

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

YARI PARAMETRİK VE PARAMETRİK OLMAYAN
EŞİTLEME YÖNTEMLERİ İLE BİR YETENEK
TESTİNİN FARKLI FORMLARININ EŞİTLENMESİ

DOKTORA TEZİ

ŞERİFE ZEYBEKOĞLU

ANKARA
HAZİRAN, 2024



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**YARI PARAMETRİK VE PARAMETRİK OLMAYAN
EŞİTLEME YÖNTEMLERİ İLE BİR YETENEK
TESTİNİN FARKLI FORMLARININ EŞİTLENMESİ**

DOKTORA TEZİ

ŞERİFE ZEYBEKOĞLU

DANIŞMAN: PROF. DR. KAAAN ZÜLFİKAR DENİZ

**ANKARA
HAZİRAN, 2024**

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Şerife Zeybekoğlu adlı öğrencinin hazırladığı “Yarı Parametrik ve Parametrik Olmayan Eşitleme Yöntemleri ile Bir Yetenek Testinin Farklı Formlarının Eşitlenmesi” başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Programı’nda jüri üyelerince oy birliği ile **Doktora Tezi** olarak kabul edilmiştir.

	<u>Jüri Üyeleri</u>	<u>İmza</u>
Başkan	Prof. Dr. Selahattin GELBAL	
Üye	Prof. Dr. Kaan Zülfikar DENİZ (Danışman)	
Üye	Prof. Dr. Hülya KELECİOĞLU	
Üye	Doç. Dr. Celal Deha DOĞAN	
Üye	Doç. Dr. Seher YALÇIN	

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim Öğretim Yönetmeliği’nin ilgili maddeleri uyarınca, jüri üyeleri tarafından 12/06/2024 tarihinde, Enstitü Yönetim Kurulu tarafından ise .../.../20... tarihinde kabul edilmiştir.

.....

Prof. Dr. Mehmet İktbal YETİŞİR
Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.

Şerife ZEYBEKOĞLU

ÖZET

YARI PARAMETRİK VE PARAMETRİK OLMAYAN EŞİTLEME YÖNTEMLERİ İLE BİR YETENEK TESTİNİN FARKLI FORMLARININ EŞİTLENMESİ

ZEYBEKOĞLU, Şerife

Doktora Tezi, Eğitim Bilimleri Anabilim Dalı

Tez Danışmanı: Prof. Dr. Kaan Zülfikar DENİZ

Haziran, 2024, xv + 173 sayfa

Araştırmada Thurstone'un Çoklu Faktörler Modeli temel alınarak dokuzuncu sınıf öğrencileri için iki paralel formu olan bir yetenek testi geliştirilip, testin paralel formlarının farklı eşitleme desenlerinde yarı parametrik ve parametrik olmayan eşitleme yöntemleriyle eşitlenmesi amaçlanmıştır. Bu test Thurstone'un Çoklu Faktörler Modelindeki bellek, sözel, sayısal, muhakeme, yer ve mekâna uyma, algısal ve kelime akıcılığı olmak üzere yedi faktörü içermektedir. Bu testle ile lise eğitimi düzeyine gelmiş öğrencilerin yeteneklerinin güçlü ve zayıf yönlerinin belirlenmesi amaçlanmaktadır. Bu amaçla geliştirilen yetenek testinin Antalya ili beş merkez ilçedeki resmi ortaöğretim kurumlarında 2022-2023 eğitim-öğretim yılında deneme ve nihai uygulaması yapılmıştır. Deneme uygulaması sonucunda toplam 706 öğrenci araştırmaya dahil edilmiş; elde edilen veri üzerinde test ve madde istatistikleri incelenmiş olup geçerlik ve güvenilirlik analizleri yapılmıştır. Test üzerinde gerekli redaksiyonlar yapıldıktan sonra test nihai uygulama için hazırlanmıştır. Nihai uygulama sonucunda toplam 4201 öğrenci araştırmaya dahil edilmiştir. Elde edilen veri üzerinde test ve madde istatistikleri incelenmiş ve eşitleme işlemi gerçekleştirilmiştir. Testin paralel formları eşit yüzdelikli, Kernel ve Parametrik olmayan Bayes (PoB) yöntemleriyle eşitlenmiştir. Eşitleme sürecinde eş değer gruplar (EG), denk olmayan gruplarda ortak madde (NEAT) ve denk olmayan gruplarda ortak değişken (NEC) desenlerinde eşitlemeler yapılmıştır. Eşitlemelerde sözde ortak maddeler ve ortak değişken olarak cinsiyet ve kategorik hale getirilen Liselere Giriş Sınavı (LGS) puan düzeyleri kullanılmıştır. Bu araştırmada ortak değişken kullanımına izin veren esnek bir yaklaşım olarak tanıtılan PoB yöntemiyle yapılan eşitlemenin farklı yöntemlerde ve

desenlerde yapılan eşitlemelerle karşılaştırması yapılmıştır. Diğer yöntemler ile PoB yönteminden elde edilen sonuçları karşılaştırmak için, eşitlenmiş puanların dağılımının hedef teste uzaklıklarını inceleyen Hellinger Uzaklığı kullanılmıştır. PoB yönteminde hedef teste en yakın dağılımlar elde edilmiştir. Bu dağılımlar içerisinde en yakın uzaklık iki ortak değişken birlikte kullanıldığında elde edilmiştir. PoB yöntemiyle elde edilen uzaklıkların diğer eşitleme yöntemleri sonucunda oluşanların yaklaşık onda biri olduğu belirlenmiştir. Bu araştırmada PoB yöntemiyle yapılan test eşitlemenin klasik yöntemlere kıyasla daha bilgilendirici olduğu ve daha iyi sonuç verebileceği görülmüştür.

Anahtar Sözcükler: Yetenek testi, Test eşitleme, Eşit yüzdelikli eşitleme, Kernel eşitleme, PoB yöntemiyle eşitleme, Hellinger Uzaklığı.

ABSTRACT

EQUATING PARALLEL FORMS OF AN APTITUDE TEST USING SEMIPARAMETRIC AND NONPARAMETRIC METHODS

ZEYBEKOGLU, Serife

Dissertation, Department of Educational Sciences

Supervisor: Prof. Dr. Kaan Zlfikar DENİZ

June, 2024, xv + 173 pages

In this study, it was aimed to develop an aptitude test with two parallel forms for ninth grade students based on Thurstone’s Multiple Factors Model and to equate the parallel forms of the test with semiparametric and nonparametric equating methods in different equating designs. This test includes the seven factors of Thurstone’s Multiple Factors Model: memory, verbal, numerical, reasoning, spatial, perceptual and word fluency. With this test, it is aimed to determine the strengths and weaknesses of the abilities of students who have reached the high school education level. For this purpose, a trial and final application of the developed aptitude test was carried out in official secondary education institutions in five central districts of Antalya province in the 2022-2023 academic year. As a result of the trial application, a total of 706 students were included in the study; test and item statistics were examined on the data obtained and validity and reliability analyses were performed. After the necessary corrections and redactions were made on the test, the test was prepared for the final application. As a result of the final application, a total of 4201 students were included in the study. Test and item statistics were examined on the obtained data and equating process was performed. Parallel forms of the test were equated with equipercentile, Kernel and Bayesian nonparametric (BNP) methods. Equivalent groups (EG), non-equivalent groups with anchor test (NEAT) and non-equivalent groups with covariates (NEC) were used in the equating process. Gender and categorised High School Entrance Examination (HSEE) score levels were used as covariates in equating. In this study, the comparison of equating with the BNP method, which is introduced as a flexible approach that allows the use of covariates, with equating with different methods and designs was made. In order to

compare the results obtained from the BNP method with the other methods, Hellinger Distance, which analyses the distances of the distribution of equated scores to the target test, was used. In the BNP method, the distributions closest to the target test were obtained. Among these distributions, the closest distance was obtained when two covariates were used together. It was determined that the distances obtained by BNP method were approximately one tenth of the distances obtained by other equating methods. In this study, it was seen that test equating with BNP method is more informative and can give better results compared to classical methods.

Key Words: Aptitude test, Test equating, Equipercentile equating, Kernel equating, BNP method to equating, Hellinger distance.

ÖNSÖZ

Doktora eğitimim boyunca akademik desteğini esirgemeyen, bilimsel bakış açısıyla araştırmamı zenginleştiren, bana yeni ufuklar açan, tez çalışmamın her aşamasında yanımda olan, takıldığım her konuda büyük bir sabır ve ilgiyle yol gösteren, hep daha iyisini yapabilmem için beni cesaretlendiren ve bana güvenen, akademik disipliniyle her zaman örnek olan çok değerli danışmanım Prof. Dr. Kaan Zülfikar DENİZ'e teşekkürlerimi sunarım.

Tez izleme komitemde yer alarak görüş ve önerileriyle çalışmama katkı sağlayan, anlayışlı ve güler yüzlü tavırlarıyla beni motive eden değerli hocalarım Prof. Dr. Hülya KELECİOĞLU ve Doç. Dr. Celal Deha DOĞAN'a; savunma jürimde yer alarak değerli katkılarını sunan Prof. Dr. Selahattin GELBAL ve Doç. Dr. Seher YALÇIN'a teşekkürlerimi sunarım.

Doktora eğitimim sürecinde ders almaktan büyük onur duyduğum, her zaman öğrencilerinin yanında olan ve değerli bilgi birikimlerini aktarmaktan mutluluk duyan Eğitimde Ölçme ve Değerlendirme Anabilim Dalındaki tüm hocalarıma teşekkürlerimi sunarım. Çalışmamın veri analizinde bilgi birikimini ve tecrübesini benimle paylaşan ve desteğini esirgemeyen değerli hocam Doç. Dr. Meltem YURTÇU'ya teşekkürlerimi sunarım.

Doktora sürecinin bana kazandırdığı, dostluklarını ve akademik desteklerini hep hissettiren değerli arkadaşlarım Ayşe BİLİCİOĞLU GÜNEŞ ve Evrim YALÇIN'a teşekkür ederim.

Son olarak her koşulda yanımda olan sevinçlerimi, üzüntülerimi, kaygılarımı ve streslerimi benimle birlikte paylaşan, bana benden çok inanan, sevgilerini tarif edemeyeceğim canım anneme, babama ve kardeşlerime teşekkür ederim.



Canım Aileme...

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
İÇİNDEKİLER.....	x
TABLolar DİZİNİ.....	xii
ŞEKİLLER DİZİNİ	xiii
KISALTMALAR	xv
BÖLÜM 1	1
GİRİŞ.....	1
Problem	6
Amaç	9
Önem	10
Sayıtlar	12
Sınırlılıklar.....	12
BÖLÜM 2.....	13
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR.....	13
Bağlama ve Eşitleme	13
Tahmin.....	13
Ölçekleme.....	14
Eşitleme	16
Eşitleme Gereksinimleri	17
Eşitleme Basamakları	19
Eşitleme Desenleri	20
Eşitleme Yöntemleri	27
Gözlenen Puan Eşitleme	27
Ortalama Eşitleme	27
Lineer (Doğrusal) Eşitleme	28
Eşit Yüzdelikli Eşitleme.....	30
Yerel (Local) Eşitleme	32
Kernel Eşitleme	32
Madde Tepki Kuramına Dayalı Eşitleme.....	41
MTK Gerçek Puan Eşitleme	45
MTK Gözlenen Puan Eşitleme.....	45
Bayes Yaklaşımı	46
PoB Yaklaşımı ile Eşitleme	49
İlgili Araştırmalar	55

BÖLÜM 3.....	67
YÖNTEM.....	67
Araştırmanın Modeli	67
Çalışma Grubu.....	67
Verilerin Toplanması.....	68
Verilerin Çözümlemesi.....	76
Deneme Uygulaması.....	77
Nihai Uygulama.....	86
Test Eşitleme Varsayımlarının Test Edilmesi	96
Karşılaştırma Kriteri	102
BÖLÜM 4.....	105
BULGULAR VE YORUMLAR	105
EG ve NEAT Desenlerinde Eşit Yüzdelikli Eşitleme Sonuçlarının İncelenmesi .	105
EG ve NEAT Desenlerinde Kernel Eşitleme Sonuçlarının İncelenmesi.....	112
NEC Deseninde Kernel Eşitleme Sonuçlarının İncelenmesi.....	119
NEC Deseninde PoB Eşitleme Sonuçlarının İncelenmesi.....	125
BÖLÜM 5.....	147
SONUÇLAR VE ÖNERİLER	147
Sonuçlar.....	147
Öneriler.....	152
Uygulayıcılara Yönelik Öneriler	152
Araştırmacılara Yönelik Öneriler	153
KAYNAKLAR.....	155
EKLER	169
EK 1. Etik Kurul Onayı.....	170
EK 2. Antalya İl Milli Eğitim Müdürlüğü Araştırma İzni.....	171
BENZERLİK BİLDİRİMİ	172
ÖZGEÇMİŞ.....	173

TABLÖLAR DİZİNİ

Tablo	Sayfa
Tablo 1. Tek Grup Deseni	21
Tablo 2. EG Deseni	21
Tablo 3. Dengelenmiş Tek Grup Deseni	23
Tablo 4. NEAT Deseni	24
Tablo 5. NEC Deseni.....	26
Tablo 6. Thurstone Çoklu Faktörler Modeli.....	70
Tablo 7. Faktörlere Ait Yeterlik İfadeleri.....	72
Tablo 8. Belirtke Tablosu	73
Tablo 9. Test Formlarında Bulunan Maddeler	77
Tablo 10. Deneme Uygulamasına Ait Test İstatistikleri	78
Tablo 11. A ve B Formlarına Ait Madde Analizleri.....	79
Tablo 12. A Formunda İkinci Düzey DFA Uyum İndeksleri.....	83
Tablo 13. B Formunda İkinci Düzey DFA Uyum İndeksleri.....	85
Tablo 14. Test Formlarında Bulunan Maddeler	86
Tablo 15. A ve B Formlarında İki Kategorili Modeller Arasında Yapılan ANOVA Sonuçları	87
Tablo 16. A ve B Formlarında Çok Kategorili Modellerin Uyum İndeksleri	87
Tablo 17. A Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri	88
Tablo 18. A Formunda Kısa Cevaplı Maddelerin MTK'ye Dayalı Madde İstatistikleri	90
Tablo 19. A Formunda Nihai Uygulama Verilerine Ait Test İstatistikleri.....	90
Tablo 20. B Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri.....	92
Tablo 21. B Formunda Kısa Cevaplı Maddelerin MTK'ye Dayalı Madde İstatistikleri	94
Tablo 22. B Formunda Nihai Uygulama Verilerine Ait Test İstatistikleri.....	94
Tablo 23. A ve B Formları ve Sözde Ortak Testler İçin Test İstatistikleri.....	97
Tablo 24. Gruplarda 2022 LGS Yerleştirme Puan Ortalaması.....	99
Tablo 25. Ortak Değişkenlerin Alt Kategorilerinde Bulunan Öğrenci Sayıları	100
Tablo 26. EG Deseninde Eşit Yüzdelikli Eşitleme Sonuçları	105
Tablo 27. NEAT Deseninde Eşit Yüzdelikli Eşitleme Sonuçları.....	107
Tablo 28. EG Deseninde Kernel Eşitleme Sonuçları	112
Tablo 29. NEAT Deseninde Kernel Eşitleme Sonuçları	115
Tablo 30. NEC Deseninde Kernel Eşitleme Sonuçları.....	120

ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 1. Dengelenmiş Tek Grup Deseninde İki EG ve İki Tek Grup Deseni	23
Şekil 2. Lineer Eşitleme ve Lineer Regresyonun Karşılaştırılması.....	30
Şekil 3. Eşit Yüzdelikli Eşitleme.....	31
Şekil 4. Formlara Ait Histogram Grafikleri	78
Şekil 5. A Formu İçin İkinci Düzey DFA	82
Şekil 6. B Formu İçin İkinci Düzey DFA.....	84
Şekil 7. A formuna Ait Histogram Grafiği.....	91
Şekil 8. B formuna Ait Histogram Grafiği	95
Şekil 9. EG Deseninde Eşit Yüzdelikli Eşitleme Puan Dağılımları	106
Şekil 10. EG Deseninde Eşit Yüzdelikli Eşitleme SEE Dağılımları	106
Şekil 11. EG Deseninde Eşit Yüzdelikli Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	107
Şekil 12. NEAT Deseninde Eşit Yüzdelikli Eşitleme Puan Dağılımları.....	108
Şekil 13. NEAT Deseninde Eşit Yüzdelikli Eşitleme SEE Dağılımları.....	109
Şekil 14. NEAT Deseninde Eşit Yüzdelikli Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	110
Şekil 15. Eşit Yüzdelikli Eşitleme Sonuçlarının Değerlendirilmesi	111
Şekil 16. EG Deseninde Kernel Eşitleme Puan Dağılımları	113
Şekil 17. EG Deseninde Kernel Eşitleme SEE Dağılımları	114
Şekil 18. EG Deseninde Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	114
Şekil 19. NEAT Deseninde Kernel Eşitleme Puan Dağılımları	116
Şekil 20. NEAT Deseninde Kernel Eşitleme SEE Dağılımları	116
Şekil 21. NEAT Deseninde Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	117
Şekil 22. EG ve NEAT Desenlerinde Kernel Eşitleme Sonuçlarının Değerlendirilmesi	118
Şekil 23. NEC Deseninde Kernel Eşitleme Puan Dağılımları.....	120
Şekil 24. NEC Deseninde Kernel Eşitleme SEE Dağılımları.....	121
Şekil 25. NEC Deseninde Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	122

Şekil 26. NEC Deseninde Kernel Eşitleme Sonuçlarının Değerlendirilmesi.....	123
Şekil 27. Cinsiyet Değişkenine Göre Formlardaki Puan Dağılımları	126
Şekil 28. Cinsiyet Değişkenine Göre Formlardaki Puan Dağılımları İçin Güven Aralıkları	127
Şekil 29. Kız ve Erkek Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi.....	128
Şekil 30. Cinsiyet Ortak Değişkenine Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları.....	129
Şekil 31. LGS Puan Düzeylerine Göre Formlardaki Puan Dağılımları.....	130
Şekil 32. LGS Puanı Yüksek Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları.....	131
Şekil 33. LGS Puanı Orta Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları.....	132
Şekil 34. LGS Puanı Düşük Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları.....	132
Şekil 35. LGS Puan Düzeylerine Göre Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi	133
Şekil 36. LGS Puan Düzeyleri Ortak Değişkenine Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	134
Şekil 37. LGS Puan Düzeylerine Göre A Formunda Kız Öğrencilerin Puan Dağılımları İçin Güven Aralıkları.....	135
Şekil 38. LGS Puan Düzeylerine Göre A Formunda Erkek Öğrencilerin Puan Dağılımları İçin Güven Aralıkları	135
Şekil 39. LGS Puan Düzeylerine Göre B Formunda Kız Öğrencilerin Puan Dağılımları İçin Güven Aralıkları.....	136
Şekil 40. LGS Puan Düzeylerine Göre B Formunda Erkek Öğrencilerin Puan Dağılımları İçin Güven Aralıkları	136
Şekil 41. LGS Puan Düzeylerine Göre Kız Öğrencilerin Formlardaki Puan Dağılımı	138
Şekil 42. LGS Puan Düzeylerine Göre Erkek Öğrencilerin Formlardaki Puan Dağılımı.....	140
Şekil 43. LGS Puan Düzeylerine Göre Kız ve Erkek Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi	142
Şekil 44. İki Ortak Değişkene Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları	143

KISALTMALAR

LGS	Liselere Giriş Sınavı
DFA	Doğrulayıcı Faktör Analizi
EG	Eş Değer Gruplar Deseni
NEAT	Denk Olmayan Gruplarda Ortak Madde Deseni
NEC	Denk Olmayan Gruplarda Ortak Değişken Deseni
KTK	Klasik Test Kuramı
MTK	Madde Tepki Kuramı
PRE	Yüzde Görelî Hata (Percent Relative Error)
SEE	Eşitlemenin Standart Hatası (Standard Error of Equating)
BIAS	Eşitleme Yanlılığı
RMSE	Eşitleme Hatası (Root Mean Squared Error)
SEED	Eşitleme Farkının Standart Hatası (Standard Error of Equating Difference)
RMSD	Eşitleme Hatası (Root Mean Squared Difference)
PoB	Parametrik Olmayan Bayes
DP	Dirichlet Süreci
DDP	Bağımlı Dirichlet Süreci
BP	Bernstein Polinomları
MCMC	Markov Zinciri Monte Carlo

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problem durumu, amaç, önem, sayılılar ve sınırlılıklara ilişkin bilgiler yer almaktadır.

Testler bir toplumun eğitim sisteminin ürünlerini değerlendirirken kullanılan önemli parametrelerden biri haline gelmiştir. Boehm ve White (1982) bu durumu şu şekilde açıklamaktadır: “Testler mükemmel olmaktan uzaktır fakat yine de öğrencileri değerlendirmek için daha adil bir yol yoktur... Test puanları çocukların okuldaki yeteneği hakkında elde edilebilecek en nesnel kanıttır.” Testler çoğu modern toplumda okullarda, endüstrilerde ve devlet kurumlarında en objektif karar verme yöntemi olarak kabul edilmektedir. Testler değerlendirme, yerleştirme, rehberlik, işe alma, terfi, araştırma ve öğretim amacıyla kullanılmaktadır. Testin özü sınava giren kişinin gizil yeteneğini ortaya çıkarmaktır (Emaikwu, 2014).

Daha önce öğrenci, öğretmen, koç veya ebeveyn olan herkes bireylerin yeni bilgileri öğrenme hızlarına göre farklılık gösterdiğini fark etmiştir. Bazı insanlar “hızlı öğrenir”, bazıları “yavaş öğrenir” ve çoğu ise bu ikisinin arasında bir yerdedir. Yeteneklerdeki bireysel farklılıkları bilimsel olarak ölçme çabalarına yol açan durum bu basit gözlemdir (Stemler ve Sternberg, 2013). Psikologlar 20. yüzyıl boyunca yetenek kavramını ölçmekten ziyade bu kavramı tanımlamak için mücadele etmişlerdir (Press, 2014).

Karmel ve Karmel (1978) bir yetenek testinin amacını şu şekilde açıklamıştır: “öğrenilenleri değil, öğrenilebilecekleri ölçmek.” Yani yetenek testleri öğrencilerin gelecekteki performanslarını tahmin etmek için kullanılır. Bir fırsat verildiğinde öğrencilerin potansiyel kapasitelerini, öğrenme, bilgi ve beceri edinme yeteneklerini ölçerler (Joint Committee on Standards for Educational Evaluation, 2003). Snow (1992) yetenek olarak kabul edilen özelliklerin ilgi uyandıran özellikler olması gerektiğini fakat kolaylıkla değiştirilebilen özellikler olmaması gerektiğini belirtmiştir. Diğer yandan yeteneklerin sabit olduğu varsayımının olmadığını eklemiştir. Snow (1992) yetenek kavramının bilişsel yetenekle sınırlı olmadığını güçlü bir şekilde savunmuştur. Kişilik, motivasyon ve benlik kavramı gibi diğer yönlerin de yeteneğin önemli bileşenleri

olduğunu belirtmiştir. Sonuç olarak, Snow’a göre yetenek sadece bilişsel faktörlerden değil duyuşsal ve konatif süreçlerden de oluşur. Duyuşsal bileşenler, duyguları ifade ederken; konatif bileşenler, hedef belirleme ve irade ile ilgili çaba gerektiren süreçlerdir. Yeteneğin bilişsel, duyuşsal ve konatif yönlerden oluştuğu fikri, en erken yetenek testi kavramlarından bile daha fazla teorik çekiciliğe sahiptir (Stemler ve Sternberg, 2013). Snow (1996) ise yeteneği bireyin hem belirli bir performans için hazır olma durumu hem de o performans için uygunluğu şeklinde açıklamıştır. Uygunluk kişi-durum etkileşimidir, yani durumların kişilerin getirdiği donanımına ne ölçüde uyduğudur. Cohen ve Swerdlik (2009) yetenek testlerinin formel olmayan öğrenmelere ve yaşam deneyimlerine yönelik olduğunu ve hem yordama hem de hazır bulunuşluk testi olarak kullanılabildiğini belirtmiştir. Emaikwu (2014) yeteneğin hem doğuştan gelen kapasiteyi hem de çevrenin birey üzerindeki etkilerini içerdiğini ifade etmiştir. Yetenek terimi yerine kullanılan “aptitude” ve “ability” kavramları arasında da bulanıklık söz konusudur. Geleneksel bir ayrım “ability” doğuştan gelen özellikler olarak tanımlanırken, “aptitude” nispeten kontrolsüz ve bilinmeyen koşullar altında gerçekleşen formel öğrenmeleri içermektedir (Muchinsky, 2009).

Psikolojik test tarihinde yetenek kavramı 17. ve 18. yüzyıl boyunca giderek zekâ kavramıyla eşitlenmiştir (Snow, 1992). Aslında zekâ testleri, Alfred Binet’in zekâ teorisine dayanan Stanford Binet Testi gibi erken yetenek testleri ile bir yüzyıldan fazla süre akademik yeteneğin standart bir ölçüsü olmuştur. Bununla birlikte “zekâ” terimi değişmez bir özellik taşıdığı için “yetenek testleri” artık “zekâ testleri” olarak adlandırılmamakta ve kesin bir şekilde “zekâ testleri” olarak yorumlanmamaktadır (Popham, 2003). Carroll (1993) hiyerarşik bir zekâ modeli tanımlamıştır. Çeşitli yetenek ölçülerinin altında yatan yapı bu çalışmada bulunabilir. Carroll üç seviyeden oluşan bir yapı tanımlamıştır. Genel zekâ hiyerarşinin en üstündedir; ikinci seviye akıcı ve kristalize zekâ, hafıza, geri getirme, görsel ve işitsel algı, bilişsel hız ve işlem hızından oluşur; bireysel yetenek ve başarı testleri ise üçüncü seviyededir.

