p$ ) değerleri 0'a yakın olan maddeler doğru cevaplandırılmaya meyilliye ve  $p$  değerleri 1'e yakın olan maddeler yanlış cevaplandırılmaya meyilliye tepki örüntülerini anormal olarak nitelendirilmektedir (Meijer ve Sijtsma, 2001). Bu indeksler bireyin puan örüntüsünü aynı örnekteki başka bir bireyin puan örüntüsü ile karşılaştırmaktadır (Meijer ve Sijtsma, 2001). Gruba dayalı birey-uyum indekslerinin çoğu ise hesaplaması nispeten kolay olan Guttman (1944, 1950) modeline dayanmaktadır. Parametrik olmayan indeksler belli bir ağırlık ( $w$ ) belirlenerek hesaplanmaktadır. Parametrik olmayan ilk birey-uyum indeksi Guttman (1944) tarafından  $G$  istatistiği olarak üretilmiştir. Bu indeksi hesaplamada maddeler güçlüklerine göre zordan kolay doğru sıralandıktan sonra (1,0), (0,0), (0,1) ve (1,1) şeklinde Guttman puan örüntüleri oluşturulur. Maddeyi doğru cevaplama 1, maddeyi yanlış cevaplama 0 ise, madde-puan örüntüsünde (1,0) şeklinde oluşan bir örüntü "hata" veya "ters" olarak adlandırılır (Meijer ve Sijtsma, 2001). Gruba dayalı istatistiklerin tümü bu ters örüntülerin ağırlıklandırılmış sayılarına bağlıdır ve sadece farklı ağırlıklandırmalardan dolayı farklılaşmaktadır. Örneğin Guttman modelinde eğer bir birey sadece kolay maddeleri doğru cevaplıyorsa bu bireyin cevap örüntüsü "Guttman Mükemmel Örüntü" olarak adlandırılır ve bu örüntüden sapmalar  $G$  istatistiğini vermektedir.

Tablo 1.2.2'de H bireyi sadece en zor madde ile en kolay maddeyi doğru yanıtlamıştır. Oysa beklenen zor madde doğru yanıtlanıyorsa 2 ve 3. maddelerin de doğru yanıtlanmasıdır. Bu durumda 2 birim sapma vardır ve  $G=2'$ dir. Test uzunluğu arttıkça  $G$  değeri de artma eğiliminde olduğundan  $G^*$  istatistiği (van der Flier, 1977) üretilmiştir.  $G^*$  istatistiği  $G$  istatistiğinin aralığını [0,1] olarak normalleştirmektedir (Karabatsos, 2003).

$G$  istatistiği kullanılarak iki birey-uyum indeksi daha türetilmiştir. Bu indekslerden biri NCI (Norm Conformity Index; Tatsuoka ve Tatsuoka, 1983), diğeri ise ICI'dir (Individual Consistency Index; Tatsuoka ve Tatsuoka, 1983).

**Tablo 1.2.2.**  
Guttman Modeline Dayalı İndekslere İlişkin Örnek Bir Uygulama

Birey	Maddeler				Tepki Çiftleri	G	G*
	1 (Zor)	2	3	4 (Kolay)			
A	1	1	1	1	(11) (11) (11) (11) (11) (11)	0	0.00
B	0	1	1	1	(01) (01) (01) (11) (11) (11)	0	0.00
C	0	0	1	1	(00) (01) (01) (01) (01) (11)	0	0.00
D	0	0	0	1	(00) (00) (01) (00) (01) (01)	0	0.00
E	0	0	0	0	(00) (00) (00) (00) (00) (00)	0	0.00
F*	0	0	1	0	(00) (01) (00) (01) (00) (10)	1	0.33
G*	0	1	0	1	(01) (00) (01) (10) (11) (01)	1	0.33
H*	1	0	0	1	(10) (10) (11) (00) (01) (01)	2	0.50
I*	1	1	1	0	(11) (11) (10) (11) (10) (10)	3	1.00
Güçlük	0.33	0.44	0.55	0.67	Madde Çiftleri (12) (13) (14) (23) (24) (34)	$G = \sum_{he} u_{nh}(1 - u_{ne})$	$G^* = \frac{G}{X_i(n - X_i)}$
(*) Anormal tepki örüntüsünü ifade etmektedir.							

Diğer bir parametrik olmayan birey-uyum indeksi ise Kane ve Brennan (1980) tarafından üretilen uyuşmaya dayalı *A* indeksidir. Bu indeks bir bireyin *j* maddesine verdiği tepki ( $u_j$ ) ile o maddenin güçlüğü ( $p_j$ ) eşleşmesine bağlıdır. Aşağıda örnek bir hesaplama için verilen tabloda anormal tepki örüntülerinin miktarını ölçmeye çalışan uyuşmazlık indeksi (Disagreement index, *D*) ile bireyin gözlenen uyuşma indeksi oranını veren güvenilirlik indeksi (Dependability index, *E*) yer almaktadır. *D* indeksinin [0,1] aralığında normalleştirilmesiyle elde edilen *E* indeksi 1 olduğunda tepki örüntüsü Guttman Mükemmel Örüntü ile de uyum göstermektedir.

**Tablo 1.2.3.**  
A, D ve E İndeksleri İçin Örnek Bir Uygulama

Birey	Maddeler				A (Uyuşma)	D (Uyuşmazlık)	E (Güvenilirlik)
	1 (Zor)	2	3	4(Kolay)			
A	1	1	1	1	$.33+.44+.55+.67=2.00$	$2.00-2.00=0.00$	1
B	0	1	1	1	$.44+.55+.67=1.67$	$1.67-1.67=0.00$	1
C	0	0	1	1	$.55+.67=1.23$	$1.23-1.23=0.00$	1
D	0	0	0	1	0.67	$0.67-0.67=0.00$	1
E	0	0	0	0	0.00	0.00	---
F*	0	0	1	0	0.55	$0.67-0.55=0.12$	0.82
G*	0	1	0	1	$.44+.67=1.11$	$1.23-1.11=0.12$	0.90
H*	1	0	0	1	$.33+.67=1.00$	$1.23-1.00=0.23$	0.81
I*	1	1	1	0	$.33+.44+.55=1.33$	$1.67-1.33=0.33$	0.80
Güçlük	0.33	0.44	0.55	0.67	$A = \sum_{j=1}^J u_j p_j$	$D = A_{max} - A$	$E = \frac{A}{A_{max}}$
(*) Anormal tepki örüntüsünü ifade etmektedir.							

Korelasyon ya da kovaryansa dayalı parametrik olmayan indeksler ise bireylerin maddelere verdikleri tepki vektörleri ( $X_i$ ) ile her bir maddede  $N$  tane bireyin doğru cevaplarının oranı ( $P$ ) arasındaki ilişkiye dayanmaktadır. Bu ilişki için Donlan ve Fischer (1968) bireysel nokta çift serili korelasyonun ( $r_{pbis}$ ) ve bireysel çift serili korelasyonun ( $r_{bis}$ ) kullanımını önermiştir. Sato (1975) ise anormal cevaplayıcıları belirlemek için Uyarı İndeksini (Caution Index,  $C$ ) önermiştir. Bu indeks yine Guttman Mükemmel Örüntüden sapan tepki örüntülerinin derecesini ortaya koymayı amaçlamaktadır. Bunun için  $X_i$  ile  $P$  arasındaki kovaryanstan faydalanmaktadır.  $C$  indeksi hesaplanırken bu kovaryansların oranından bir çıkarılmaktadır.

Harnisch ve Linn (1981) ise uyarı indeksini normalleştirerek MCI (Modified Caution Index) üretmişlerdir. Bu indeks Guttman Mükemmel modelinin cevaplarının vektörlerinin kovaryansı ile çok zor  $x$  maddeleri için Guttman tepkilerine mükemmel zıtlıkta olan tepki vektörünün kovaryansı arasındaki oranı vermektedir. Karabatsos (2003) anormal tepki

örüntülerini belirlemede  $MCI \geq 0.26$  kritik değerini kullanmayı önermektedir. Tabloda tüm anormal tepki örüntüsü sergileyen bireylerin bu değerin üzerinde kaldığı görülmektedir.

**Tablo 1.2.4.**

Korelasyon ya da Kovaryansa Dayalı İndeksler İçin Örnek Bir Uygulama

Birey	Maddeler				$r_{pbis}$	MCI	$H_i^T$
	1 (Zor)	2	3	4 (Kolay)			
A	1	1	1	1	---	0	---
B	0	1	1	1	0.765	0	.273
C	0	0	1	1	0.890	0	.250
D	0	0	0	1	0.788	0	.273
E	0	0	0	0	---	0	---
F*	0	0	1	0	0.240	0.36*	-.091*
G*	0	1	0	1	0.455	0.50*	.000*
H*	1	0	0	1	0.020	0.50*	-.250*
I*	1	1	1	0	-0.788	0.82*	-.818*
Güçlü k	0.33	0.44	0.55	0.67	$r_{pbis} = Corr(X_n, p)$	$MCI = \frac{Cov(X_n^*, p) - Cov(X, p)}{Cov(X_n^*, p) - Cov(X_n', p)}$	$H_i^T = \frac{\sum_{j \neq i} \sigma_{ij}}{\sum_{j \neq i} \sigma_{ij}^{max}}$
(*) Anormal tepki örüntüsünü ifade etmektedir.							

van der Flier (1980) uyarı indeksi için iki dönüştürme önermiştir. Bunlardan biri MCI ile aynı şekilde hesaplanan; ancak sadece kovaryans yerine logaritma kullanılan  $U3$  indeksidir; diğeri ise bu indeksin standardize edildiği  $ZU3$  indeksidir.

$$ZU3 = \frac{U3 - E(U3)}{[Var(U3)]^{1/2}} \quad (15)$$

Diğer bir korelasyona dayalı birey uyum indeksi ise Sijtsma (1986; Sijtsma ve Meijer, 1992) tarafından önerilen  $H_i^T$  istatistiğidir. Bu istatistik bir bireyin cevap vektörü ile diğer bireylerin cevap vektörleri arasındaki uyuşmayı incelemektedir.  $\beta_i$  yerel bağımsız olan tekrarlı ölçümlerde  $k$  sabit maddeler dizisi için  $i$  bireyinin maddelere verdiği doğru cevapların beklenen oranı olsun.  $\beta_{ij}$  ise  $i$  ve  $j$  bireylerinin aynı şekilde beklenen oranı olsun.  $\sigma_{ij} = \beta_{ij} - \beta_i\beta_j$  ise  $i$  ve  $j$  bireylerinin puanları arasındaki kovaryanstır. Eğer bireyler arasında  $i < j$  ilişkisi varsa; dolayısıyla  $\beta_i \leq \beta_j$ , maksimum kovaryans ise  $\beta_{ij} = \beta_j$  olduğunda elde edilmektedir.

$$\sigma_{ij}^{max} = \beta_i(1 - \beta_j) \quad (16)$$

Tek bir birey için  $H_i^T$  istatistiğinin hesaplanmasında kullanılan eşitlik şu şekildedir:

$$H_i^T = \frac{\sum_{j \neq i} \sigma_{ij}}{\sum_{j \neq i} \sigma_{ij}^{max}} \quad (17)$$

$i$  ve  $j$  bireyine ait madde-puan örüntüleri arasındaki kovaryansların her biri maksimum değerini aldığı anda  $H_i^T$  istatistiği de maksimum değeri olan 1'i almaktadır ve ortalama kovaryans (pay) ise 0 olduğunda  $H_i^T = 0$  olmaktadır, negatif olduğunda da  $H_i^T < 0$  olmaktadır (Meijer ve Sijtsma, 2001).

Karabatsos'a (2003) göre  $H_i^T \leq 0.22$  kritik değerinde anormal tepki örüntüleri sergileyen bireyler en iyi şekilde belirlenmektedir.

#### 1.2.6.2. Madde Tepki Kuramına Dayalı Birey-Uyum İndeksleri

MTK'ye dayalı birey-uyum indeksleri Rasch modeli, 2PL ve 3PL modeline bağlı olarak farklılaşmaktadır. Meijer ve Sijtsma (2001) parametrik olan bu istatistikleri artıklara dayalı ve olabirliğe dayalı olmak üzere iki sınıfta incelemiştir.

##### 1.2.6.2.1. Artıklara Dayalı İndeksler

Artıklara dayalı birey-uyum indeksleri genellikle testteki maddelerde beklenen ve gözlenen çıktıların ağırlıklandırılmış farklarının toplamı olarak gösterilmektedir (Cook, 2013). Yapılan işlemlerde artıkların birbirini yok etmesini engellemek için kareleri alınarak toplam yapılmaktadır. Ortalama artıkların karesine dayalı bu indeksleri kısaca şöyle tanımlayabiliriz:

##### **U ve W istatistikleri:**

Wright ve Stone (1979) tarafından önerilen ***U istatistiği*** standardize edilmiş artıkların karelerine dayanmaktadır. Bu indeks *uyum dışı ortalama kare* (Outfit Mean Square) olarak adlandırılmaktadır. Bir  $a$  bireyinin  $i$  maddesine verdiği yanıt ile bu maddeye beklenen doğru yanıt verme olasılığı arasındaki tutarsızlık şöyledir:

$$Y_{ia} = (u_{ia} - P_{ia}) \quad (18)$$

Bu  $Y_{ia}$  değerinin büyük olması  $i$  maddesine  $a$  bireyinin anormal tepki verdiğini göstermektedir. Artıkların standardize edilmesi ile karesi  $\chi^2$  dağılımı sergileyen Eşitlik 19 elde edilmektedir;

$$z_{ia} = \frac{u_{ia} - E[u_{ia}]}{\sqrt{Var[u_{ia}]}} = \frac{u_{ia} - P_{ia}}{\sqrt{P_{ia}(1 - P_{ia})}} \quad (19)$$

$U$  istatistiğine ait eşitlik şöyledir:

$$U = OMS = \sum_{i=1}^n \frac{z_{ia}^2}{n} \quad (20)$$

$U$  istatistiği,  $n$  maddeye dayalı standardize edilmiş artıkların karelerinin ortalaması şeklinde yorumlanmaktadır (Meijer ve Sijtsma, 2001).

$U$  istatistiğindeki her bir karelerin standartlaştırılmış artıkların karesi, her bir maddenin varyansına dayanmaktadır. Bu varyans ise  $P_{ia}(1 - P_{ia})$  şeklinde gösterilmektedir. Bu varyanslar ise bir bireyin yeteneği ile madde güçlüğü arasındaki tutarsızlığa dayanmaktadır.

Wright (1980) diğer bir artıklara dayalı indeks olan uyum içi ortalama kare (Infit Mean Square, IMS) olarak adlandırılan  **$W$  istatistiğini** türetmiştir. Bu istatistik varyans-artıkların karelerinin ağırlıklandırılmış ortalaması olarak hesaplanmaktadır;

$$W = IMS = \frac{\sum (u_{ia} - P_{ia})^2}{\sum P_{ia}(1 - P_{ia})} \quad (21)$$

Wright ve Stone (1979)  $W$  istatistiğinin, bireyin yeteneğinden farklı güçlüğü sahip bir maddeye verilen beklenmeyen tepkilere  $U$  istatistiğine göre daha az hassas olduğunu varsaymıştır (Meijer ve Sijtsma, 2001). Karabatsos (2003) ise çalışmasında 1.3 değerinin  $U$  ya da  $W$  istatistiği için minimum kritik değer olduğunu vurgulamıştır. Wright (1995) çalışmasında ise bu istatistikler için en düşük limitin 0.8 değerine ayarlandığında cevap örüntülerinin modelle en iyi şekilde uyum gösterdiğini ortaya koymuştur.

Bu artıklara dayalı indekslerin Guttman Mükemmel Örüntüden sapmaları inceleyen parametrik olmayan indekslerden farkı bireylerin tepki örüntülerinin olasılık modellerinden beklenen örüntülerden sapmaları incelemesidir.

#### **UB ve UW İstatistikleri:**

Smith (1995) tarafından geliştirilen  $UB$  ve  $UW$  istatistikleri Rasch modele dayanmaktadır. Smith (1995), testi  $S$  tane örtüşmeyen madde alt kümelerine ayırarak kümeler arası ağırlıklandırılmamış uyum istatistiği hesaplamıştır (Meijer ve Sijtsma, 2001).

$$UB = \frac{1}{S-1} \sum_{s=1}^S \frac{\{\sum_{i \in A_s} [X_i - P_i(\theta)]\}^2}{\sum_{i \in A_s} P_i(\theta)[1 - P_i(\theta)]} \quad (22)$$

Alt grup içindeki madde sayısı  $m_s$  olup, kümeler için birey indeksi Eşitlik 23 ile elde edilmektedir.

$$UW = \frac{1}{m_s} \sum_{i \in A_s} \frac{[X_i - P_i(\theta)]^2}{k P_i(\theta)[1 - P_i(\theta)]} \quad (23)$$

#### 1.2.6.2.2. Olabilirliğe Dayalı İndeksler

Olabilirliğe dayalı birey-uyum indekslerinin tümü log-olabilirlik fonksiyonuna bağlıdır. Olabilirlik fonksiyonunun logaritmasını maksimum yapan yetenek düzeyinin kestirimi (MLE) ile beklenen tepkiler ortaya koyularak anormal tepki örüntüleri belirlenmeye çalışılmaktadır. Bu sınıfa giren birey-uyum indeksleri şu başlıklar altında özetleyebiliriz.

##### $l_0$ , $l_z$ ve $l_{zm}$ İndeksleri:

İlk olarak Levine ve Rubin (1979) tarafından birey-uyum indeksi olarak önerilen  $l_0$  log-olabilirlik fonksiyonu ile bireylerin MTK modeli tarafından üretilen olasılık fonksiyonun maksimum değerlerine uymayan tepkileri ortaya koymaya çalışılmaktadır. Bu istatistiğe ait eşit şöyledir:

$$l_0 = \sum_{i=1}^n \{X_i \ln P_i(\theta) + (1 - X_i) \ln [1 - P_i(\theta)]\} \quad (24)$$

Bu istatistiğin kullanımında şu problemler ortaya çıkmaktadır (Meijer ve Sijtsma, 2001)

1. Standardize değildir.
2. Bir madde-puan örüntüsünün uyumsuz şeklinde sınıflandırılabilmesi için yokluk hipotezine ihtiyaç vardır. Bu yokluk hipotezi ise  $l_0$  için bilinmemektedir.

$l_0$  istatistiğinin bu sınırlılıkları nedeniyle Drasgow, Levine ve Williams (1985) standartlaştırılmış hali olan  $l_z$  indeksini önermiştir. Bu indeks sade ve gerçek yetenek düzeyleri ele alındığında asimptotik standart normal dağılıma sahiptir (Meijer ve Sijtsma, 2001),

$$l_z = \frac{l_0 - E(l_0)}{[Var(l_0)]^{1/2}} \quad (25)$$

$l_0$  istatistiğinin beklenen değeri ile varyansına ilişkin eşitlikler ise;

$$E(l_0) = \sum_{i=1}^n \{P_i(\theta) \ln[P_i(\theta)] + (1 - P_i(\theta)) \ln [1 - P_i(\theta)]\} \quad (26)$$

ve

$$Var(l_0) = \sum_{i=1}^n P_i(\theta)(1 - P_i(\theta)) \left[ \ln \frac{P_i(\theta)}{1 - P_i(\theta)} \right]^2 \quad (27)$$

Yüksek negatif değerler (-2'den küçük) bireyin tepki örüntülerinin anormal olduğu yönünde yorumlanırken, yüksek pozitif değerler ise (+2'den büyük) tepkilerin model ile oldukça iyi bir şekilde uyum gösterdiği yönünde yorumlanabilmektedir.

$l_z$  indeksi sadece gerçek yetenek düzeyleri kullanıldığında asimptotik normal dağılım sergilemekte olup (Magis, Raîche ve Beland, 2012), gerçek yetenek yerine kestirilen yetenek düzeyi kullanıldığında dağılımın negatif çarpık olduğu ve varyansının beklenenden daha düşük değerler aldığı ortaya çıkmıştır (Molenaar ve Hoijtink, 1990; Reise, 1995). Çalışmalar  $l_z$  indeksinin test uzunluğundan oldukça etkilendiğini göstermektedir. Test uzunluğu azaldıkça anormal tepkilerin belirlenmesi de zorlaşmaktadır (Schmitt ve diğ., 1999).

Molenaar ve Hoijtink, (1990, s.96)  $l_0$  istatistiğini Rasch modeli altında Eşitlik 28'deki gibi ifade edilebileceğini ortaya koymuştur;

$$d_0 = - \sum_{i=1}^n \ln [(1 + \exp(\theta - \delta_i)) + r\theta] \quad (28)$$

olmak üzere;

$$l_0 = d_0 + M \quad (29)$$

ve

$$M = - \sum_{i=1}^n \delta_i X_i \quad (30)$$

dır.

Hesaplanmasının kolay olması nedeniyle Molenaar ve Hoijtink (1990)  $l_0$  istatistiğinin yerine  $M$  istatistiğinin kullanımını önermektedir.

Drasgow, Levine ve McLaughlin, (1991) tarafından  $S$  alt testten oluşan bir tek boyutlu test için önerdiği  $l_z$ 'nin genelleştirilmiş hali olan  $l_{zm}$  sadece  $S$  alt testlerin varyanslarını içermesinden ötürü farklılaşmaktadır.

$$l_{zm} = \frac{\sum_{s=1}^S \{l_0^{(s)} - E[l_0^{(s)}]\}}{\sum_{s=1}^S \{Var[l_0^{(s)}]\}^{1/2}} \quad (31)$$

Bu istatistik her ne kadar madde-puan örüntülerinde uyumsuzluğu belirlemede etkili olsa da bu belirleme oranları uzun testlerde birbirine yakın olmakta, bununla birlikte  $l_z$ 'nin sınırlılıklarını içermektedir.

### 1.2.6.3. Birey Tepki Fonksiyonu (Person Response Function, PRF)

Trabin ve Weiss (1983) uyumsuz madde puan örüntülerini belirleyebilmek için birey tepki fonksiyonunu (BTF) önermiştir (Meijer ve Sijtsma, 2001). Trabin ve Weiss (1983) sabit  $\theta$  değeri için BTF'yi, maddeyi doğru cevaplama olasılığını maddenin yerinin ( $\delta$ , *item location*) fonksiyonu olarak tanımlamıştır.

Gözlenen BTF'yi belirlemede öncelikle maddeler güçlüklerine göre  $\delta$  değerleri düşükten yükseğe doğru sıralanır. Sonrasında maddeler  $\delta$  parametrelerine göre gruplandırılarak maddelerin alt kümeleri oluşturulur. Oluşturulan her alt kümedeki doğru cevap oranları belirlenerek sabit yetenek parametresi kestirilir ve bu şekilde gözlenen BTF oluşturulur.

Beklenen BTF ise doğru cevap olasılıklarının ortalaması 3PL modele göre kestirilerek oluşturulur. Elde edilen gözlenen BTF ile beklenen BTF arasındaki farka bakılır ve bu farkın büyük olması durumunda uyumsuz tepkilerin olduğu yorumu yapılmaktadır.

$k$  maddenin sıralandığı ( $\delta_1 < \delta_2 < \dots < \delta_k$ ) ve  $m$  tane madde içeren  $S$  tane alt testin sıralandığını ( $A_s$ ) düşünelim. Dolayısıyla  $A_1 = (1, 2, \dots, m)$ ,  $A_2 = (m + 1, \dots, 2m)$ , ...,  $A_S = (k - m + 1, \dots, k)$  ve  $S \times m = k$  olmaktadır. 3PL model altında beklenen BTF için doğru cevapların beklenen oranının kestirimine ait eşitlik;

$$m^{-1} \sum_{i \in A_s} P_i(\hat{\theta}), \quad s = 1, 2, \dots, S \quad (32)$$

elde edilen beklenen BTF gözlenen BTF ile karşılaştırılır.

$$m^{-1} \sum_{i \in A_s} X_i, \quad s = 1, 2, \dots, S \quad (33)$$

Belirli bir  $\hat{\theta}$  için her bir alt testte beklenen BTF ile gözlenen BTF arasındaki fark ile bunun madde sayısına oranı;

$$D_s(\hat{\theta}) = m^{-1} \sum_{i \in A_s} [X_i - P_i(\hat{\theta})], \quad s = 1, 2, \dots, S \quad (34)$$

olarak hesaplanmaktadır. Sonrasında her alt test için  $D_s$  değerleri toplanır ve

$$D(\hat{\theta}) = \sum_{s=1}^S D_s(\hat{\theta}) \quad (35)$$

elde edilir.

$D(\hat{\theta})$  bireyin model ile uyumunu veren bir ölçüdür (Meijer ve Sijtsma, 2001). Örneğin zor maddelerde diğer bireyden kopya çeken bireylerin, bu alt testlerden aldıkları puanlar, beklenen BTF yardımıyla kestirilenden daha yüksek olacaktır (Meijer ve Sijtsma, 2001, s.122).

Klauer ve Retting (1990), Trabin ve Weiss'in (1973) yöntemlerini kullanarak üç standartlaştırılmış birey-uyum indeksi önermişlerdir (Meijer ve Sijtsma, 2001). Uzun testlerde asimptotik olarak  $\chi^2$  dağılımı sergileyen bu indekslerden en bilineni  $\chi_{sc}^2$  indeksidir.

$$\chi_{sc}^2 = \sum_{s=1}^S \frac{V_s^2(\hat{\theta})}{I_s(\hat{\theta})} \quad (36)$$

İndekse ait Eşitlik 35'te yer alan  $V_s^2(\hat{\theta})$  ise;

$$V_s^2(\hat{\theta}) = \sum_{g \in A_z} [X_i - P_i(\hat{\theta})] w_i(\hat{\theta}) \quad (37)$$

ve

$$w_i(\hat{\theta}) = \frac{dP_i(\theta)/d\theta}{P_i(\theta)(1 - P_i(\theta))} \quad (38)$$

dır.

Klauer ve Retting (1990) çalışmalarında en az 80 test uzunluğunda bu istatistiğin  $\chi^2$  dağılımı sergilediğini göstermişlerdir.

#### 1.2.6.4. Optimal Birey-Uyum İndeksleri

Levine ve Drasgow (1988) uyumsuz madde-puan örüntülerini belirlemek amacıyla bir olabilirlik oranı istatistiği önermiştir. Bu istatistiğin aynı I. Tip hata oranında diğerlerine göre daha yüksek güce sahip olduğunu ortaya koymuşlardır. Bu istatistik;

$$\lambda(X) = \frac{P(X = x)_{uyumsuzluk}}{P(X = x)_{uyumluluk}} \quad (39)$$

şeklinde hesaplanmaktadır. Eğer bu istatistik büyük değerlere sahipse tepki örüntülerinin modelle uyumsuz olduğu söylenebilmektedir.

Aynı zamanda Klauer (1991, 1995) Rasch modelini içeren yokluk hipotezine karşı alternatif hipotezi test ederek  $T(X)$  istatistiğinin uyumsuz tepkileri ortaya koyabileceğini bulmuştur. Bunun için bir üstel Rasch modeli ortaya koymuştur;

$$P(X = x|\theta) = \mu(\theta)h(x) \exp[\theta R(x)] \quad (40)$$

Eşitlik 40'da bilinmeyen terimler şöyledir;

$$\mu(\theta) = \prod_{i=1}^n [1 + \exp(\theta - \delta_i)]^{-1} \quad (41)$$

$$h(x) = \exp\left(-\sum x_i \delta_i\right) \quad (42)$$

Klauer (1995) iki parametrelili (ek olarak yeni bir birey parametresi,  $\eta$ , kullanılmakta) üstel model ile uyumsuz tepki örüntüsünü belirlemeye çalışmaktadır;

$$P(X = x|\theta, \eta) = \mu(\theta, \eta)h(x) \exp[\eta T(x) + \theta R(x)] \quad (43)$$

Bu şekilde alternatif modele dayanan  $T(X)$ ,  $H_0: \eta = 0$  yokluk hipotezine karşı  $H_1: \eta \neq 0$  hipotezi test edilerek bulunmaya çalışılmaktadır.

#### 1.2.6.5. Uyarı (Caution) İndeksine Dayalı İstatistikler

Korelasyon/kovaryansa dayalı indekslerden olan uyarı indeksinden ( $C$ ) yola çıkarak Tatsuoka (1984) MTK modellerine dayalı altı birey-uyum indeksi geliştirmiştir. Uyarı indeksinin genel eşitliği Eşitlik 44'teki gibidir:

$$C_i = 1 - \frac{Cov(X_n, p)}{Cov(X_n^*, p)} \quad (44)$$

Eşitlikte  $X_n$ , bireyin madde puan vektörü,  $X_n^*$ , teorik Guttman vektörü,  $p$ , bireyler boyunca doğru cevaplanan madde puan vektörüdür.

Genişletilmiş ilk indeks olan  $ECI1$  indeksi, bireyin puan vektörü ile bireyler boyunca doğru cevaplanan madde puan vektörü ( $p$ ) arasındaki kovaryansın, MTK modelleri altındaki beklenen doğru cevaplama olasılığı ile bireyler boyunca doğru cevaplanan madde puan vektörü ( $p$ )

arasındaki kovaryansa oranıdır (Meijer ve Sijtsma, 2001).  $ECI1$  indeksi Eşitlik 45 ile hesaplanmaktadır;

$$ECI1 = 1 - \frac{Cov(X_n, p)}{Cov[P(\theta), p]} \quad (45)$$

Diğer beş indeks ise aynı mantığa dayanmakta olup tümü (1) bireyin tepki vektörü ( $X$ ), (2) maddenin  $p$  değerlerinin vektörü ( $P$ ) ya da (3) tüm bireyler boyunca bir doğru cevabın ortalama olasılığına ait vektörün ( $G$ ) kovaryansına ya da korelasyonuna bağlıdır.

$ECI2$  indeksi ise  $n$  tane bireyin bir maddeyi doğru cevaplama olasılığının ortalaması ile madde-puan vektörü arasındaki kovaryansa eşittir (Meijer ve Sijtsma, 2001). Bireyin maddeyi doğru cevaplama olasılığının ortalaması ile  $ECI2$  indeksine ait eşitlikler sırasıyla şu şekildedir;

$$G = \frac{1}{n} \sum_{i=1}^n P_i(\theta) \quad (46)$$

$$ECI2 = 1 - \frac{Cov(X_n, G)}{Cov[P(\theta), G]} \quad (47)$$

$ECI3$  indeksinin ise  $ECI2$ 'den farkı kovaryans yerine korelasyonun kullanılmasıdır. Bu indeks,  $n$  tane bireyin bir maddeyi doğru cevaplama olasılığının ortalaması ile madde-puan vektörü arasındaki korelasyona eşittir (Meijer ve Sijtsma, 2001).  $ECI3$  indeksi Eşitlik 48 yardımıyla hesaplanmaktadır.

$$ECI3 = 1 - \frac{Corr(X_n, G)}{Corr[P(\theta), G]} \quad (48)$$

$ECI4$ ,  $ECI5$  ve  $ECI6$  indeksleri,  $X_n$  tepki vektörü ve  $P(\theta)$  arasındaki korelasyon veya kovaryans ile elde edilmektedir. Bu indekslere ait eşitlikler sırasıyla Eşitlik 49, 50 ve 51'deki gibidir:

$$ECI4 = 1 - \frac{Cov[X_n, P(\theta)]}{Cov[G, P(\theta)]} \quad (49)$$

$$ECI5 = 1 - \frac{Corr[X_n, P(\theta)]}{Corr[G, P(\theta)]} \quad (50)$$

$$ECI6 = 1 - \frac{Cov[X_n, P(\theta)]}{Var[P(\theta)]} \quad (51)$$

Tüm bu altı indeks genel bir ifadeyle (Eşitlik 52) standart normal dağılıma dönüştürülebilmektedir;

$$ECI_{b_z} = \frac{ECI_b - E(ECI_b)}{SE(ECI_b)} \quad (52)$$

İndekslerin değerleri anormal tepkiler arttıkça artmaktadır.

#### 1.2.6.6. Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT) İçin Birey-Uyum İndeksleri

Yukarıdaki bölümlerde de değinildiği gibi BOBUT uygulamalarının son yıllardaki artışıyla maddelerin teşhiri ya da bireylerin önceden maddeler hakkında bilgi sahibi olması ve hatırlaması gibi olumsuz durumları belirleme çalışmaları hız kazanmıştır. Bunu için bazı birey-uyum indeksleri üretilmiştir. Var olan indekslerin dağılım özellikleri BOBUT uygulamalarında beklenen teorik dağılımla uyuşmadığından, üretilen yeni indekslerin daha güçlü olması beklenmektedir (Meijer ve Sijtsma, 2001).

Maddeleri hatırlamayı belirleme adına McLeod ve Lewis (1999)  $Z_c$  istatistiğini önermiştir. Bu istatistik hesaplanmadan önce madde bankası maddelerin güçlüklerine göre kolay-orta-güç olmak üzere üç bölüme ayrılmaktadır. Kolay maddeler için artıkların ortalaması  $Kolay[P_i(\theta) - X_i]$ , çok zor maddeler için artıkların ortalaması  $Zor[P_i(\theta) - X_i]$  olarak ifade edildiğinde  $Z_c$ , Eşitlik 53 yardımıyla hesaplanmaktadır.

$$Z_c = \frac{Kolay[P_i(\theta) - X_i] - Zor[P_i(\theta) - X_i]}{\{\sum_{Kolay}\{P_i(\theta)[1 - P_i(\theta)]\}/n_{Kolay}^2\} + \{\sum_{Zor}\{P_i(\theta)[1 - P_i(\theta)]\}/n_{Zor}^2\}} \quad (53)$$

Bir birey kolay bir maddeyi yanlış, zor bir maddeyi doğru cevapladığında  $Z_c$  pozitif değerler almaktadır. Bu durum da anormal tepki örüntüsü için bir ipucu olarak görülmektedir. Ancak 14 hatırlanmış maddenin yer aldığı bir BOBUT uygulamasında bu istatistiğin düşük güce sahip olduğu ortaya konmuştur (Meijer ve Sijtsma, 2001).

Testin sonlarında kestirilen yetenek düzeyinin, yetenek düzeyine yaklaştığında birbirini izleyen doğru ve yanlış tepkilerden oluşan madde puan örüntüsünün modele uyumlu olup

olmadığını belirlemek için Bradlow Weiss ve Cho (1998) ve van Krimpen-Stoop ve Meijer (2000) bir birey-uyum indeksi önermişlerdir (Meijer ve Sijtsma, 2001). Bu indeks ardışık doğru veya yanlış cevaplanmış maddeleri incelemektedir. Ardışık olan bu bir dizi doğru veya yanlış cevapların olması uyumsuzluğun olduğunu göstermektedir. Böylece çalışmalarında bu pozitif veya negatif artıkların toplamının  $([P_i(\theta) - X_i])$  birikimli toplam işlemi (cumulative sum procedure-CUSUM) olarak kullanılabilceğini bulmuşlardır (Meijer ve Sijtsma, 2001, s.123).

Testteki her  $i$  maddesi için  $([P_i(\theta) - X_i])$ 'nin ağırlıklandırılmış hali olan  $T_i$  istatistiği hesaplanıp, bu  $T_i$  değerlerinin toplamı olan  $T_i$ s artıklar yardımıyla hesaplanmaktadır.

$C_i^+$  i pozitif ardışık artıkların toplamı ile  $C_i^-$  negatif ardışık artıkların toplamı sırasıyla;

$$C_i^+ = \max[0, T_i + C_{i-1}^+] \quad (54)$$

$$C_i^- = \min [0, T_i + C_{i-1}^-] \quad (55)$$

olarak hesaplanmakta ve dolayısıyla  $C_0^+ = C_0^- = 0$  'dır.

$AS$ 'yi alt sınır ve  $ÜS$ 'yi üst sınır olarak düşündüğümüzde,  $C^+ > ÜS$  veya  $C^- < AS$  olduğunda madde-puan örüntüsünün modele uymadığını söyleyebiliriz (Meijer ve Sijtsma, 2001).