Yeteneğin tanımlanmasına ve değerlendirilmesine, pek çok meslek ve alan katkıda bulunmuş olsa da özellikle üç alan bu yapıya kalıcı bir ilgi göstermiştir (Stemler ve Sternberg, 2013). Bu alanlar (a) üniversiteye giriş sınavları, (b) yetenek belirleme ve personel seçimi konusunda istihdam testleri ve (c) yetenek-müdahale etkileşimi (Aptitude-treatment interactions) ve dinamik değerlendirme teknikleri ile ilgili olan sınıf değerlendirmesidir.

Amerika Birleşik Devletleri'nde (ABD), 1925 yılında Kolej Kurulu o zamana kadar ortaya çıkan zekâ testlerinden yararlanarak, temel olarak genel akıl yürütme yeteneğinin test edilmesine dayanan standartlaştırılmış bir Akademik Yetenek Testi'ni (Scholastic Aptitude Test – SAT) kolejlere sunmuştur. SAT, bir kişide bulunan bilişsel yetenekler yelpazesinin dar bir kısmına (sözel ve sayısal muhakeme) dayanmaktadır (Stemler ve Sternberg, 2013). Testin “aptitude”, “ability”, başarı ve bunların kombinasyonunu ölçüp ölçmediği konusunda birkaç yıl süren tartışmalardan sonra testin başlığı Scholastic Assessment Test olarak ve daha sonra önceki adları terk edilerek sadece SAT olarak değiştirilmiştir (Katz ve Brown, 2019). Üniversiteye kabul sürecinde yaygın olarak kullanılan diğer sınav Amerikan Kolej Testi'dir (American College Testing – ACT). Kapsam geçerliği üzerine yapılan önemli çalışmalardan sonra “No Child Left Behind Act” (2001) mevzuatı sonrasında çok sayıda eyalet tarafından benimsenmiştir (Lissitz ve Samuelson, 2007). SAT, üç çeyrek yüzyılı aşkın bir süredir lise öğrencilerinin sözel ve sayısal yeteneklerini ölçmeye odaklanmıştır. Yarım yüzyıldan fazla bir süredir var olan ACT, genel yetenekten ziyade konu bilgisine yaptığı vurguyla SAT'tan farklıdır (Popham, 2006). Sternberg ve meslektaşları (Sternberg ve the Rainbow Project Collaborators, 2006) standart SAT'a ek olarak bu özelliklerin en azından bazılarını içeren yeni bir değerlendirme geliştirmiştir. Amaçları, üniversiteye kabul için teorik temelleri olan ve şu anda SAT tarafından değerlendirilmekte olandan daha geniş beceri yelpazesini ölçen bir yetenek testi geliştirmektir. Ayrıca, bireylerin test bağlamında güçlü yanlarından yararlanmalarına ve zayıflıklarını telafi etmelerine olanak tanıyacaktır. Testleri, Rainbow Projesi olarak bilinen projenin bir parçası olarak geliştirmişlerdir. Rainbow Projesinin amacı SAT'ın yerini almak değil SAT'ın ölçmediği bilişsel becerileri ölçen SAT'ı destekleyecek testler tasarlamaktır.

SAT ve ACT'nin kullanımıyla birlikte yıllar içinde ortaya çıkan geçerlilik endişeleri ve test adaleti konusu hararetli bir şekilde tartışılmaktadır. Lisansüstü Kayıt Sınavı (Graduate Record Examination – GRE), Lisansüstü Yönetim Giriş Sınavı (Graduate Management Admission Test – GMAT), Tıp Fakültesi Giriş Sınavı (Medical College Admission Test – MCAT) ve Hukuk Fakültesi Giriş Sınavı (Law School Admission Test – LSAT) gibi üniversiteye giriş değerlendirmelerinin geçerliliği ve adilliği ile ilgili endişeler devam etse de bu standartlaştırılmış değerlendirmeler lisansüstü ve profesyonel programlar da dahil olmak üzere yüksek öğretim kurumlarının büyük çoğunluğu tarafından kullanılmaya devam etmektedir (Katz ve Brown, 2019).

Yetenek deęerlendirmesinin gl bir tarihsel baęa sahip olduęu bir dięer alan ise istihdam testleri, yani personel seimidir. Bu baęlamda, zellikle dikkat ekici olan iki alan ordu ve zel sektrdr. Birinci Dnya Savaşı lme hareketinde bymeyi teŗvik eden nemli bir faktrdr. Byk insan gc seferberlięi ile ordu, hangi adamların hangi hizmet dallarına uygun olduęunu hızlı ve kolay bir řekilde belirlemenin bir yoluna ihtiya duymuŗtur. Grup zek testleri bu ihtiyaı karŗılamak iin geliŗtirilmiŗtir. ABD Ordusu o zamanki Amerikan Psikoloji Derneęi baŗkanı Robert M. Yerkes'i (1876-1956), subay eęitim programında baŗarılı olma olasılıęı en yksek olan acemileri belirlemeye yardımcı olacak bir test tasarlaması iin grevlendirmiŗ ve "Ordu Alfa" testi geliŗtirilmiŗtir. Alfa hem bir grup zek testi hem de bir yetenek testi veya "tahmin testi"dir. nk subay eęitim programında iŗe alınanların olası baŗarısını doęru bir řekilde tahmin edebilmektedir (Press, 2014). ABD'de 1917 yılında geliŗtirilen Ordu Alfa ve Beta testlerinden sonra daha farklı bir oklu yetenek lm olan Silahlı Hizmetler Mesleki Yetenek Bataryası'na (Armed Services Vocational Aptitude Battery – ASVAB) geilmiŗtir. ASVAB 1970'lerde tanıtılmıŗ ve Kara, Deniz ve Hava Kuvvetleri tarafından kullanılmıŗtır. Bylece kiŗi ile mesleęinin talepleri arasındaki uyuma dikkat etmeyi gerektiren bir yapı olarak yetenek kavramı, yeteneęin askeri tanımına egemen olmuŗtur (Katz ve Brown, 2019). Mesleki hazırlık ve personel seimi ile ilgili olarak; okullar, rehabilitasyon merkezleri ve insan kaynakları ynetimi tarafından eŗitli yetenek ltleri kullanılmaktadır. Yetenek belirlemenin temel amacı, bir organizasyonda alıŗanların gl ve zayıf ynlerini tespit ederek farklı iŗ rollerini yerine getirmelerini saęlamaktır. Bylece sistemin genel potansiyelini artırmak amalanmaktadır (Stemler ve Sternberg, 2013).

Yetenek deęerlendirmesinin zerinde durduęu nc alan sınıf deęerlendirmesi olmuŗtur. Bu alanda yetenekle ilgili iki araŗtırma yetenek-mdahale etkileŗimi ve dinamik deęerlendirmelerdir. ęrenciler analitik dŗnme, yaratıcılık ve eleŗtirel dŗnme gibi farklı biliŗsel becerilerin farklı kombinasyonlarına sahiptir. ęrencilerin farklı ęrenme yntemleri bulunduęu gibi ęretmenlerin de farklı ęretme yntemleri bulunmaktadır. Sternberg, Grigorenko, Ferrari ve Clinkenbeard'ın (1999) araŗtırmasında yetenek kalıplarına daha iyi uyan ęretim koŗullarına yerleŗtirilen ęrencilerin, ęretim koŗulları uymayan ęrencilerden daha iyi performans gsterdięi bir yetenek-mdahale etkileŗimi olduęu grlmŗtir. Dięer bir deyiŗle, ęrencilere en azından ęrenme zamanının bir blmnde kendi ęrenme stillerine uygun bir řekilde ęretildięinde okulda daha baŗarılı olmuŗlardır. Dinamik deęerlendirmenin temel dayanaęı ise, kiŗiye

sadece bir test uygulayıp bu test puanını yorumlayarak bu kişinin yeteneğinin gerçekten anlaşılamayacağı veya değerlendirilemeyeceğidir. Genel olarak, bu alandaki araştırmacılar yeteneği, büyük ölçüde şekillendirilebilir ve duruma özgü bilişsel bir kapasite olarak tanımlama eğilimindedir (Stemler ve Sternberg, 2013). Dinamik değerlendirmede psikologların ve eğitimcilerin başlıca ilgisi, sınava girenlerin önceki eğitim deneyimleri göz önüne alındığında şimdi nerede oldukları değil, bundan sonra kendilerine yeterli eğitimsel müdahale yapıldığında yarın nerede olabilecekleridir (Grigorenko, 2009).

Ülkemizde de tanı ve yönlendirme amaçlarıyla sıklıkla kullanılan yetenek testleri mevcuttur. Bunlardan Temel Kabiliyetler Testi (TKT) Thurstone tarafından geliştirilmiş olup şekil, kavrama, sözcüklerle akıl yürütme, sayı ve sözcük çağrışım testlerinden oluşmaktadır. İlk uyarılama çalışması 1953 yılında yapılan testin son uyarlaması 2003 yılında yapılmıştır (Kıran, 2020). Milli Eğitim Bakanlığı Test ve Araştırma Bürosu'nda tercüme edilmiş olan aracın 5 – 7 yaş formu Rehberlik Araştırma Merkezlerinde büyük gruplara uygulanarak, çalışmayan maddeler çıkarılmış ve Türk çocuklarına uygun hale getirilip normları belirlenmiştir. Bu testin 7 – 11 yaş formu ilköğretim, 11 – 17 yaş formu liselerde rehberlik çalışmalarında kullanılabilecek nitelikte ise de bugüne kadar sadece bazı araştırmalarda kullanılmıştır (Kuzgun, 2017). Ayırıştırıcı Yetenek Testi (AYT) Bennet ve meslektaşları tarafından geliştirilmiştir. Sözel akıl yürütme, sayısal beceri, soyut düşünme, uzay ilişkileri yeteneği, mekanik akıl yürütme, büro işleri yeteneği, imla ve dil becerisi olmak üzere sekiz alt testten oluşmaktadır. Test bazı araştırmacılar tarafından küçük gruplar üzerinde güvenilirlik ve geçerliği saptandıktan sonra meslek rehberliği araştırmalarında kullanılmıştır. Ancak testin Türkiye için geçerli normları bulunmamaktadır (Kuzgun, 2017). Genel Yetenek Test Bataryası (GYTB) 1947 yılında ABD iş ve işçi bulma kurumu tarafından geliştirilmiştir (Kıran, 2020). Batarya dokuz yeteneği ölçen on iki bölümden oluşmaktadır. İş Kurumu şubelerinde iş için başvuranlardan meslek eğitimi görmesi gerekenleri uygun meslek programına yönleltmek amacıyla bu test bataryasından yararlanılabileceği düşünülmüştür. Aracın Türkiye'ye kazandırılmasına yönelik çalışmalar 1996 yılında başlamıştır. Kendini Değerlendirme Envanteri (KDE) öğrencilerin yetenekleri, ilgileri ve meslek değerleri ile ilgili algılarını ölçmek amacıyla Kuzgun tarafından geliştirilmiştir. Envanterin üç alt ölçeği sözel, sayısal ve şekil uzay yeteneğini ölçen yetenek testidir. Lise öğrencileri için geliştirilmiştir (Öner, 1997). KDE üniversite eğitimi düşünen lise öğrencileri için geliştirildiğinden Kuzgun tarafından daha küçük sınıflardaki öğrencilere yönelik Akademik Benlik Kavramı Ölçeği

(ABKÖ) geliştirilmiştir. Bu ölçekte KDE'nin üç yetenek ölçeğine el-göz koordinasyonu yeteneği de eklenmiştir (Kuzgun, 2017).

Boehm ve White (1982) testlerin mükemmel olmadığını ama öğrencileri adil bir şekilde değerlendirmemizi sağladığını belirtmiştir. Bu sebeple testler geçmişten günümüze kadar hep kullanılmıştır. Önemli olan testlerin hem geliştirilmesinde hem de uygulanmasında gereken önemin verilmesidir. Hazırlanan bir ölçme aracının geçerli ve güvenilir olmasının yanında uygulamada test güvenliğinin de sağlanması gerekmektedir. Uygulanan testlerde güvenliği sağlamak amacıyla farklı önlemler alınmaktadır. Örneğin SAT uygulanırken testteki soruların sırası değiştirilir, böylece arka arkaya oturan hiçbir cevaplayıcıya aynı kitapçık verilmez. Uygulanan bir testte güvenliği sağlamak amacıyla eşdeğer olacak şekilde farklı formlar da kullanılabilir. Fakat bir testin paralel formlarını oluşturmak oldukça güçtür. Test formlarında bulunan soruların farklı güçlükte olması, içeriğin farklılaşması ve psikometrik özelliklerinin değişmesi paralelliği bozan etkenlerdir. Bu farklılık ise test sonuçlarına göre bireyler hakkında karar verirken adaletsizliğe sebep olmaktadır. Bu durumda, benzer özellikleri olan test formlarını inceleyerek, bunlardan elde edilen puanların birbirlerinin yerine kullanılmasına imkân sağlayan test eşitleme süreci (Kolen ve Brennan, 2014) gerekmektedir. Eğer test geliştirme süreci mükemmel olsaydı eşitleme gerekli olmazdı (von Davier, 2011). Bu sebeple puanlar arasında adil bir şekilde karşılaştırılabilirliği sağlamak amacıyla test formları arasında eşitleme yapılması gerekmektedir.

Problem

Livingston (2014) birçok insana, test puanlarını eşitlemek gibi alışılmadık bir konunun mantığını anlamak için para gibi tanıdık bir konuyla başlamalarını önermiştir. Bir şeyin bir yıldaki fiyatını başka bir yıldaki fiyatıyla nasıl anlamlı bir şekilde karşılaştırabilirsiniz? Ekonomistler “sabit dolar” denilen bir parametre kullanmaktadır. Amerika’da her yıl hükümet ekonomistleri, tipik bir Amerikan ailesinin bir yılda satın aldığı şeyleri temsil eden belirli bir ürün seçiminin maliyetini hesaplar. Ekonomistler bu ürün karışımını “pazar sepeti” olarak adlandırmaktadır. Referans yılı olarak bir yılı seçerler. Daha sonra diğer yıllardaki pazar sepetinin maliyetini referans yıldaki maliyetiyle karşılaştırırlar. Bu analiz, onların diğer yıllardaki ücretleri ve fiyatları referans yılındaki dolar cinsinden ifade etmelerini sağlar. Ortalama öğretmen maaşlarında

1958'dekini 1998'deki ile karşılaştırmak için her iki maaşı da referans yılı dolarına çevirmeleri gerekmektedir. Şimdi, tüm bunların eğitim testiyle ne ilgisi var?

Eğitimde testlerin güvenliğini sağlamak amacıyla farklı formlar kullanılmaktadır. Testin uygulandığı aynı sınav salonunda farklı formlar ya da yıl içerisinde birden fazla uygulanan sınavlarda farklı sürümler kullanılmaktadır. Bu farklı formlar aynı yapıyı ölçmek için tasarlansa da psikometrik özellikleri bir şekilde farklılık gösterir. Örneğin bir form diğerinden daha zorsa bu durumda zor olan formda öğrencilerin daha düşük puan alması beklenir. Puan eşitleme test formlarındaki bu istenmeyen farklılıkların puanlar üzerindeki etkilerini ortadan kaldırmayı amaçlar. Puan eşitleme sınava girenlere karşı adil olmak ve puan kullanıcılarına testin farklı formları veya sürümlerinde aynı anlama gelen puanlar sağlamak için gereklidir (von Davier, 2011).

Testleri eşitleme nedenleri durumdan duruma değişiklik gösterse de genellikle üç kategoriye ayrılabilir (Zhu, 1998). Testleri eşitlemenin ana nedenlerinden biri aynı testi tekrar tekrar uygulamak zorunda kalmadan, paralel testlerden elde edilen puanların doğruluğunu incelemek ve test geliştirmektir. İkinci neden test puanlarının birbiri ile değiştirilebilirliğini sağlamaktır. Pratikte aynı değişkeni ölçmek için birden fazla test kullanılırken, bu testlerden alınan puanlar, farklı ölçeklerde oluşturuldukları için genellikle birbirleriyle değiştirilemez. Sınava girenleri karşılaştırmak veya eşdeğer kriterler belirlemek için öncelikle test puanlarının aynı ölçeğe yerleştirilmesi gerekir. Test eşitlemenin üçüncü nedeni ise test sürekliliğini sağlamaktır. Uygulamada bir yetenek veya özelliğin büyümesini veya değişimini ölçmek için farklı yetenek seviyelerinde birden fazla test kullanılmaktadır. Bu durumda, testlerin zorluğu genellikle kasıtlı olarak farklı tasarlanır. Bununla birlikte, sadece test zorluğunu değiştirmek büyümeyi veya değişimi ölçmeyi garanti etmez. Bunu yapmak için, testlerin yine aynı ölçekte farklı seviyelere ayarlanması gerekir.

Eşitleme işlemindeki en önemli adım, eşitleme deseninin ve yönteminin en az hatayı oluşturacak şekilde seçilmesidir. Test puanlarındaki farklılıklar yalnızca test formlarının zorluğundaki farklılıklardan kaynaklanmaz. Bu zorluğun sınava giren bireylerin yetenek farklılıklarıyla karıştırılmasını önlemek gerekmektedir. Bir testin iki formu dikkate alındığında veriyi toplamanın farklı yolları vardır: Tek grup deseninde aynı popülasyondan sınava giren örneklem grubuna her iki test formu da uygulanır. Eş değer gruplar (EG) deseninde aynı popülasyondan alınan iki gruba test formunun yalnızca biri uygulanır. Dengelenmiş tek grup deseninde aynı popülasyondan alınan iki gruba her iki test formu farklı sırayla uygulanır. Bu desenlerin ortak noktası aynı popülasyondan

örneklem alınmış olmasıdır. Ancak farklı popülasyonlardan örneklem almanın mümkün olduğu desenler de vardır. Denk olmayan gruplarda ortak madde (NEAT) deseninde iki farklı popülasyondan iki örnek alınır. Her grup sadece bir test formu alır ve her iki gruba da ortak bir test formu uygulanır. Ortak test, testin kısaltılmış bir versiyonudur ve sınava girenlerin yeteneklerindeki farklılıkları ölçmek ve kontrol etmek için kullanılır. Bu desene yönelik potansiyel bir tehdit, sınava girenlerin ortak maddeleri tanımaları ve bu tür maddeleri testin geri kalanı kadar ciddiye almamaları gerçeğidir (Gonzalez ve Wiberg, 2017). Ayrıca standart testler kullanan bütün ölçme programlarının ortak testlere sahip olmaması bu desenin bazı durumlarda kullanılmasını imkânsız hale getirmektedir. Testlerde ortak madde bulunmadığı zaman denk olmayan gruplarda ortak değişken (NEC) deseni önemli bir alternatiftir. Bu desende ortak maddelerin yokluğunda testleri bağlamak için ortak değişkenler kullanılır.

Cevapların doğru veya yanlış olarak kodlandığı çoktan seçmeli testler için ölçme programlarında en sık kullanılan test puanlarından biri, testteki doğru cevap sayısına karşılık gelen toplam puanlardır. Toplam puanların alabileceği olası değerler ardışık tamsayılar olduğundan eşitleme fonksiyonu ile ilgili bariz bir sorun puan dağılımlarının ayrık olması ve bunlara karşılık gelen puanları kullanılmaz hale getirmesidir. Eşitleme literatüründe bu sorunun ortak çözümü ayrık puanları eşitleme için kullanılmak üzere “süreklileştirme (continuize)”dir. Farklı süreklilik yöntemleri eşitleme fonksiyonu için parametrik, yarı parametrik veya parametrik olmayan istatistiksel çıkarımlar üreten farklı eşitleme yöntemleri tanımlar (Gonzalez ve von Davier, 2013).

Yapılan sınavlarda test puanlarının yanı sıra çoğu zaman sınava giren öğrencilerin cinsiyet, okul türü, uygulama zamanı vb. gibi arka plan değişkenlerine ulaşılabilmesine rağmen geleneksel eşitleme yöntemlerinde eşitleme fonksiyonunu tahmin etmek için bu bilgiler kullanılmamaktadır. Bununla birlikte faktör değişkenleri, bir eşitleme fonksiyonunu tahmin etmek için verilerin (puanların) toplandığı teste girenlerin gruplarını tanımlamak için yaygın olarak kullanılır. Örneğin EG deseninde, sınava giren grup ikiye bölünür ve her yarıya iki farklı test formundan biri uygulanır, böylece faktör değişkeni uygulanacak formu tanımlar. Ayrıca ortak değişkenlerin kullanımı, eşitleme literatüründe keşfedilmemiş bir konu gibi görünmektedir (Gonzalez, Barrientos ve Quintana, 2015a).

Test puanlarını eşitlemek amacıyla kullanılan yöntemler farklı kuram ve varsayımlara dayanmaktadır. Yaygın uygulamalarda kullanılan klasik eşitleme yöntemleri eşitleme dönüşüm fonksiyonunu tahmin etmek için test puanları dışında hiçbir

ek bilginin kullanılmadığı parametrik veya yarı parametrik yöntemlerdir. Ancak son yıllarda eşitleme dönüşümüne yol açan puan dağılım fonksiyonlarının tahmininde ortak değişkenlerin de kullanımına izin veren yöntemler geliştirilmiştir. Bu bağlamda test eşitlemede ortak değişkenlerin kullanımına izin veren esnek bir Parametrik olmayan Bayes (PoB) modeli önerilmiştir (von Davier, 2011). Bu eşitleme yönteminden elde edilen eşitlenmiş puanların, klasik yöntemlere göre farklılık gösterip göstermeyeceği araştırma konusudur.

Amaç

Bu araştırmada dokuzuncu sınıf öğrencileri için paralel formları olan bir yetenek testi geliştirilmesi ve testin paralel formlarının farklı eşitleme desenlerinde farklı eşitleme yöntemleri kullanılarak eşitlenmesi amaçlanmaktadır. Ayrıca eşitleme sürecinde klasik yöntemlerle yapılan eşitlemelerin PoB yöntemiyle yapılan eşitlemeyle karşılaştırılması amaçlanmaktadır. Alt amaçlar şu şekilde ifade edilebilir:

1. Yetenek testinin paralel formlarından elde edilen puanların,
 - a) EG deseninde eşit yüzdelikli eşitleme yöntemi ile eşitlenmesi sonucunda elde edilen hata değerleri ne düzeydedir ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
 - b) NEAT deseninde eşit yüzdelikli frekans tahmini ve eşit yüzdelikli zincirleme eşitleme yöntemleri ile eşitlenmesi sonucunda elde edilen hata değerleri ne düzeydedir ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
2. Yetenek testinin paralel formlarından elde edilen puanların,
 - a) EG deseninde Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri ile eşitlenmesi sonucunda elde edilen hata değerleri ne düzeydedir ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
 - b) NEAT deseninde Kernel eşitleme yöntemlerinden son tabakalama eşit yüzdelikli, zincirleme eşit yüzdelikli, son tabakalama lineer ve zincirleme lineer yöntemleri ile eşitlenmesi sonucunda elde edilen hata değerleri ne düzeydedir ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?

3. Yetenek testinin paralel formlarından elde edilen puanların NEC deseninde, cinsiyet; Liselere Giriş Sınavı (LGS) puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri ile eşitlenmesi sonucunda elde edilen hata değerleri ne düzeydedir ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
4. Yetenek testinin paralel formlarından elde edilen puanların NEC deseninde, cinsiyet; LGS puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak PoB modeli ile eşitlenmesi sonucunda;
 - a) Cinsiyet değişkeninin alt kategorilerinden elde edilen eşitlenmiş puanlar arasında; eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
 - b) LGS puanlarının alt kategorilerinden elde edilen eşitlenmiş puanlar arasında; eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?
 - c) Cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alındığında değişkeninin alt kategorilerinden elde edilen eşitlenmiş puanlar arasında; eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunmaktadır?

Önem

Akademik Personel ve Lisansüstü Eğitim Giriş Sınavı (ALES), Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS) ve Yükseköğretim Kurumları Yabancı Dil Sınavı (YÖKDİL) gibi ülkemizde ulusal düzeyde yapılan sınavların farklı zamanlarda farklı formları ile tekrarları yapılmaktadır. Test puanı eşitleme, bir testin farklı formlarından alınan test puanlarının karşılaştırılabilir olmasını sağlamak için kullanılan istatistiksel süreçtir. Test formlarında ortak test katılımcılarına veya ortak maddelere erişim olup olmadığına bağlı olarak bir dizi veri toplama deseni ve test eşitleme yöntemi bulunmaktadır (Wiberg, 2015). Bu çalışmada ortak değişken kullanılarak eşitlemenin yapılması amaçlanmaktadır. Ortak değişken kullanımına izin veren esnek bir yaklaşım olarak tanıtılan PoB yöntemiyle yapılan eşitlemenin farklı yöntemlerde ve desenlerde yapılan eşitlemelerle karşılaştırılması amaçlanmaktadır.

Farklı şekillerde eşitlenen test puanlarında, ortak değişkenleri kullanan üç bağımsız çalışma 2015 yılında yayımlanmıştır. Wiberg ve Branberg (2015) tarafından yapılan ilk çalışmada, denk olmayan gruplarda ortak değişkenler adı verilen yeni bir desen aracılığıyla test eşitlemeyi geliştirmek için ortak değişkenler kullanılmıştır. Bu desende, sınava girenler sınav puanlarıyla yüksek düzeyde ilişkili olan bir dizi ortak değişken aracılığıyla kategorize edilir ve bu kategoriler ortak maddelerle benzer şekilde kullanılır. Haberman (2015) tarafından yapılan ikinci çalışmada, teste girenlerin arka plan bilgileri kullanılarak sözde eşdeğer gruplar oluşturulmuştur. Üçüncü çalışmada Longford (2015) tarafından nedensel çıkarım üzerine kurulu bir test puanı eşitleme yöntemi önerilmiştir. Tüm bu yöntemlerde ana bileşen, ortak değişkenlerin kullanılmasıdır. İlginç bir şekilde, bu üç çalışmanın tümü 2015 yılının başında birbirine atıfta bulunmadan yayımlanmıştır. Schultz ve Schultz (2007) bir döneme damgasını vuran düşüncelerin, inançların, fikirlerin veya entelektüel eğilimlerin oluşturduğu genel çerçeveyi bir başka deyişle çağın genel düşünsel iklimini “Zeitgeist” kavramı olarak açıklamıştır. Toplum ve çevresel koşullar hazır olduğunda dönemin entelektüel eğilimlerinin ortaya çıktığını belirtmiştir. Test eşitlemede aynı dönemde birbirinden habersiz çalışmalarda ortak değişkenlerin kullanılmaya başlanması eşitlemenin Zeitgeist’i olarak nitelendirilebilir. Wiberg (2015) gelecek yıllarda test puanının ortak değişkenlerle eşitlenmesi konusunda daha fazla araştırma yapılmasını beklediğini belirtmiştir.

Geleneksel eşitleme yöntemlerinde parametrik ve yarı parametrik yöntemler kullanır (Gonzalez vd., 2015a). Bu çalışmada esnek bir yaklaşım olan PoB modeli ile eşitleme yapılmıştır. Eşitleme için modelin avantajı da diğer eşitleme yöntemlerinde olduğu gibi bazı parametrelerin üretilme zorunluluğunun olmamasıdır (Gonzalez, Barrientos ve Quintana, 2015b). Bu araştırma, diğer istatistiksel kestirim yöntemlerine göre daha kesin kestirimler yaptığı belirtilen Bayes yaklaşımının test eşitlemede ve eğitimde kullanılması açısından önemlidir.

Literatürde gerçek veri kullanılan ve farklı örneklem büyüklüklerine sahip eşitleme çalışmalarının genellikle Klasik Test Kuramına (KTK) dayandığı görülmektedir (Kazanç, 2019; Özdemir, 2017; Skaggs, 2005; Tan, 2015). Kuramdaki eşitleme koşullarının ve varsayımlarının sağlanması daha zor olduğu için daha büyük örneklemlemlerle yapılan ya da simülasyon verisi kullanılan eşitleme çalışmalarının ise Madde Tepki Kuramına (MTK) dayandığı görülmektedir (Brossman ve Lee, 2013; Gündüz, 2015; Kilmen, 2010; Uysal, 2014). Literatür taraması sonucunda test eşitlemede PoB modelini tanıtan Gonzalez ve diğerlerinin (2015a; 2015b) çalışmaları; bu modelin

performansını Madde Tepki Kuramına dayalı eşitleme yöntemleriyle karşılaştıran Yurtçu'nun (2018) çalışması incelenmiştir. Bu araştırmada geliştirilen bir yetenek testi üzerinde bu modelin performansı incelenmiş ve klasik eşitleme yöntemleriyle karşılaştırılmıştır. Bu araştırma, gerçek veri kullanılması ve alana formları eşitlenmiş bir yetenek testi sunması açısından önemlidir.

Sayıtlılar

Bu araştırmada;

- Öğrencilerin LGS puanları için verdikleri bilginin doğru olduğu varsayılmıştır.
- LGS puanlarının yeteneği ölçtüğü ve yetenek düzeyleri için geçerli bir ölçüt olduğu varsayılmıştır.
- Öğrencilerin testleri yanıtlarken en yüksek performanslarını sergilemeye çalıştıkları varsayılmıştır.

Sınırlılıklar

- Araştırma Antalya ilinde üç lise türünden (Anadolu Lisesi – Fen Lisesi – Meslek Lisesi) elde edilen veriler ile sınırlıdır.
- Araştırma eşit yüzdelikli, Kernel ve PoB eşitleme yöntemleri ile sınırlıdır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde eşitleme, eşitleme gereksinimleri, eşitleme basamakları, eşitleme desenleri ve eşitleme yöntemleri ile ilgili kuramsal açıklamalar ve araştırma problemlerine yönelik ilgili araştırmalar yer almaktadır.

Bağlama ve Eşitleme

Bağlama terimi, bir testteki puandan başka bir testteki puana dönüşümü tanımlamak için kullanılır ve bağlamanın özel bir türü eşitlemedir (Dorans, Moses ve Eignor, 2010). Holland ve Dorans (2006) test puanlarının bağlanması için tarihsel bir arka plan oluşturulmuştur. Holland ve Dorans (2006), Linn (1993) ve Mislevy (1992) daha önceki çalışmalara dayanarak bağlama yöntemleri için bir çerçeve belirlemiştir. Farklı bağlama yöntemleri arasındaki ayrımlar; testler arasındaki benzerlikler ve farklılıklar ile bağlama puanlarını kullanma ve yorumlama biçimleriyle ilgilidir.

Holland ve Dorans (2006) bağlama yöntemlerini tahmin, ölçekleme ve eşitleme olarak adlandırılan üç temel kategoriye ayırmıştır. Bu kategoriler arasında ayrım yapmak önemlidir çünkü aynı olmadıkları halde genellikle benzer veya özdeş olarak görülürler. Test uzmanlarının bu farklılıkları anlaması ve gerektiğinde bunları test kullanıcılarına iletebilmesi gerekir.

Tahmin

Tahmin, puan bağlamanın en eski şeklidir. Psikometrinin ilk günlerinden beri tahmin etme, eşitleme ile karıştırılmıştır. Embretson ve Reise (2000) “Lineer eşitlemede, bir test formundaki puanlar diğer test formuna bir regresyon modeli aracılığıyla dönüştürülür.” şeklinde yazmıştır. Burada amaç, sınava girenlerin diğer bilgilerine dayanarak sınavdaki puanlarını tahmin etmektir. Bu bilgi; diğer testlerden alınan puanlar, demografik bilgiler veya diğer bilişsel bilgi türleri (örneğin, seçilen derslerdeki notlar) dahil olmak üzere doğası gereği çok değişkenli olabilir (Dorans vd., 2010).

On dokuzuncu yüzyılın sonunda lineer regresyon yöntemleri çeşitli sosyal ve psikolojik olaylara da uygulanmıştır. Regresyonun etkilerini ilk kez gözlemleyen Galton'dur (Stigler, 1986). Bir testteki puanları tahmin etmek için lineer regresyon yöntemlerinin kullanılması puanları bağlamanın en eski yaklaşımıdır. Tahmin yöntemlerinin karşılaştırılabilir puanlar yaratmanın yolu olmadığı çok erken fark edilmiştir. Thorndike (1922) lineer regresyonun karşılaştırılabilir puanlar bulmak için neden tatmin edici bir yöntem olmadığına dair ilk argümanları vermiştir. Daha sonra Flanagan (1951) regresyon fonksiyonlarının simetri eksikliğini vurgulayarak bu yöntemlerin simetriklik şartını sağlamadığını belirtmiştir. Bu yöntemlerde tahmin edilen ile tahmin yapmak için kullanılan puanlar arasında her zaman bir asimetri vardır. Tahmin ve eşitleme arasındaki ayrım yıllar içinde tekrar tekrar doğrulanmıştır (Angoff, 1971; Holland ve Dorans, 2006; Linn, 1993; Mislevy, 1992).

Ölçekleme

Ölçekleme yöntemleri, bağlama yöntemlerinin en eski ikinci grubudur. Farklı testler üzerinde karşılaştırılabilir puanlar (yani ölçekleme) oluşturma ihtiyacı ve bunu yapmak için yöntemlerin icadı neredeyse psikometri alanının kendisi kadar eski bir tarihe sahiptir. Ölçekleme prosedürleri, başlangıçta karşılaştırılabilir puanlar oluşturma yöntemleri olarak adlandırılmıştır (Holland, 2007). Ölçeklemenin amacı, iki farklı testten alınan puanları ortak bir ölçeğe dönüştürmektir (Dorans, Moses ve Eignor, 2009). Ölçekleme eşitleme ile genellikle karıştırılır çünkü ölçekleme için kullanılan istatistiksel prosedürler testleri eşitlemek için de kullanılabilir (Dorans vd., 2010). Ölçekleme; batarya ölçekleme, çapa ölçekleme, dikey ölçekleme, kalibrasyon ve uyum gibi birçok alt kategoriye sahiptir. Ölçeklemenin alt kategorileri; testlerin farklı yapıları ölçtüğü durumlardan başlayarak testlerin benzer yapıları ölçtüğü durumlara kadar bir süreklilik oluşturur (Holland, 2007).

Batarya Ölçekleme (Farklı Yapılar ve Ortak Grup). Farklı yapıları ölçen iki veya daha fazla test ortak bir gruba uygulandığında her test için ölçek puanları ortak bir dağılıma sahip olacak şekilde dönüştürülebilir. Kolen (2004) bu durumu batarya ölçekleme olarak ifade etmiştir. SAT-Sözel (SAT-V) ve SAT-Sayısal (SAT-M) puanları, 1990 yılında test edilen bir öğrenci grubunda aynı dağılımı verecek şekilde yeniden

ölçeklenmiştir. Yeni ölçekleme ile 1990 yılında test edilen ve SAT-M puanı SAT-V puanından daha yüksek olan bir öğrencinin matematiksel kısımda sözel kısımdan daha iyi performans gösterdiği sonucuna varılabilir. Testler bu şekilde ölçeklenmediğinde bu tür çıkarımlar yapmak doğru olmaz (Holland, 2007).

Çapa Ölçekleme (Farklı Yapılar ve Farklı Gruplar). Farklı gruplara farklı yapıları ölçen iki veya daha fazla test uygulandığında ve bu testlerde ortak bir çapa ölçüsü bulunduğunda batarya ölçeklemesine önemli bir alternatif oluşturur. Bu yöntem çapa ölçekleme olarak adlandırılır. Dil sınavları bunun için iyi örnekler sunar. Diğer durumlarda sınava girenler okulda aldıkları derslere göre hangi sınava gireceklerini seçebilirler. Sınava girip farklı testleri alan gruplar genellikle eşdeğer değildir. Bu durumda çapa ölçüsü, ölçekleme için kullanılan bilgidir (Holland, 2007).

Dikey Ölçekleme (Benzer Yapılar ve Benzer Güvenilirlik, Ancak Farklı Güçlükler ve Farklı Gruplar). Okullarda farklı sınıflar için hedeflenen akademik konuların testleri, güçlük açısından farklı olması amaçlanan benzer yapıların testleri olarak görülebilir. Matematik veya okuma gibi konular birçok alt konuya ve boyuta sahiptir. Bu nedenle, farklı düzeyler için testlerle ölçülen yapılar biraz farklılık gösterebilir ancak testler genellikle güvenilirlik açısından benzerdir. Belirli bir konudaki ilerlemenin zaman içinde izlenebilmesi için bu tür testlerden elde edilen puanların ortak bir ölçeğe konulması istenir. Bu tür ölçeklemeye dikey ölçekleme denir (Kolen ve Brennan, 2014). Farklı şekilde de isimlendirilmiştir. Örneğin, Angoff (1971) farklı yetenek seviyelerinde testleri kalibre etme olarak adlandırmıştır ve dikey eşitleme terimi de kullanılmaktadır.

Kalibrasyon (Aynı Yapı, Farklı Güvenirlik ve Ortak Grup). Kolen ve Brennan (2014) test bağlama literatüründe kalibrasyon teriminin çeşitli anlamlarda kullanıldığını belirtmiştir. Kalibrasyonun MTK'daki standart kullanımı madde parametre tahminlerinin ortak bir ölçekte konumlanmasıdır (Lord, 1980). Linn (1993) kalibrasyonu tanımlarken, aynı yapıları ölçen ancak farklı istatistiksel özelliklere, özellikle farklı güvenilirlik veya güçlüğü sahip testler için bağlama yöntemlerine atıfta bulunmuştur. Bu çerçevede eşitleme, eşit veya en azından benzer güvenilirliğe sahip testler için ayrılmıştır. Holland'a (2007) göre kalibrasyon bir testin kısa formunun puanlarını uzun forma ölçeklemektir.

Uyumlar (Benzer Yapılar, Güçlük ve Güvenilirlik). Bazen birbirine bağlanacak testlerin tümü benzer yapıları ölçer ancak bunlar farklı özelliklere göre oluşturulur. Çoğu durumda test uzunluğu ve güvenilirliği benzerdir. Ayrıca genellikle benzer kullanımları vardır ve sınava giren kişiler tarafından aynı amaç için alınabilirler. Bağlamanın kullanımı, her iki testteki puanları sanki diğer testteki puanlarmış gibi ifade etmektir. Uyumlar, birbirine çok benzeyen ancak puanlarının birbirinin yerine kullanılacağı düşüncesiyle oluşturulmayan testlerin ölçeklenmesini temsil eder (Holland, 2007). Birçok kolej ve üniversite, kabul kararları için ACT veya SAT puanlarını dikkate alır ve bu kurumlar genellikle bu testlerden birinin sonuçlarını yorumlama konusunda diğerinden daha fazla deneyime sahiptir. Dorans, Lyu, Pommerich ve Houston (1997) bu iki testin puanlarını birbirine bağlayan bir uyum tablosu oluşturmuştur. Bu uyum, her iki testi de sınırlı bir zaman çerçevesi içinde almış olan 100.000'den fazla kişiden alınan verilere dayanmaktadır. Başvuranlar buna göre değerlendirilmiştir.

Eşitleme

Ordu Alfa ve Beta testleri kullanılarak 1918 yılının sonunda 1.7 milyondan fazla kişi test edilmiştir. Alfa İngilizce okuyup yazabilenler ve Beta ise bunları yapamayanlar için geliştirilmiştir. Test güvenliği sorunları alternatif formların kullanılmasına yol açmıştır ve Alfa'nın beş farklı formu oluşturulmuştur. Test maddelerinin formlara rastgele atanmasıyla güçlük bakımından benzer ancak aynı olmayan neredeyse paralel test formlarının kullanılması düşünülmüştür. Fakat Yoakum ve Yerkes bu beş formun tam olarak eşdeğer olmadığını ve B formunun daha zor olduğunu belirtmiştir. Ordu istatistikçilerini daha fazla endişelendiren şey, testin Alfa ve Beta sürümlerinde elde edilen puanların karşılaştırılabilirliğidir. Özel bir askeri personel örneği her ikisiyle de test edilmiş ve bu veriler Alfa ve Beta'yı ortak bir ölçeğe yerleştirmek için kullanılmıştır. Bu iki test, format ve sorulan sorular açısından oldukça farklı olduğundan bu test eşitlemeden ziyade batarya ölçekleme durumudur. SAT sınavının 1938 yılında iki formu oluşturulmuş ve 1940'ta bunları eşitleme ihtiyacı ortaya çıkmıştır. SAT için formların güçlüğüne etkisini ortadan kaldırmak amacıyla, 1941 yılında ortak testle eşitlemenin ilk sürümleri kullanılmıştır. SAT sözel ve matematik testleri 1942 yılında, Nisan 1941'de oluşturulan sözel teste bağlanmıştır. Tüm bağlama işlemlerinde, 1942 yılından sonrakiler eşitleme şeklinde olmuştur (Holland, 2007).

Eşitleme, bu çerçevedeki bağlama yöntemlerinin üçüncü kategorisidir. Tüm bağlama çerçeveleri, eşitlemeyi iki testteki puanlar arasında en güçlü bağlantı şekli olarak tanımlar. Eşitleme, bağlanan testler arasındaki ilişkiler hakkında hiçbir varsayımda bulunmayan yöntemlerle (tahmin ve batarya ölçekleme) başlayan ve çok benzer testleri birbirine bağlamak için uygun yöntemlere (uyumlar ve eşitleme) kadar ilerleyen bir sürekliliğin son noktasını temsil eder (Holland ve Dorans, 2006).

Eşitleme, bağlanan testlere çok güçlü gereksinimlerin yerleştirildiği bir ölçekleme biçimi olarak görülebilir (Dorans vd., 2009). Hambleton ve Swaminathan (1985) bir testten elde edilen puanları diğerinden elde edilen puanlara dönüştürmek veya ilişkilendirmek için kullanılan istatistiksel prosedür şeklinde tanımlamıştır. Angoff (1971) bir test formunun birim sistemini diğer test formunun birim sistemine dönüştürme süreci olarak tanımlamıştır. Angoff (1971) bu tanımlamayı açıklayabilmek için Celsius ve Fahrenheit ölçeklerinde kullanılan birim dönüştürmesini örnek olarak göstermiştir. Celsius ve Fahrenheit ölçekleri benzer bir yapıyı (sıcaklığı) ölçmektedir. Fakat, farklı ölçekler kullanıldığında farklı sonuçlar elde edilmektedir. Bu şekilde farklı sonuçlar elde edilmesinin sebebi ölçek birimlerinin farklı olmasıdır. Bu nedenle her ikisi de sıcaklığı ölçen ölçekler arasında doğrusal bir dönüşüm yapılmaktadır.

Eşitleme Gereksinimleri

Test eşitlemeyi diğer bağlama biçimlerinden ayıran şey amacıdır. Eşitlemenin amacı, iki testten alınan puanların herhangi bir amaç için birbirinin yerine kullanılmasına izin vermektir. Bu çok zorlu bir hedeftir ve deneyimler göstermiştir ki iki test ve bunları birbirine bağlamak için kullanılan yöntemler çok güçlü gereksinimleri karşılamalıdır (Holland, Dorans ve Petersen, 2007).

Teorik olarak test formları aynı psikometrik özelliklere sahip olacak şekilde oluşturulmalıdır. Böyle bir durum söz konusu olduğunda eşitleme gereksiz olur. Gerçekte paralel olan birden fazla test oluşturmak neredeyse imkansızdır ve bu sebeple eşitleme gereklidir (Dorans vd., 2010). Bir bağlamanın eşitleme olması için yaygın olarak bazı gerekliliklerin sağlanması gerekir (Hambleton ve Swaminathan, 1985; Holland ve Dorans, 2006; Kolen ve Brennan, 2014; Lord, 1980).

- a) **Eşit Yapı:** Eşitleme yapılacak testler için en önemli kural bu testlerin aynı özellikleri, aynı davranışları, aynı beceriyi, aynı yetenekleri ölçüyor olmasıdır. Bir

testin ölçtüğü yapıyı tanımlamak o kadar kolay değildir. Bununla birlikte bu gerekliliğin kontrol edilmesinin en yaygın yolu, test sorularının içeriğinin incelenmesidir (Dorans ve Holland, 2000).

- b) **Eşit Güvenirlik:** Testlerin güvenilirliklerinin eşit olması, bireylerin alacakları puanların hatadan arınık olarak ölçülebilmesi anlamına gelmektedir. Eğer eşitlenen testlerin güvenilirlikleri eşit değilse, daha fazla hata bulunduran testin puanları daha az hata bulunduran testin puanlarına eşit görünecektir. Bu durum güvenilir bir şekilde test geliştirmeye aykırıdır (Dorans ve Holland, 2000; Lord, 1980). Güvenirlik düzeyleri farklı olan sınavların puanlarını eşitlemede, yanlış puanlar eşitlenmeye çalışıldığından elde edilen sonuç doğru olmayacaktır.
- c) **Simetriklik:** Eşitleme işleminin hangi formun X, hangi formun Y olarak isimlendirildiğine bakılmaksızın aynı olmasıdır (Lord, 1980). Dorans ve Holland (2000) bu özelliği; eşitleme sayesinde Y puanından X puanını elde ettiğimiz gibi, X puanından Y puanını elde edebilmemiz şeklinde açıklamaktadır. Eşitlemede, yapılan dönüşüm işlemi ters çevrilebilmelidir. Aynı yapıyı ölçen iki test formu X ve Y olmak üzere, X formundan Y formuna puanları dönüştürürken kullanılan fonksiyon, Y formundan X formuna dönüştürürken de kullanılabilmelidir (Kolen ve Brennan, 2014).
- d) **Eşitlik:** Bir testin X veya Y formlarından hangisinin alındığının bir fark oluşturmaması gerekir (Dorans vd., 2010). Bu gereklilik Lord (1980) tarafından tanıtılmıştır ve tüm eşitleme girişimi üzerine kara bir bulut oluşturan Eşitleme Teoremini tartışmak için kullanılmıştır (Dorans ve Holland, 2000). Bu özelliğe göre, X ve Y test formlarından alınan puanlara ait ortalama, standart sapma, varyans gibi istatistikler birbirine çok yakın olduğunda test formlarının birbirine eşit ve birbirine dönüştürülebilir olduğu ifade edilebilir (Lord, 1980). Lord'un Eşitleme Teoremi; "Eşitleme imkânsız ya da gereksizdir" şeklinde yorumlanabilir (Dorans ve Holland, 2000).
- e) **Gruptan Bağımsızlık:** X ve Y testlerinin puanları arasındaki eşitleme fonksiyonunu hesaplamak için kullanılan örneklemin seçimi önemli olmamalıdır. Başka bir deyişle, X ve Y puanlarını birbirine bağlamak için kullanılan eşitleme fonksiyonu gruptan bağımsız olmalıdır (Dorans ve Holland, 2000; Kolen ve Brennan, 2014).

Dorans ve Holland (2000) bu beş gereksinimi “. . . belirsiz, alakasız, pratik olmayan, önemsiz veya katı olarak eleştirilebilir” şeklinde açıkça tartışmıştır. Lord (1980) (d)’yi en temel gereksinim olarak görürken, Livingston (2014) (d) ve (e) gereksinimlerinin pratikte ulaşılamaz olduğunu savunmuştur. Bu görüş farklılıklarına bakılmaksızın bu beş gereksinim iki testin başarılı bir şekilde eşitlenip eşitlenemeyeceğinin değerlendirilmesi için bir kriter olarak öneme sahiptir (Holland vd., 2007).

Uygulamada; gereksinimler (a) ve (b), testlerin aynı içerik ve istatistiksel özelliklere göre oluşturulması gerektiği anlamına gelir. Gereksinim (c), test eşitleme için regresyon yöntemlerinin kullanılamayacağını belirtir. Gereksinim (d), hem (a) hem de (b)’nin neden gerekli olduğunu açıklamak için kullanılabilir. Ancak gereksinim (d)’nin deneysel olarak değerlendirilmesi zordur ve kullanımı öncelikle teoriktir (Lord, 1980). Gereksinim (e)’nin pratikte uygulanması kolaydır ve (a) ve (b)’nin neden gerekli olduğunu açıklamak için kullanılabilir (Holland ve Dorans, 2006).

Hambleton ve Swaminathan (1985) ile Crocker ve Algina’ya (2008) göre eşitlenecek testler güçlükleri bakımından ikiye ayrılmaktadır. Aynı yapıda ve benzer güçlük düzeyindeki iki testi eşitleme, **yatay eşitleme** olarak adlandırılmaktadır. Yatay eşitleme, genellikle farklı dönemlerde uygulanan ve farklı formları olan bir testin gizliliğinin korunması gerektiğinde kullanılır. Ayrıca bu formların birbirine paralel olması beklenir. Yatay eşitleme, benzer yetenek düzeyindeki bireylere uygulanır (örn. aynı sınıf düzeyi). Aynı yapıda ve farklı güçlük düzeyinde olan, farklı yetenek düzeyindeki kişilere uygulanan iki testi (örn. farklı sınıflardaki kişilere uygulanan testler) eşitleme ise **dikey eşitleme** olarak tanımlanmaktadır. Yıllara göre öğrenci başarı gelişiminin izlenmesi bu kapsamda yer almaktadır.

Test eşitleme **doğrudan** ve **dolaylı** olmak üzere iki şekilde uygulanabilmektedir. Doğrudan test eşitlemede bir testten elde edilen puanlar doğrudan diğer testteki puanlara çevrilirken, dolaylı test eşitlemede üçüncü bir test kullanılarak çevrilmektedir. Kolen ve Brennan (2014) dolaylı eşitlemeyi *zincirleme eşitleme* olarak da isimlendirmektedir.