### 1.3. Silme ve Anormal Puan Artışı Belirleme Yöntemleri

Silme olarak adlandırılan bu tür hilede bireyler cevap kağıtlarında maddelere verdikleri kendi cevaplarını, başka bireylerden aldıkları ipuçlarıyla silerek değiştirir ve bu değişim esnasında cevap kağıtlarında bir miktar kazıntı oluşur. Bu şekilde hile ile bireyde anormal bir puan artışı meydana gelir. Bunu belirlemek için geliştirilen yöntemler iki kısımda incelenmektedir (Bishop ve Egan, 2017).

#### 1.3.1. Sıkça Görülen Yanlıştan Doğruya Cevap Değişimini Belirlemeye Yönelik Yaklaşımlar

Bu amaca yönelik olarak iki farklı yöntemden söz edilebilir. Bunlardan birincisi Van der Linden ve Jeon (2012) tarafından önerilen "Van der Linden ve Jeon indeksi" adını alan yöntemdir. Diğer yöntem ise Wollack, Cohen ve Eckerly (2015) tarafından önerilen EDI (Erasure Detection Index) yöntemidir. Bu indeksin grup düzeyinde araştırma yapmak için olan uzantısı da mevcuttur.

### **1.3.2. Test Uygulamaları Arasında Oluşan Radikal Puan Farklılaşmasını Belirlemeye Yönelik Yaklaşımlar**

Bu amaca hizmet eden ve ulaşılabilen yöntemler, Jacob ve Levitt (2003,2004) yöntemleri; Skorupski ve Egan (2011, 2014) tarafından önerilen Bayesli Hiyerarşik Lineer Modelleme (Bayesian Hierarchical Linear Modeling) yöntemi; Liu, Liu ve Simon (2014) Bayesli polinom karışık-etki büyüme modeli (Bayesian polynomial mixed-effect growth model) ve Clark, Skorupski ve Murphy (2013) tarafından önerilen birikimli lojit regresyon (cumulative logit regression) yöntemidir.

Tüm bu yöntemler iş birliği yapan birey ve grupları belirlemede kullanmak adına araştırmacılar tarafından uyarlanmaktadır. Ancak alanyazında bu tür hileyi belirlemede sınırlı sayıda araştırma mevcuttur. Bireyler arasındaki iş birliği bazı değişkenlere bağlıdır. Aynı okulda öğrenim görmek, aynı öğretmenin öğrencisi olmak, aynı sosyal etkinlikte ya da grupta yer almak gibi. Aralarında ilişki olan bireyler topluluğundan belli zamanlarda elde edilen ölçümler de dolayısıyla iç içe geçmiş yani belli bir hiyerarşik ya da kümelenmiş yapı sergilemektedir. Öğrencilerin ya da bireylerin ortak paylaşımları ile testlerde yaptıkları iş birliğini ortaya koymada bu yapıyı inceleyen istatistiksel yöntemlerden faydalanılabilmektedir. Bu yöntemler arasında kümeleme analizi, hiyerarşik lineer modelleme, faktör analizi, makine öğrenmesi sayılabilir.

## **1.4. İş Birliği Yapan Grupları Belirlemede Kullanılabilecek Yöntemler**

### **1.4.1. Hiyerarşik Lineer Modelleme**

Bryk ve Raudenbush (1992) tarafından ortaya konan hiyerarşik lineer modelleme (HLM), hiyerarşik bir yapı sergileyen veri setindeki her bir hiyerarşik düzeydeki değişkenler arasındaki ilişkileri araştırmaktadır. HLM'de hiyerarşik yapı içerisindeki her bir düzey, kendi içerisindeki alt modelle temsil edilmektedir. Bu alt modeller, verilen bir düzey içerisindeki değişkenler arasındaki ilişkileri açıklar ve diğer düzeyin bir düzeyde meydana gelen etkisini belirler.

HLM'de, ilk olarak üst düzeydeki birimlerden (gruplardan), daha sonra ise alt birimlerden (gruplar içindeki birimlerden) örnek çekildiği varsayılır (Heck ve Thomas, 2000). Gruplardaki birimlerin benzer özelliklere sahip olmaları bu birimlerden elde edilen gözlemlerin birbirine bağımlı olmasına neden olur. Böylece regresyon analizlerinde istenen değişkenlerin birbirinden bağımsızlığı varsayımının gerekliliği ortadan kalkmış olur ve HLM regresyon modellerinin göz ardı ettiği değişkenler arasındaki bağımlılığı dikkate alınmasıyla diğer yöntemlere göre avantaj sağlamaktadır.

HLM, düzeyler arasındaki (öğrenci ve okul gibi) değişkenlerin uygun ve doğru tahminlerini yapabilmek için kullanılmaktadır. Bu açıdan bakıldığında, HLM test güvenliğini ihlal edecek iş birliği yapan bireyler arasındaki ilişkiyi belirlemede kullanılacak bir yöntem olarak görülebilir.

#### **1.4.2. Faktör Analizi**

Faktör analizi (FA), ilk olarak Karl Pearson, Charles Spearman ve diğer zekayı tanımlamaya ve ölçmeye çalışanlar tarafından 20. yüzyılın başlarında ortaya konmuştur (Johnson ve Wichern, 2007). Faktör analizi, özellikle çok boyutlu olan psikolojik yapılarda birbiriyle ilişkili örtük değişkenleri tanımlamaya yarayan çok değişkenli istatistiksel bir yöntemdir. Bu yöntemin temel amaçlarından biri veri indirgemek, diğeri ise değişkenleri sınıflandırmaktır. Faktör analizinin işlevi kısaca, aralarında ilişki bulunan p sayıdaki değişkenle açıklanan yapıyı, kendi içlerinde ilişkili; ancak aralarında ilişki bulunmayan daha az sayıdaki ( $k < p$ ) yeni değişkenle (faktör) açıklamaktır (Alpar, 2013). Bu açıklamayı yaparken kümeleme analizinde kullanılan birimler arası yakınlık yerine verilerdeki değişim yani varyans-kovaryans bilgilerini kullanmaktadır.

Faktör analizi uygulama şekline ve amacına göre birçok türü vardır. Bunlar arasında sıklıkla kullanılan yöntemlerden biri bir grup gözlenen değişken arasındaki ilişkiden yararlanılarak ilgilenilen yapıyı açıklayan sürekli örtük değişkenleri (faktör) belirlemek için kullanılan açıklayıcı faktör analizi (AFA), diğeri ise açıklayıcı faktör analizi ile ortaya konan faktörlerin varsayımsal ya da kuramsal faktör yapılarına uygunluğunu test etmek için kullanılan doğrulayıcı faktör analizidir (Özdamar, 2013). Diğeri türleri ise R tipi, Q tipi, O tipi, T tipi ve S tipi faktör analizidir.

Faktör analizi türlerinden de görüldüğü gibi bu analizde amaç kısaca ilgilenilen yapıyı açıklayabilen değişkenleri ortaya koymaktır. Bu bakımdan iş birliği yapan bireyler arasındaki ilişkiyi ortaya koymada bu analizden faydalanılabilir.

#### **1.4.3. Makine Öğrenmesi**

Makine öğrenmesi, makinelerin örnek veriler ya da geçmiş deneyimlerden elde ettiği kurallarla etkili ve düşük hatalı tahmin yapabilen algoritmaların tasarlandığı bir öğrenme şeklidir. Deneyimlerle performansını arttıran makinenin ise öğrenmiş olduğu kabul edilmektedir (Harrington, 2012). Makine öğrenmesi, veriden kural öğrenebilen, değişikliklere uyum sağlayabilen ve deneyim ile performansını geliştirebilen yazılımları tasarlamakla ilgilenmektedir (Blum, 2007).

Dolayısıyla makine öğrenmesi algoritmalar yardımıyla elde var olan veri setinden faydalanarak gelecek için tahminde bulunmayı sağlayan bir yöntemdir. Kısaca geçmiş bilgilerden geleceği tahmin etmeye yardımcı olmaktadır. Özellikle çok büyük miktardaki veri setlerini bilgisayar programları ile bu algoritmalar kullanılarak analiz edip yeni veri için en uygun modeli bulmaya çalışmaktadır. Bu sayede testi alan bireylerin ortak karakteristik özellikleri belirlenebilmektedir. Makine öğrenmesi yüz tanıma, istenmeyen e-posta filtrelemesi, optik karakter algılama, konuşma dilini anlama, tıbbî teşhis, kümeleme ve hava durumu tahmini gibi birçok amaçla kullanılmaktadır. Bu amaçlar doğrultusunda makine öğrenmesi alanyazında farklı sınıflandırılmalarla birlikte genellikle denetimli ve denetimsiz öğrenme olarak ikiye ayrılmaktadır (Harrington, 2012; Donalek, 2011; Hastie, Tibshirani ve Friedman, 2008; Witten ve Frank, 2005). Aynı zamanda günümüzde Bishop (2006), Brownlee (2013) ve Murphy (2012) gibi araştırmacılar tarafından denetimli (supervised), denetimsiz (unsupervised) ve pekiştirmeli (reinforcement) öğrenme olarak yapılan sınıflandırmalarla beraber buna yarı-denetimli (semi-supervised) öğrenme türü de eklenerek sınıflandırmanın genişletildiği görülmektedir. Bu araştırmada iş birliği yapan bireylerin tespit edilmesinde yarı-denetimli ve denetimsiz öğrenme türleri kullanılmıştır.

#### **1.4.3.1. Denetimli (Gözetimli/Öğreticili/Danışmanlı) Öğrenme**

Eğitim (training) verisi adı verilen bir veri setini, girdi vektörlerinin örneklerini ve bunlara karşılık gelen hedef çıktılarını içermektedir (Bishop, 2006). Bu tür makine öğrenmelerinin sınıflama ve regresyon olmak üzere iki türü vardır. Sınıflamada, her bir girdi vektörünü sonlu olan kesikli bir kategoriye atamak amaçlanmaktadır. Eğer çıktılar bir ya da daha fazla sürekli değişken içeriyorsa bu durumda regresyon kullanılmaktadır. Denetimli öğrenmede sıklıkla kullanılan yöntemler ise en yakın komşu yöntemleri, basit ve çoklu doğrusal regresyon, lojistik regresyon, Poisson regresyonu, Bayes sınıflandırıcıları, karar ağaçları veya rastgele orman (C4.5, C5, Gini, Random Forest), yapay sinir ağları, SVM (destek vektör makinesi)'dir.

#### **1.4.3.2. Yarı-Denetimli (Yarı-Gözetimli/Yarı-Öğreticili/Yarı-Danışmanlı) Öğrenme**

Yarı-denetimli öğrenme denetimli (tüm veriler etiketli) ve denetimsiz (tüm veriler etiketsiz) öğrenme arasında bir yaklaşımdır. Kendi tahminlerinin doğru olduğunu varsayan bu öğrenme türünde, elde var olan etiketlenmemiş verilere ek olarak tüm verilerin yerine sadece bir miktar verinin denetimli bilgisine ihtiyaç duyulmaktadır. Böylece yarı-denetimli öğrenme etiketlenmiş verinin elde edilmesindeki zaman ve maliyet problemine yönelik, büyük bir

etiketlenmemiş veri ile az miktardaki etiketlenmiş veri kullanılarak sınıflandırma yapma imkanı sağlamaktadır (Amini ve Gallinari, 2002; Zhu, 2005). Yarı-denetimli öğrenme araştırmacılara bu yönden kolaylık sağlamakta ve yüksek doğruluk vermesi nedeniyle teoride ve pratikte büyük ilgi görmektedir (Zhu, 2005). Kopya durumlarında ise yarı-denetimli öğrenme ile sınav sırasında iş birliği yapmış bireylerin ihbar edilmesi ya da bir kısmının yakalanması sonucunda elde edilen bilgiler etiketlenmiş veriyi oluşturarak, diğer iş birliği yapmış ancak bilinmeyen bireyleri tespit etmede kullanılabilir. Yarı-denetimli öğrenmede birçok yöntem kullanılmaktadır. Bunlardan sıklıkla kullanılanları kendi kendine öğrenme (self-learning, self-training, self-labeling, decision-directed learning), birlikte eğitim (co-training), grafik tabanlı yöntemler, TSVM (transductive support vector machines), EM algoritmalarının kullanıldığı karışık yöntemlerdir. Doğasında iki sınıfa ayrılan verilerde birlikte eğitim yöntemi; benzer özelliklere sahip iki nokta aynı sınıfta olma eğiliminde olduğunda grafik tabanlı yöntemler; sınıflar iyi kümelmiş veri üretiyorsa EM algoritmalarının kullanıldığı karışık yöntemler; elde var olan denetlenen sınıflandırıcı karmaşık ve değiştirilmesi zor olduğunda kendi kendine öğrenme yöntemi kullanılmaktadır (Zhu, 2005). Bu çalışmada ise iş birliği yapan bireyleri ve bu bireylerin bulunduğu grupları belirlemede kendi kendine öğrenme yöntemi kullanılmıştır.

#### 1.4.3.2.1. Kendi Kendine Öğrenme (Self-Learning)

En sık kullanılan yarı-denetimli öğrenme yöntemlerden biri olan kendi kendine öğrenme yönteminde (self-learning/self-training/self-labeling/decision-directed learning), tekrar eden denetimli öğrenme yöntemleri kullanılmaktadır. Süreç ilk olarak az miktardaki etiketli verinin eğitilmesi ile başlamaktadır. Her bir etiketlenmemiş veri setleri eğitilen bu etiketli veriler yardımı ile tekrarlanan algoritma süreçleri (EM, Estimation-Maximization) sayesinde etiketlenirler. Her bir adım bir önceki adımda öğrenilen sınıflandırma bilgisi kullanılarak geliştirilir. Bu süreç her bir adımda bir önceki adımda kullanılan denetimli yöntemin ürettiği tahminleri kullanarak yeniden eğitilmesiyle devam eder (Chapelle, Schölkopf ve Zien, 2006). Kendi kendine öğrenme yöntemi için basit bir algoritma aşağıdaki gibidir;

- Etiketlenmiş veri  $L$ , etiketlenmemiş veri  $U$ , maksimum tekrar  $M$  olsun;
  - $M$  sayıda tekrar et;
    - Eğitilmiş  $L$  verisi ile bir  $h$  sınıflandırıcısı eğit,
    - $h$  ile  $U$  içinde veriyi sınıflandır,
    - $U$  içinde en güvenilir alt veri olan  $U'$  verisini bul,
  - $L + U' \rightarrow L$
  - $U - U' \rightarrow U$
  - Tekrarı sonlandır.

Alanyazında bu süreçlerde önem taşıyan güven aralığı seçiminin (confidence selection) farklı yöntemlerle yapılarak yarı-denetimli öğrenme yöntemini geliştiren araştırmaların yapıldığı görülmüştür. Bunlardan biri Wang, Spencer, Ling ve Zhang'in (2008) ortaya koyduğu karar ağaçları sınıflandırıcılarına dayalı uyarlanmış Değer Farkı Metriği (value difference metric) ile güven aralığı seçimidir.

Li ve Zhou (2005) ise kendi kendine öğrenme yönteminin performansını arttırmak için yanlış etiketlenmiş verileri tanımlayarak veri setinden çıkararak SETRED (Self Training with Editing) yöntemini geliştirmişlerdir. SETRED sınıflandırıcısı, kendi kendine öğrenme yönteminde olduğu gibi ilk olarak sürece etiketlenmiş veriyi eğitmekle başlanmaktadır. Bu süreç bir önceki adımdaki tahminlerle etiketlenmemiş verilerin eğitilmesiyle yinelemeli olarak devam etmektedir. SETRED, genişletilmiş etiketli veriye yanlış sınıflandırılmış örneklerin girmesini önlemek için bir değişiklik şeması kullanır. Bunun için her yinelemede yanlış etiketlenmiş örnekleri önceki adımda oluşturduğu komşuluk bilgileriyle tanımlayarak veri setinden çıkarmaktadır. Aşağıda SETRED yöntemine ait basit bir algoritmaya yer verilmiştir;

- Etiketlenmiş veri  $L$ , etiketlenmemiş veri  $U$ , maksimum tekrar  $M$  olsun;
  - $M$  sayıda tekrar et;
    - Eğitilmiş  $L$  verisi ile bir  $h$  sınıflandırıcısı eğit,
    - $h$  ile  $U$  içinde veriyi sınıflandır,
    - $U$  içinde en güvenilir alt veri olan  $U'$  verisini bul,
    - $L + U' \rightarrow L'$
  - $L$  ile  $G$  komşuluk grafiği oluştur;
  - Her bir  $x_i \in L'$  için;
    - gözlenen değerleri hesapla,
    - $G$  ile  $x_i$  komşularını bul,
    - gözlenen değerlerin dağılımını hesapla,
    - sınıfı yanlış tahmin edilen verileri çıkar ( $L' \leftarrow L' - \{(x_i, \hat{y}_i)\}$ ).
  - $U - U' \rightarrow U$
  - Tekrarı sonlandır (Li ve Zhou, 2005).

Bu algoritma yardımıyla kendi kendine öğrenme yönteminde her bir tekrarda yanlış tahminlerden elde edilen bilgilerin kullanımı önlenmiş olmaktadır.

Gu, Zhang ve Lui (2011) ise ISBOLD (instance selection method based on the original labeled) yöntemi ile orijinal etiketlenmiş veriler için Naive Bayes sınıflandırıcısı kullanarak kendi kendine öğrenme ve birlikte eğitim yöntemlerinin performansını arttırmayı hedeflemektedir.

### 1.4.3.3. Denetimsiz (Gözetimsiz/Öğreticisiz/Danışmansız) Öğrenme

Denetimsiz öğrenmede, girdi vektörleri olmasına rağmen bu vektörlere karşılık gelen hedef çıktıları bulunmamaktadır. Oysa denetimli ve yarı-denetimli öğrenmelerde etiketlenmiş veriler bulunmaktadır. Denetimli öğrenmede tahminlerde bulunurken koşullu yoğunluk kestiriminden faydalanılırken, denetimsiz öğrenmede koşulsuz yoğunluk kestirimi yapılmaktadır. Aynı zamanda denetimsiz öğrenmede tek değişkenli olasılık modelleri yerine çok değişkenli olasılık modelleri ortaya konmaya çalışılmaktadır (Murphy, 2012). Denetimli öğrenmede ise genellikle sadece tek bir değişken kestirilmeye çalışılır. Denetimli öğrenmeye karşın birçok alanda kullanılan denetimsiz öğrenmede verileri elle etiketlemek için bir uzmana ihtiyaç duyulmamaktadır (Murphy, 2012). Bu tür öğrenmelerde, *kümeleme*, *yoğunluk (density) kestirimi* ve *görüntüleme (visualization) yöntemleri* kullanılmaktadır. Eğer amaç veri içinde benzer özellikler taşıyan grupları bulmaksa *kümeleme*, girdi içinde verinin dağılımını belirlemekse *yoğunluk kestirimi*, veriyi çok boyutlu uzaydan iki ya da üç boyuta düşürerek göstermekse *görüntüleme yöntemleri* kullanılmaktadır (Bishop, 2006). Bu araştırmada amaç örneklem içindeki iş birliği yapan birey ve grupları ortaya koymaktır; bu nedenle kümeleme yöntemi kullanılmıştır.