Eşitleme Basamakları

Test eşitleme Kolen ve Brennan’a (2014) göre yedi basamaktan oluşan bir işlemdir. Bu basamaklar aşağıdaki gibidir:

1. Eşitleme amacına karar verilmesi
2. Alternatif formların oluşturulması (test formları benzer içeriğe sahip olmalıdır)
3. Veri toplama deseninin seçimi
4. Veri toplama deseninin uygulanması (testin veri toplama desenine göre uygulanması ve puanların elde edilmesi)
5. Eşitlemenin bir veya daha fazla işlevsel tanımının seçilmesi (eşitleme, kestirim yapılacak test formları arasındaki ilişki türlerine ilişkin olarak bir seçim yapmayı gerektirir. Örneğin, doğrusal veya doğrusal olmayan eşitleme yöntemlerinin uygulanıp uygulanmayacağına karar verilmesi)
6. Bir veya daha fazla eşitleme ilişkilerini kestirim yönteminin seçilmesi (eşitlemede çeşitli yöntemler bulunmaktadır. Örneğin doğrusal eşitleme ilişkileri Tucker ve Levine yöntemleri kullanılarak kestirilebilir)
7. Eşitleme sonuçlarının değerlendirilmesi.

Eşitleme Desenleri

Bazı test programları gelişigüzel bir şekilde bağlantı kurar ve sihirli bir test eşitleme yönteminin puanlarını düzenlemede “makineden tanrı (deus ex machina)” rolünü oynamasını bekler (Dorans vd., 2010). Halbuki sürecin başından itibaren düzgün planlanması gerekir. Test eşitlemede sonuçların yorumlanmasını karmaşık hale getirebilecek iki faktör vardır. Birincisi iki testin güçlüğüdür (test eşitleme ile ilgili olan budur), diğeri ise sınava giren grubun bu testler üzerindeki yeteneğidir (eşitleme sürecinde ortadan kaldırılması gereken kafa karıştırıcı bir faktördür). Test eşitlemede sadece test özelliklerindeki farklılıkları ayarlamakla ilgilenilir ve bu ayarlamalar yapılırken sınava girenlerin yeteneklerindeki farklılıkların kontrol edilmesi istenir. Uygulamada sınav güçlüğü ve yetenek ayırımına değinmek için iki farklı prosedür vardır. Birincisi sınava giren ortak bir grup kullanmak, diğeri ise bir ortak test kullanmaktır (Holland vd., 2007). Literatürde yaygın olarak kullanılan eşitleme desenleri aşağıda verilmiştir.

Tek Grup Deseni

Tek grup deseni, en kolay veri toplama desendir. Tek grup deseninde, P popülasyonundan seçilen tek bir örneklemdaki tüm sınava girenler her iki test formunu

(X-Y) da alır. Tek grup deseni, muhtemelen test eşitleme için kullanılan en eski veri toplama desendir (Dorans vd., 2010). Teste giren grubun hedef popülasyonun temsili bir örneği olması gerekli değildir. Teste giren grup, her iki formda da hedef gruptan aynı derecede güçlü ya da zayıf olabilir (Livingston, 2014). Tek grup deseni, sınava girenlerin her iki formu da almasını sağlayarak yetenek farklılıklarını kontrol eder (Holland vd., 2007).

Tablo 1

Tek Grup Deseni

Popülasyon	Grup	X	Y
P	1	✓	✓

Tek grup deseninin temel avantajı, sınava girenlerin her iki test formunu da alması nedeniyle istatistiksel olarak güçlü olmasıdır. Nispeten küçük örneklem boyutlarında doğru eşitleme sonuçları sağlayabilir. Tek grup deseninin temel dezavantajı, sınav katılımcılarının ikinci formdaki performanslarının, muhtemelen ilk formu alma deneyiminden etkilenmesidir. Tek grup deseni; sıra etkilerine, uygulama etkilerine veya bazı durumlarda yorulma etkilerine karşı oldukça hassastır. Bu etkilerin ihmal edilebilir olduğunu varsaymak gerekir (Livingston, 2014).

Eş Değer Gruplar Deseni

Çoğu eşitleme durumunda, sınava giren kişinin her iki test formunu da alması için yeterli sınav süresi ayarlamak imkansızdır. En basit çözüm, bir grubun sadece bir test formunu almasıdır. Bu sebeple EG deseninde, ortak bir P popülasyonundan iki eşdeğer grup alınır; biri X formu ve diğeri Y formu ile test edilir (Dorans vd., 2009).

Tablo 2

EG Deseni

Popülasyon	Grup	X	Y
P	1	✓	
P	2		✓

Sınava giren bir grup yalnızca bir test formu aldığından, EG deseni için sıra etkileri sorunu ortaya çıkmaz. Sorun, testle ölçülen her şeyde eşdeğer olan iki grup seçmektir. Pratikte bu iki şekilde yapılır. İlk olarak, P'den iki rastgele grup seçilir ve her biri tek bir formla test edilir. Bu olasılığı yansıtmak için bu desene bazen “rastgele gruplar deseni” denir (Kolen ve Brennan, 2014). Bu durumda iki grup aynı popülasyondan alınan iki rastgele örneğin olabileceği kadar “eşdeğer” olur. İkincisi ve daha yaygın olan durum, iki test formu için test kitapçıklarını spirallemeektir. Bu uygulamada paketleme sürecinde kitapçıklar dönüşümlü olarak düzenlenir. Böylece testler sınava giren birinci kişi X, ikinci kişi Y, üçüncü kişi tekrar X, dördüncü kişi tekrar Y formunu alacak şekilde dağıtılır. İyi yürütüldüğünde, spirallenmiş örnekler genellikle rastgele örneklerden biraz daha “eşdeğer” (daha az farklı)dir (Holland vd., 2007).

Spirallemenin uygulanabilir olması için bazı varsayımların geçerli olması gerekir. Örneğin, iki test formu için zaman sınırları aynı olmalıdır (Holland vd., 2007). Ayrıca sınava girenlerin süreci bozacak şekilde oturmaması gerekir. Örneğin, sınava giren kişiler sistematik olarak oturmuş (erkek-kız) ise, erkeklerin hepsine X formu ve kızlara Y formu verilebilir. Diğer bir durumda, birçok test odası olduğunu varsayalım. Her odadaki ilk sınava giren kişiye X formu uygulandıysa, tek sayıda kişinin sınava girdiği odalara X kitapçığı Y kitapçığından daha fazla verilecektir (Kolen ve Brennan, 2014).

Bu yöntemin avantajı, uygulama ve yorgunluk etkilerinden kaçınmasıdır. Bununla birlikte, gruplar aynı olmadığından yetenek dağılımlarındaki farklılıklar eşitleme prosedüründe bilinmeyen bir sapmaya neden olur (Hambleton ve Swaminathan, 1985). EG deseninin temel sınırlılığı doğru eşitleme sonuçları elde edebilmek için çok sayıda test katılımcısı gerektirmesidir (Livingston, 2014).

Dengelenmiş Tek Grup Deseni

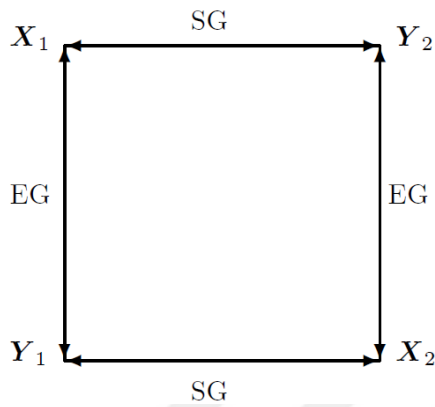
Eşitlemede sıra etkileri sorunu, tek gruplu eşitleme desenini uygunsuz hale getirir. Sorunun üstesinden gelmenin bir yolu, sınava girenleri iki gruba ayırmak ve grupların iki test formunu alma sırasını dengelemektir. X formu ilk alındığında X_1 ve ikinci alındığında X_2 , Y formu da ilk alındığında Y_1 ve ikinci alındığında Y_2 olarak tanımlandığında; bir grup önce X_1 formunu sonra Y_2 formunu alır, diğer grup önce Y_1 formunu, sonra X_2 formunu alır (Holland vd., 2007).

Tablo 3

Dengelenmiş Tek Grup Deseni

Popülasyon	Grup	X1	X2	Y1	Y2
P	1	✓			✓
P	2		✓	✓	

Dengelenmiş tek grup deseni hem tek grup deseni hem de EG desenini içerir. Bu desenin içerisinde X_1 - Y_2 ve X_2 - Y_1 için iki tek grup deseni; X_1 - Y_1 ve X_2 - Y_2 için iki EG deseni vardır.



Şekil 1. Dengelenmiş Tek Grup Deseninde İki EG ve İki Tek Grup Deseni

Kaynak. von Davier, Holland ve Thayer, 2004

Teste girenler iki formu, testin ölçtüğü bilgi ve beceri seviyelerinde bir değişiklik olmayacak kadar yakın zamanda almak zorundalar. İdeal olarak, iki sınav katılımcısı grubu mümkün olduğunca benzer olmalıdır. Bu desen gruplar biraz farklı olsa bile, farklılıklar büyük olmadığı sürece genellikle iyi sonuçlar verir (Livingston, 2014). Bu desende de tek grup deseninde olduğu gibi gruplar hedef evrenden daha başarısız ya da daha başarılı olabilirler, yani örneklem grubunun hedef evreni temsil etmek gibi bir zorunluluğu yoktur (Kolen ve Brennan, 2014).

Dengelenmiş tek grup deseninin temel avantajı tek grup deseniyle aynıdır; nispeten küçük örneklem gruplarında doğru eşitleme sonuçları verir. Başlıca dezavantajı ise uygulanabilirliğidir. Bu desen genellikle veri toplamak için özel bir çalışma gerektirir (Dorans vd., 2010).

Denk Olmayan Gruplarda Ortak Madde Deseni

Bu desende P ve Q olmak üzere iki popülasyon vardır; P'den X testine giren ve Q'dan Y testine giren bir örneklem alınır. Ek olarak, her iki örneklem de bir ortak test (A) alır. von Davier ve diğerleri (2004) “Denk Olmayan Gruplarda Ortak Test” (Non-Equivalent groups with Anchor Test – NEAT); Kolen ve Brennan (2014) ise “Denk Olmayan Gruplarda Ortak Madde” (Common-İtem Nonequivalent Groups – CINEG) veya basitçe “ortak madde” veya “ortak test” deseni olarak atıfta bulunmuştur. Test eşitleme ve bağlama için ortak maddelerin veya ortak testlerin kullanımı diğer veri toplama desenlerinden daha sonra ortaya çıkmıştır. NEAT deseni Holland ve Dorans'ta (2006) belirtildiği gibi hem eşitleme hem de bazı ölçekleme biçimleri için kullanılmıştır.

Tablo 4

NEAT Deseni

Popülasyon	Grup	X	Y	A
P	1	✓		✓
Q	2		✓	✓

NEAT deseni, içinde iki tek grup deseni bulundurur. Bu desen EG deseninden daha esnektir, çünkü X ve Y formunu alan iki grubun farklı veya eşdeğer olmamasına izin verir. Ayrıca tek grup deseninden daha verimlidir, çünkü sınava girenlerin hem X hem de Y formunu almasını gerektirmez. Ortak testlerin kullanımı, önceki veri toplama desenlerinin küçük bir varyasyonu gibi görünmektedir. Fakat ortak maddelerin kullanımı; tek grup deseni, EG deseni ve dengelenmiş tek grup deseninde gerekmeyen yeni varsayımları içerir. NEAT deseninde, Q'daki sınava girenler için X formunun gözlemlenmediği ve P'deki sınava girenler için Y formunun gözlemlenmediği gerçeğini telafi etmek için bir tür varsayım gereklidir (Dorans vd., 2009).

Ortak testin, eşitlenen iki testin “mini versiyonu” olması gerektiği sıklıkla tavsiye edilir (Angoff, 1971). Ortak testler iç ortak test ve dış ortak test olarak kullanılabilir. Dış ortak testteki maddeler, ayrı bir zamanda uygulanır ve genellikle sınava girenlerin puanına dahil edilmez. Ortak testin tüm testin mini bir versiyonu olması, dış ortak testin gizlenme ihtiyacıyla çelişebilir. Bu testi, testin başka bir bölümü gibi görünecek şekilde gizlemek önemlidir. Bunun bir nedeni, bazılarının ortak testi belirleyebilmesi ve bunun nihai puanlarına dahil edilmediğini bilerek (yapmamaları talimatı verilmiş olsa bile)

atlamaları veya gereken önemi vermemeleridir. Bu tür davranışlar onlara fayda sağlıyor gibi görünse de eğer birçok kişi tarafından bu yapılırsa, eşitlemede ortak testin kullanılma şekli nedeniyle tüm sınava girenlerin puanlarının düşmesine neden olabilir (Holland vd., 2007).

İç ortak testlerde ise maddeler, değerlendirmenin bir parçasıdır ve her bir sınava girenin puanına dahil edilir. İç ortak maddeler genellikle test boyunca yayılır. İç ortak testteki maddeler puana dahil edildiğinden, sınava girenlerin bunları atlamaları pek olası değildir. İç ortak testle ilgili başlıca sorunlar; bağlam etkileri ve güvenlik ihlalleridir. Bağlam etkileri, ortak maddeler farklı yerlerde (örn. bir formda 10. maddede ve diğer formda 20. maddede), farklı test koşulları altında (örn. kâğıt- kalem ve bilgisayar tabanlı) veya iki testte farklı türdeki maddelere bitişik olduklarında ortaya çıkabilir. Ortak test maddeleri bir test uygulamasında kullanıldığında, güvenlik ihlallerine sebep olabilir ve gelecekteki uygulamalarda sınava girenler tarafından bilinir hale gelebilir (Dorans vd., 2009).

Ortak maddelerin etkili olması için, farklı formlarda istatistiksel özelliklerini korumaları gerekir. Ortak testin en önemli özelliği eşitlenen iki testteki puanlarla korelasyonunun mümkün olduğunca yüksek olmasıdır (Livingston, 2014). İç ortak testlerin dış ortak testlere göre bir avantajı, ortak maddeler toplam puana dahil edildiğinden eşitlenen testlerle korelasyonların genellikle yüksek olmasıdır. Ortak testin uzunluğu testin en az %20'si kadar olmalıdır (Angoff, 1971; Cook ve Eignor, 1991). En iyi uygulamalar için yüksek korelasyona ihtiyaç duymanın anlamı, uzun ortak testlerin genellikle eşitleme için kısa olanlardan daha iyi sonuçlar vermesidir. Daha uzun ortak testler genellikle daha güvenilirdir ve testlerle daha yüksek korelasyona sahiptir. Uygulamada hem ortak testin hem de eşitlenen testlerin yüksek güvenilirliğe sahip olması arzu edilir (Holland vd., 2007).

İki grup ortak bir popülasyondan ($P = Q$) çekildiği zaman, ortak testin rolünün farklı olduğu bir durum vardır. NEAT deseninin bu özel durumu “eş değer gruplarda ortak test deseni”dir. Bu durumda ortak testin rolü, deneysel bir desendeki bir ortak değişkenin rolü ile aynıdır. İlgili parametrelerin tahmininde kesinlik kazanmak için kullanılır. Bu özel durumda, A'nın X ve Y formunun ölçtüğü aynı yapıyı ölçmesi, ayrıca bir test puanı olması gerekli değildir (Holland vd., 2007).

Denk Olmayan Gruplarda Ortak Değişken Deseni

NEAT deseninde, ortak maddeler kullanılarak gruplar arasındaki yetenek farklılıkları ayarlanır ancak her testte ortak maddeler bulunmayabilir. Eşdeğer olmayan gruplarda ortak madde yoksa, bunların yerine anlamlı ortak değişkenler kullanılabilir ve bu desene “Denk Olmayan Gruplarda Ortak Değişken” (Non-Equivalent groups with Covariates – NEC) deseni denir (Wiberg ve Branberg, 2015). NEAT deseninde olduğu gibi, iki farklı P ve Q popülasyonundan örneklem alınır ve her birine X ve Y test formlarından biri uygulanır (Gonzalez ve Wiberg, 2017).

Tablo 5

NEC Deseni

Popülasyon	Grup	X	Y	C
P	1	✓		✓
Q	2		✓	✓

Bu desende bir ortak test kullanılmaz. Bunun yerine test puanlarıyla yüksek oranda korelasyon gösteren değişkenler kullanılır. Bu yaklaşımla ilgili bir sorun, kullanılan ortak değişken sayısı ile birlikte ortak değişken kategorilerinin sayısının hızla artması ve böylece her bir kategori içindeki gözlemlerin sayısının azalmasıdır. Bu sorunu önlemek için, NEC deseninde eğilim puanları (propensity scores) kullanılabilir (Wallin ve Wiberg, 2017).

Daha öncesinde Holland ve diğerleri (2007) sınava giren kişiler hakkında ek bilgi kullanmanın mantıklı olup olmadığı sorusunu gündeme getirmiş ve bunun ileri araştırmalar için önemli bir alan olduğunu öne sürmüştür. Test eşitleme ve ortak değişkenlerin kullanımıyla bağlantılı gelecekte birçok zorluk vardır. Bununla birlikte, iyi ortak değişkenler olmadan ortak değişkenlerle iyi çalışan bir eşitleme gerçekleştiremeyeceği vurgulanmalıdır. Bu nedenle, hangi ortak değişkenlerin kullanılabileceğini ve bunların düzenli olarak nasıl toplanacağını incelemek araştırmaların önemli bir bölümüdür. Test puanlarıyla yüksek oranda korelasyon gösteren ortak değişkenler eşitleme durumunda iyi çalışma eğiliminde olduğundan, ortak değişkenleri seçerken ortak değişkenler ve test puanları arasındaki korelasyonu incelemek gerekmektedir (Wiberg, 2015).

Eşitleme Yöntemleri

Test eşitleme Gözlenen Puan Eşitleme ve MTK'ye Dayalı Eşitleme yöntemleri olarak iki sınıfa ayrılabilir (Gonzalez, 2014; Hambleton ve Swaminathan, 1985; Kolen ve Brennan, 2014). Gözlenen Puan Eşitleme yöntemlerine ortalama, lineer ve eşit yüzdelikli eşitleme; Tucker ve Levine gözlenen puan eşitleme ve Levine gerçek puan eşitleme; Kernel (Gauss) eşitleme yöntemleri örnek verilebilir. MTK'ye dayalı yöntemler MTK gerçek puan ve MTK gözlenen puan eşitleme yöntemleri; ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemleri gibi parametre bağlama yöntemlerini içerir (Kolen ve Brennan, 2014).

Bu iki geleneksel eşitleme yöntemi standart yöntemler olarak adlandırılmaktadır. Yeni teorik ve karmaşık eşitleme yöntemlerinin geliştirilmesi, günümüzde eşitleme araştırmalarında yaygındır. Bazıları tamamen yeniyken diğerleri standart yöntemlerin uzantılarıdır. Bunlar istatistik kitaplarında tipik olarak bulunan konuların (üstel aileler, PoB modelleri, zaman serisi analizi vb.) açıkça bir eşitleme çerçevesine yerleştirildiği yöntemleri içerir. Ayrıca yerel eşitleme ve lineer olmayan Levine yöntemi gibi hibrit yöntemler eşitlemede yeni olanaklar olarak ortaya çıkmıştır. Yeni, daha teorik ve karmaşık eşitleme yöntemleri standart olmayan yöntemler olarak adlandırılmaktadır (Gonzalez, 2014).

Gözlenen Puan Eşitleme

Bu yöntem zayıf teorik varsayımlara sahiptir ve kolay uygulanabilir. Bu teori örnekleme bağlıdır, madde parametrelerinin kestirimleri bireylere ve bireylerin parametrelerinin kestirimleri teste bağlıdır. Test eşitlemede yapılan tartışmalar testlerin ham puanlara dayalı olarak eşitlenmesinin eşitlik, simetri ve grup değişmezliği varsayımlarını tam olarak sağlayamaması nedenleriyle istenmediğini göstermektedir (Hambleton ve Swaminathan, 1985). Gözlenen puan eşitlemede kullanılan bazı yöntemler şunlardır (Gonzalez, 2014): ortalama, lineer, eşit yüzdelikli, yerel ve kernel eşitleme.

Ortalama Eşitleme. Gözlenen puan eşitlemeye dayalı olarak geliştirilen test eşitleme yöntemleri içerisinde en az parametre kullanılan ortalama eşitlemedir. Ortalama eşitleme yöntemi X ve Y testleri için puan dağılımlarının varyansı eşit olduğu

varsayıldığında ortaya çıkar. Ortalama eşitleme X ve Y test formlarının puan ölçeği boyunca yalnızca ortalamalarına göre farklılık gösterdiği varsayımına dayanır (Gonzalez ve Wiberg, 2017).

Ortalama eşitlemede puan ölçeği boyunca X formunun Y formundan güçlük bakımından sabit bir miktarda farklı olduğu kabul edilir. Sabit bir fark, birçok test durumunda aşırı kısıtlayıcı olsa da ortalama eşitleme bazı önemli eşitleme kavramlarını göstermek için yararlıdır (Kolen ve Brennan, 2014). Ortalama eşitleme aşağıdaki eşitliklerle ifade edilebilir (Gonzalez, 2014).

$$x - \mu_x = y - \mu_y \quad (1)$$

$$y = x - \mu_x + \mu_y \quad (2)$$

Ortalama eşitlemede eşitleme dönüşümü parametrelere bağlı olduğundan parametrik bir yöntemdir (Gonzalez, 2014).

Lineer (Doğrusal) Eşitleme. Lineer eşitleme, X formu ve Y formundaki puan dağılımlarının ortalamalar ve standart sapmalardaki farklılıklar dışında bütün özelliklerinin aynı olduğu varsayımına dayanmaktadır (Crocker ve Algina, 2008). Lineer eşitleme, iki form arasındaki güçlük farklarını sabit olarak düşünmek yerine bunun puan ölçeği boyunca değişmesine izin verir. Örneğin lineer eşitleme, X formunun düşük başarılı öğrenciler için Y formundan daha zor olmasına, ancak yüksek başarılı öğrenciler için daha az zor olmasına izin verir (Kolen ve Brennan, 2014).

İki testten elde edilen puanların ortalamadan eşit uzaklıkta olduğu varsayılır ve standart sapma birimleri eşitlenir. Lineer eşitleme aşağıdaki eşitliklerle ifade edilebilir (Gonzalez, 2014).

$$\frac{x - \mu_x}{\sigma_x} = \frac{y - \mu_y}{\sigma_y} \quad (3)$$

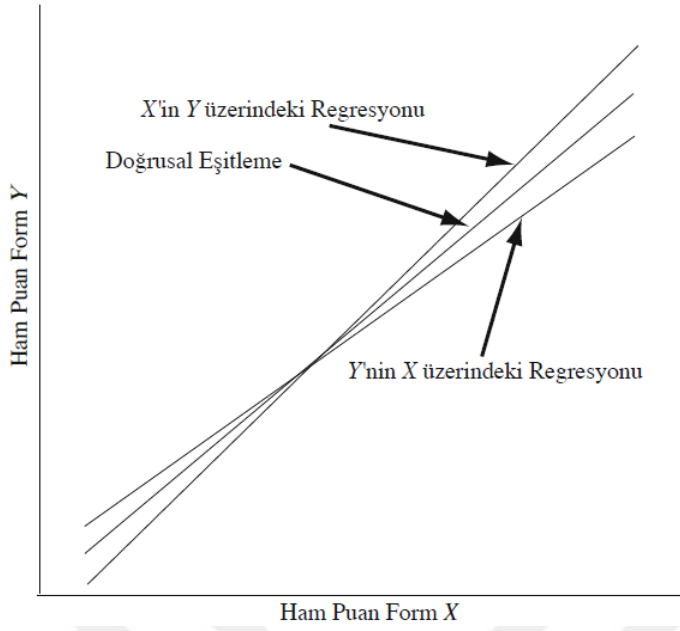
$$y = \frac{\sigma_x}{\sigma_y}(x - \mu_x) + \mu_y \quad (4)$$

Eğer iki test formunun standart sapmaları eşit olsaydı Eşitlik (3), Eşitlik (2)'ye eşit olacak şekilde basitleştirilebilirdi. Bu nedenle iki formun standart sapmaları eşitse, ortalama ve lineer eşitleme aynı sonucu verir (Kolen ve Brennan, 2014). Lineer

eşitlemede eşitleme dönüşümü parametrelere bağlı olduğundan parametrik bir yöntemdir (Gonzalez, 2014).

Lineer eşitleme yönteminin birtakım sınırlılıkları bulunmaktadır. Yukarıdaki eşitliği uyguladığımızda ham puan tam sayıysa, düzeltilmiş puan neredeyse hiçbir zaman tam sayı olmaz. Bu durumda test formunda düzeltilmiş puan için mümkün olan bir puan bulunmayacaktır. Bu soruna “ayrıklık sorunu” veya “arada puan sorunu” denir. Bu sorunun yaşanacağı test türleri, puan aralığındaki herhangi bir sayının olası bir puan olabileceği testlerdir. Böyle bir durumda enterpolasyon (öteleme) yapılması gerekir (Livingston, 2014).

Lineer eşitleme denklemi aldatıcı bir şekilde lineer bir regresyon denklemi gibidir. Aradaki fark, lineer regresyon için σ_y / σ_x terimlerinin X ve Y arasındaki korelasyon ile çarpılmasıdır. Bununla birlikte lineer regresyon denklemi, bir eşitleme fonksiyonu olarak nitelendirilmez çünkü korelasyon katsayısı 1 olmadıkça, X’in Y üzerindeki regresyonu, Y’nin X üzerindeki regresyonundan farklıdır (Kolen ve Brennan, 2014). Bu nedenle, regresyon denklemleri genel olarak eşitleme fonksiyonları olarak kullanılamaz (Hambleton ve Swaminathan, 1985). Lineer regresyon ve lineer eşitleme arasındaki karşılaştırma Şekil 2’de gösterilmektedir. Y’nin X üzerindeki regresyonu, X’in Y üzerindeki regresyonundan farklıdır. Ayrıca şekildeki grafikte gösterilen yalnızca bir lineer eşitleme ilişkisi olduğuna dikkat edilmelidir. Bu ilişki, X formunun puanlarını Y formunun ölçeğine dönüştürmek veya Y formunun puanlarını X formunun ölçeğine dönüştürmek için kullanılabilir (Kolen ve Brennan, 2014).