#### 1.4.3.3.1. Kümeleme

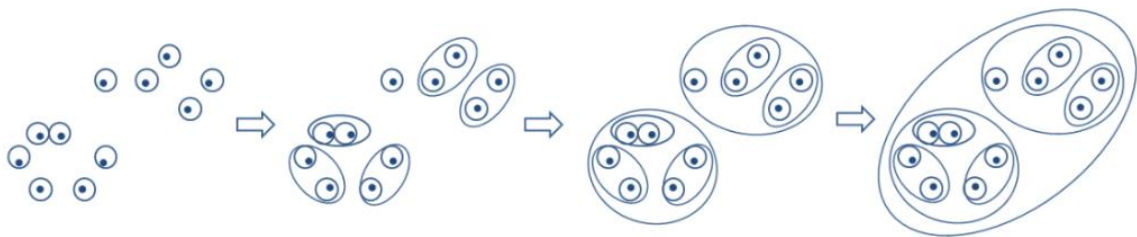
Kümeleme analizi, veri matrisindeki gözlemlerin içinde yer alan grupları birbirleriyle benzer alt kümelere ayırmaya yarayan yöntemler topluluğudur. Başka bir ifadeyle kümeleme analizi, bir veri kümesindeki bilgileri belirli yakınlık derecelerine göre gruplara ayırma işlemidir. Bu ayırma işlemi ile kümeler oluşturulmaktadır.

Kümelemenin amacı, öncelikle birey ya da nesne arasındaki benzerliklere göre az sayıda birbiriyle ilişkili özel kümeleri oluşturmak, daha sonra bu kümelere giren birey ya da nesnelerin profilini ortaya koymaktır. Kümeleme işleminde küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır. Denetimli ve yarı-denetimli öğrenmede araştırmacının veriyi etiketlediği kadar kümeler mevcuttur; ancak denetimsiz öğrenmede farklı sayıda kümeler de seçilebilmektedir. Bu ise araştırmacının amacına bağlıdır. İkinci olarak veri seti içindeki her bir noktanın hangi kümeye ait olduğu tahmin edilmektedir. Tahmin yapabilmek için verilerin benzerliklerini ortaya koymada uzaklık ölçüleri kullanılmaktadır. Birimler arası bu uzaklıkları belirleyebilmek için alternatif ölçü ve yöntemler bulunmaktadır. Öklit, kare Öklit, korelasyon, Manhattan Mahalanobis, Minkowski gibi ölçüler sıklıkla kullanılan uzaklık ölçüleridir. Verinin özelliklerine göre kümeleme analizinin uygulanmasında bu uzaklıklardan uygun olanı seçilir. Son aşamada ise elde edilen uzaklık matrisleri ile amaç doğrultusunda kümeleme yöntemleri

belirlenmektedir. Alanyazında birçok kümeleme algoritmasından söz edilse de iki temel algoritma daha hakimdir (Çokluk, Şekercioğlu ve Büyüköztürk, 2014; Güzeller, 2016). Bunlardan biri araştırmacının önceden küme sayısını bilmediği durumlarda kullanabileceği *hiyerarşik kümeleme analizi*, diğeri ise araştırmacının önceden kümeleri öngörüp küme sayısını belirlediği durumlarda kullanabileceği *hiyerarşik olmayan kümeleme analizidir*. Hiyerarşik kümeleme analizinde sıklıkla kullanılan yöntemler, birleştirici yöntemlerden *tek bağlantı* (en yakın komşuluk), *tam bağlantı* (en uzak komşuluk), *ortalama bağlantı*, *merkezi* ve *Ward yöntemleri*; ayırıcı yöntemlerden ise *bölünmüş ortalamalar* ve *otomatik etkileşim belirleme yöntemleridir*. Hiyerarşik olmayan kümeleme analizinde ise sıklıkla k-ortalama, metoid parçalama, yığılma ve bulanık kümeleme yöntemleri kullanılmaktadır. Bu çalışmada denetimsiz öğrenme algoritması için hile yoluyla testte iş birliği yapan bireyleri ve grupları belirlemede hiyerarşik kümeleme yöntemi kullanılmıştır.

#### 1.4.3.3.1.1. Aşamalı (Hiyerarşik) Kümeleme Yöntemleri

Aşamalı bir şekilde kümeleme yapılan bu yöntemde bir alt aşamadaki küme alt grupları bir sonraki aşamadaki kümeleri oluşturmak için bir araya getirilmektedir (Alpar, 2013). Bu kümeleme yönteminde kümelerin oluşturulmasında iki tür aşamadan söz edilebilir. Bunlardan birincisi olan birleştirici (agglomerative) aşamalı kümeleme yöntemlerinde ilk olarak gözlemlerin her biri bir küme olarak ele alınıp aşamalı bir şekilde birbirine en yakın kümeler birleştirilerek küme sayısı azaltılır. Ayırıcı (divisive) aşamalı yöntemlerde ise birleştirici kümeleme sürecinin tam tersi yönde bir aşama izlenir. Analizin başında tüm gözlemler bir küme olarak ele alınır ve sonrasında farklı ya da benzemeyen gözlemler birbirinden ayrılarak küme sayısının arttığı bir aşama izlenir. Her iki yöntemde de girdi olarak gözlemler arasındaki benzerlik (similarity) ya da benzeşmezlik/farklılık (dissimilarity) matrisi kullanılmaktadır. Bu analizler sonucunda elde edilen kümeleri görsel bir şekilde ortaya koyabilmek için ağaç diyagramları ya da buz saçağı grafiklerinden yararlanılmaktadır.



Şekil 1.4.1. Hiyerarşik Kümeleme Yöntemlerinin Yapılanması

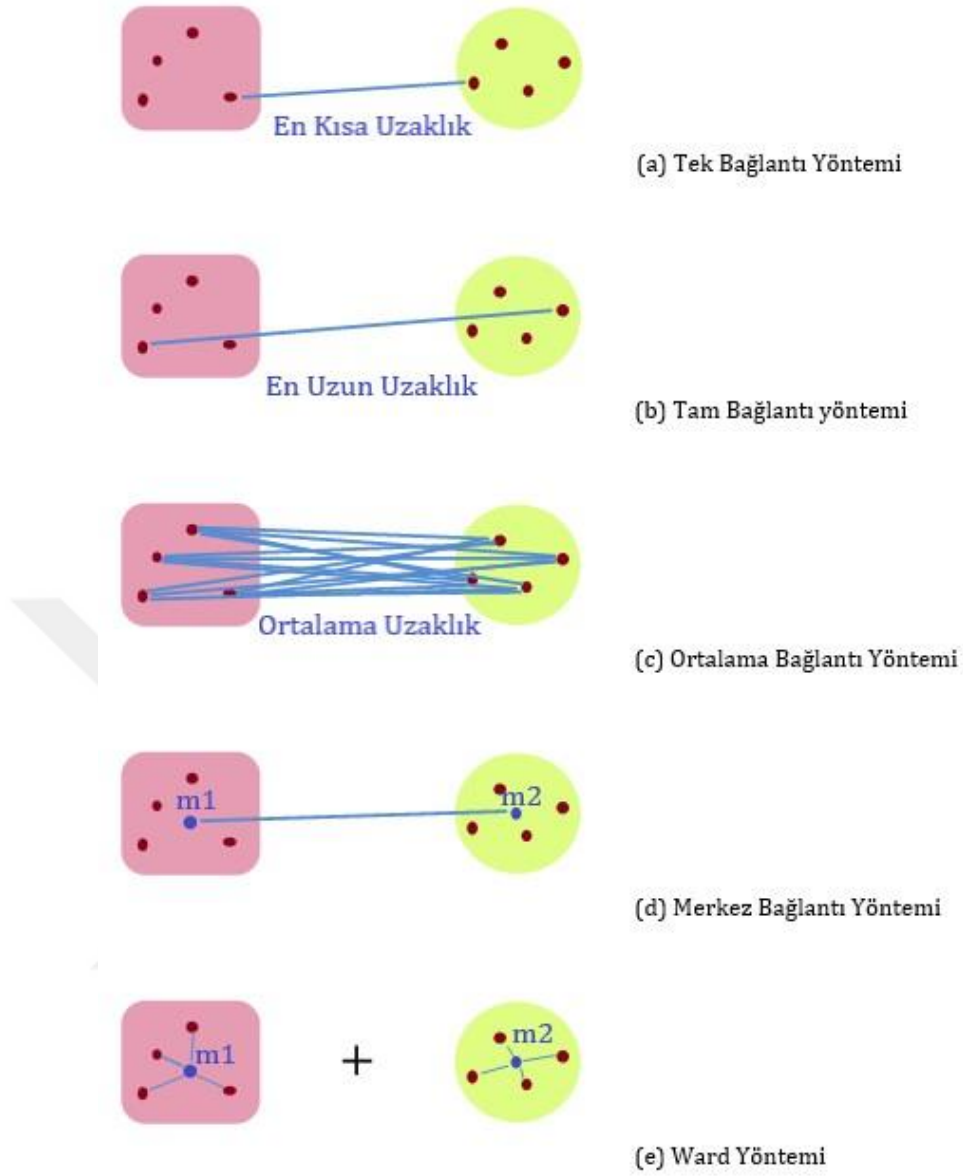
**En Yakın Komşuluk (Tek Bağlantı) Yöntemi:** En basit kümeleme yöntemlerinden biri olan  $K$  en yakın komşuluk yönteminde, eğitim setinde test verilerine en yakın olan  $K$  noktalarını ele almaktadır. İlk olarak uzaklık matrisindeki en yakın iki gözlem alınarak birinci küme oluşturulur ve ardından aşamalı bir şekilde bu süreç devam eder. Yeni gözlemler önceki kümelerden birine dahil edilmekte ya da yeni iki elemanlı başka bir küme oluşturulmaktadır. Süreç sonunda tüm gözlemlerin her biri bir kümeye dahil edilmiş olunur.

**Tam Bağlantı Yöntemi:** Tek bağlantı yönteminden farklı olarak birinci aşamadan sonra yeni kümelerin oluşturulduğu ikinci aşamada en uzak gözlemlerin dikkate alınmasıdır. Bu yönteme aynı zamanda en uzak komşuluk yöntemi adı verilmesinin nedeni de budur. En uzak gözlemlerin dikkate alınmasıyla tek bağlantı yönteminde açığa çıkan zincirleme sorunu ortadan kalkmıştır; ancak aşırı değerlerden etkilenme sorunu ortaya çıkmıştır (Rokach ve Maimon, 2005; Alpar, 2013).

**Ortalama Bağlantı Yöntemi:** Bu yöntemde gözlemler kümelere dahil edilirken gözlemlerin yakınlığı ya da uzaklığı yerine ortalamaları dikkate alınmaktadır. Böylece bu yöntemde uç değerlerden daha az etkilenilmektedir. Aynı zamanda bu yöntem, uzun kümelerin bölünmesi ve küçük küme-içi değişkenliğe sahip kümeleri birleştirme eğilimine sahiptir (Rokach ve Maimon, 2005; Alpar, 2013).

**Merkez Bağlantı Yöntemi:** Aşırı değerlerden en az etkilenen kümeleme yöntemidir (Alpar, 2013, s. 333). Bu yöntemde, kümelerin merkezleri arasındaki uzaklıklar dikkate alınmaktadır. Bunun için bir kümeyi oluşturan gözlemlerin ortalamaları alınmaktadır.

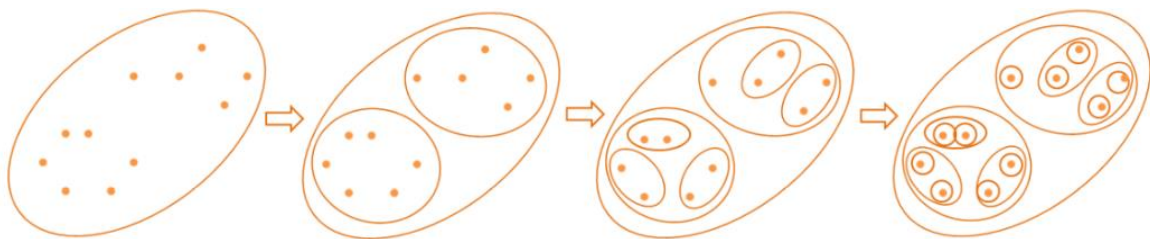
**Ward Yöntemi:** Bu yöntem, ilk olarak Ward Jr. (1963) ile Ward Jr. ve Hook (1963) tarafından ortaya konmuştur (Gan, Ma ve Wu, 2007). Ward yönteminde gözlemleri kümelemede, gözlemler arasındaki en küçük varyans kriteri kullanılmaktadır. Bunun için bir kümenin merkezinde yer alan elemanın, kümenin içinde bulunan diğer elemanlardan ortalama uzaklığı dikkate alınmaktadır. Bu yöntemde her aşamada iki alt küme bir sonraki seviyeyi oluşturmak için birleştirilir. Ward yöntemi küçük kümeleri birleştirme eğilimindedir ve aynı zamanda aşırı değerlerden etkilenmektedir. Yapılan çalışmalar (Hands ve Everitt, 1987; Ferreira ve Hitchcock, 2009) ise hiyerarşik kümeleme yöntemleri arasında Ward yönteminin diğerlerine göre çoğunlukla daha iyi sonuç verdiğini ortaya koymuştur. Araştırmalar sonucunda Ward yöntemi daha çok tercih edilmektedir (Murtag ve Contreras, 2017, s.7).



**Şekil 1.4.2.** Hiyerarşik Kümeleme Yöntemlerinin Şematik Gösterimi

#### 1.4.3.3.2. Aşamalı Olmayan Kümeleme Yöntemleri

Küme sayısının önceden belirlendiği bu yöntemlerde, kendi içinde benzer ancak kendi aralarında farklı yapıda olan kümeleri ortaya koymak amaçlanmaktadır. Bu yöntemler arasında sıklıkla kullanılan k-ortalama kümeleme (k-means clustering) yöntemidir.



**Şekil 1.4.3.** Aşamalı Olmayan Kümeleme Yöntemlerinin Yapılanması

### 1.5. Yöntemlerin Performansını Belirlemede Kullanılabilecek Ölçütler

Hile yapan bireyleri belirlemede kullanılan indeks ve istatistiklerin performansını ortaya koymada alanyazında sıklıkla kullanılan ölçütler, I. Tip hata oranları ve güçleridir. I. Tip hata oranı, gerçekte hile yapmayan bireyleri, yöntemin analiz sonucunda hile yapmış olarak kabul etmesidir. İndeksin gücü ise hile yapan bireyleri doğru olarak belirleyebilmesidir. Bu bilgiler yöntemin performansı için oldukça önemlidir; ancak hile yaptığı halde yöntemin gözden kaçırarak “iş birliği yapmadı” olarak belirlediği II. Tip hata değeri ve bu ölçülerin birlikte ele alınarak ortaya konduğu değerler de oldukça öneme sahiptir. Özgüllük (specificity), duyarlılık (sensitivity), pozitif tahmin değeri, negatif tahmin değeri, F ölçütü ve doğruluk (accuracy) değerleri gibi bu değerler özellikle sağlık alanında tanı koymada sıklıkla kullanılmaktadır. Bu değerlerin hesaplanması ise aşağıda yer alan Tablo 1.5.1’deki gibi hazırlanmış kontenjans tablosu yardımıyla kolaylıkla yapılabilmektedir.

**Tablo 1.5.1.**  
Kontenjans Tablosu

		Gerçek	
		İş Birliği Yapan	İş Birliği Yapmayan
Karar	İş Birliği Yaptı	A	B
	İş Birliği Yapmadı	C	D

Tablo 1.5.1 incelendiğinde iş birliği yapıp analiz sonucuna göre iş birliği yaptığı ortaya konan bireylerin sayısı A olup bu değer doğru pozitif (true positive) olarak adlandırılmaktadır. B değeri ise gerçekte iş birliği yapmayan ancak analiz sonucunda iş birliği yaptığı öngörülen yanlış pozitif (false positive) değerdir. Gerçekte iş birliği yapmış ancak analiz sonucu iş birliği yapmadığı ortaya çıkan bireylerin sayısı olan C, yanlış negatif (false negative) değerdir. D ise gerçekte iş birliği yapmayan bireyleri doğru olarak belirlediği birey sayısıdır ve bu değer doğru negatif (true false) olarak adlandırılmaktadır. Alanyazında kopya belirleme indeksleri ile çalışılan birçok araştırmada indekslerin performansının ölçütü olarak güç ve I. Tip hata oranları elde edilmektedir. Bu oranlar ise yine sırasıyla A ve B değerleri elde edilerek hesaplanmaktadır.

Duyarlılık, yöntemin gerçek iş birlikçi bireyleri ayırma gücüdür.

$$\text{Duyarlılık} = A / (A+C)$$

Özgüllük ise yöntemin gerçekteki iş birliği yapmayan bireyleri ortaya koyma gücüdür.

$$\text{Özgüllük} = D / (B + D)$$

Pozitif tahmin değeri (PTD), analiz sonucunda belirlenen iş birlikçi bireyler içerisindeki doğru tahminlerin oranıdır.

$$\text{PTD} = A / (A+B)$$

Negatif tahmin değeri (NTD) ise analiz sonucunda belirlenen iş birlikçi olmayan bireyler içerisindeki doğru tahminlerin oranıdır.

$$\text{NTD} = D / (C+D)$$

F ölçütü ise duyarlılık ve pozitif tahmin değerinin harmonik ortalaması olup bu iki değer birlikte ele alınıp yorumlanmasını sağlamaktadır.

$$F \text{ ölçütü} = \frac{2(\text{Duyarlılık})(\text{PTD})}{(\text{Duyarlılık}) + (\text{PTD})}$$

Doğruluk ise duyarlılık ve özgüllük değerleri birlikte ele alınarak tek bir ölçü ile analiz sonucunun doğru olup olmadığını ortaya koyan bir değerdir. Bu değer, gerçekte iş birliği yapan ve yapmayan bireylerin yöntemin ne oranda doğru olarak belirlediği bilgisini vermektedir.

$$\text{Doğruluk} = (A+D) / (A+B+C+D)$$

Bir yöntemin duyarlılığının, özgüllüğünün, pozitif ve negatif tahmin değerinin ve doğruluğunun yüksek olması beklenir. Bu değerler 1'e yaklaştıkça yöntemin performansı da artmaktadır.

## 1.6. Araştırmanın Amacı

Bu çalışmada amaç, çoktan seçmeli maddelerden oluşan testlerde hile yapan bireylerin ve grupların iş birliğini belirlemede M4 benzerlik istatistiği kullanılarak, makine öğrenme yöntemlerinin performansını belirlemektir. Bu amaç doğrultusunda uzaklık ölçüsü olarak M4 istatistiğinden elde edilen uzaklık matrisi yardımıyla denetimsiz makine öğrenmesi yöntemlerinden Ward hiyerarşik kümeleme yöntemi ile yarı-denetimli makine öğrenmesi türü

olan kendi kendine öğrenme ve SETRED yönteminin çeşitli koşullar altında performanslarını ortaya koymak hedeflenmiştir.

### **1.7. Problem Cümlesi**

Çalışmada problem cümlesi “Çeşitli koşullarda çoktan seçmeli maddelerde hile yapan bireylerin iş birliğini belirlemede M4 istatistiği ile elde edilen benzerlik matrisini kullanan makine öğrenmesi yöntemlerinin performansları nasıldır?” şeklinde belirlenmiştir. Bu temel probleme dayalı olarak çalışmanın alt problem cümleleri şöyle sıralanabilir?

#### **1.7.1. Alt Problemler**

Çalışmada şu sorulara cevap aranmaktadır:

1. Hile yapan bireylerin iş birliğini belirlemede çeşitli koşulların Ward hiyerarşik kümeleme yönteminin performansına olan etkisi nasıldır?

1.1. Hile yapan grup sayısının yöntemin performansına olan etkisi nasıldır?

1.2. Hile yapan grupta yer alan bireylerin sayısının yöntemin performansına olan etkisi nasıldır?

1.3. Kopya çekilmiş madde sayısının yöntemin performansına olan etkisi nasıldır?

2. Hile yapan bireylerin iş birliğini belirlemede çeşitli koşulların kendi kendine öğrenme ve SETRED yöntemlerinin performansına olan etkisi nasıldır?

2.1. Hile yapan grup sayısının yöntemlerin performanslarına olan etkisi nasıldır?

2.2. Hile yapan grupta yer alan bireylerin sayısının yöntemlerin performanslarına olan etkisi nasıldır?

2.3. Kopya çekilmiş madde sayısının yöntemlerin performanslarına olan etkisi nasıldır?

2.4. Kendi kendine öğrenme ve SETRED yöntemlerinin her bir koşula göre performansları arasındaki farklar nasıldır?

### **1.8. Araştırmanın Kapsam ve Sınırlılıkları**

Bu çalışma, araştırmaya dahil edilen benzerlik istatistiği ve yöntemler, seçilen değişkenler ve düzeyleri ile sınırlıdır.

## 2. KAYNAK ARAŞTIRMALARI

Alanyazında testlerde bireyler arasında cevapların paylaşılarak yapıldığı iş birliğini belirlemeye yönelik çalışmalar sınırlı sayıdadır. Bu çalışmalar kısaca şöyle özetlenebilir:

Jacob ve Levitt'in (2003) ortaya koydukları JL yöntemi öğretmenlerin sınıf içinde yaptıkları hilelere odaklanmıştır. Bu yöntem, bir sınıf içinde normal olmayan cevapların tümüne ve beklenmeyen puanlara ait iki bilginin katışık (joint) dağılımına dayanmaktadır. Çalışmalarında yaptıkları simülasyondan birinde bireyleri seçkisiz olarak sınıflara atayarak öğretmen belli öğrencilere blok halinde aynı maddelerin doğru yanıtlarını vermekte, diğerinde ise sıftaki seçkisiz olarak belirlenen bazı öğrencilerin yanlış cevaplarını doğru cevaplarla değiştirmektedir. Ancak yöntem bu hileleri belirlemede zayıf kalmıştır. Hile yapılan üç maddenin sınıfın %25'inde değiştirildiğinde, sadece %4'ünü; altı maddenin sınıfın yarısında değiştirildiğinde ise bu hile durumlarının ise %60'ından daha azını belirleyebilmiştir.

Zhang, Searcy ve Horn (2011), çalışmalarında iş birliği yapan grup ve içindeki bireyleri belirlemede öncelikle birey uyum indekslerinden *Iz* (Drasgow, Levine ve Williams, 1985) ve *U* (Wright ve Stone, 1979) indeksleri ile faktör analizini kullanmıştır. Araştırmacılar, birey-uyum indekslerini anormal cevaplar veren bireyleri belirlemede kullandıktan sonra bu bireylerden aynı madde setlerinde iş birliği yapanları gruplamada *Q* tipi faktör analizinden faydalanmaktadır. Çalışmalarında hile yapan bireylerin yetenek dağılımı, hile yapan grup sayısını, bu grupların büyüklüğünü ve kirlenmiş madde sayısını değişimlemiştir. İlk olarak birey-uyum analizinde 2000 üzerindeki bireyden %4.2'sini hile yapmış olarak belirlemiş, faktör analizinde ise bu bireylerden sadece %18 ile %20'sini gruplayabilmiştir.

Belov (2013) çalışmasında test öncesinde paylaşılmış maddeleri alan bireyleri belirlemede dağılımlar arası farkları belirleyen Kullback-Leibler divergence istatistiğini kullanmıştır. Algoritma 1 olarak adlandırdığı iki istatistiksel test sürecini ortaya koyarak hileyi belirlemeye çalışmaktadır. Birinci test hile yapan grubu, diğeri ise grup içindeki bireyleri belirlemektedir. Çalışmasında Algoritma 1 süreci ile birey-uyum istatistikleri olan *G* (Guttman, 1944) ve *G\** istatistiğini (van der Flier, 1977) 0.0025, 0.005, 0.0075, 0.01, 0.025 ve 0.05 alfa düzeylerinde karşılaştırmıştır. Grup büyüklüğünü (5, 10, 15, 20, 30) ve grup içinde yer alan bireylerin sayısını (5, 10, 15, 20, 30) değişimlemiştir. Araştırması sonucunda bu iki istatistiğe göre Algoritma 1'in eşzamanlı olarak I. Tip ve II. Tip hatayı azalttığı görülmüştür. Ancak grup büyüdükçe gücünde azalma olduğu ortaya çıkmıştır.

Wollack ve Maynes (2017) ise araştırmalarında benzerlik istatistiklerinden *M4* ile en yakın komşuluk (tek bağlantı) kümeleme analizini birlikte kullanarak hile yapan bireylerin iş birliğini belirlemeye çalışmışlardır. Çalışmalarında 50 madde kullanılmış ve üç bağımsız değişken

olan grup büyüklüğü (5, 10, 20), kirlenmiş maddelerin (contamination items) sayısı (25, 35, 50) ve kirlenme etkisi (contamination effect) (2.0, 2.5, 3.0) çaprazlanarak 27 koşul üretilmiştir. Her bir koşulda 100 bağımsız hile yapan grup için cevap üretilmiştir. Böylece iş birlikçi birey sayısı 500 ile 2000 arasında değişmiştir. Aynı zamanda örnekleme 5,000 iş birliği yapmayan birey bulunmaktadır. I. Tip hata çalışması sonucunda hata oranının tüm koşullarda birbirine yakın olduğu ve düşük olduğu görülmüştür. Kümelemenin I. Tip hata oranına bakıldığında kümelerin %46.3'ünün yanlışlıkla belirlenen hiçbir birey içermediği ve genel olarak I. Tip hatayı iyi kontrol ettiği görülmüştür. Güç çalışmasında ise en çok etkinin kirlenmiş madde sayısı değişkenine ait olduğu görülmüştür. Bu maddeler 25 olduğunda etki büyük olsa da gücün düşük olduğu; tüm maddeler kirlenmiş olduğunda ise etki küçük olsa da gücün oldukça yüksek olduğu ortaya konmuştur. Diğer değişkenler sabit tutulduğunda, kirlenme etkisi arttıkça güç de artmaktadır. Grup büyüklüğünün ise tek başına güç üzerinde etkisinin oldukça düşük olduğu ortaya çıkmıştır.

## **2.1. Araştırmanın Önemi**

Alanyazın araştırması sonucunda test güvenliğini tehdit eden bireyler arası iş birliğini belirlemeye yönelik çalışmaların (Jacob ve Levitt, 2003; Zhang, Searcy ve Horn, 2011; Belov, 2013; Wollack ve Maynes, 2017) oldukça az olduğu görülmektedir. Bireyleri gruplandırma Zhang, Searcy ve Horn (2011) çalışmalarında faktör analizini; Wollack ve Maynes (2017) ise tek grubu belirlemek için en yakın komşuluk kümeleme analizini kullanmışlardır; ancak Ward yönteminin ve yarı-denetimli makine öğrenmesinin kullanıldığı bir çalışmaya ve bununla birlikte bu yöntemlerle farklı iki ve daha fazla iş birlikçi grubu belirlemeye yönelik bir çalışmaya rastlanmamıştır.

Aynı zamanda Türkiye'de yapılan araştırmalar arasında test güvenliğini tehdit eden bireyler arası iş birliğini belirlemeye yönelik herhangi bir çalışmanın yapılmadığı görülmüştür. Bu nedenle bu çalışmanın Türkiye'de yapılan ölçme ve değerlendirme çalışmalarına katkıda bulunacağı düşünülmektedir.

### 3. YÖNTEM

#### 3.1. Araştırmanın Türü

Bu araştırmada simülasyon verisi kullanılarak çeşitli koşullar altında M4 istatistiği ile makine öğrenmesi yöntemlerinin hile yapan bireylerin iş birliğini belirlemedeki performansları hakkında bilgi elde edilmeye çalışılmıştır. Bu çerçevede yapılmış olan bu çalışma, benzetimli-deneysel bir çalışmadır.

#### 3.2. Veri Üretimi ve Analizi

Çalışmada GEN3PL\_RawDATA\_V2 (Luecht, 2011) programı kullanılarak, yeteneklerinin standart normal dağıldığı  $N(0,1)$  10,000 bireyin beş seçenekli 80 maddeye verdikleri tepkilerinden oluşan test verisi üretilmiştir. Bu ham veriye ait  $a$  madde parametresi 0.8 – 1.5 ve  $b$  madde parametresi -3.0 – 3.0 aralığında değişmektedir. Analizler için kullanılacak olan madde ve yetenek parametrelerinin kestirimi ise R 3.6.1 programı kullanılarak araştırmacı tarafından yazılan kod ve mirt (Chalmers, 2019) paketi yardımı ile Madde Tepki Kuramı (MTK) modellerinden nominal tepki modeline (Nominal Response Model, NRM) dayalı olarak gerçekleştirilmiştir.

**Simülasyon Çalışması:** Çalışmada R 3.6.1 bilgisayar programında yazılan simülasyon kodları ile hileli veriler oluşturularak, analizler yapılmıştır.

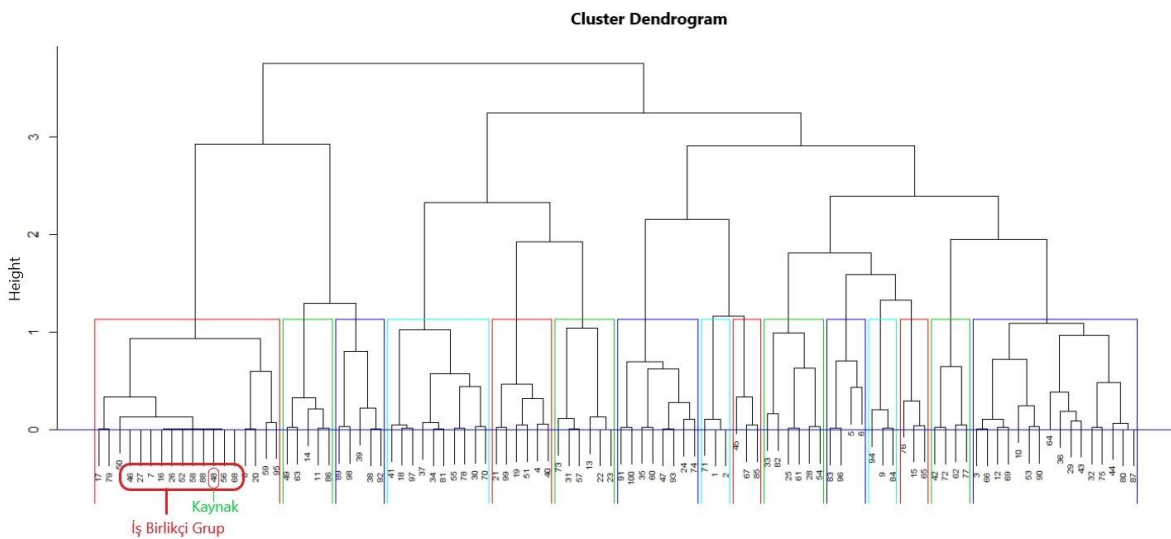
Öncelikle tüm yöntemlerin performanslarına ait analizlerde kullanılacak olan 10,000 birey arasından 500 iş birliği yapmadığı bilinen bireyler seçkisiz olarak seçilmiştir. Analizlerin yapılacağı örneklem büyüklüğünün 500 olarak seçilmesi M4 istatistiğinin hesaplanmasında kullanılan algoritmalarla benzerlik matrislerinin oluşturulmasının oldukça uzun sürmesidir. Bu örneklemin maddelere verdikleri cevaplara ait benzerlik matrisi M4 benzerlik istatistiği kullanılarak elde edilmiştir. Sonrasında yöntemlerin performansının belirlenmesine yönelik çalışmada ilgilenilen değişkenler ve düzeyleri için analizler yapılmıştır. Bunun için sırasıyla şu adımlar izlenmiştir:

#### **Çalışma 1: Ward hiyerarşik kümeleme yönteminin performansının belirlenmesi;**

1. Bir ve iki farklı iş birlikçi gruplarda yer alacak iş birlikçi sayısı (5, 10, 20) koşulunun her bir düzeyi için yeteneği 2.5 – 3.0 arasında değişen bir kaynak ve diğerleri iş birlikçi olması istenen bireyler 9,500 birey arasından seçkisiz olarak seçilmiştir.

2. Kopya oranı koşulunun her düzeyi (%10, %15, %20, %40) (araştırmacının yazdığı algoritmalarla çeşitli kopya oranı düzeylerinde yöntemin performansının sınanması sonucunda değişimin bu düzeylerde en fazla olduğu görülmüş ve bu düzeyler araştırmaya dahil edilmiştir) için seçilen kaynağın tepkilerine diğer bireylerin maddelere verdikleri tepkilerin eşleşmesi sağlanmıştır. İki iş birlikçi grupta ise grupların birbirinden farklı iş birliği yapmış olması için bu maddelerin en az %75'inin gruplar arasında farklı maddelerden eşleşmiş olmasına dikkat edilmiştir.
3. Tepkilerin eşleşmesi sağlandıktan sonra oluşturulan her bir hileli durum için iş birliği yaptığı bilinen bireyler ile tüm bireylerin maddelere verdikleri tepkileri arasındaki benzerlikleri ortaya koyan M4 istatistiği hesaplanarak her bir birey çiftleri için benzerlik matrisi oluşturulmuştur.
4. Benzerlik matrisi uzaklık matrisi olarak kullanılarak Ward kümeleme yöntemi ile kümeleme analizi yapılmıştır. Elde edilen ağaç diyagramları (dendrogram) uzaklık ölçülerine göre incelenmiştir. Ağaç diyagramlarında 0 (sıfır) yüksekliği iş birliği yapmış bireyler hakkında bilgi vermektedir. İncelemeler ise bu doğrultuda yapılmıştır.
5. 100 replikasyon ile bu adımlar tekrar edilerek Ward kümeleme yönteminin performans ölçütleri hesaplanmıştır.

80 maddenin %40'unda hile yapan 10 birey ve 1 kaynak ile toplamda 100 bireyin yer aldığı örnek bir veri için hiyerarşik kümeleme analizi sonucuna ait ağaç diyagramı Grafik 3.1'de yer almaktadır.



**Grafik 3.1.** Ward Kümeleme Analizi İçin Ağaç Diyagramı Örneği

## **Çalışma 2: Kendi kendine öğrenme ve SETRED yöntemlerinin performansının belirlenmesi;**

1. Çalışma 1'in ilk 3 adımında olduğu gibi her bir koşul ve düzeyleri için birey çiftleri için M4 istatistiği hesaplanarak benzerlik matrisleri oluşturulmuştur.
2. Yarı-denetimli öğrenme için bu matrise ait bilginin %20'si eğitim verisi olarak seçkisiz olarak seçilmiştir. Seçilen bu bireylerden iş birlikçi olanlara "1", olmayanlara "2" olarak sınıf atanmıştır. İki farklı iş birlikçi grubun yer aldığı veride ise birinci grupta yer alan iş birlikçi bireylere "1", ikinci grupta yer alan bireylere "2", iş birlikçi olmayan bireylere ise "3" olarak sınıf atanmıştır.
3. Eğitim verisi kullanılarak, kendi kendine öğrenme ve SETRED yöntemleri ile yapılan test verisine ait sınıf tahminlerine göre yöntemlerin sınıflandırma performansları belirlenmiştir.
4. 2. ve 3. adımlar 100 replikasyonla yinelenmiştir.