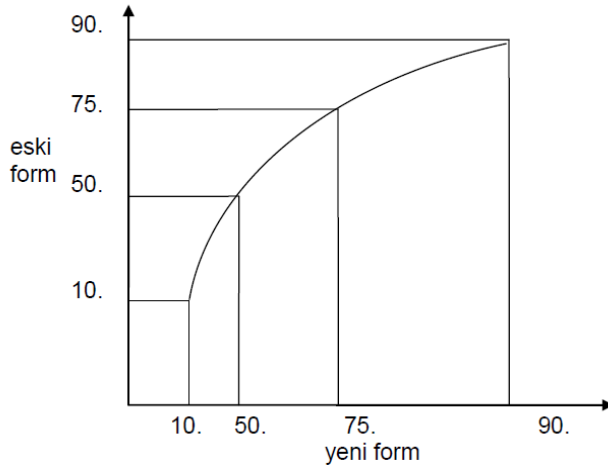
Şekil 2. Lineer Eşitleme ve Lineer Regresyonun Karşılaştırılması
Kaynak. Kolen ve Brennan, 2014

Eşit Yüzdelikli Eşitleme. Eşit yüzdelikli eşitleme yönteminde X ve Y formlarında aynı yüzdelik sıralara sahip puanlar eşitlenir (Gonzalez ve Wiberg, 2017). Braun ve Holland (1982), X ve Y sürekli rasgele değişkenler olduğunda Eşitlik 5'in eşit yüzdelikli bir eşitleme fonksiyonu olduğunu belirtmiştir. X puanlarının Y ölçeğine dönüştürülmesinden kaynaklanan puanların dağılımı; G'ye (popülasyonda Y üzerindeki puanların dağılımı) eşitse, bir $e_Y(x)$ fonksiyonu eşit yüzdelikli dönüşüm olarak adlandırılır (Gonzalez, 2014).

$$\varphi(x) = e_Y(x) = G^{-1}(F(x)) \quad (5)$$

φ dağılım fonksiyonlarından oluşturulduğundan, eşit yüzdelik eşitleme yöntemi parametrik değildir (Gonzalez, 2014).

Şekil 3'te eşit yüzdelikli eşitlemede puan dönüşümü gösterilmektedir. Yeni formda 10. yüzdelik dilimdeki puan, referans formdaki 10. yüzdelik dilimdeki puana eşittir. Bu durum diğer yüzdelik dilimler için de geçerlidir. Ayrıca eğrinin eğiminin, düşük puanlar için fazla ve yüksek puanlar için daha az olduğuna dikkat edilmelidir. Eğimdeki bu varyanslar, yöntemin hem düşük başarılı hem de yüksek başarılı öğrencilere uygulanmasını mümkün kılar (Livingston, 2014).



Şekil 3. Eşit Yüzdelikli Eşitleme

Kaynak. Livingston, 2014

Eşitlik 5'teki dönüşüm test puanlarının sürekli rastgele değişkenler olduğu varsayımına dayanmaktadır. Bununla birlikte, tipik olarak test puanları ayrıktır. Öteleme ve düzgünleştirme yöntemleri kullanılarak dağılım sürekli hale getirilmekte ve bu sorun ortadan kaldırılmaktadır (Cook ve Petersen, 1987).

Eğitsel ve psikolojik ölçmede, yüzdelikler ve yüzdelik sıralar kullanılarak ayrık test puanlarını süreklileştirmek için bir yaklaşım mevcuttur. Bu yaklaşımda, örneğin 28'lik bir tamsayı puanının 27.5 – 28.5 aralığındaki puanları temsil ettiği ve 28 puan alan adayların bu aralıkta eşit olarak dağıldığı kabul edilir. Yirmi sekiz puanın yüzdelik sıralaması, 28'in altındaki puanların yüzdesi olarak tanımlanır. Örneğin; sınava girenlerin %18'inin Y formundan 27.5'in altında, %5'inin ise 27.5 ile 28.5 arasında bir puan aldığını varsayalım. Bu durumda, Y formundaki 28 puanın yüzdelik sıralaması $\%(18 + 1/2(5)) = \%20.5$ olacaktır. Tipik olarak kullanılan terminolojide, bir tamsayı puanının yüzdelik sıralaması o puanı içeren aralığın orta noktasındaki yüzdelik sıralamasıdır (Kolen ve Brennan, 2014).

Holland ve Thayer (1989) yüzdelik sıraları kullanmak için istatistiksel bir gerekçe sunmuştur. Yaklaşımlarında süreklileştirme ve Kernel düzgünleştirme süreci olarak adlandırdıkları yöntemi kullanmışlardır. Kesikli bir rasgele değişken X ve $-1/2$ ile $+1/2$ aralığında eşit olarak dağılmış bir rasgele değişken U verildiğinde, yeni bir rasgele değişken tanımlanmıştır: $X^* = X + U$. Bu yeni rastgele değişken sürekli dir.

Eşit yüzdelikli eşitleme yöntemi, ortalama eşitleme ve lineer eşitleme yöntemleri ile karşılaştırıldığında daha büyük örneklemlemlere ulaşılmasını gerektirmektedir. Test

formlarının dağılımı, küçük örneklerde düzensizlik gösterebilirken, bu dağılımın daha büyük örneklerde daha düzenli olacağı varsayılır. Eğer testlerin puan dağılımları benzerse, lineer eşitleme ve eşit yüzdellikli eşitleme aynı sonucu verir (Kolen, 1988). Ayrıca eşit yüzdellikli eşitleme lineer eşitlemeden daha az varsayımda bulunur ve bu nedenle çok çeşitli araştırma ve testlerde kullanılabilir (Woldbeck, 1998).

Yerel (Local) Eşitleme. Yerel eşitleme, eşitlik ve grup değişmezliği gereksinimleri karşılanmadığında kullanılabilen bir yöntemidir (van der Linden, 2010). φ dönüşümünü elde etmek için eşit yüzdellikli eşitlemeler puan dağılımlarını kullanırken, yerel eşitleme yöntemleri puanların yeteneğe bağlı koşullu dağılımlarını kullanır. Eşitleme fonksiyonu aşağıdaki eşitlikte verilmiştir (Gonzalez, 2014).

$$\varphi(x; \theta) = G_{Y/\theta} - 1(F_{x/\theta}(x)) \quad \theta \in R \quad (6)$$

θ 'yı tahmin etmek için, En Çok Olabilirlik (Maximum Likelihood – ML), Maksimum Sonsal Kestirim (Maximum A Posteriori – MAP) veya Beklenen Sonsal Kestirim (Expected A Posteriori – EAP) tahminleri hesaplanır ve bireylerin yanıt modelleri kullanılarak bir MTK modelinin geçerli olduğu varsayılarak elde edilir. Ortaya çıkan ayrık koşullu dağılımları süreklileştirmek için lineer enterpolasyon kullanılır. Hem vektör hem de fonksiyon parametreleri söz konusu olduğundan Yerel eşitleme aynı zamanda yarı parametrik bir yöntemdir (Gonzalez, 2014).

Test eşitlemenin, test formlarını alan popülasyonların yeteneklerine göre ayarlanan puan dağılımlarına dayanması gerektiği fikri eşitleme literatüründe mevcuttur. Yerel eşitleme fikri, bu eşitleme mantığını bir adım öteye taşıyor: Eski ve yeni test formlarını alan gruplar için aynı şekilde dağılmış yetenekler elde etmeye çalışmak yerine, sınava giren kişilerin gözlenen puan dağılımlarını iki formda aynı yeteneklerle eşitlemeye çalışır. Geleneksel ve yerel eşitleme arasında göze çarpan bir fark, sınava girenlerin yetenek seviyeleri için bir dönüşüm ailesi yerine tek bir dönüşümün kullanılmasıdır (van der Linden ve Wiberg, 2010).

Kernel Eşitleme. Kernel eşitleme yöntemi Holland ve Thayer (1989) tarafından tanıtılmış ve daha sonra von Davier ve diğerleri (2004) tarafından geliştirilmiştir. Eşit yüzdellikli eşitleme ve Kernel eşitlemenin farklı olduğu nokta süreklileştirme biçimleridir

(von Davier vd., 2004). Kernel eşitleme; geleneksel eşit yüzdelikli eşitleme yöntemindeki gibi lineer enterpolasyonun kullanılmasının aksine, Gauss Kernel kullanılarak puan dağılımlarının ayırık dağılımlardan sürekli dağılımlara dönüştürüldüğü eşit yüzdelikli gözlenen puan eşitleme prosedürüdür (von Davier vd., 2006). Kernel eşitleme yarı parametrik eşitleme yönteminin bir örneği olarak düşünülebilir (Gonzalez ve Wiberg, 2017).

Kernel eşitlemenin avantajları şu şekilde ifade edilebilir: Birinci olarak, bu yöntem diğer yöntemlerle karşılaştırıldığında daha düşük standart hata vermektedir ve örneklem büyüklüğüne daha az bağlıdır. Çünkü ön düzgünleştirmenin kullanılmasıyla hem veriler uyumlu hale getirilir hem de verilerin dönüşümü düzgünleştirilir. İkinci olarak Kernel eşitleme tüm desenlere ve eşitleme işlemlerine uygulanabilmektedir (von Davier vd., 2004).

Bu yöntem beş adımdan oluşmaktadır (von Davier vd., 2004): ön düzgünleştirme, puan dağılımlarının kestirilmesi, süreklileştirme, eşitleme ve eşitlemenin standart hatasının hesaplanması.

1. Ön Düzgünleştirme. Bu adımda bir eşitleme desende toplanan veriler, gerçek puan dağılımlarını tahmin etmek için istatistiksel prosedürler kullanılarak düzeltilir. Çeşitli tekniklerin kullanıldığı ön düzgünleştirme, eşit yüzdelikli eşitlemeye yönelik çeşitli yaklaşımlarda standart bir araç haline gelmiştir. Tek değişkenli ve iki değişkenli puan dağılımları için loglineer modellerin kullanılması, aşırı esneklikleri ve puan dağılımlarının pek çok alışılmadık özelliğini barındırma yetenekleri nedeniyle savunulmaktadır (von Davier vd., 2006).

Bu ön düzgünleştirme işleminin sonuçları iki yönlüdür. İlk olarak, eşitleme işleminin geri kalanı için gerekli olan düzgünleştirilmiş puan dağılımları elde edilir. İkinci olarak, işlemin ilerleyen aşamalarında eşitlemenin standart hatasını hesaplamak için kullanılacak bir matris hesaplanır. Her ön düzgünleştirme yönteminin böyle bir matrisi vardır, ancak loglineer yöntemlerde bunu verimli bir şekilde bulmanın standart bir yolu mevcuttur (von Davier vd., 2006).

Puan olasılık tahminlerinin uygun derecelerde düzgünlük göstermesi gerekir ve bu çeşitli şekillerde yapılabilir. Ön düzgünleştirme, ilgili puan olasılıklarının istatistiksel tahmin sorunu olarak görülmektedir. Tahmini puan olasılıklarını seçerken en az dört istatistiksel özellik dikkate alınmalıdır (von Davier vd., 2004):

- **Tutarlılık:** Örneklem büyüklükleri arttıkça, tahminlerin popülasyon değerlerine yakınlaşması gerekir.
- **Verimlilik:** Örneklem büyüklükleri göz önüne alındığında, tahmini puan olasılıklarının popülasyon değerlerinden sapmaları mümkün olduğunca küçük olmalıdır. Bu sapmalar beklenmektedir, önemli olan bunu minimumda tutmaktır.
- **Pozitiflik:** Tahmini puan olasılıkları her olası puan için pozitif olmalıdır. Bir puan olasılığının sıfır olduğunu tahmin etmek çoğu test için mantıksızdır. İç ortak test ile NEAT deseninde ve tek grup deseninin bazı uygulamalarında istisnai durumlar vardır.
- **Bütünlük:** Düzgünleştirilmiş puan dağılımlarıyla gözlenen puan dağılımlarının bütünlüğünü koruması yani eşleşmesi gerekir. Bu özellik, lineer eşitlemedeki ortalama ve varyans eşleştirme özelliği ile yakın ilişki içindedir ve eşit yüzdelikli yöntemler için de istenilen özelliktir.

Kernel eşitleme NEAT desenindeki iç ortak ve dış ortak testte yapıldığında, diğer aşamalar aynıdır fakat ön düzgünleştirme aşaması farklılaşmaktadır. Dış ortak test ayrı bir bölüm olarak alınır ve bu testten alınan puan toplam puana dahil edilmez. Elbette sınava giren kişi dış ortak testten yüksek puan alırsa testten de yüksek puan alması beklenir ancak bu ilişki belirleyici değildir. İç ortak test bu açıdan farklıdır. Örneğin X ve Y testinde A iç ortak testin puanı, X^* ve Y^* ise testlerin birbirinden farklı olan bölümlerinin puanı olsun. Bu bölümlerden alınacak puanlar doğru sayı puanlarıdır. Bu durumda $X = X^* + A$ ve $Y = Y^* + A$ olacaktır. Örneğin, X testinde A formundan alınacak puan X testinden alınacak puanı belirli değerlerden uzak tutar. A formundan alınan 5 puanla X^* negatif olmayacağı için toplam puan X daima 5 veya daha yüksek ($X \geq 5$) olacaktır. Diğer bir ifadeyle iç ortak test kullanıldığı zaman, testten alınan toplam puan ortak teste eşit ya da ondan büyük olmalıdır asla küçük olamaz. Toplam test puanı ile ortak test arasındaki bu belirleyici ilişki, ortak olasılıklar sisteminde “yapısal sıfırlar” ile sonuçlanır (von Davier vd., 2004).

2. Puan Dağılımlarının Kestirilmesi. Ön düzgünleştirme yapıldıktan sonra r ve s olarak adlandırdığımız T üzerindeki puan olasılık dağılımlarının kestirilmesi için düzgünleştirilmiş puan dağılımlarını kullanan formüller üretilir. r ve s puan olasılık dağılımlarını ifade etmek üzere bunlara ait eşitlikler aşağıda verilmiştir;

$$r_j = P\{X = x_j|T\}, \quad s_k = P\{Y = y_k|T\} \quad (7)$$

$$r = (r_1, \dots, r_j) \text{ ve } s = (s_1, \dots, s_k) \quad (8)$$

X için puan olasılıkları X-ham puanları $\{x_j\}$ ile ilişkilidir ve Y için Y-ham puanları $\{y_k\}$ ile ilişkilidir. Kullanılan eşitleme desenine göre kestirilen puan dağılımları farklılık gösterir (Mao, von Davier ve Rupp, 2006).

Bu adımda önemli bir kısım tasarım fonksiyonunun (Design Function – DF) tanımlanması ve kullanılmasıdır. DF her bir eşitleme desenini karakterize eder ve hem eşitlemenin standart hatasını (Standart Error of Equating – SEE) hem de eşitleme farkının standart hatasını (Standart Error of Equating Difference – SEED) hesaplamak için kullanılır (Gonzalez ve Wiberg, 2017). Veri toplama desenlerine göre DF'nin tanımlanması aşağıda verilmiştir (von Davier vd., 2004).

EG Deseni. Bu desende DF'nin tanımlanması çok basittir çünkü r ve s 'yi elde etmek için verilerin daha fazla dönüştürülmesine gerek yoktur yani doğrudan tahmin edilir. DF'nin tanımlanması aşağıda verilmiştir,

$$\begin{pmatrix} r \\ s \end{pmatrix} = DF(r, s) = \begin{pmatrix} I_J & 0 \\ 0 & I_K \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix}. \quad (9)$$

Tek Grup Deseni. İncelenen kişiler tek bir popülasyondan seçildiğinde hedef popülasyon EG deseninde olduğu gibi P 'dir. $v(P)$, P 'nin sütunlarının üst üste istiflenmesiyle oluşur.

$$P = (p_1, p_2, \dots, p_K) \quad (10)$$

$$v(P) = \begin{pmatrix} p_1 \\ \vdots \\ p_K \end{pmatrix} \quad (11)$$

Tek grup deseni için DF, $v(P)$ 'nin r ve s 'ye dönüştürülmesidir. Bunu $v(P)$ 'nin doğrusal bir fonksiyonu olarak ifade etmek için M ve N olmak üzere iki 0-1 matrisi tanımlanmıştır.

$$r = Mv(P) = \sum_k p_k \quad (12)$$

$$s = Nv(P) = \begin{pmatrix} 1_j^t p_1 \\ \vdots \\ 1_j^t p_K \end{pmatrix} \quad (13)$$

DF'nin tanımlanması aşağıda verilmiştir,

$$\binom{r}{s} = DF(P) = \binom{Mv(P)}{Nv(P)} = \binom{M}{N} v(P). \quad (14)$$

Dengelenmiş Tek Grup Deseni. w_x ve w_y sıra etkilerine tabi olmayan verilerde ağırlığı göstermek üzere ($0 \leq w \leq 1$), iki bağımsız tek grup yaklaşımı kullanan dengelenmiş tek grup deseni için DF aşağıda tanımlanmıştır,

$$\binom{r}{s} = DF(P_{(12)}, P_{(21)}) = \begin{pmatrix} w_x M & (1 - w_x)M \\ (1 - w_y)N & w_y N \end{pmatrix} \begin{pmatrix} v(P_{(12)}) \\ v(P_{(21)}) \end{pmatrix}. \quad (15)$$

NEAT Deseni – Zincirleme Eşitleme. Zincirleme eşitleme NEAT deseni için kullanılan en eski yöntem olabilir. Tek grup deseninde zaten mevcut olan fikirlerin basit bir uzantısıdır. Zincirleme eşitleme, X puanlarının Y puanlarına iki aşamalı bir dönüşümünü kullanır. Önce X formunu P popülasyonu üzerinde A ortak teste bağlar ve ardından Q popülasyonu üzerinde A ortak testi Y formuna bağlar.

Zincirleme eşitlemede ikisi X ve Y formuna (F_P, G_Q) ve diğer ikisi ise ortak teste (H_P, H_Q) ait olmak üzere 4 adet yığılmalı dağılım fonksiyonu vardır. P popülasyonu için F_P , X formuna ait ve H_P , ortak teste ait yığılmalı dağılım fonksiyonunu; Q popülasyonu için G_Q , Y formuna ait ve H_Q , ortak teste ait yığılmalı dağılım fonksiyonunu göstermektedir.

$$r_P = (r_{Pj}), t_P = (t_{Pl}), t_Q = (t_{Ql}), s_Q = (s_{Qk}) \quad (16)$$

$$j = 1, \dots, J, \quad k = 1, \dots, K, \quad l = 1, \dots, L \quad (17)$$

P popülasyonundaki X ve A formları ve Q popülasyonundaki Y ve A formları için marjinal olasılıklar şu şekilde verilir,

$$r_{pj} = Prob\{X = x_j|P\} = \sum_l p_{jl}, \quad (18)$$

$$t_{pl} = Prob\{A = a_l|P\} = \sum_j p_{jl}, \quad (19)$$

$$t_{ql} = Prob\{A = a_l|Q\} = \sum_k q_{kl}, \quad (20)$$

$$s_{Qk} = Prob\{Y = y_k|Q\} = \sum_l q_{kl}. \quad (21)$$

Tek grup deseninde olduğu gibi, X ve Y'nin marjinal olasılıklarını $v(P)$ ve $v(Q)$ vektörleştirilmiş diziler cinsinden yazmak için iki 0-1 matrisi M ve N kullanılmıştır.

$$r_P = M_P v(P) \quad \text{ve} \quad t_P = N_P v(P), \quad (22)$$

$$s_Q = M_Q v(Q) \quad \text{ve} \quad t_Q = N_Q v(Q). \quad (23)$$

Bu doğrusal dönüşümler, zincirleme eşitleme aracılığıyla yapıldığında ilk tasarım fonksiyonu DF_P ve ikinci tasarım fonksiyonu DF_Q aşağıdaki gibi tanımlanır,

$$\begin{pmatrix} r_P \\ t_P \end{pmatrix} = DF_P(P) = \begin{pmatrix} M_P \\ N_P \end{pmatrix} v(P), \quad (24)$$

$$\begin{pmatrix} t_Q \\ s_Q \end{pmatrix} = DF_Q(Q) = \begin{pmatrix} N_Q \\ M_Q \end{pmatrix} v(Q). \quad (25)$$

NEAT Deseni – Son Tabakalama Eşitleme: Ön düzgünleştirme adımından sonra, son tabakalama eşitlemede P ve Q popülasyonlarının ağırlıklandırılması ile oluşan sentetik evren T üzerinde X ve Y formlarına ait marjinal dağılımlar belirlenir.

$$T = wP + (1 - w)Q, \quad (0 \leq w \leq 1) \quad (26)$$

T üzerindeki r_j ve s_k olasılıkları, T'nin iki katmanındaki (yani, P ve Q) koşullu olasılıklara dayalı olarak hesaplanır.

$$r_j = Prob\{X = x_j|T\} = wr_{pj} + (1 - w)r_{Qj}, \quad (27)$$

$$s_k = Prob\{Y = y_k|T\} = ws_{Pk} + (1 - w)s_{Qk}, \quad (28)$$

$$t_l = Prob\{A = a_l|T\} = wt_{Pl} + (1 - w)t_{Ql} \quad (29)$$

P popülasyonunda A verilen X'in ve Q popülasyonunda A verilen Y'nin koşullu dağılımları aşağıdaki gibi gösterilmektedir.

$$r_P(x_j|a_l) = Prob\{X = x_j|A = a_l, P\}, \quad (30)$$

$$s_Q(y_k|a_l) = Prob\{Y = y_k|A = a_l, Q\} \quad (31)$$

$l = 1, \dots, L$ için p_l ve q_l vektörleri sırayla P ve Q'nun l . sütunları olarak alındığında DF aşağıdaki gibi ifade edilebilir,

$$\binom{r}{s} = DF(P, Q) \quad (32)$$

$$r = r(P, Q, w) = \sum_l \left[w + \frac{(1 - w)(t_{Ql})}{t_{Pl}} \right] p_l \quad (33)$$

$$s = s(P, Q, w) = \sum_l \left[(1 - w) + \frac{w(t_{Pl})}{t_{Ql}} \right] q_l. \quad (34)$$

3. Süreklileştirme. Bu adım, eşit yüzdelik eşitleme yöntemleri tartışmalarında genellikle göz ardı edilir, ancak hepsinde bulunur. T üzerinde X ve Y için ayrı puan dağılımları ile başlanır ve bunlar sürekli puan dağılımlarına çevrilir. Günlük bir istatistiksel yöntem gibi görünen bir adımdır ancak aslında olağandışıdır çünkü tüm ayrık dağılım, orijinaline yakın sürekli bir dağılıma dönüştürülür. Diğer eşit yüzdelikli eşitleme yöntemleri, ayrık puan dağılımlarını yüzdelik sıralara dayalı parçalı doğrusal kümülatif dağılım fonksiyonlarıyla değiştirir. Kesikli dağılımları süreklileştirmek amacıyla Kernel eşitleme Gauss Kernel düzgünleştirme kullanır. r 'yi süreklileştirmenin Gauss Kernel yöntemi aşağıdaki eşitliklerde verilmiştir (von Davier vd., 2006);

$$\mu_{XT} \sum_j x_j r_j, \sigma_{XT}^2 = \sum_j (x_j - \mu_{XT})^2 r_j, \alpha_X = \sqrt{\sigma_{XT}^2 / (\sigma_{XT}^2 + h_X^2)} \quad (35)$$

$$F_T(x; h_x) = \sum_j r_j \phi \left(\frac{x - \alpha_X x_j - (1 - \alpha_X) \mu_{XT}}{h_x \alpha_X} \right). \quad (36)$$

Sürekli $G_T(y, h_Y)$, puan olasılıkları s ve Y -puanları $\{y_k\}$ kullanılarak benzer şekilde hesaplanır. Gauss Kernel süreklileştirmenin önemli bir özelliği süreklilikteki düzgünlük derecesini kontrol eden h_X ve h_Y bant genişliklerinin seçimidir. Bant genişliği (h), tahmin yapılan yoğunluk fonksiyonları düzgün olacak ancak orijinal ayırık puan dağılımlarının özellikleri de korunacak şekilde seçilmelidir. Bant genişliği birkaç farklı şekilde seçilebilir. İdeal bant genişliğinin seçiminde von Davier ve diğerleri (2004) ceza (penalty) fonksiyonunun kullanılmasını önermiştir. Ceza fonksiyonunun hesaplanması aşağıdaki eşitliklerde verilmiştir;

$$Ceza_1(h) = \sum_j \left[\left(\frac{r_j}{d_j} \right) - f_T(x_j; h) \right]^2 \quad (37)$$

$$Ceza_2(h) = \sum_j A_j(1 - B_j) . \quad (38)$$

Ceza₁ ve Ceza₂, K katsayısıyla ağırlıklandırılıp birleştirilir ve aşağıdaki gibi bir eşitlik elde edilir.

$$Ceza_1(h) + K * Ceza_2(h) \quad (39)$$

X formunda kullanılan ceza fonksiyonu aynı şekilde Y formunda da kullanılabilir. Dağılımda düzgünleştirilmesi gereken boşlukların olduğu çeşitli uygulamalarda K ağırlıklandırma katsayısının eşitlikte 1 olarak alınması önerilmektedir (Mao vd., 2006).