#### 4. BULGULAR

Yöntemlerin olası iş birliğini belirlemedeki performanslarına ait elde edilen bulguların ortaya konduğu bu bölümde, Ward kümeleme yönteminin ve kendi kendine öğrenme ile SETRED yönteminin performans ölçütleri ayrı ayrı tablolaştırılarak yer almıştır. Aynı zamanda anlaşılması ve yorumlanmasının kolaylığı açısından bu ölçütler iş birliği yapan bireylerin grup sayısı temel alınarak R programı yardımıyla çok boyutlu olarak grafiklendirilmiştir.

##### 4.1. Ward Yönteminin Kümeleme Performansı

Ward yönteminin 500 iş birliği yapmamış birey içerisine simülasyon yardımıyla değişen kopya oranı ve değişen sayıda iş birliği yapan bireylerin yerleştirildiği toplam örneklem içinden iş birliği yapan ve yapmayan bireyleri kümeleme performansına ait bulgular aşağıdaki Tablo 4.1.1 ve Tablo 4.1.2’de ve bunların çok boyutlu görselleri Grafik 4.1.1 ve Grafik 4.1.2’de yer almaktadır.

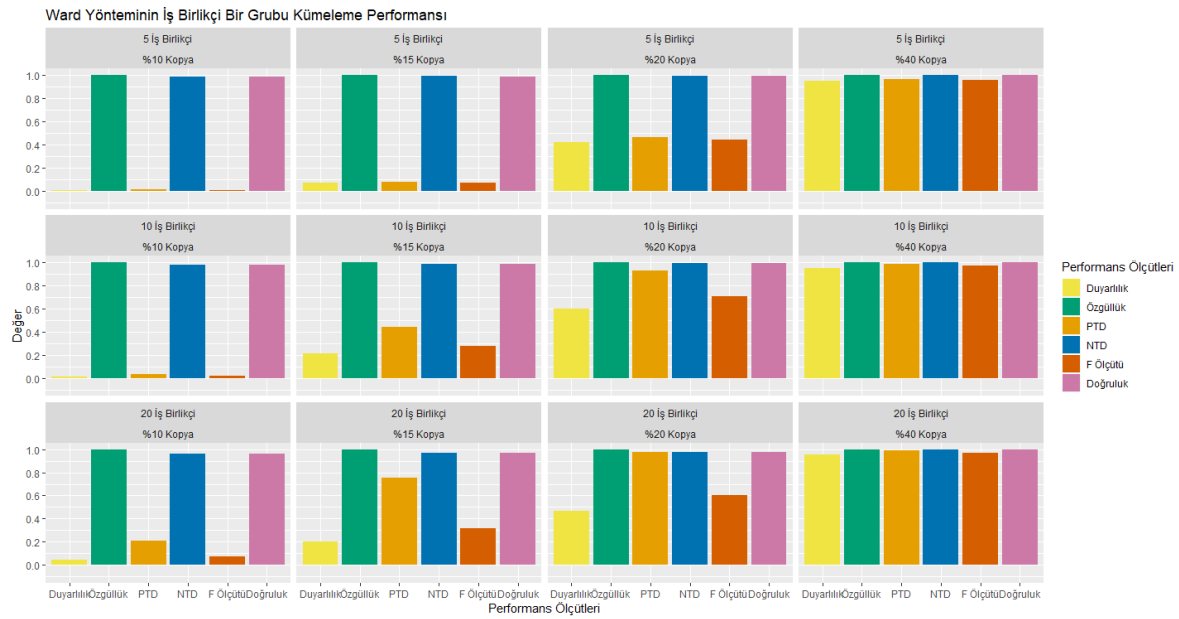
**Tablo 4.1.1.**

Ward Yönteminin İş Birliği Yapan Bir Grup Bireyi Kümeleme Performansına Ait Değerler

Grup Sayısı	İş Birlikçi Sayısı	Kopya Oranı	Duyarlılık	Özgüllük	PTD	NTD	F Ölçütü	Doğruluk
1	5	%10	0.0080	0.9995	0.0100	0.9902	0.0089	0.9897
		%15	0.0700	0.9995	0.0767	0.9908	0.0727	0.9903
		%20	0.4240	0.9997	0.4635	0.9943	0.4416	0.9940
		%40	0.9520	0.9999	0.9680	0.9995	0.9581	0.9994
	10	%10	0.0140	0.9990	0.0323	0.9806	0.0195	0.9797
		%15	0.2100	0.9994	0.4411	0.9845	0.2794	0.9839
		%20	0.5950	0.9997	0.9255	0.9920	0.7068	0.9918
		%40	0.9520	0.9997	0.9879	0.9990	0.9682	0.9988
	20	%10	0.0480	0.9982	0.2082	0.9633	0.0771	0.9617
		%15	0.2015	0.9986	0.7567	0.9690	0.3138	0.9680
		%20	0.4690	0.9994	0.9741	0.9793	0.6031	0.9790
		%40	0.9530	0.9996	0.9912	0.9981	0.9710	0.9978

Tablo 4.1.1 incelendiğinde yöntemin özgüllük, negatif tahmin değeri (NTD) ve doğruluğunun tüm koşullarda 1’e yakın değerler aldığını görmekteyiz. Bu değerlerden gerçekte iş birliği yapmayan bireyleri diğerlerinden ayırma gücü olan özgüllük ve tüm iş birliği yapmayanlar arasından doğru tahminlerin oranı olan NTD değerlerinin 1’e yakın olması yöntemin iş birliği yapmayan bireyleri doğru olarak kümelediğini göstermektedir. Yöntemin gerçekte iş birliği yapan ve yapmayan bireyleri ayırma gücünü birlikte ele alarak ortaya koyduğu performans ölçütü olan doğruluğun da yüksek olduğu görülmektedir. Ancak yöntemin gerçekteki

iş birlikçi bireyleri belirleme performansını ortaya koymak için incelenen duyarlılık, PTD ve F ölçütüne ait değerler koşullara göre değişkenlik göstermektedir.



**Grafik 4.1.1.** Ward Yönteminin İş Birlikçi Yapan Bir Grup Bireyi Kümeleme Performansı

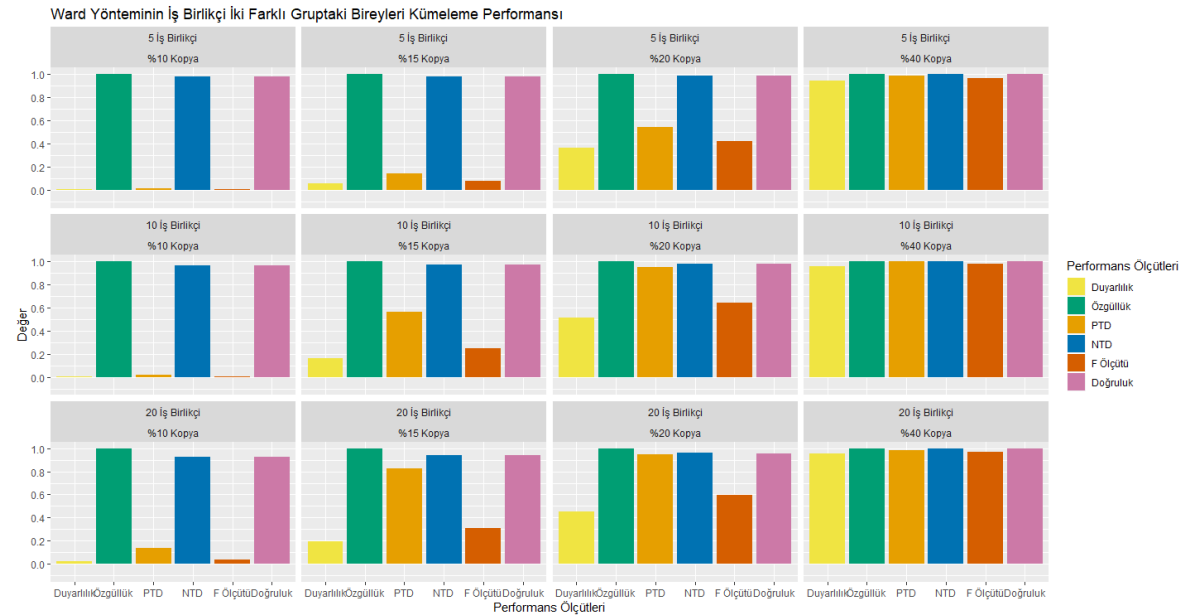
Grafik 4.1.1'de de görüldüğü gibi özellikle %20'nin altındaki kopya oranında Ward yöntemine ait duyarlılık, PTD ve F ölçütlerinin 0.5'in altında değerler aldığı görülmektedir. Özellikle %10 kopya oranında ve iş birlikçi sayısının 5 ve 10 olduğu durumlarda bu ölçütler 0'a yakın değerler almaktadır. Bu koşullarda yöntemin iş birliği yapan bireyleri ayırma gücünün çok düşük olduğu ortaya çıkmıştır. Bu koşullar içinde sadece maddelerin %15'inde hile yapan 20 iş birlikçi arasından iş birliği yapanların tahmininin oranını ortaya koyan PTD değerinin 0.7567 olduğu görülmektedir. Aynı zamanda iş birlikçi sayısının ve kopya oranının artmasıyla bu değerlerin de yükseldiği görülmektedir. %20 kopya oranında 10 iş birliği yapan bireyi belirlemenin, 20 iş birliği yapan bireyi belirlemeye göre duyarlılık, PTD ve F ölçütü değerlerinin 0.1'e yakın miktarda daha fazla olduğu ortaya çıkmıştır. Bununla birlikte yöntemin genellikle PTD değeri duyarlılığından daha yüksektir. Bu da yöntemin iş birliği yapmış olarak belirlediği bireyler içerisindeki doğru tahmininin, iş birliği yapan bireyleri ayırt etme gücünden daha yüksek olduğunu göstermektedir. Kopya oranının %40 olduğu durumlarda ise iş birlikçi sayısının her koşulunda tüm performans değerlerinin 1'e yaklaştığı görülmektedir.

**Tablo 4.1.2.**

Ward Yönteminin İş Birlikçi Yapan İki Farklı Grubu Kümeleme Performansına Ait Değerler

Grup Sayısı	İş Birlikçi Sayısı	Kopya Oranı	Duyarlılık	Özgüllük	PTD	NTD	F Ölçütü	Doğruluk
2	5	%10	0.0040	1	0.0100	0.9805	0.0057	0.9805
		%15	0.0570	1	0.1400	0.9815	0.0810	0.9815
		%20	0.3660	0.9997	0.5424	0.9875	0.4231	0.9873
		%40	0.9460	0.9998	0.9872	0.9989	0.9637	0.9987
	10	%10	0.0040	1	0.0200	0.9617	0.0067	0.9617
		%15	0.1645	0.9995	0.5655	0.9677	0.2482	0.9674
		%20	0.5115	0.9992	0.9491	0.9809	0.6426	0.9804
		%40	0.9535	1	1	0.9981	0.9757	0.9982
	20	%10	0.0220	0.9995	0.1414	0.9274	0.0376	0.9271
		%15	0.1942	0.9979	0.8230	0.9394	0.3067	0.9384
		%20	0.4515	0.9976	0.9465	0.9580	0.5966	0.9572
		%40	0.9515	0.9987	0.9856	0.9961	0.9675	0.9952

Tablo 4.1.2 ve Grafik 4.1.2 incelendiğinde de yöntemin iki farklı iş birlikçi grubu belirlemedeki performansı ile bir grubu belirlemedeki performansının paralellik gösterdiği görülmektedir. Aynı şekilde yöntemin iş birlikçi olmayan bireyleri belirlemede kullanılan özgüllük ve NTD ile doğruluk ölçütlerinin tüm koşullarda 1 ve 1'e yakın değerler aldığı görülmektedir.



**Grafik 4.1.2.** Ward Yönteminin İş Birlikçi Yapan İki Farklı Gruptaki Bireyleri Kümeleme Performansı

İki gruba ait iş birliği yapan bireyleri belirlemede ise kopya oranı ve birey sayısı arttıkça duyarlılık, PTD ve F ölçüsüne ait değerler de artmaktadır. Ancak %10 kopya oranında ve iş birliği sayısının 5 ve 10 olduğu durumlarda yine bu değerlerin oldukça düşük olduğu ortaya çıkmıştır. Yöntemin iş birliği yapan bireyleri ayırmada yetersiz olduğu durumlarda duyarlılığının yine PTD'den düşük olduğu görülmektedir. Analizler sonucunda maddelerin %40'ının hile ile paylaşıldığı her koşulda Ward yönteminin tüm kümeleme performans ölçütlerinin 1'e yakın değerler aldığı ortaya konmuştur.

#### 4.2. Kendi Kendine Öğrenme ve SETRED Yöntemlerinin Sınıflandırma Performansları

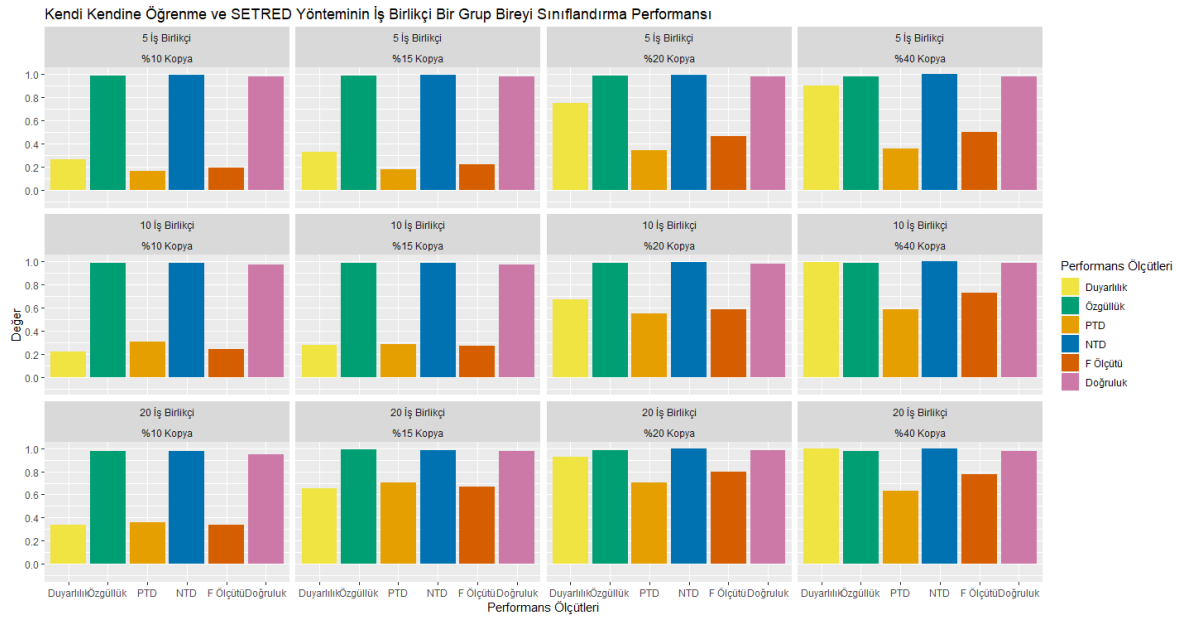
Yarı-denetimli öğrenme türü olan kendi kendine öğrenme ve SETRED yönteminin çeşitli koşullar altında iş birliği yapan ve yapmayan bireyleri sınıflandırma performansına ait bulgular aşağıdaki Tablo 4.2.1 ve Tablo 4.2.2'de ve bunların çok boyutlu görselleri Grafik 4.2.1 ve Grafik 4.2.2'de yer almaktadır.

**Tablo 4.2.1.**

Kendi Kendine Öğrenme ve SETRED Yöntemlerinin İş Birliği Yapan Bir Grup Bireyi Sınıflandırma Performansına Ait Değerler

Yöntem	Grup Sayısı	İş Birlikçi Sayısı	Kopya Oranı	Duyarlılık	Özgüllük	PTD	NTD	F Ölçütü	Doğruluk	
Kendi Kendine Öğrenme	1	5	%10	0.2650	0.9884	0.1648	0.9926	0.1923	0.9813	
			%15	0.3250	0.9866	0.1752	0.9932	0.2187	0.9801	
			%20	0.7525	0.9848	0.3420	0.9975	0.4637	0.9825	
			%40	0.8975	0.9821	0.3536	0.9990	0.4981	0.9812	
	10	%10	0.2225	0.9879	0.3070	0.9845	0.2382	0.9729		
		%15	0.2750	0.9860	0.2824	0.9855	0.2684	0.9721		
		%20	0.6675	0.9871	0.5499	0.9933	0.5845	0.9808		
		%40	0.9950	0.9845	0.5826	0.9999	0.7286	0.9847		
	20	%10	0.3369	0.9747	0.3625	0.9736	0.3367	0.9502		
		%15	0.6525	0.9879	0.7024	0.9862	0.6654	0.9750		
		%20	0.9281	0.9836	0.7010	0.9971	0.7954	0.9815		
		%40	1	0.9761	0.6345	1	0.7739	0.9770		
	SETRED	1	5	%10	0.2650	0.9884	0.1648	0.9926	0.1923	0.9813
				%15	0.3250	0.9866	0.1752	0.9932	0.2187	0.9801
				%20	0.7525	0.9848	0.3420	0.9975	0.4637	0.9825
				%40	0.8975	0.9821	0.3536	0.9990	0.4981	0.9812
10		%10	0.2225	0.9879	0.3070	0.9845	0.2382	0.9729		
		%15	0.2750	0.9860	0.2824	0.9855	0.2684	0.9721		
		%20	0.6675	0.9871	0.5499	0.9933	0.5845	0.9808		
		%40	0.9950	0.9845	0.5826	0.9999	0.7286	0.9847		
20		%10	0.3369	0.9747	0.3625	0.9736	0.3367	0.9502		
		%15	0.6525	0.9879	0.7024	0.9862	0.6654	0.9750		
		%20	0.9281	0.9836	0.7010	0.9971	0.7954	0.9815		
		%40	1	0.9761	0.6345	1	0.7739	0.9770		

Analiz sonuçlarına göre hazırlanmış olan Tablo 4.2.1 incelendiğinde kendi kendine öğrenme yöntemi ve bu yöntemin geliştirilmiş türü olan SETRED yönteminin sınıflandırma performansına ait değerlerin birbiri ile aynı olduğu görülmektedir. Bu durum her iki yöntemin iş birliği yapan ve yapmayan bireyleri ayırt etme gücünün birbiri ile aynı olduğunu göstermektedir. Her iki yöntemin performans ölçütlerine ait değerlerin aynı olması nedeniyle yöntemler için aynı grafikler elde edileceğinden bu bölümde ayrı ayrı grafiklere yer vermek yerine, her iki yöntem için de geçerli ortak grafiklere yer verilmiştir.



**Grafik 4.2.1.** Kendi Kendine Öğrenme ve SETRED Yönteminin İş Birlikçi Bir Grup Bireyi Sınıflandırma Performansı

Yöntemlerin bir grup iş birlikçi bireyi sınıflandırma performansına ait tablo ve grafikler incelendiğinde tüm koşullarda özgüllük, NTD ve doğruluk ölçütlerinin 1 ve 1'e yakın değerler aldığı görülmektedir. Bu da yöntemlerin iş birliği yapmayan bireyleri tüm bireyler arasından ve iş birliği yapmayan bireyler içinden doğru olarak sınıflandırdığını göstermektedir. Diğer ölçütlere bakıldığında genel olarak iş birlikçi birey ve kopya oranı arttıkça iş birlikçi bireyleri sınıflandırma performansı da artmaktadır. Ancak %20'nin altındaki kopya durumlarında bu değerlerin genel olarak 0.5'ten küçük olduğu görülmektedir. 20 işbirlikçinin maddelerin %15'ini hile ile paylaştığı durum ile iş birlikçi sayısının 10 ve üzeri olduğu diğer koşullarda bu değerlerin 0.5'in üzerine çıktığı görülmektedir. Kopyanın %10 ve %15 olduğu durumlarda çoğunlukla PTD ve duyarlılık değerleri birbirine yakındır. Ancak %20 ve üzeri kopyada duyarlılığın PTD'den daha fazla arttığı görülmektedir. Özellikle iş birlikçi bireyin 5 olduğu %40 kopya durumunda PTD'nin diğer ölçütlere göre oldukça düşük olduğu görülmektedir. Bu durum yöntemlerin iş birliği yapan

bireyleri tahmin etmedeki doğruluğunun diğer performans ölçütlerine göre daha düşük olduğunu göstermektedir. Grup içindeki iş birlikçi sayısının 10 ve üzeri olduğunda ve kopya çekilen maddelerin oranının %20 ve üzeri olduğunda tüm ölçütler 0.5'ten büyük değerler almaktadır. %40 kopya oranında ise yöntemlerin duyarlılığı da 1ve 1'e yakın değerler almaktadır. Bu koşulda tüm iş birlikçi bireyleri ayırt etmede yöntemlerin performansının yüksek olduğu ortaya çıkmıştır. Ancak genel olarak kopya oranı sabit tutulduğunda gruptaki iş birlik yapan birey sayısındaki artış yöntemlerin performansına olan etkisi düşüktür.

**Tablo 4.2.2.**

Kendi Kendine Öğrenme ve SETRED Yöntemlerinin İş Birliği Yapan İki Farklı Gruptaki Bireyleri Sınıflandırma Performansına Ait Değerler

Yöntem	Grup Sayısı	İş Birlikçi Sayısı	Kopya Oranı	Duyarlılık	Özgüllük	PTD	NTD	F Ölçütü	Doğruluk
Kendi Kendine Öğrenme	2	5	%10	0.1125	0.9784	0.1125	0.9715	0.1125	0.9512
			%15	0.1312	0.9857	0.1312	0.9723	0.1312	0.9589
			%20	0.1188	0.9898	0.1188	0.9720	0.1188	0.9625
			%40	0.7750	0.9905	0.7750	0.9927	0.7750	0.9838
	2	10	%10	0.1994	0.9656	0.1994	0.9485	0.1994	0.9185
			%15	0.2638	0.9477	0.2638	0.9516	0.2638	0.9056
			%20	0.4275	0.9633	0.4275	0.9626	0.4275	0.9303
			%40	0.9106	0.9586	0.9106	0.9940	0.9106	0.9557
	2	20	%10	0.3275	0.9231	0.3275	0.9109	0.3275	0.8525
			%15	0.5741	0.9284	0.5741	0.9421	0.5741	0.8864
			%20	0.8475	0.9454	0.8475	0.9790	0.8475	0.9338
			%40	0.9434	0.8947	0.9434	0.9918	0.9434	0.9005
SETRED	2	5	%10	0.1125	0.9784	0.1125	0.9715	0.1125	0.9512
			%15	0.1312	0.9857	0.1312	0.9723	0.1312	0.9589
			%20	0.1188	0.9898	0.1188	0.9720	0.1188	0.9625
			%40	0.7750	0.9905	0.7750	0.9927	0.7750	0.9838
	2	10	%10	0.1994	0.9656	0.1994	0.9485	0.1994	0.9185
			%15	0.2638	0.9477	0.2638	0.9516	0.2638	0.9056
			%20	0.4275	0.9633	0.4275	0.9626	0.4275	0.9303
			%40	0.9106	0.9586	0.9106	0.9940	0.9106	0.9557
	2	20	%10	0.3275	0.9231	0.3275	0.9109	0.3275	0.8525
			%15	0.5741	0.9284	0.5741	0.9421	0.5741	0.8864
			%20	0.8475	0.9454	0.8475	0.9790	0.8475	0.9338
			%40	0.9434	0.8947	0.9434	0.9918	0.9434	0.9005

Tablo 4.2.2 yöntemlerin iki farklı iş birlikçi grup olduğu durumlardaki bireyleri sınıflandırma performansını vermektedir ve tablo incelendiğinde iş birlikçi tek gruptan oluşan verilerin analiz sonucunda olduğu gibi yöntemlerin aynı koşullarda aynı performans değerlerine sahip olduğu görülmektedir. Bu da yöntemlerin ilgilenilen koşullar altında iş birliği yapan ve

yapmayan bireyleri sınıflandırmadaki performanslarının farklılık göstermediğini ortaya koymaktadır.



**Grafik 4.2.2.** Kendi Kendine Öğrenme ve SETRED Yönteminin İş Birlikçi İki Farklı Gruptaki Bireyleri Sınıflandırma Performansı

Kendi kendine öğrenme ve SETRED yöntemlerine ait grafik ve tablolar incelendiğinde iki farklı gruptaki iş birlikçi bireyleri sınıflandırma performansları ile bir gruptaki iş birlikçi bireyleri sınıflandırma performansları benzerlik göstermektedir. Özgüllük, NTD ve doğruluk ölçütleri Tablo 4.2.1’de görüldüğü gibi Tablo 4.2.2’de de 1 ve 1’e yakın değerler almaktadır. Aynı şekilde duyarlılık, PTD ve F ölçütü gruplardaki iş birlikçi sayısı ve kopya oranı arttıkça artmaktadır. Bu ölçütler iş birlik yapan tek gruplu simülasyon verilerinin analiz sonuçlarından farklı olarak tüm koşullarda birbirine yakın değerler almakta ve değerler koşullara bağlı olarak birbirine yakın oranlarda değişmektedir. Bu değerlerin en düşük olduğu koşullar ise iş birlikçi sayısının 5 olduğu %10, %15 ve %20 kopya oranlarıdır. Bu değerler iş birlikçi birey sayısı arttıkça artmaktadır. Özellikle 20 iş birlikçi bireyin maddelerin %20’sinde hile yaptığı durumda ve %40’ının hile yapıldığı tüm koşullarda bu ölçütler oldukça yüksektir. Aynı zamanda %40 kopya oranında yöntemlerin sınıflandırma performanslarının iş birlikçi sayısı arttıkça 1’e yakın değerler aldığı ortaya çıkmıştır.

## 5. TARTIŞMA, SONUÇ VE ÖNERİLER

Bu çalışmada M4 benzerlik istatistiği kullanılarak testlerde hile yoluyla iş birliği yapan birey ve grupları belirlemede denetimsiz öğrenme türü olan Ward hiyerarşik kümeleme yöntemi ile yarı-denetimli öğrenme türü olan kendi kendine öğrenme ve geliştirilmiş hali SETRED yöntemlerinin performansları incelenmiştir. Denetimsiz ve her iki yarı-denetimli öğrenme türü için tek grup ve iki farklı gruptaki bireyleri belirlemeye yönelik koşullar değiştirilerek ayrı ayrı analizler yapılmıştır.

Ward yönteminin iş birliği yapan bir gruptaki bireyler ile iki farklı gruptaki iş birliği yapan bireyleri kümelemek için yapılan analizler sonucunda yöntemin performans ölçütlerinin oldukça benzer sonuçlar ürettiği ortaya çıkmıştır. Bu sonuca göre iş birliği yapan bireyler bir grup da olsa iki farklı grup da olsa yöntemin kümeleme performansının benzer olduğu söylenebilmektedir. Ward yönteminin her iki durumda da her koşulda özgüllük, NTD ve doğruluk değerlerinin 1 ve 1'e yakın değerler aldığı, dolayısıyla gerçekte iş birliği yapmayan bireyleri doğru olarak belirlediği ortaya çıkmıştır. Ancak %20'nin altındaki kopya oranındaki durumlarda yöntemin, iş birliği yapan bireyleri ve grupları ayırt etmede duyarlılık, PTD ve F ölçütü değerlerinin düşük olması nedeniyle yetersiz kaldığı görülmüştür. Özellikle %10 kopya oranında ve iş birlikçi sayısının 5 ve 10 olduğu durumlarda bu ölçütlerin 0'a yakın değerler alması nedeniyle yöntemin iş birlikçi bireyleri kümeleyemediği söylenebilmektedir. İş birlikçi sayısının ve kopya oranının artmasıyla Ward yönteminin hile yoluyla iş birliği yapan bireyleri ve grupları kümeleyebilme performansı da artmaktadır. Ward yönteminin testteki maddelerin %40'ında yapılan hileli tüm durumlarda iş birliği yapan ve yapmayan bireyleri doğru olarak kümelediği ortaya çıkmıştır. Bu durum Ward yönteminin hile yoluyla iş birliği yapan bireyleri ayırt etmede oldukça güçlü olduğunu ortaya koymaktadır.

Kendi kendine öğrenme ve SETRED yönteminin sınıflandırma performansına ait analiz sonuçlarına göre her iki yöntemin performans ölçütlerine ait değerler tüm koşul ve durumlarda aynı çıkmıştır. Dolayısıyla her iki yöntemin sınıflandırma performansının aynı olduğu söylenebilir. Aynı zamanda analiz sonuçlarına göre tek grup ya da iki farklı grupta iş birliği yapan bireylerin yer aldığı durumlarda özgüllük, NTD ve doğruluk ölçütlerine ait değerlerin 1 ve 1'e yakın olduğu görülmüş, bu nedenle yöntemlerin iş birliği yapmayan bireyleri doğru olarak sınıflandırdığı ortaya çıkmıştır. Yine her iki durumda yöntemlerin çoğu koşulda kopya oranı arttıkça duyarlılığı, PTD ve F ölçütünün değerleri de artmaktadır. Ancak %10 kopya durumunda ve 5 ile 10 iş birlikçi bireylerin maddelerde %15 oranında hile yaptığı durumlarda yöntemlerin performansı oldukça düşüktür; bu nedenle bu koşullarda yöntemlerin iş birlikçi bireyleri sınıflandırmada yetersiz kaldığı ortaya çıkmıştır.

İki iş birlikçi grubun yer aldığı örneklemedeki analiz sonuçlarına göre iş birlikçi birey sayısı arttıkça yöntemlerin duyarlılığı, PTD ve F ölçütünün değerleri artmakta, ancak bir iş birlikçi gruplu örneklemlerde yapılan analizlerde aynı kopya oranlarında bu değerlerin çok büyük oranda değişmediği ortaya çıkmıştır. Bu nedenle iş birlikçi bir gruptaki birey sayısının bu yarıdenetimli öğrenme yöntemlerinin sınıflandırma performansına etkisinin düşük olduğu söylenebilir.

Bir iş birlikçi grubun olduğu durumlar ile iki farklı iş birlikçi grubun olduğu durumlara ait analiz sonuçları %20 ve üzeri kopya oranında farklılık göstermektedir. İki iş birlikçi grubun yer aldığı örnekleme ait sonuçlara göre 5 iş birlikçi, testteki maddelerin %20'sinde hile yaptığında yöntemlerin iş birlikçi bireyleri sınıflandıramadığı, iş birlikçi sayısının 10'a çıktığında ise yetersiz kaldığı ortaya çıkmıştır. Tek gruplu durumlardan farklı olarak yöntemlerin %20 kopya oranı ve 20 iş birlikçinin olduğu ve %40 kopyanın olduğu tüm koşullarda iki gruptaki iş birlikçi bireyleri sınıflandırmada oldukça güçlü olduğu ortaya çıkmıştır. Ancak iş birlikçi bir grubun yer aldığı örneklemede yapılan analiz sonuçlarına göre bu koşullardaki yöntemlerin PTD ölçütüne ait değerler diğer ölçütlere göre daha düşüktür; bu durum yöntemlerin iş birliği yapmış olarak belirlediği bireylerde hatalı tahminler yaptığını göstermektedir. Yöntemlerin bu sınıflandırma performansının örneklemede iki iş birlikçi grup olduğunda daha yüksek olmasının nedeni yarıdenetimli makine öğrenmesine verilen eğitim verisinin daha ayrıntılı olması (3 sınıf) ve bunun sonucunda tahminlerde hatayı azaltması olarak görülebilir.

Sonuç olarak Ward kümeleme yöntemi, testteki maddelerin %40'ı ve üzerinde hile ile iş birliği yapan bireyleri belirlemede oldukça güçlü bir yöntemdir. %20 ve altında ise yöntemin performansı yetersiz kalmaktadır. Gruplardaki iş birlikçi birey sayısı arttıkça yöntemin kümeleme performansı ise artmaktadır. Kendi kendine öğrenme ve SETRED yöntemlerinin ise iş birliği yapmayan bireyleri sınıflandırma performansı her koşulda yüksek olsa da iş birlikçi bir gruplu örneklemlerde iş birliği yapan bireylerin sınıfını tahmin etmede performansının istenilen düzeyde olmadığı; gruptaki birey sayısının yöntemlerin performansına etkisinin düşük olduğu ortaya çıkmıştır. İki farklı iş birlikçi gruplu örneklemlerde ise %40 ve üzeri kopyada ve 20 iş birlikçi bireyin maddelerin %20'si oranında hile yaptığı durumlarda bu yöntemlerin iş birlikçi bireyleri ve grupları oldukça iyi ayırt ettiği; gruplardaki iş birliği yapan birey sayısı arttıkça bu performansın da arttığı sonucuna ulaşılmıştır. Kopya oranı arttıkça tüm yöntemlerin iş birlikçileri belirleme performansının artacağı düşünülmektedir.

Günümüzde özellikle teknolojinin gelişmesi ve sosyal medyanın kullanımının artması ile kolaylıkla paylaşılabilen bilgilerle teste girecek bireyler testlerde kullanılan maddeler hakkında ön bilgilere sahip olabilmekte ya da testin uygulama sürecinde birbirleriyle iş birliği yapabilmektedir. Özellikle bireyler karşılıklı olarak yardımlaşabileceği gibi bu yardımlaşmayı grup halinde de yapabilmektedirler. Karşılıklı yapılan türden hileleri belirlemede alanyazında

birçok çalışmaya rastlanmaktadır; ancak grup halinde yapılan hileyi kümeleme veya sınıflandırma yoluyla belirlemeye çalışan çalışma oldukça azdır. Bu işlemlere günümüz teknolojisinin ulaştığı en uç noktalardan olan makine öğrenmesi süreçlerini dahil eden çalışmalar ise çok daha enderdir. Bu çalışmada makine öğrenmesi yöntemlerinden olan denetimsiz ve yarı-denetimli yöntemler kullanılarak grup veya gruplar halinde yapılan iş birliği belirlenmeye çalışılmıştır. Çalışma sonucunda iki ya da daha fazla birey arasında gerçekleşen bu tür iş birliğini belirlemede, uzaklık ölçüsü olarak M4 istatistiğinden elde edilen benzerlik matrisinin kullanıldığı Ward kümeleme yönteminin oldukça güçlü bir yöntem olduğu görülmüştür. Yarı-denetimli makine öğrenmesi yöntemlerinin ise algoritmalarının yazımı esnasında araştırmacı tarafından belirlenen eğitim verisindeki etiketlenmiş verilerin oranı, etiketlenmemiş verilerin oranı gibi değişkenlerden oldukça etkilendiği düşünülmektedir. Bu çalışmada yarı-denetimli yöntemlerin özellikle yüksek kopya oranlarında yapılan iş birliğini belirlemede performanslarının yüksek olduğu; ancak değişkenlerin düzeyleri değişimlenerek performanslarının yeni çalışmalarla sınanması gerektiği düşünülmektedir.

Wollack ve Maynes'in (2017) en yakın komşuluk yöntemini kullanarak hile yoluyla iş birliği yapan bireyleri kümelemeye yönelik yaptıkları çalışma dışında, bu yöntemlerin kullanıldığı alanyazında ulaşılabilen başka bir çalışmaya rastlanmamıştır. Wollack ve Maynes'in (2017) kullandıkları hiyerarşik kümeleme yönteminin bu çalışmada kullanılan Ward yönteminden farklı olması nedeniyle çalışmalar sonucunda elde edilen bulgular karşılaştırılamamaktadır.

Bu çalışma, ilgilenen yöntemler, koşullar ve örneklem sayısı ile sınırlıdır. Aynı çalışma farklı benzerlik istatistikleri, farklı denetimsiz ve yarı-denetimli öğrenme türleri kullanılarak yapılabilir. Kendi kendine öğrenme türü için yapılan analizlerde eğitim verisinin oranı değişimlenerek yöntemin performansı belirlenebilir. Koşulların düzeyleri ve örneklem sayısı değişimlenerek yöntemlerin performanslarına olan etkileri araştırılabilir.