Kernel eşitlemede, bant genişlikleri geniş olarak alındığında ortaya çıkan özel durum lineer eşitleme fonksiyonlarını içerir. Nispeten küçük bant genişlikleri bile neredeyse lineer Kernel eşitleme fonksiyonlarını oluşturabilir. O halde süreklileştirme belirli bir şekilde yapıldığında bir anlamda Kernel eşitleme lineer eşitlemedir (von Davier vd., 2004).

4. Eşitleme. Yukarıdaki adımlardan sonra Kernel eşitleme için eşitleme fonksiyonları hesaplanır. Bu eşitlikler aşağıda verilmiştir (von Davier vd., 2004);

X formunun Y formuna eşitlenmesi için Kernel eşitleme fonksiyonu,

$$\hat{e}_Y(x) = G_{h_Y}^{-1}(F_{h_X}(x; r); s) = G_{h_Y}^{-1}(F_{h_X}(x)) . \quad (40)$$

Y formunun X formuna eşitlenmesi için Kernel eşitleme fonksiyonu,

$$\hat{e}_X(y) = F_{hx}^{-1}(G_{hy}(y; s); r) = F_{hx}^{-1}(G_{hy}(y)). \quad (41)$$

5. Eşitlemenin Standart Hatasının Hesaplanması. Bu yöntemde SEE ve SEED hesaplamak mümkündür. SEE yukarıdaki dört adıma karşılık gelen üç faktöre bağlıdır. Bunlar: ön düzgünleştirme, düzgünleştirilmiş verilerden r ve s 'yi hesaplama ve 4. adımdaki eşitleme fonksiyonunun matematiksel formunun birleşimidir. Analitik formüllere dayanan Kernel eşitleme, çeşitli eşitleme desenleri için SEE'yi hesaplamak üzere Taylor açılımı veya delta yönteminin kullanılmasına izin verir (von Davier vd., 2006).

Kernel eşitleme için SEE'nin hesaplanması söz konusu olduğunda, çeşitli eşitleme desenleri arasındaki temel fark 2. adımdır. Her desen, önceden düzgünleştirilmiş verileri puan olasılıklarına (r ve s) eşitlemek için farklı bir formül (bir tasarım fonksiyonu) gerektirir, ancak diğer adımların SEE'ye katkıları tüm desenler için aynıdır (Mao vd., 2006).

Kernel eşitlemenin etkililiğini belirlemek amacıyla SEE ve SEED kullanılır. SEE örneklem varyansını hesaplamaktadır ve büyük örneklemelere bağlıdır (von Davier vd., 2004). SEE'nin eşitlikleri aşağıda verilmiştir.

X'in Y'ye eşitlenmesindeki SEE;

$$SEE_Y(x) = \hat{\sigma}_Y(x) = \sqrt{Var(\hat{e}_Y(x))}. \quad (42)$$

Y'nin X'e eşitlenmesindeki SEE;

$$SEE_X(y) = \hat{\sigma}_X(y) = \sqrt{Var(\hat{e}_X(y))}. \quad (43)$$

Kernel eşitleme yöntemlerini karşılaştırılmak için, iki Kernel eşitlemeye ait SEE arasındaki farkı gösteren SEED kullanılır (von Davier vd., 2004).

$$SEED_Y(x) = \sqrt{Var(\hat{e}_1(x) - \hat{e}_2(x))}. \quad (44)$$

Madde Tepki Kuramına Dayalı Eşitleme

Gözlenen puan eşitlemeye dayalı yöntemlerin bazı güçsüz noktalarından dolayı MTK'ye dayalı test eşitleme önem kazanmıştır. Gözlenen puan eşitlemede eşitlik, simetri ve gruptan bağımsızlık varsayımlarının karşılanması sorunlar çıkmaktadır. Bunun yanı sıra MTK'de madde parametreleri gruptan, yetenek parametreleri ise cevaplanan madde setinden bağımsız olarak kestirilebilmektedir. Bu nedenle, sınava giren bir kişinin zor bir sınava veya kolay bir sınava girmesinin bir önemi yoktur. Yetenek tahminleri, örnekleme dalgalanmalarına tabi olarak karşılaştırılabilir olacaktır (Hambleton ve Swaminathan, 1985).

MTK ile eşitleme yapılabilmesi için karşılanması gereken varsayımlar bulunmaktadır. Bu varsayımlardan birincisi tek boyutluluktur. Tek boyutluluk, bireyin test performansının altında yalnızca tek bir gizil yeteneğinin (örn. matematik yeteneği) olduğunu varsayar (De Ayala, 2009). Ancak bazı bilişsel ya da kişisel faktörler (test kaygısı, motivasyon düzeyi vb.) test performansı üzerinde etkili olabildiği için bu varsayım tamamen karşılanmayabilir. Bu varsayımın karşılanması için, test performansını etkileyen baskın bir faktörün olması yeterli görülmektedir. Testle ölçülmek istenen özellik, baskın faktör olarak ifade edilmektedir (Hambleton ve Swaminathan, 1985). İkinci varsayım ise yerel bağımsızlıktır. Yerel bağımsızlık, testle ölçülen özellik sabit tutulduğunda bir bireyin iki maddeye verdiği tepkilerin istatistiksel olarak birbirinden bağımsız olması şeklinde tanımlanmaktadır. Diğer bir deyişle, farklı maddelere verilen tepkilerin arasında ilişki olmaması anlamına gelir (Lord, 1980).

MTK modelleri tek boyutlu ve çok boyutlu modeller şeklinde ikiye ayrılmaktadır. Tek boyutlu modeller iki kategorili ve çok kategorili olarak; çok boyutlu modeller ise telafi edici ve telafi edici olmayan modeller olarak sınıflandırılır (Embretson ve Reise, 2000).

MTK'ye dayalı eşitleme 3 adımda ifade edilebilir (Cook ve Eignor, 1991):

- Verilere uygun bir MTK modelinin seçildiğini varsayarsak ilk adım, eşitleme için bir veri toplama desenini seçmektir. MTK ile eşitleme kullanırken, model varsayım ihlallerinin etkilerini en aza indiren bir eşitleme deseni seçmek önemlidir (Kolen ve Brennan, 2014).
- İkinci adım kalibrasyon işleminin gerçekleştirilmesidir. Bazı eşitleme desenlerinde bu adım gerekli değildir. Tek bir kalibrasyon çalışması kullanılırsa,

eşitlenecek iki test formu için madde parametresi tahminleri otomatik olarak aynı ölçekte olacaktır. Bu şekilde **eş zamanlı kalibrasyon** yetenek ve madde parametresinin tahminlerini ortak bir ölçeğe yerleştirir ve eşitleme ihtiyacını ortadan kaldırır (Hambleton ve Swaminathan, 1985). Yani MTK’de iki testin madde parametreleri aynı ölçekteyse eşitleme gerekli olmayabilir. Bu nedenle MTK’de eşitleme ortak bir ölçek geliştirmeye indirgenmiştir (Kim ve Cohen, 1998). İki testin eş zamanlı kalibrasyonunun mümkün olmaması durumunda, yetenekleri ilişkilendirmek için eşitleme gerekli olabilir. Bu durumda, her iki test formu için ayrı ayrı madde ve yetenek parametreleri kestirilir. Bu **ayrı kalibrasyon** işleminde kestirilen parametreler için doğrusal bir dönüşüm işlemi yapılır (Cook ve Eignor, 1991).

MTK’de ölçeklerin dönüşümü benimsenen eşitleme desenine bağlı olarak değişecektir. Örneğin tek grup deseni ve EG deseninde; yetenekler her iki grupta da aynı ortalama ve standart sapmaya sahip olacak şekilde ölçeklendirildiğinden başka bir dönüşümün gerekli olduğu varsayılmaz. NEAT deseninde, X formu ve Y formunun parametreleri ayrı ayrı tahmin edilir. Bu formları alan gruplar eş değer kabul edilmediğinden parametre tahminleri aynı ölçekte değildir. Bununla birlikte iki form için ortak olan bir dizi madde vardır. Bu ortak maddeler için madde parametre tahminleri, ölçek dönüşümü için kullanılabilir (Gonzalez ve Wiberg, 2017).

Yetenek ve madde parametreleri için doğrusal dönüşüm denklemleri aşağıdaki eşitliklerde verilmiştir (Cook ve Eignor, 1991).

$$\theta^* = A\theta + B \quad (45)$$

θ : Kişinin yetenek düzeyi

θ^* : Kişinin girmedığı sınavdan alacağı kuramsal yetenek düzeyi

A : Eşitleme denkleminin eğimi

B : Eşitleme denkleminin sabiti

$$a_i^* = \frac{1}{A} a_i \quad (46)$$

$$b_i^* = Ab_i + B \quad (47)$$

$$c_i^* = c_i \quad (48)$$

a_i : Ayırıcılık parametresi

a_i^* : Dönüştürülmüş ayırıcılık parametresi

b_i : Güçlük parametresi

b_i^* : Dönüştürülmüş güçlük parametresi

c_i : Şans parametresi

c_i^* : Dönüştürülmüş şans parametresi

Görüldüğü gibi sadece a ve b parametreleri dönüştürülür, şans parametresi ölçek dönüştürmeden bağımsızdır.

Madde ve yetenek parametrelerini ortak bir ölçeğe yerleştirmek amacıyla farklı yöntemler geliştirilmiştir. Bu yöntemler eşitleme için gerekli olan katsayıları (A ve B) elde etmeyi amaçlamaktadır. Literatürde bu yöntemler; moment yöntemleri ve karakteristik eğri yöntemleri olmak üzere iki başlık altında incelenir.

Moment Yöntemleri. Ortak maddelerin madde parametre tahminlerinin ortalamalarını ve standart sapmalarını değiştirmek, ölçekleri dönüştürmenin en kolay yoludur. Literatürde moment yöntemleri ortalama-ortalama ve ortalama-standart sapma yöntemleri olarak iki kategoride değerlendirilmektedir. Bu katsayılar aşağıdaki gibi tanımlanır (Kolen ve Brennan, 2014).

Ortalama-Ortalama Yöntemi.

$$A = \frac{\mu(a_i)}{\mu(a_j)} \quad (49)$$

$$B = \mu(b_j) - A\mu(b_i) \quad (50)$$

Ortalama-Standart Sapma Yöntemi.

$$A = \frac{\sigma(b_j)}{\sigma(b_i)} \quad (51)$$

$$B = \mu(b_j) - A\mu(b_i) \quad (52)$$

Bazen ortalama-ortalama yöntemi ortalama-standart sapma yerine tercih edilmektedir. Bunun sebebi b parametresi tahminlerinin a parametresi tahminlerinden daha kararlı olmasıdır. Bununla birlikte Baker ve Al-Karni (1991) ortalamaların tipik olarak standart sapmalardan daha kararlı olduğunu ve bu sebeple ortalama-ortalama yönteminin tercih edilebileceğini belirtmiştir.

Karakteristik Eğri Yöntemleri. Şimdiye kadar ele alınan yöntemlerle ilgili potansiyel bir sorun; a , b ve c parametresi tahminlerinin çeşitli kombinasyonları hemen hemen aynı madde karakteristik eğrileri ürettiğinde ortaya çıkar. Örneğin, iki tahminde çok farklı b parametresi tahminlerine sahip bir madde çok benzer madde karakteristik eğrilerine sahip olabilir. Bu durumda iki tahmindeki maddeler için madde karakteristik eğrileri çok benzer olmasına rağmen, ortalama-standart sapma yöntemi b parametresi tahminleri arasındaki farktan aşırı derecede etkilenebilir. Bu sorun şimdiye kadar açıklanan ölçek dönüştürme yöntemlerinin bütün madde parametre tahminlerini eş zamanlı olarak ele almamasından kaynaklanmaktadır. Bu soruna cevaben aşağıdaki yöntemler geliştirilmiştir (Kolen ve Brennan, 2014).

Haebara Yaklaşımı. Her bir madde için oluşturulan karakteristik eğriler arasındaki farkın karelerinin toplamıdır.

$$Hdiff(\theta_i) = \sum_{j:V} [p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - p_{ij}\left(\theta_{ji}; \frac{\hat{a}_{lj}}{A}, Ab_{lj} + B, c_{lj}\right)]^2 \quad (53)$$

$$Hcrit = \sum_i Hdiff(\theta_i) \quad (54)$$

$p_{ij}(\theta_{ji}; a_{jj}, b_{jj}, c_{jj})$: Madde karakteristik fonksiyonu

$p_{ij}\left(\theta_{ji}; \frac{\hat{a}_{lj}}{A}, Ab_{lj} + B, c_{lj}\right)$: Eşitlenmiş madde karakteristik fonksiyonu

Stocking-Lord Yaklaşımı. Her bir madde için oluşturulan karakteristik eğriler arasındaki farkın toplamının karesidir.

$$SLdiff(\theta_i) = \left[\sum_{j:V} p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - \sum_{j:V} p_{ij}\left(\theta_{ji}; \frac{\hat{a}_{lj}}{A}, Ab_{lj} + B, c_{lj}\right) \right]^2 \quad (55)$$

$$SLcrit = \sum_i SLdiff(\theta_i) \quad (56)$$

$p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj})$: Madde karakteristik fonksiyonu

$p_{ij}\left(\theta_{ji}; \frac{\hat{a}_{lj}}{A}, Ab_{lj} + B, c_{lj}\right)$: Eşitlenmiş madde karakteristik fonksiyonu

Ortalama-ortalama ve ortalama-standart sapma yöntemleriyle karakteristik eğri yöntemlerini karşılaştıran araştırmalar, genellikle karakteristik eğri yöntemlerinin daha kararlı sonuçlar ürettiğini bulmuştur (Baker ve Al-Karni 1991; Hanson ve Beguin 2002).

- Üçüncü adım ise eşitlemedir. Eğer test puanları yetenek ölçeğindeki puanlar olarak bildirilirse eşitleme; kalibrasyon ve daha sonra parametre tahminlerinin aynı ölçeğe yerleştirilmesi yoluyla gerçekleştirilir ve bu adım gerekli değildir (Cook ve Eignor, 1991). Bunun dışında MTK gerçek puan eşitleme ve MTK gözlenen puan eşitleme yöntemleri kullanılabilir (Hambleton ve Swaminathan, 1985).

MTK Gerçek Puan Eşitleme. Test puanları tahmini gerçek puanlar açısından raporlanacaksa, gerçek puanlar çeşitli yetenek seviyeleri için tahmin edilir ve tablo halinde veya grafikte gösterilir (Hambleton ve Swaminathan, 1985). Bir formda belirli bir yetenek düzeyine karşılık gelen gerçek puanın diğer formda o yetenek düzeyindeki gerçek puana eşit olduğu kabul edilir. Ancak, sınava girenlerin gerçek puanları hiçbir zaman bilinemez (Kolen ve Brennan, 2014). Bu noktada gerçek puan eşitleme ile ilgili pratik bir sorundan söz edilmelidir. Bu sorun iki test formunda eşitlenecek puanların tahmini gerçek puanlar değil gözlenen puanlar olması gerçeğiyle ilgilidir (Cook ve Eignor, 1991). Fakat puanları bu şekilde ele almanın teorik bir sebebi yoktur. Lord ve Wingersky (1984) bu durumu gerçek puan dönüşümlerinin gözlemlenen puan dönüşümlerine benzer olduğunu göstererek gerekçelendirilmiştir.

MTK Gözlenen Puan Eşitleme. Başka bir prosedür olan gözlenen puan eşitlemede kestirilen parametrelerle formlar için gözlenen puan dağılımları oluşturulur. Daha sonra her iki formda elde edilen dağılımlar eşit yüzdelikli eşitleme ile eşitlenir (Hambleton ve Swaminathan, 1985). Bu yöntem sınava giren popülasyondaki yetenek dağılımının açık bir şekilde belirtilmesini gerektirir. MTK gözlenen puan eşitleme ile karşılaştırıldığında MTK gerçek puan eşitleme; (a) daha kolay hesaplama ve (b) yetenek dağılımına bağlı olmayan dönüştürme avantajlarına sahiptir (Kolen ve Brennan, 2014). Gözlenen puan eşitleme teorik bir dayanağı olmasına rağmen, daha karmaşık bir süreç gerektirir. Bu sebeple uygulamada gerçek puan eşitlemeye göre daha az kullanılır.

Bayes Yaklaşımı

Bayesçi yöntemlere ihtiyaç duyulmasının temel nedenlerinden biri sıfır hipotezi anlamlılık testine dayanan geleneksel veri analizi yöntemlerinin (örneğin, maksimum olabilirlik tahmini) bilinen bazı sorunlarının olmasıdır (Cashen ve Geiger, 2004; Cortina ve Landis, 2011; Lance, 2011). Bu sorunların başında, bir sonucun “istatistiksel olarak anlamlı” olduğunu belirtmenin temelini iyi tanımlanmamış olması gelmektedir. Geleneksel veri analizinde p değeri olarak adlandırılan değerin herhangi bir veri kümesi için özgün bir değeri yoktur. Bu değer araştırmanın amacına ve örneklem büyüklüğüne göre farklılaşabilmektedir. Bir diğer sorun da parametrelerle ilgili yetersiz tahminler üretmeleridir. Geleneksel yöntemler çoğu zaman birçok hesaplama kısıtlamasına sahiptir ve birçok varsayıma dayanmaktadır. Bayes analizi geleneksel yöntemlerin sorunlarını çözmekte ve birçok avantaj sağlamaktadır. Bayes analizinde p değeri yoktur, çıkarımlar tüm parametrelerle ilgili zengin ve eksiksiz bilgi sağlar ve modeller farklı veri türleri için kolayca özelleştirilebilir. Bayesçi yöntemler ayrıca bir deneyin hedefine ulaşma olasılığını (yani istatistiksel güç veya tekrarlama olasılığı) tutarlı bir şekilde tahmin eder (Kruschke, 2010).

Geleneksel istatistiksel tekniklerde, sıfır hipotezi her zaman ilgilenilen değişkenler arasında hiçbir ilişki olmadığını varsayacak şekilde kurulur. Bu durum değişkenler arasındaki ilişki hakkında hiçbir fikir olmadığında mantıklıdır. Bununla birlikte araştırmacıların değişkenlerle ilgili daha önceki çalışmalara dayanan bir önsel bilgiye sahip olmaları sıklıkla karşılaşılan bir durumdur. Bayesçi yöntemlerde, modelin tahmin edilmesine yardımcı olmak için bu arka plan bilgisi (“önsel” olarak adlandırılan) kullanılır. Bu yöntemde verilerden öğrenilebilir ve gelecekteki araştırmalara yeni bilgiler dahil edilebilir. Geleneksel (yani frekansçı) çerçevede olduğu gibi bir olayın sonsuz kez tekrarlanması kavramına güvenilmez. Bayesçi yaklaşımlarda önceki araştırmaların sonuçlarını göz ardı ederek temelde aynı sıfır hipotezini tekrar tekrar test etmek yerine, parametrelerin tahminine yardımcı olmak için önsel bilgiler sürece dahil edilir (van de Schoot vd., 2013; van de Schoot ve Depali, 2014).

Bayesçi ve geleneksel yaklaşımların her ikisi de bilinmeyen parametrelerle ilgilenmektedir (Kaplan ve Depaoli, 2012; van de Schoot ve Depaoli, 2014). Bayesçi istatistiksel çıkarım ve geleneksel yöntemler arasındaki temel fark, bilinmeyen parametrelerin doğası ile ilgilidir. Geleneksel yaklaşımda ilgilenilen bir parametrenin bilinmediği ancak sabit olduğu varsayılır. Yani popülasyonda yalnızca bir gerçek

popülasyon parametresi, örneğin bir ortalama veya bir regresyon katsayısı olduğu varsayılır. Bayesçi yaklaşımda tüm bilinmeyen parametreler belirsiz olarak ele alınır. Bu nedenle de parametreler bir olasılık dağılımı ile tanımlanmalıdır (van de Schoot vd., 2013).

Popülasyon parametresi hakkında taşıdıkları kesinlik miktarına göre farklılık gösteren üç tür önsel bilgi bulunmaktadır: (1) bilgilendirici olmayan önseller (2) bilgilendirici önseller (3) zayıf bilgilendirici önseller. Bilgilendirici olmayan önseller, popülasyon parametresinin neye benzediğine dair büyük ölçüde belirsizliği yansıtmak için kullanılır. Zayıf bilgilendirici önseller modele bazı bilgiler dahil eder ve bilgilendirici olmayan önsellere kıyasla popülasyon parametresi hakkında daha fazla kesinlik yansıtır. Bu önsel bazı faydalı bilgiler içerir fakat tipik olarak nihai parametre tahmini üzerinde çok fazla etkisi yoktur. Son olarak popülasyon parametresi hakkında en fazla kesinliği içeren önseller bilgilendirici önsellerdir. Bilgilendirici önseller modelin tahmini için çok önemli olan ve nihai tahminler üzerinde büyük bir etkiye sahip olabilen kesin sayısal bilgiler içerir (van de Schoot ve Depali, 2014).

Bayes istatistiğinin neden seçilebileceğine dair dört sebep bulunmaktadır: (1) karmaşık modeller bazen geleneksel yöntemler kullanılarak tahmin edilemez, (2) güven aralıklarından olasılık tanımı tercih edilebilir, (3) arka plan bilgisi analizlere dahil edilebilir ve (4) yöntem büyük örneklemelere bağlı değildir (van de Schoot ve Depali, 2014). İlk olarak, bazı karmaşık modeller geleneksel istatistiksel yöntemler kullanılarak tahmin edilemez. Bayes tahmini, istatistikte yaygın olarak karşılaşılan bazı sorunların üstesinden gelebilir. Örneğin; zorlu parametre tahminleri elde etmek, model tanımlamaya yardımcı olmak, daha doğru parametre tahminleri üretmek ve küçük örneklem boyutlarının olduğu durumlarda çıkarımlar yapmak. İkinci olarak, Bayes istatistiklerinde güven aralıkları yorumlanmaktadır. Bayesçi güven aralıkları sezgisel olarak daha çekici olabilir. Üçüncü olarak, Bayes istatistiğinde modeldeki parametreler hakkında önceki yayınlara dayalı olarak halihazırda bilinen her şey bilgilendirici önselleri belirlemek için kullanılabilir. Dördüncü olarak, Bayes istatistiği büyük örneklemelere (yani merkezi limit teoremine) dayanmaz ve bu nedenle arka plandaki matematiğin çalışması için büyük örneklemeler gerekmez. Birçok makale küçük veri seti bağlamında Bayes istatistiğinin faydalarını göstermiştir (örn. Zhang, Hamagami, Lijuan Wang, Nesselroade ve Grimm, 2007). Ayrıca bazı analizler için karşılanması gereken varsayımlar (eşitleme için süreklileştirme veya düzgünleştirme, örneklem büyüklüğü gibi) Bayes için söz konusu değildir (Kruschke, Aguinis ve Joo, 2012; Kruschke, 2013; van de Schoot vd., 2013).

Bir kiři herhangi bir olaya tanık olmadan, yazı tura gibi řans olaylarında önceden inançlara sahip olabilir. Bunlara “önsel inançlar” ya da yazı tura ile ilgili “olasılıklar” veya “oranlar” denir. Bazı yazı tura olayları gözlemlendikten sonra, gelecekteki olayların olasılığı hakkındaki inanç “sonsal inançlar” olarak değıştirilebilir. Madeni paranın on kez tura geldiğini gördükten sonra, madeni paranın hilesiz olduğuna olan inanç değıştirilebilir. Bu revize edilmiş inanç sonsal inanç haline gelir ve buna koşullu olasılık denir. Çünkü başlangıçta inanılan olasılık arka arkaya on tura görme koşulu ile değıştirilmiştir (Brown, 1981). Bayes teoremi, bir olayın olma olasılığının diğeri olayın olma olasılığına bağılı olduğu koşullu olasılık teoreminden ortaya çıkmaktadır. Bayesçi yöntemlerde önsel bilgi, iki olaydan gerçekleşmiş olandır.

Bayesçi çıkarım üç adımda özetlenebilir: (1) olasılıksal modelin oluşturulması, (2) sonsal dağılımın hesaplanması, (3) modelin kontrol edilmesi. İlk olarak, gözlenen y değerleri ve θ parametreleri için bileşik olasılık yoğunluk dağılımı tanımlanır. Bu iki nicelik için olasılık yoğunluk fonksiyonu řu şekilde ifade edilebilir:

$$p(\theta, y) = p(y | \theta)p(\theta) = p(\theta | y)p(y) \quad (57)$$

Burada $p(\theta)$ önsel dağılım olarak adlandırılır ve nicelik hakkındaki önsel bilgimizi temsil eder. $p(y | \theta)$ ise gözlemlediğimiz verinin dağılımıdır. Bayes teoremi ile sonsal dağılım aşağıdaki şekilde ifade edilebilir:

$$p(\theta | y) = \frac{p(\theta)p(y | \theta)}{p(y)} \quad (58)$$

Bayesçi istatistikte amaç θ parametresinin bir dağılımını elde etmek olduğundan, bu terim ihmal edilirse sonsal dağılım aşağıdaki gibi ifade edilebilir:

$$p(\theta | y) \propto p(y | \theta)p(\theta) \quad (59)$$

Bu denklem başka bir deyişle;

$$\text{Sonsal olasılık} \propto \text{Olabilirlik} * \text{önsel Olasılık} \quad (60)$$

şeklinde ifade edilebilir (Gelman, vd., 2014).

PoB Yaklaşımı ile Eşitleme

Cevapların doğru veya yanlış olarak kodlandığı çoktan seçmeli testler için, ölçme programlarında testteki doğru cevap sayısına karşılık gelen toplam puanlar kullanılır. Toplam puanların alabileceği olası değerler ardışık tamsayılar olduğundan, diğer testte bunlara karşılık gelen puanları kullanılamaz hale getiren dağılımların ayrıklığı önemli bir sorun oluşturmaktadır. Eşitleme literatüründe bu soruna verilen ortak çözüm, ayırık puan dağılımlarını süreklileştirmektir. Farklı süreklileştirme yöntemleri; parametrik, parametrik olmayan veya yarı parametrik istatistiksel çıkarımlar üreten farklı eşitleme yöntemlerini tanımlar (Gonzalez ve von Davier, 2013).

PoB yaklaşımında Karabatsos ve Walker (2009), iki testten alınan puanların dağılımları arasındaki pozitif bağımlılığı açıklayan bir test eşitleme yaklaşımı önermiştir. Bu yaklaşımda eşitleme fonksiyonu, parametrik olmayan bir şekilde tahmin edilerek elde edilmektedir. Parametrik Bayes, model oluşturabilmek için sınırlı sayıda parametre kullanırken, PoB modellerinde bu şekilde bir sınırlama yoktur. Parametre sayısının esnek olması ve kullanılan parametre uzayının sonsuz boyutlu olması PoB modelinin temelini teşkil etmektedir (Müller ve Quintana, 2004).

Çoğu sınavda test puanlarının yanı sıra tamamlayıcı bilgiler (örneğin sınava giren kişinin cinsiyeti, okul türü vb. gibi arka plan değişkenleri) mevcut olsa da şu anda kullanımda olan eşitleme yöntemlerinden çoğu eşitleme fonksiyonunu tahmin etmek için bu tür bilgileri kullanmamaktadır. Bununla birlikte faktör değişkenleri, bir eşitleme fonksiyonunu tahmin etmek için verilerin (puanların) toplandığı sınav katılımcılarını tanımlamak için yaygın olarak kullanılmaktadır. Örneğin EG deseninde, sınava girenlerden oluşan grup yarıya bölünür ve her yarıya iki farklı sınav formu verilir, böylece faktör değişkeni uygulanacak formu tanımlar. Bununla birlikte, ortak değişkenlerin kullanımı eşitleme literatüründe oldukça keşfedilmemiş bir konu gibi görünmektedir (Gonzalez vd., 2015a).

Branberg ve Wiberg (2011) puan dağılımlarının ortalamasının ortak değişkenlere bağlı olduğu gözlenen puan lineer eşitlemede ortak değişkenlerin kullanılmasını önermiştir. Bayesçi istatistiklerde önsel bilgiler genellikle ortak değişkenlerden oluşmaktadır (Boone, Ye ve Smith, 2009). Testleri birbirine eşitlemede çoğunlukla parametrik ve yarı parametrik yöntemler tercih edilmektedir (Gonzalez vd., 2015a). Fakat geleneksel eşitleme yöntemleri eşitleme dönüşümlerinde arka plan değişkenlerinin kullanılmasında sınırlı kalmaktadır. PoB modeli ise eşitleme fonksiyonlarının

kestirilmesinde “ortak değişkenlerin” kullanılmasına izin vermektedir. Bu yöntem dağılım fonksiyonlarının konumlarını ve şeklini, regresyon modellerindeki gibi sadece ortalamalar veya dağılım fonksiyonlarıyla değil ortak değişkenlerle (z) de değiştirmektedir.

Bayesçi eşitleme modelinin önemli bir avantajı EG desenlerinde, daha önceki gözlenen puan eşitleme yöntemlerinde sıklıkla varsayıldığı gibi test puanlarının dağılımının bağımsız değil aslında ilişkili olduğunu açıklamasıdır. Pratikte eşitlenecek olan iki testin aynı psikolojik yapıyı ölçmek için tasarlandığı göz önüne alındığında bu bağımlılık makul görünmektedir (Karabatsos ve Walker, 2011). PoB yönteminde dağılımların süresizliği ve düzensizliği giderildiğinden, geleneksel yöntemlerde yapılan ön düzgünleştirme, bant genişliği parametresi seçme ve SEE üretilmesine gerek duymamaktadır (Gonzalez vd., 2015b). Ayrıca PoB modelinde, eşitleme fonksiyonu %95 güvenilir aralıkta tahmin edilmektedir. Bu nedenle Bayes modeli, herhangi bir örneklem boyutu için eşitlenmiş puanlardaki belirsizliği tam olarak hesaba katmak için bir yol sağlamaktadır. Buna karşılık gözlenen puan eşitlemeye yönelik önceki tüm yaklaşımlar, eşitlenen puanın güven aralığını tahmin etmek için yalnızca büyük örneklem yaklaşımlarına dayanmaktadır (Karabatsos ve Walker, 2011).

PoB modelleri, geleneksel parametrik alternatiflere kıyasla sağladıkları esneklik nedeniyle son yıllarda büyük ilgi görmüştür. Herhangi bir PoB modelinin ana bileşeni, dağılım fonksiyonları üzerindeki önsel süreçlerdir. Bu nedenle, genel olarak rastgele olasılık modelleri olarak adlandırılmaktadır. Parametrik olmayan olasılık dağılımları ile ilgili farklı modeller mevcuttur (Barrientos, Jara ve Quintana, 2012). En popüler parametrik olmayan olasılık modeli Dirichlet Sürecidir (Dirichlet Process – DP) ve elde edilen olasılık fonksiyonu ayrık bir yapıya sahiptir. Önerilen bir diğer yöntem Bernstein Polinomlarıdır (Bernstein Polynomials – BP) ve bu yöntemde olasılık fonksiyonu sürekli bir yapıya sahiptir. Daha sonra Barrientos ve diğerleri (2012) Bağımlı Bernstein Polinom Süreci (Dependent Bernstein Polynomial Process – DBPP) modelini geliştirmiştir. Bu model, normal koşullarda hatta sınırlanma problemi olmayan koşullarda daha etkili bir performans göstermektedir.

Dirichlet Süreci ve Bernstein Polinomları. En popüler PoB olasılık modeli, Ferguson (1973) tarafından tanımlanan DP’dir. DP’yi tanımlamanın uygun bir yolu,

$G \sim DP(m, G^*)$ 'nin aşağıdaki gibi ifade edilebileceğini belirten Sethuraman'ın (1994) temsildir;

$$G(\cdot) = \sum_{i=1}^{\omega} \omega_i \delta_{\theta_i}(\cdot) \quad (61)$$

Burada $\theta_i, \theta_i \sim^{iid} G^*, \omega_i = v_i \prod_{j < i} (1 - v_j)$ ve $v_j \sim^{iid} Beta(1, m); \delta_{\theta_i}(\cdot)$ üzerinde bir noktasal kümedir. Bu eşitlik stick-breaking temsili olarak adlandırılır. Tanım gereği, ortaya çıkan rastgele olasılık fonksiyonu G ayrıktır. DP'nin ayrık doğası, yoğunluklar için bir olasılık modeli olarak kullanılmaması gerektiğini gösterir. Bu problemle başa çıkmak için standart bir yaklaşım, genellikle DPM (DP Mixture) olarak adlandırılan, DP'ye dayalı düzgün yoğunlukların bir karışımını tanımlamaktır (Gonzalez vd., 2015a).

İstatistik literatüründe PoB yöntemlerinin uygulamalarındaki artış, büyük ölçüde DPM modellerinin sonsal hesaplama için basit ve verimli yöntemler içermesi ile açıklanabilir (Ferguson, 1983). DPM modelleri, Bayes hiyerarşik modellerindeki bileşenler için DP önsellerini içerir ve bu da son derece esnek bir model sınıfıyla sonuçlanır.

Bir DPM modeli f , şu şekilde tanımlanır;

$$f(\cdot | G) = \int_{\theta} \psi(\cdot, \theta) G(d\theta). \quad (62)$$

Burada her $\theta \in \Theta$ için, $\psi(\cdot, \theta)$ bir olasılık yoğunluk fonksiyonudur; $\Theta \subseteq R^q$ ve G, θ üzerinde tanımlanan bir DP'dir. Esneklikleri ve uygulamadaki kolaylıkları nedeniyle, DPM modelleri artık çok çeşitli uygulamalarda kullanılmaktadır. Ayrıca sonsal tutarlılık ve yakınsama oranları hakkında zengin bir teorik literatür, tekil (single) yoğunluk tahminlerinde çıkarım için DPM modellerinin kullanımını önerir (Gonzalez vd., 2015a).

Önerilen yöntemin temelini oluşturacak bir diğer önemli durum önsel BP'dir. H , $[0,1]$ aralığında bir fonksiyon olsun; H 'nin k 'inci dereceden BP'si aşağıdaki şekilde tanımlanır (Lorentz, 1986);

$$B(t; k, H) = \sum_{j=0}^k H\left(\frac{j}{k}\right) \binom{k}{j} t^j (1-t)^{k-j}. \quad (63)$$

Bu polinom, k sonsuza gittiğinde ve H sürekli bir fonksiyon olduğunda, H 'ye düzgün bir şekilde yaklaşma özelliğine sahiptir. Ayrıca H , desteği birim aralık olan bir dağılım fonksiyonu ise $B(t; k, H)$ 'de $[0,1]$ üzerinde bir dağılım fonksiyonudur ve ayrıca $H(0) = 0$ ise, $B(t; k, H)$ yoğunluk fonksiyonuna sahiptir;

$$b(t; k, H) = \sum_{j=1}^k \omega_{jk}^* \beta(t|j, k-j+1). \quad (64)$$

Burada $\omega_{jk}^* = H(j/k) - H((j-1)/k)$ ve $\beta(\cdot | a, b)$ pozitif parametreler a ve b ile beta yoğunluğunu gösterir. Petrone (1999) bu tanımdan yola çıkarak hem H hem de k 'yı rastgele nicelikler olarak kabul ederek, $[0,1]$ aralığında tanımlanan dağılımlar uzayında $b(\cdot; k, H)$ 'ye dayalı olarak ortaya çıkan rastgele yoğunluğu önermiştir. Bu nedenle eğer $H \sim DP(m, G^*)$ ise, rastgele bir $B(\cdot; k, H)$ dağılımının DP önselini takip ettiği söylenir. Bu son durumda, $b(t; k, H)$ 'nin aşağıdaki gibi bir DPM modeli olarak yazılabileceğinin gözlemlenmesi önemlidir:

$$\begin{aligned} b(t; k, H) &= \int_{(0,1]} \beta(t | \lceil \theta k \rceil, k - \lceil \theta k \rceil + 1) H(d\theta) \\ &= \sum_{i=1}^{\omega} \omega_i \beta(t | \lceil \theta_i k \rceil, k - \lceil \theta_i k \rceil + 1). \end{aligned} \quad (65)$$

Burada θ_i ve ω_i Eşitlik 61'deki gibi tanımlanır. Dolayısıyla Eşitlik 65'teki model sonsuz bir beta yoğunlukları karışımına karşılık gelir.

Bağımlı Rastgele Olasılık Modelleri. Birçok uygulamada, verilerin ortak değişkenlere bağlı olmasına izin verilmesi istenir. Bu fikir genişletilerek, PoB yaklaşımı altında bir dizi ortak değişkene bağlı sürekli dağılım $\{G_z : z \in Z\}$ 'yi içeren bir olasılık modeli tanımlamak istenmiştir. Buradaki problem, sonucun iyi teorik özelliklere sahip

olması ve kolayca uygulanabilmesi amacıyla bir dizi ortak değişkene bağlı olasılık dağılımları için olasılık modeli oluşturmaktır (Gonzalez vd., 2015a).

MacEachern (1999), bağımlı DP (Dependent Dirichlet Process – DDP) olarak adlandırılan önsel bağımlı modellerden birini önermiştir. Burada bağımlılık, G 'nin stick-breaking gösteriminde ortak değişkenleri z ile göstererek, ω ve θ elemanlarını değiştirerek elde edilir;

$$G_z(\cdot) = \sum_{i=1}^{\omega} \omega_i(z) \delta_{\theta_i(z)}(\cdot). \quad (66)$$

Diğer önemli bir önsel durum, ortak değişkene bağlı puan dağılımlarını modellemek için kullanılacak olan bağımlı BP sürecidir (Dependent Bernstein Polynomials Processes – DBPP). Bu süreç, Barrientos, Jara ve Quintana (2017) tarafından tanıtılmış, Eşitlik 65'teki Dirichlet süreci H 'nin, Eşitlik 66'nın bağımlı bir işlemiyle değiştirilmesiyle tanımlanmıştır.

Barrientos ve diğerleri (2012) çalışmalarında, DBPP'nin iki özel durumunu belirtmektedir. DBPP'nin özel koşullarından biri “ortak ağırlıklar” ve “kestirici bağımlı destek noktaları” ile bağımlı stick-breaking süreci bir arada kullanıldığı zaman “tekil ağırlıklı DBPP (wDBPP)” modelini oluşturmaktadır.

Her $z \in Z$ için, F_z 'nin Lebesgue ölçüsüne göre aşağıdaki formda yoğunluk fonksiyonuna sahip olduğu varsayılmaktadır. Bu fonksiyon wDBPP'nin bir versiyonudur;

$$f(z)(\cdot) = \sum_{j=1}^{\omega} \omega_j \beta(\cdot \mid \lceil k\theta_j(z) \rceil, k - \lceil k\theta_j(z) \rceil + 1). \quad (67)$$

Burada $\omega_j = v_j \prod_{i < j} [1 - v_i]$, $v_1, v_2, \dots$ ortak dağılım $Beta(1, a)$ ile rastgele değişkenlerdir; k sonlu boyutlu bir parametre λ tarafından indekslenen dağılıma sahip ayrık bir rastgele değişkendir; $\theta_j(z) = h_z(r_j(z))$, $r_1, r_2, \dots$ bağımsız ve gerçek değerli rastlantısal süreçlerdir (ψ tarafından indekslenmiş kural ile); h_z , R üzerinde tanımlanan $[0; 1]$ değerli çift yönlü sürekli fonksiyondur.

Önsel Çıkarım. wDBPP modelinin hesaplamalı uygulaması, Markov Zinciri Monte Carlo (Markov Chain Monte Carlo – MCMC) yöntemlerine dayanmaktadır. wDBPP modeli için $h_z(\cdot) = \exp\{\cdot\}/(1 + \exp\{\cdot\})$, $r_j(z) = z^T Y_j$ ve $Y_j|\mu, S \sim^{iid} N_p(\mu, S)$, $j = 1, 2, \dots$ olduğu varsayılmaktadır. Model aşağıdaki gibi tamamlanmıştır;

$$\begin{aligned} v_j|a &\sim^{iid} \text{Beta}(1, a), & k|\lambda &\sim \text{Poisson}(\lambda) \mathbb{I}_{\{k>1\}} \\ \mu|m_0, S_0 &\sim N_p(m_0, S_0), & S|v, \psi &\sim IW_p(v, \psi). \end{aligned} \quad (68)$$

Burada $IW_p(v, A)$, serbestlik derecesi v olan p boyutlu İverted-Wishart dağılımını belirtmektedir. Gonzalez ve diğerleri (2015a) yapmış oldukları çalışmalarında önsel dağılımı oluşturmak için $\lambda = 25$, $m_0 = 0_p$, $S_0 = 2.25 * I_p$, $v = p + 2$ ve $a = 1$ değerlerini kullanmışlardır.

Sonsal Çıkarım. wDBPP modelinin sonsal dağılımını keşfetmek için bir MCMC algoritması uygulanmıştır. Hesaplamalı uygulama, bağımlı stick-breaking işleminde sonlu bir boyutsal yaklaşıma dayanmaktadır. Sonsal çıkarım için $\{v_j^{(l)}, Y_j^{(l)}\}_{j=1}^N$ ve $k^l, l = 1, 2, \dots, M$ örneklenen parametreleri göstermek üzere, M sonsal örnek boyutudur ve N stick-breaking işlemine yaklaşmak için kullanılan bileşenlerin sayısıdır. Sonlu stick-breaking yaklaşımının tüm j için $v_j^{(N)} = 1$ gerektirdiğine dikkat edin. Puan dağılımlarının sonsal çıkarımı aşağıdaki şekilde ifade edilebilir (Gonzalez vd., 2015a);

$$F^{(l)}(z)(\cdot) = \sum_{j=1}^N \omega_j^{(l)} \bar{\beta}(\cdot | \lceil k^{(l)} \theta_j^{(l)}(z) \rceil, k^{(l)} - \lceil k^{(l)} \theta_j^{(l)}(z) \rceil + 1). \quad (69)$$

Burada $\omega_j^{(l)} = v_j^{(l)} \prod_{i<j} [1 - v_i^{(l)}]$, $\bar{\beta}(\cdot | a, b)$; pozitif parametreler a ve b ve $\theta_j^{(l)} = h_z(Z^T Y_j^{(l)})$ ile beta kümülatif dağılım fonksiyonunu belirtir.

İlgili Araştırmalar

Bu bölümde araştırmanın problem durumu kapsamında ilk olarak Bayeşçi yöntemlerle eşitlemenin, daha sonra Kernel eşitlemenin son olarak da eşit yüzdelikli eşitlemenin öncelikli olduğu çalışmalara yer verilmiştir. Bu çalışmalar içerisinde de öncelikle NEC deseninin kullanıldığı daha sonra NEAT ve EG desenlerinin kullanıldığı çalışmalar yıllara göre sıralanmıştır.

Karabatsos ve Walker (2009) yapmış oldukları çalışmada test eşitleme için PoB modelini tanıtmıştır. Bu modelin tüm eşitleme desenlerine uygulanabileceği ve gözlenen puan eşitlemesine yönelik önceki tüm yaklaşımlara göre önemli teorik ve pratik avantajlara sahip olduğu belirtilmiştir. Bayeşçi eşitleme modelinde EG desenlerinde önceki yöntemlerde sıklıkla varsayıldığı gibi test puanlarının dağılımlarının bağımsız olmak yerine ilişkili olduğu gerçeği açıklanmıştır. Pratikte eşitlenecek iki testin aynı psikolojik özelliği ölçmek için tasarlandığı düşünüldüğünde bu bağımlılık makul görünmektedir. Bayeşçi eşitleme modeli; Kernel, lineer, ortalama ve eşit yüzdelikli eşitleme modellerinden daha fazla hesaplama çabası ve matematiksel uzmanlık gerektirse de modelin temel avantajları göz önüne alındığında bu çaba göz ardı edilebilir. Önceki dört modelin veriler hakkında oldukça gerçekçi olmayan varsayımlarda bulunduğu ve asimetri, ayırık puanlar gibi diğer teknik sorunları taşıdığı iddia edilmiştir. Bayes modeli, bu tür sorunları önlemesi açısından diğer modellerden daha iyi performans göstermiştir. Ayrıca ortalama, lineer ve eşit yüzdelikli eşitleme yöntemlerinin PoB modelinin özel durumları olduğu kanıtlanmıştır.

Gonzalez ve diğerleri (2015a) yapmış oldukları çalışmada eşitleme dönüşümüne yol açan puan dağılım fonksiyonlarının tahmininde ortak değişkenlerin kullanımına izin veren, test eşitleme için esnek bir yaklaşım olan PoB modelini kullanmıştır. Bu yöntemde eşitleme dönüşümleri ortak değişken değerlerine göre oluşturulur. Önerilen modelin gerçek ve simüle edilmiş verilere uygulamaları tartışılmış, elde edilen sonuçlar diğer mevcut eşitleme yöntemleriyle karşılaştırılmıştır. Bu modelle yapılan eşitleme; Kernel ve eşit yüzdelikli eşitleme yöntemlerine göre ya daha iyi ya da onlarla eşit performans göstermiştir. Bu yöntem uç değerlerde ve sıfır frekanslı değerlerde eşitleme fonksiyonunun kestirimini kolaylaştırmıştır. PoB yönteminin esnek bir yöntem olduğu, eşitleme için gerekli olan birçok varsayımın karşılanmasına gerek olmadığı ve ortak maddenin olmadığı durumlarda da eşitlemenin yapılabileceği belirtilmiştir.

Yurtçu, Kelecioğlu ve Boone (2021) yapmış oldukları çalışmada PoB modelini kullanarak puanların eşitlenmesini temel almıştır. Araştırmanın örneklemini PISA 2012 verisindeki Kanada ve İtalya örneklemleri oluşturmuştur. Araştırmada; cinsiyet, matematik öz yeterlik puanları ve ortak madde puanları ortak değişken olarak kullanılmıştır. Eşitlenmiş puanların dağılımının hedef teste uzaklıkları incelenmiştir. Bu puanlar MTK'ye dayalı eşitleme yöntemlerinden elde edilen eşitlenmiş puanların hedef teste uzaklıklarıyla karşılaştırılmıştır. PoB modelinde, eşitlenmiş puanların kullanılan ortak değişkenlere göre değiştiği belirlenmiştir. Modelde cinsiyet, matematik özyeterlik puanları ve ortak madde puanları birlikte ortak değişken olarak alındığında, eşitlenmiş puanların hedef teste olan uzaklıkları birbirine yakın ancak dağılımları farklı çıkmıştır. Araştırma sonucunda bu modelin klasik yöntemlere göre daha fazla bilgi verdiği, ortak madde olmadığı durumlarda ortak değişkenlerin kullanılabileceği görülmüştür. Ayrıca bu yöntemin bazı durumlarda daha iyi sonuç verebileceği ve eşitlenmiş puanların hedef teste daha yakın olduğu belirlenmiştir.

Varas, Gonzalez ve Quintana (2020) yapmış oldukları çalışmada test eşitlemede puan dağılımları için Parametrik olmayan gizil Bayes yaklaşımını önermiştir. Eşitleme dönüşümünü tahmin etmek için farklı parametrik, yarı parametrik ve parametrik olmayan modeller önerilmiştir. Tüm bu yöntemlerde eşitleme dönüşümü, puanların aslında ayrık bir ölçekte tanımlandığı gerçeğini göz ardı ederek eşitleme sonucunda sürekli puanlar vermektedir. Bu çalışmada, orijinal ayrık ölçeğe uygun şekilde eşitlenmiş puanlar üreten bir eşitleme yöntemi tanıtılmıştır. Bu yöntemde puanlar rastgele sıralı değişkenler olarak alınır ve geleneksel eşitleme yöntemlerinde yaygın olarak kullanılan süreklileştirme adımından kaçınılır. Teklif edilen yöntemin performansı hem bir simülasyon çalışmasında hem de gerçek bir veri üzerinde gösterilmiştir. Parametrik olmayan gizil Bayes yaklaşımıyla elde edilen ayrık eşitlenmiş puanlar, eşit yüzdelikli eşitleme ve Gauss Kernel eşitleme yöntemleriyle karşılaştırılmıştır. Geleneksel eşitleme yöntemleriyle elde edilen eşitlenmiş puanların ayrık versiyonlarıyla karşılaştırıldığında, bu yaklaşım ölçeğin tüm aralığı dikkate alındığında daha iyi performans göstermiştir. Ele alınan hemen hemen tüm simüle edilmiş senaryolarda, önerilen yöntem her bir olası test puanındaki gerçek ayrık puanları doğru bir şekilde tahmin etmiştir. Ayrıca bu yöntem neredeyse tüm puan aralıkları için düşük SEE kestirimi yapmıştır.

Kim, Livingston ve Lewis (2011) yapmış oldukları çalışmada küçük örneklemelere uygulanan testlerde, puanların eşitlenmesini iyileştirmek için tanımlayıcı bilgilerin kullanılmasına yönelik deneysel Bayes yaklaşımını açıklamıştır. Aynı testin iki farklı

formu üzerinde yeniden örnekleme çalışmaları yapılmıştır. Her çalışmada zincirleme ortalama ve zincirleme lineer eşitleme yöntemlerinin deneysel ve deneysel olmayan Bayes sürümleri, ortak maddeli bir eşitleme için toplanan büyük bir veri kümesinden alınan küçük örneklemelere uygulanmıştır. Eşitleme kriteri büyük örneklemdeki zincirleme lineer eşitlemedir. Bir testin farklı formlarının eşitlenmesi tanımlayıcı bilgilerle sağlanmıştır. Hedef formun örneklem büyüklüğü 10 ile 200 arasında değişirken, temel formun örneklem büyüklüğü 200’de sabitlenmiştir. Çalışmada örneklem sayısının değişimi ve testlerin zorluk düzeyleri açısından farklılıkları ele alınmıştır. Hedef form için 50 veya daha az örneklemde deneysel Bayesin eşitlemenin doğruluğunu artırdığı gözlemlenmiştir. Ancak hedef testin temel testten daha zor olduğu durumlarda eşitlemenin doğruluğunu azalttığı görülmüştür.