## 6. KAYNAKLAR

- [1]Alpar, R. (2013). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Ankara: Detay Yayıncılık.
- [2]Amini, M. R., & Gallinari, P. (2002). Semi-supervised logistic regression. *In Fifteenth European Conference on Artificial Intelligence* (pp.390–394).
- [3]Angoff, W. H. (1974). The development of statistical indices for detecting cheaters. *Journal of the American Statistical Association*, 69, 44-49.
- [4]Armstrong, R. D., & Shi, M. (2009). A parametric cumulative sum statistic for person fit. *Applied Psychological Measurement*, 33(5), 391-410.
- [5]Belov, D. I. (2011). Detection of answer copying based on the structure of a high-stakes test. *Applied Psychological Measurement*, 35(7), 495-517.
- [6]Belov, D. I., & Armstrong, R. D. (2010). Automatic detection of answer copying via Kullback-Leibler Divergence and K-Index. *Applied Psychological Measurement*, 34(6), 379-392.
- [7]Belov, D. I., Pashley, P. J., Lewis, C., & Armstrong, R. D. (2007). Detecting aberrant responses with Kull- back-Leibler distance. In K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 7-14). Tokyo: Universal Academy Press.
- [8]Bird, C. (1927). The detection of cheating in objective examinations. *School and Society*, 25, 261-262.
- [9]Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- [10]Bishop, C. M., & Egan, K. (2017). Detecting erasures and unusual gain scores: understanding the status quo. In G. J. Cizek and J. A. Wollack, Eds., *Handbook of quantitative methods for detecting cheating on tests*. New York, NY: Routledge.
- [11]Blum, A. (2007). Machine learning theory. Erişim Adresi: <http://www.cs.cmu.edu/afs/cs/user/avrim/www/Talks/mlt.pdf>.
- [12]Bradlow, E. T., Weiss, R. E., & Cho, M. (1998). Bayesian identification of outliers in computerized adaptive testing. *Journal of the American Statistical Association*, 93, 910-919.
- [13]Brownlee, J. (2013). A tour of the most popular machine learning algorithms. Erişim Adresi: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>.
- [14]Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- [15]Chalmers, R.P. (2019). mirt: A multidimensional item response theory package for the R environment. Erişim Adresi: <https://cran.r-project.org/web/packages/mirt/mirt.pdf>
- [16]Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-Supervised Learning*. London: The MIT Press Cambridge.
- [17]Cizek, G. J. (1999). *Cheating on tests: how to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.
- [18]Cizek, G. J., & Wollack, J. A. (2017). *Handbook of quantitative methods for detecting cheating on tests*. New York, NY: Routledge.
- [19]Clark, J. M. (2010). *Aberrant response patterns as a multidimensional phenomenon: Using factor-analytic model comparison to detect cheating*. ProQuest LLC. University of Kansas.
- [20]Clark, J. M., Skorupski, W. P., & Murphy, S. T. (2013, October). *Using nonlinear regression to identify unusual performance level classification rates*. Paper presented at the Conference on Statistical Detection of Potential Test Fraud, Madison, WI.
- [21]Cook, R. J. (2013). *Application of item response theory models to the algorithmic detection of shift errors on paper and pencil tests*. (Doctoral Dissertation). University of Massachusetts.
- [22]Çokluk, Ö. S., Şekercioğlu, G. ve Büyüköztürk, S. (2014). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi Yayıncılık.
- [23]Deng, W. (2008). *An innovative use of the standardized log-likelihood statistic to evaluate person fit*. Dissertation Abstracts International Section A: Humanities and Social Sciences. Rutgers State University of New Jersey.
- [24]Donalek, C. (2011). Supervised and unsupervised learning. Erişim Adresi: [http://www.astro.caltech.edu/~george/aybi199/Donalek\\_Classif.pdf](http://www.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf)

- [25]Donlan, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group determined item difficulties. *Educational and Psychological Measurement*, 28, 105–113.
- [26]Drasgow, F., & Levine, M. V. (1986). Optimal detection of certain forms of inappropriate test scores. *Applied Psychological Measurement*, 10, 59–67.
- [27]Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement*, 11, 59–79.
- [28]Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1991). Appropriateness measurement for some multidimensional test batteries. *Applied Psychological Measurement*, 15, 171–191.
- [29]Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38, 67-86.
- [30]Embretson, S. E., & Reise, S. P. (2000) *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Inc.
- [31]Ferrando, P. J., & Lorenzo-Seva, U. (2007). A measurement model for Likert responses that incorporates response time. *Multivariate Behavioral Research*, 42, 675-706.
- [32]Ferrando, P. J. (2009). Multidimensional factor-analysis-based procedures for assessing scalability in personality measurement. *Structural Equation Modeling*, 16, 109-133.
- [33]Ferreira, L., & Hitchcock, D. B. (2009). A comparison of hierarchical methods for cluster functional data. *Communications in Statistics-Simulation and Computation*, 38(9), 1925-1949.
- [34]Frery, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education*, 6(2), 153–165.
- [35]Gan, G., Ma, C., & Wu, J. (2007). *Data clustering theory, algorithms, and applications society for industrial and applied mathematics*. Philadelphia: SIAM.
- [36]Gu, Y., Zhang, H., & Lui, X. (2011). *Instance selection in semi-supervised learning*. In C. Butz and P. Lingas (Eds), In advances in artificial intelligence: 24th Canadian Conference on Artificial Intelligence, Canadian AI, LNAI 6657 (pp. 158-169). Berlin: Springer-Verlag.
- [37]Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- [38]Guttman, L. A. (1950). The basis for scalogram analysis. In Stouffer, S. A., Guttman, L.A., and Schuman, E. A. (Eds.) *Measurement and prediction. Volume 4 of Studies in social psychology in World War II*. Princeton: Princeton University Press.
- [39]Güzeller, C. O. (Ed.). (2016). *SPSS Uygulamalı çok değişkenli istatistik teknikleri* (2. Baskı). Ankara: Maya Akademi Yayın Dağıtım Eğitim Danışmanlık.
- [40]Hands, S., & Everitt, B (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical cluster techniques. *Multivariate Behavioral Research*. 22, 235-243.
- [41]Haney, W. M., & Clarke, M. J. (2007). Cheating on tests: Prevalence, detection, and implications for on-line testing. In E. M. Anderman and T. B. Murdock (Eds.), *Psychology of academic cheating* (pp. 255-287). San Diego: Elsevier Academic Press.
- [42]Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18, 133–146.
- [43]Harpp, D. N., & Hogan, J. J. (1993). Crime in the classroom: Detection and prevention of cheating on multiple-choice exams. *Journal of Chemical Education*, 70(4).
- [44]Harrington, P. (2012). *Machine Learning in Action*. USA: Manning Publications Co.
- [45]Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- [46]Heck, R.H., & Thomas, S. L., (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York: Routledge.
- [47]Holland, P.W. (1996). *Assessing unusual agreement between the incorrect answers of two examinees using the K-index: statistical theory and empirical support* (Research Report RR-94-4). Princeton: Educational Testing Service.
- [48]Hui, H.-fai. (2010). *Stability and sensitivity of a model-based person-fit index in detecting item pre-knowledge in computerized adaptive test*. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. University of Hong Kong.

- [49] Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118, 843-877.
- [50] Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis (6th Edition)*. New Jersey: Pearson Education Inc.
- [51] Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement*, 4, 105-126.
- [52] Klauer, K. C. (1991). An exact and optimal standardized person test for assessing consistency with the Rasch model. *Psychometrika*, 56, 535-547.
- [53] Klauer, K. C. (1995). The assessment of person fit. In G. H. Fischer and I. W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications*, (pp. 97-110). New York: Springer-Verlag.
- [54] Klauer, K. C., & Rettig, K. (1990). An approximately standardized person test for assessing consistency with a latent trait model. *British Journal of Mathematical and Statistical Psychology*, 43, 193-206.
- [55] Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277-298.
- [56] Levine, M. V., & Drasgow, F. (1983). Appropriateness measurement: Validating studies and variable ability models. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 109-131). New York: Academic Press.
- [57] Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika*, 53, 161-176.
- [58] Levine, M. V. & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 4, 269-290.
- [59] Li, M., & Zhou, Z. H. (2005). SETRED: Self-training with editing. *In Advances, in Knowledge Discovery and Data Mining* (pp.611-621). Berlin: Springer.
- [60] Liu, X., Liu, F., & Simon, M., (2014, April). *Are the score gains suspicious?-A Bayesian growth analysis approach*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.
- [61] Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- [62] Magis, D., Raïche, G., & Beland, S. (2012). A didactic presentation of snijders's index of person fit with emphasis on response model selection and ability estimation. *Journal of Educational and Behavioral Statistics*, 37(1), 57-81.
- [63] Maynes, D. D. (2009, April). *Combining statistical evidence for increased power in detecting cheating*. Presented at the annual conference of the National Council on Measurement in Education, San Diego, CA.
- [64] Maynes, D. D. (2017). Detecting potential collusion among individual examinees using similarity analysis. In G. J. Cizek and J. A. Wollack, Eds., *Handbook of quantitative methods for detecting cheating on tests*. New York, NY: Routledge.
- [65] McLeod, L. D., & Lewis, C. (1999). Detecting item memorization in the CAT environment. *Applied Psychological Measurement*, 23, 147-160.
- [66] McManus, I. C., Lissauer, T., & Williams, S. E. (2005). Detecting cheating in written medical examinations by statistical analysis of similarity of answers: Pilot study. *BMJ: British Medical Journal (International Edition)*, 330(7499), 1064-1066.
- [67] Meijer, R. R. (1996). Person-Fit research: An introduction. *Applied Measurement in Education*, 9, 3-8.
- [68] Meijer, R. R. & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107-135.
- [69] Molenaar, I. W., & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, 55, 75-106.
- [70] Murphy, K. P. (2012). *Machine Learning: A probabilistic perspective*. London: The MIT Press.
- [71] Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview II. *WIREs Data Mining and Knowledge Discovery*, 7 (6), 1-16.

- [72] Nering, M. L., & Meijer, R. R. (1998). A comparison of the person response function and the lz person fit statistic. *Applied Psychological Measurement*, 22, 53–69.
- [73] Özdamar, K. (2013). *Paket programlar ile çok değişkenli veri analizi (Cilt 2)*. Ankara: Nisan Kitapevi.
- [74] Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, 14 (2), 127-137.
- [75] Reise, S. P., & Due, A. M. (1991). The influence of test characteristics on the detection of aberrant response patterns. *Applied Psychological Measurement*, 15(3), 217-226.
- [76] Rokach, L., & Maimon, O. (2005). Clustering Methods. In *Data Mining And Knowledge Discovery Handbook, Chapter 15* (pp.321-352).
- [77] Saupe, J.L. (1960). An empirical model for the corroboration of suspected cheating on multiple-choice tests. *Educational and Psychological Measurement*, 20, 475-489.
- [78] Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. *Applied Psychological Measurement*, 23, 41-53.
- [79] Shu, Z. (2011). *Detecting test cheating using a deterministic, gated item response theory model*. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. University of North Carolina at Greensboro.
- [80] Sato, T. (1975). *The construction and interpretation of S-P tables*. Tokyo: Meiji Tosho.
- [81] Sijtsma, K. (1986). A coefficient of deviance of response patterns. *Kwantitatieve Methoden*, 7, 131-145.
- [82] Sijtsma, K., & Meijer, R. R. (2001). The person response function as a tool in person-fit research. *Psychometrika*, 66, 191-208.
- [83] Skorupski, W. P., & Egan, K. (2011, April). *Detecting cheating through the use of hierarchical growth models*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- [84] Skorupski, W. P., & Egan, K. (2014). A Bayesian hierarchical linear modeling approach for detecting cheating and aberrance. In N. M. Kingston & A. Clark (eds.) *Test fraud: Statistical detection and methodology*. New York: Routledge.
- [85] Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and Psychological Measurement*, 45, 433-444.
- [86] Snijders, T. A. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, 66, 331-342.
- [87] Sotaridona, L. S., & Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, 40, 53-69.
- [88] Sotaridona, L. S., van der Linden, W. J., & Meijer, R. R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement*, 30(5), 412-431.
- [89] Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, 40, 331-352.
- [90] Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, 49, 95-110.
- [91] Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement*, 20, 221-230.
- [92] Trabin, T. E., & Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 83-108). New York: Academic Press.
- [93] van der Ark, L. A., Emons, W. H. M., & Sijtsma, K. (2008). Detecting answer copying using alternate test forms and seat locations in small-scale examinations. *Journal of Educational Measurement*, 45(2), 99-117.
- [94] van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology*, 13, 267-298.
- [95] van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, 31, 181-204.

- [96]van der Linden, W. J., & Jeon, M. (2012). Modeling answer changes on test items. *Journal of Educational and Behavioral Statistics*, 37, 180–199.
- [97]van der Linden, W. J., & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics*, 31(3), 283-304.
- [98]van Krimpen-Stoop, M. L. A., & Meijer, R. R. (2001). CUSUM-Based person-fit statistics for adaptive testing. *Journal of Educational and Behavioral Statistics*, 26 (2), 199-217.
- [99]Wang, B., Spencer, B., Ling, C.X., & Zhang, H. (2008). Semi-supervised self-training for sentence subjectivity classification. In *The 21st Canadian Conference on Artificial Intelligence* (pp. 344–355).
- [100]Weiss, D. J. (1973). *The stratified adaptive computerized ability test*. Minneapolis MN: University of Minnesota, Department of Psychology.
- [101]Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. USA: Morgan Kaufmann Publishers.
- [102]Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- [103]Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- [104]Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. *Applied Psychological Measurement*, 21(4), 307-320.
- [105]Wollack, J. A. (2003). Comparison of answer copying indices with real data. *Journal of Educational Measurement*, 40, 189-205.
- [106]Wollack, J. A., Cohen, A. S., & Eckerly, C. A. (2015). Detecting test tampering using item response theory. *Educational and Psychological Measurement*, 75, 931–953.
- [107]Wollack, J. A., & Maynes, D. D. (2017). Detection of test collusion using cluster analysis. In G. J. Cizek and J. A. Wollack, Eds., *Handbook of quantitative methods for detecting cheating on tests*. New York, NY: Routledge.
- [108]Zhang, Y., Searcy, C. A., & Horn, L. (2011, April). *Mapping clusters of aberrant patterns of item responses*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- [109]Zopluoglu, C., & Davenport, E.C., Jr. (2012). The empirical power and type I error rates of the GBT and  $\omega$  indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement*, 1-26.
- [110]Zhu, X. (2005). *Semi supervised learning literature survey*. Technical Report, University of Wisconsin Madison.

## 7. ÖZGEÇMİŞ

### 7.1. Kişisel Bilgiler

**Adı Soyadı** : Seha YORMAZ  
**Doğum Tarihi:** 06/03/1983  
**E-Posta** : sehayormaz@gmail.com

### 7.2. Eğitim Bilgileri

Derece	Alan	Okul/Üniversite	Yıl
İlkokul	-	Aliye Pozcu İlkokulu	1991-1995
Lise	-	MTSO Anadolu Lisesi	1995-2001
Lisans	Fen Bilgisi Öğretmenliği	Hacettepe Üniversitesi	2001-2005
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2011-2014
Doktora	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2014-2019

### 7.3. Yabancı Dil: İngilizce

### 7.4. Çalıştığı Kurumlar

Unvanı	Görev Yeri	Yıl
Fen Bilimleri Öğretmeni	Tarsus Karadirlik İlköğretim Okulu	2006-2011
Fen Bilimleri Öğretmeni	Akdeniz Karacailyas İlköğretim Okulu	2011-2012
Fen Bilimleri Öğretmeni	Yenişehir Çavak Günferi Karagenç Ortaokulu	2012- Devam Ediyor

### 7.5. Bilgisayar Becerileri

**7. 5.1. İşletim Sistemi:** Windows 10 ve öncesi

**7. 5.2. Kullanılan Programlar:** Microsoft Office Programları, SPSS, ITEMAN, R Programı

### 7.6. Tezleri

**Yüksek Lisans:** “Kopya Belirleme İndekslerinin Çeşitli Koşullar Altında I. Tip Hata Oranlarının ve Güçlerinin Belirlenmesi”. Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü, 2014. (Danışman: Yrd. Doç. Dr. Önder SÜNBÜL)

## 7.7. Yayınlar

### 7.7.1. Uluslararası hakemli dergilerde yayınlanan makaleler

Yormaz, S., & Sünbül, Ö. "Determination of type I error rates and power of answer copying indices under various conditions". *Educational Sciences: Theory & Practice*, 17, 5-26 (2016).

Sünbül, Ö., & Yormaz, S. "Effects of test level discrimination and difficulty on answer-copying indices." *International Journal of Evaluation and Research in Education*, 7 (1), 32-38 (2018).

Uzun, N. B., Aktaş, M., Aşiret, S., & Yormaz, S. "Using generalizability theory to assess the score reliability of communication skills of dentistry students." *Asian Journal of Education and Training*, 4(2), 85-90 (2018).

Sünbül, Ö., & Yormaz, S. "Investigating the performance of omega index according to item parameters and ability levels." *Eurasian Journal of Educational Research*, 74, 207-226 (2018).

### 7.7.2. Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında (Proceeding) basılan bildiriler.

Sünbül, Ö., & Yormaz, S. "Type I Error Rates and Powers Study of K Indices for Detecting Answer Copying". The International Congress on Education for the Future: Issues and Challenges, 367-368, Ankara, 2015.

Sünbül, Ö., & Yormaz, S. "Investigating the performance of  $\omega$  index in detecting answer copying". Conference on Test Security, Madison, 2017.

Sünbül, Ö., & Yormaz, S. "M4 istatistiğinin cevaplama benzerliği belirleme performansının incelenmesi." 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Kosova, 2018.

### 7.7.3. Ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler

Sünbül, Ö., & Yormaz, S. "Kopya Belirleme İndekslerinin Madde Ayırt Edicilik ve Güçlük Düzeylerine Göre Performanslarının İncelenmesi". Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Antalya, 2016.

### 7.7.4. Ulusal yayınevlerince basılan kitaplar veya kitaplarda bölümler

Erkuş, A., Sünbül, Ö., Sünbül Ömür, S., Yormaz, S. ve Aşiret, S. (2017). *Psikolojide ölçme ve ölçek geliştirme-II: Ölçme araçlarının psikometrik nitelikleri ve ölçme kuramları*. Ankara: Pegem Akademi.