Kim, Livingston ve Lewis (2009) yapmış oldukları çalışmada küçük örneklemelerde eşitlemenin doğruluğunu artırmak için tamamlayıcı bilgilerin kullanılmasının etkinliğini incelemiştir. Değerlendirme, dokuz farklı testi temsil eden 10 farklı test formunun büyük örneklem eşitlemelerinden elde edilen verileri kullanan yeniden örnekleme prosedürlerinden oluşmuştur. Her büyük örneklem eşitlemesi, yeniden örnekleme dayalı küçük örneklem eşitlemeleri için veri sağlamıştır. Diğer testlerin eşitlenmesinden elde edilen tanımlayıcı bilgiler deneysel Bayes kestirimine dahil edilmiştir. Küçük örneklem eşitlemeleri, tanımlayıcı bilgi olarak diğer dokuz test formunun eşitlenmesini kullanan iki deneysel Bayes yöntemi dahil olmak üzere beş yöntemle yapılmıştır. Bu yöntemler zincirleme lineer eşitleme, zincirleme ortalama eşitleme, tanımlayıcı bilgiyle zincirleme lineer eşitleme (deneysel Bayes zincirleme lineer), tanımlayıcı bilgiyle zincirleme ortalama eşitleme (deneysel Bayes zincirleme ortalama) ve bir sentetik eşitleme (ağırlıklandırılmış zincirleme lineer eşitleme)’dir. Eşitleme için örneklem ne kadar küçük olursa, diğer test formlarının eşitlenmesinden elde edilen tanımlayıcı bilgilerin değeri o kadar büyük olmuştur. Araştırmada eşitlenen form, tanımlayıcı bilgilerin elde edildiği formlarla oluşturulan modeli takip ederse tanımlayıcı bilginin, eşitlemenin doğruluğunu artıracak sonucuna ulaşılmıştır. Aksi takdirde, tanımlayıcı bilgilerin yardımcı olmayacağı hatta eşitlemenin doğruluğunu azaltabileceği belirtilmiştir. Bu araştırmada şaşırtıcı olan 25 ve 50 kişilik örneklemelerde tanımlayıcı bilgi olmadan ortalama eşitlemenin güçlü performansıdır. Zincirleme ortalama eşitleme yöntemi, bu küçük örneklemeler için sürekli olarak diğer yöntemlerden daha küçük eşitleme hatası (Root Mean Squared Error – RMSE) üretmiştir. Tanımlayıcı bilgilerin belirlenmesinin zorlu bir görev olduğu göz önüne alındığında, küçük örneklemelerde

tanımlayıcı bilgi olmadan eşitleme yapan zincirleme ortalama eşitlemenin tercih edilebileceği belirtilmiştir.

Oh, Guo ve Walker (2009) yapmış oldukları çalışmada farklı örneklem büyüklüklerinde çeşitli deneysel Bayes yaklaşımlarını karşılaştırmayı ve deneysel Bayes yönteminin küçük örneklemelerde geleneksel yöntemlerden daha kesin güvenilirlik kestirimleri üretilip üretilmeyeceğini araştırmayı amaçlamıştır. Bayes yaklaşımında tamamlayıcı bilgiler kullanılmıştır. Geleneksel olarak kullanılan güvenilirlik kestirimleri (örneğin, Cronbach'ın α) küçük örneklem için düşük hassasiyete sahip olabilir. Bu çalışmada birkaç Bayes kestirimi hem yanlılık hem de varyans açısından diğer yöntemlerle (örneğin geleneksel olarak kullanılan Cronbach'ın α katsayısı) elde edilen kestirimlerle karşılaştırılmıştır. Araştırmada deneysel Bayes yaklaşımının birkaç istisna dışında daha fazla yanlılık ürettiği ancak daha az hata ürettiği sonucuna ulaşılmıştır. Örneklem büyüklüğü arttıkça geleneksel yöntemler ile deneysel Bayes yönteminden elde edilen dağılımlar birbirine yaklaşmış ve RMSE azalmıştır. Deneysel Bayes yaklaşımlarına dayalı güvenilirliği tahmin etmek küçük örneklem grupları için umut vericidir. Fakat küçük örneklemelerde alt grupların örneklem sayısı azalacağından deneysel Bayes yöntemine dayalı güvenilirlik kestiriminin çok işlevsel olmayabileceği ifade edilmiştir.

Branberg ve Wiberg (2011) yapmış oldukları çalışmada eş değer ve eş değer olmayan gruplar desenine ortak değişkenler dahil edildiğinde gözlenen puan lineer eşitleme yöntemini incelemiştir. Çalışmanın temel amacı, ortak değişkenlerle bu ek bilgileri dahil etmenin yanlılık, varyans ve ortalama hata kareleri gibi kestirimler üzerindeki etkisini incelemektir. İlk olarak ortak değişkenlerle doğrusal olarak eşitlenen gözlenen puan eşitleme için bir model önerilmiştir. İkinci adım olarak model bir simülasyon çalışmasında kullanılmıştır. Burada amaç cinsiyet ve eğitim gibi ortak değişkenlerin kullanımının kestirimlerin ortalama hata karelerini azaltarak eşitleme doğruluğunu artırabileceğini göstermektir. Son olarak, modelin kullanımı İsveç Skolastik Değerlendirme Testinin iki uygulamasından elde edilen veriler üzerinde gösterilmiştir. Araştırma sonucunda eşitlemede ortak değişkenlerin kullanılması ile ortalama hata karelerinin azaldığı ve eşitleme kesinliğinin arttığı gözlenmiştir.

Wiberg ve von Davier (2017) yapmış oldukları çalışmada dış ortak test içeren büyük test programı İsveç Skolastik Yetenek Testi (Swedish Scholastic Assessment Test – SweSAT)'nin kalitesini incelemek ve izlemek için farklı prosedürler sunmuştur. Odak noktası, test sonuçlarının istikrarının ve kalitesinin ve dolaylı olarak ortak testlere ve

maddelere dayanan eşitlemenin zaman içinde incelenmesidir. Çalışmada test sonuçlarının farklı sınav katılımcıları için farklılık gösterip göstermediğini ve ortak test sonuçlarında zaman içerisinde beklenmedik bir davranış olup olmadığını araştırmak için ortak değişkenlerle bağlantılı olarak ortak maddelerin incelenmesi önerilmiştir. Test adilliğini potansiyel olarak tehlikeye atabileceği için bu zamansal etki araştırılmıştır. Çalışmada betimsel istatistikler, ANOVA ve karma lineer modelleri ortak değişkenlerin ortak test üzerindeki etkisini zaman içinde incelemek için kullanılmıştır. Sonuçlar, kullanılan yöntemler ve sonuçların diğer büyük test programlarında nasıl kullanılabileceğine ilişkin pratik öneriler ışığında tartışılmıştır. Çalışmada farklı yönetimlerdeki sınav katılımcıları için ortak değişkenlerin etkisinin incelenmesinin, adil testler oluşturma sürecinde yardımcı olabileceği sonucuna varılmıştır.

Wiberg ve Branberg (2015) yapmış oldukları çalışmada ortak değişkenleri NEC deseni aracılığıyla Kernel eşitleme çerçevesine dahil etmenin bir yolunu sunmuştur. NEAT deseninde eşitleme yöntemlerinden biri olan son tabakalama eşitlemede bir ortak test yerine test puanlarıyla ilişkili diğer değişkenlerin yani ortak değişkenlerin (örneğin, okul notları ve diğer test puanları) kullanılması önerilmiştir. Desenin kullanımını göstermek için bir üniversite kabul testinden elde edilen gerçek veriler kullanılmıştır. Bu çalışmada NEC deseninde elde edilen eşitlenmiş puanlar, EG deseninden elde edilen eşitlenmiş puanlarla yani ortak değişken olmadan yapılan eşitleme ve yapılandırılmış bir ortak test kullanılarak elde edilen eşitlenmiş puanlarla karşılaştırılmıştır. Sonuçlar NEC deseninin EG desenine göre daha düşük SEE üretebileceğini göstermiştir. Ayrıca ortak değişkenler bir ortak test ile birlikte kullanıldığında, geniş puan aralığında en küçük SEE'ler elde edilmiştir.

Wallin ve Wiberg (2017) yapmış oldukları çalışmada NEC deseninin Kernel eşitlemede eğilim puanlarıyla kullanımını göstermek için SweSAT verisini kullanmıştır. Test eşitlemede NEC deseni, ortak test olmadığında test puanlarıyla yüksek korelasyona sahip ortak değişkenleri ortak testin yerine kullanır. Bununla birlikte ortak değişken sayısı arttıkça, her ortak değişken kombinasyonu için gözlem sayısı azalır. Bunun yerine hem son tabakalama hem de zincirleme eşitlemeyle Kernel eşitleme çerçevesine dahil edilen eğilim puanlarını kullanmayı önermişlerdir. Test grupları eşdeğer olmadığında NEC deseni altında son tabakalama eşitlemenin EG desenine kıyasla daha iyi bir seçim olduğu sonucuna varılmıştır. NEC deseni altında zincirleme eşitlemenin o kadar başarılı olmadığı ve sonuçların NEAT deseni sonuçlarından nispeten uzak olduğu belirlenmiştir. Bu araştırma ortak maddeler ve ortak test katılımcıları olmadığında NEC deseni altında son

tabakalama eşitlemenin çok güçlü bir alternatif olduğunu göstermiştir. Ayrıca ortak değişkenleri doğrudan kullanarak eşitlemeye kıyasla, eğilim puanları kullanmak sürekli ortak değişkenlerin kullanımını kolaylaştırmıştır. Bu durumda ortak değişkenlerin eğilim puanlarıyla uygulanmasının çok kolay olduğu belirtilmiştir.

Akın Arıkan (2020) yapmış olduğu çalışmada NEAT ve NEC desenleri altında Kernel eşitleme yöntemleri ile yaptığı eşitleme sonuçlarını karşılaştırmıştır. Bu çalışmada Türkiye 2016 Akademik Becerilerin İzlenmesi ve Değerlendirilmesi Projesi'nin verileri kullanılmıştır. NEC deseni için ortak değişken olarak cinsiyet değişkeni ve sosyoekonomik düzey indeksi kullanılmıştır. Eşitlemeler sonucunda eşitleme hatası (Root Mean Squared Difference – RMSD) ve SEE değerleri hesaplanmıştır. Araştırma sonuçları, NEC deseninin NEAT desenine göre daha düşük SEE ürettiğini; en az RMSD'nin NEAT son tabakalama eşitleme deseni ve NEC deseni tarafından sağlandığını göstermiştir. Bu araştırmanın genel sonucu, test formlarının ortak maddeler olmadığında ortak değişkenler kullanılarak eşitlenebileceğidir.

Özsoy (2021) yapmış olduğu çalışmada TIMSS 2019 verisinde 1 ve 14 numaralı kitapçıklarda ikili puanlanan maddelerde NEAT ve NEC desenlerinde Kernel eşitleme yöntemlerini karşılaştırmıştır. NEAT deseninde Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli, Kernel son tabakalama lineer ve Kernel zincirleme lineer eşitleme yöntemleri kullanılmıştır. NEC deseninde ortak değişken olarak cinsiyet ve bilgisayar/tablet değişkenleri seçilmiştir. Araştırma sonuçları incelendiğinde NEC deseninden elde edilen eşitlenmiş puanların ham puanlara daha yakın olduğu belirlenmiştir. Fakat NEC deseninden elde edilen hata değerlerinin NEAT desenine göre daha yüksek olduğu görülmüştür. Sonuçta en az hata NEAT deseninde son tabakalama eşitlemeden elde edilmiştir. NEC deseninde en düşük hata ise bilgisayar/tablet değişkeni kullanılan Kernel lineer eşitleme yönteminde üretilmiştir.

Altıntaş ve Wallin (2021) yapmış oldukları çalışmada farklı dönemlerden elde edilen ve aynı psikolojik yapıları ölçen puanların istatistiksel olarak eşdeğer olup olmadığını belirlemek için NEC deseninde Kernel eşitleme yöntemini uygulamıştır. Çalışma 2017-2018 yıllarında uygulanan Ankara Üniversitesi Yabancı Uyruklu Öğrenci Sınavından elde edilen gerçek veri üzerinde yapılmıştır. Özellikle NEC deseni kullanılmıştır. Çünkü üniversite kabul testinin test grupları eşdeğer değildir ve ortak maddeleri yoktur. Analizlerden elde edilen sonuçlar, test formlarının puan dağılımlarının farklı olduğunu ve ilişkinin doğrusal olmadığını göstermiştir. Böylece eşitleme prosedürü

bu farklılıkları ortadan kaldıracak ve testlerin birbirinin yerine kullanılmasına izin verecek şekilde ayarlanmıştır.

Yıldırım Seheryeli (2022) yapmış olduğu çalışmada NEC deseninde nasıl bir ortak değişken kullanılması gerektiğini incelemiştir. Bunun için; Kernel eşitleme ile eşitlenmiş puan ortalamaları ve eşitleme hatalarının içsel, ortak değişkenlerin özelliklerinin dışsal, eşitlemedeki regresyon modelinde açıklanan varyans oranının hem içsel hem de dışsal değişken olduğu yol modeli oluşturulmuştur. NEAT deseninde KTK ve MTK'ye dayalı Kernel eşitleme yapılmıştır. Ortak maddelerden elde edilen toplam puan kategorik hale getirilmiş ve NEC deseninde ortak değişken olarak kullanılmıştır. Araştırma sonucunda NEAT deseninde KTK'den elde edilen sonuçlarda standart hataların daha düşük, MTK'de ise puan ortalamalarının daha yüksek olduğu görülmüştür. Özellikle ortak madde üç ve beş kategorili olduğunda, NEC desenindeki sonuçlar NEAT deseni ile benzerdir. Sonuçlar birlikte ele alındığında frekans dağılımları arasındaki farkın, test puanları ile ortak değişken arasındaki ilişkinin mutlak değerinin; ortak değişkenin çarpıklığının ve basıklığının mutlak değerinin, P ve Q evrenlerinde ortak değişken ile kurulan regresyon modelinin açıklanan varyans oranının anlamlı olarak eşitleme sonuçlarını etkilediği; maddelerin ortak değişkene göre DMF içerip içermemesinin ise etkilemediği görülmüştür.

Sezer Başaran, Mutluer ve Çakan (2023) yapmış oldukları çalışmada KTK'ye ve Kernel eşitlemeye dayalı yöntemlerle yapılan eşitlemelerin NEAT ve NEC desenleri kullanılarak karşılaştırılmasını amaçlamıştır. TIMSS 2019 fen testi puanları KTK'ye dayalı Tucker ve Levine gerçek puan, Levine gözlenen puan ve eşit yüzdelikli eşitleme (ön düzgünleştirme ve son düzgünleştirme) yöntemleri; Kernel eşitlemeye dayalı lineer ve eşit yüzdelikli eşitleme yöntemleriyle eşitlenmiştir. NEC desenindeki ortak değişkenler şu şekildedir: “öğrenme için ev kaynakları”, “öğrencinin fen ve matematik konusunda güveni”, “fen öğrenmeyi sevmek”, “fen derslerinde öğretimsel netlik”, “matematik başarısı”, “cinsiyet” ve “evde sınav dilini konuşmak”. NEC desenindeki eşitleme sonuçları NEAT (Kernel ve KTK) ve EG desenleriyle (Kernel) karşılaştırılmıştır. Sonuçlar eşitleme hataları ve eşitlenen puanlar arasındaki farklara göre analiz edilmiştir. Araştırmada NEAT deseninde eşitlemenin yapılamaması durumunda fen testini eşitlemek için matematik başarısının ve öğrenmeye yönelik ev kaynaklarının NEC deseninde ortak değişken olarak kullanılabileceği sonucuna varılmıştır. Ancak NEC deseninde diğer değişkenler ortak değişken olarak kullanıldığında, eşitlenen puanların EG desenine çok benzer olduğu belirlenmiştir. Ayrıca Tucker ve Kernel son tabakalama

yöntemleri lineer eşitlemede benzer sonuçlar vermiştir ve bu yöntemler EG desenindeki Kernel lineer eşitlemeden benzer şekilde farklıdır. Eşit yüzdelikli eşitlemede son düzgünleştirme ve Kernel EG deseninden elde edilen eşitlenmiş puanlar birbirine yakındır ancak son tabakalama puanlarından biraz farklıdır.

Sezer Başaran (2023) yapmış olduğu çalışmada TIMSS 2019 sekizinci sınıf matematik alt testlerini Kernel lineer ve Kernel eşit yüzdelikli eşitleme yöntemleriyle eşitlemede NEC deseninden elde edilen eşitleme hatalarının, NEAT ve EG desenlerinden farklılaşma düzeyinin belirlenmesini amaçlamıştır. Bunun için üç farklı grupta matematik dersiyle ilişkili olan öğrenci ve okul özelliklerini içeren 13 farklı ortak değişken kullanarak NEC deseninden elde ettiği eşitleme sonuçlarını NEAT ve EG desenleriyle karşılaştırmıştır. Elde edilen bulgulara NEC desenlerinin NEAT ve EG desenleriyle karşılaştırma sonuçları Kernel lineer ve Kernel eşit yüzdelikli eşitlemelerde birbirine benzerdir. NEAT son tabakalama ve zincirleme eşitleme yöntemlerine en yakın ve EG deseniyle en fazla farklılaşan desen, tüm gruplarda matematik başarısıyla yüksek düzeyde korelasyona sahip fen başarısının kullanıldığı NEC desendir. Fakat matematik başarısıyla zayıf ve çok zayıf düzeyde korelasyona sahip ortak değişkenlerin kullanıldığı NEC desenleri ile EG desenlerine ait eşitlenmiş puanların birbirine çok benzer olduğu görülmüştür. Öğrencilerin duyuşsal özelliklerinin ve evdeki eğitim kaynaklarının ortak değişken olarak kullanıldığı NEC desenlerinin NEAT ve EG desenlerine olan benzerliğinin gruplara göre değiştiği belirlenmiştir.

von Davier ve diğerleri (2006) yapmış oldukları çalışmada gözlenen puan eşitleme yöntemlerinden eşit yüzdelikli ve lineer eşitleme sonuçlarının, Kernel eşitleme yöntemine ne kadar yakın olduğunu incelemiştir. Bu çalışma, bir eşitleme kriterinin mevcut olduğu durumlarda test eşitlemenin Kernel yöntemini değerlendirmiştir. Çalışmada EG deseni, biri iç ve biri dış ortak testli iki NEAT deseni olmak üzere üç eşitleme deseni simüle etmek için gerçek bir testteki madde yanıtlarından oluşan sözde testler kullanılmıştır. Sonuçları karşılaştırmak için EG desenindeki eşitleme fonksiyonu eşitleme kriteri olarak belirlenmiştir. Bu örneklerde Kernel eşitleme sonuçlarının diğer eşitleme yöntemlerinden elde edilen sonuçlarla çok yakın olduğu belirlenmiştir. Ayrıca araştırılan hemen hemen tüm durumlarda, Kernel eşitleme sonuçları eşitleme kriterine daha yakın sonuçlar vermiştir.

Mao ve diğerleri (2006) yapmış oldukları çalışmada gerçek veri kullanarak Kernel eşitleme yöntemini zincirleme eşit yüzdelikli eşitleme, zincirleme lineer eşitleme ve Tucker lineer eşitleme yöntemleriyle karşılaştırmıştır. Çalışmada bu yöntemler, bir EG

deseni ve iki NEAT deseni olmak üzere farklı örneklem boyutları ve farklı özelliklere sahip üç araştırma desenine uygulanmıştır. EG deseninde, ideal bant genişliği kullanıldığında Kernel eşitleme sonuçları geleneksel eşit yüzdelikli eşitleme yöntemiyle benzer sonuçlar vermiştir. Geniş bant genişliği kullanıldığında Kernel eşitleme geleneksel lineer eşitleme yöntemiyle neredeyse aynı sonuçları vermiştir. NEAT deseninde; ideal bant genişliğine sahip Kernel son tabakalama eşitleme, puan aralığının çoğunda frekans kestirimi sonuçlarına benzer sonuçlar üretmiştir, geniş bant genişliği kullanıldığında ise Tucker yöntemine daha yakın sonuçlar üretmiştir. Diğer eşitleme yöntemleriyle Kernel eşitleme arasında sadece düşük puan aralıklarında farklılıklar oluşmuştur.

Ricker ve von Davier (2007) yapmış oldukları çalışmada NEAT deseninde dış ortak test uzunluğunun eşit yüzdelikli (frekans tahmini) eşitleme, zincirleme eşit yüzdelikli eşitleme ve ideal bant genişliğine sahip Kernel son tabakalama eşitleme üzerindeki etkilerini araştırmıştır. Bu çalışmada gerçek veriden alınan madde yanıtlarından oluşturulan sözde testler kullanılmıştır. EG eşit yüzdelikli eşitleme yöntemi kriter olarak seçilmiş ve eşitleme yöntemleri bu eşitleme kriteri kullanılarak değerlendirilmiştir. Araştırma sonucunda ortak test uzunluğu azaldıkça eşitleme dönüşümlerinde yanlılığın artma eğiliminde olduğu görülmüştür. Ancak eşit yüzdelikli eşitleme ve ideal bant genişliğine sahip Kernel son tabakalama eşitleme yöntemlerinin bu değişime diğer yöntemlere göre daha duyarlı olduğu görülmüştür. Ayrıca geniş bant genişliğine sahip Kernel son tabakalama eşitleme ortak testin bütün uzunluklarında zayıf performans göstermiştir.

Moses ve Holland (2007) yapmış oldukları çalışmada log lineer ön düzgünleştirmenin eşitleme fonksiyonu ve SEE doğruluğu üzerindeki etkisini incelemiştir. Bu değerlendirme; örneklem büyüklüğü, NEAT eşitleme deseni (zincirleme ve son tabakalama/frekans tahmini eşit yüzdelikli) ve süreklileştirme yöntemi (Kernel ve yüzdelik sıra) üzerinden yapılmıştır. Örneklem büyüklüğü 100, 200 ve 1000 olmak üzere üç ön düzgünleştirme yöntemi kullanılarak; Kernel zincirleme, zincirleme eşit yüzdelikli, Kernel son tabakalama ve son tabakalama eşit yüzdelikli eşitleme yöntemlerinde SEE ve SEED kestirimlerinin doğrulukları ve farklılıkları değerlendirilmiştir. En basit loglineer model testlerin ve ortak testlerin ortalamalarını, varyanslarını ve test ile ortak test arasındaki korelasyonu korumuştur. Başka bir loglineer model, testlerin ve ortak testlerin tek değişkenli dağılımlarının ilk altı momentini ve test ile ortak test arasındaki korelasyonu korumuştur. Üçüncü model ise gerçek popülasyon modelidir. Araştırmada ön düzgünleştirme derecesinin eşitleme yanlılığı (BIAS) ile doğrudan bir ilişkisi olduğu,

yalnızca en basit modellerin eşitleme fonksiyonu üzerinde pratik etkilere sahip yanlılık ürettiği sonucuna ulaşılmıştır. Ön düzgünleştirme derecelerinin süreklileştirme yöntemleriyle nasıl çalıştığı karşılaştırıldığında, Kernel süreklileştirme yönteminin yüzdelik sıra yöntemine göre bazı önemli avantajlara sahip olduğu görülmüştür. Kernel yöntemi sıfır frekanslı puanlar olsa bile süreklileştirebilirken, yüzdelik sıralama yöntemi bu tür puanları işlemek için ek prosedürler gerektirmiştir.

Grant, Zhang ve Damiano (2009) yapmış oldukları çalışmada SAT verisini kullanarak Tucker, Levine gözlenen puan, zincirleme lineer ve zincirleme eşit yüzdelikli eşitleme sonuçlarını Kernel eşitlemeyle karşılaştırmıştır. Kernel son tabakalama ve Kernel zincirleme eşitleme yöntemlerinde ideal ve geniş olarak iki tür bant genişliği seçilmiştir. Araştırmada geniş bant genişliği kullanılan Kernel zincirleme eşitleme ile zincirleme lineer eşitlemeden elde edilen sonuçların neredeyse aynı olduğu görülmüştür. Fakat Kernel son tabakalama eşitleme hem Tucker hem de Levine eşitleme sonuçlarıyla karşılaştırıldığında az da olsa farklılık göstermiştir. Tucker lineer eşitleme Levine eşitlemeye göre, geniş bant genişliğine sahip Kernel son tabakalama eşitlemeye daha yakın sonuçlar vermiştir. Geniş bant genişliğine sahip Kernel son tabakalama eşitleme ve Tucker eşitleme benzer varsayımlara dayandığından bu beklenen bir sonuçtur. İdeal bant genişliği kullanıldığında Kernel zincirleme ve Kernel son tabakalama eşitleme sonuçları zincirleme eşit yüzdelikli eşitleme ile karşılaştırıldığında puan aralığının çoğu için farkın küçük olduğu görülmüştür.

Akın Arıkan ve Gelbal (2018) yapmış oldukları çalışmada 2009 ÖBBS verisi üzerinde Kernel eşitleme sonucunda elde edilen eşitlenmiş puanları geleneksel eşitleme yöntemlerinden eşit yüzdelikli ve lineer eşitleme yöntemleriyle karşılaştırmıştır. Kernel eşitlemede ideal bant genişliği seçildiğinde, geleneksel eşit yüzdelikli eşitleme ile hemen hemen aynı sonuçları vermiştir. Geniş bant genişliği seçildiğinde ise bu kez eşitlenen puanlar geleneksel lineer eşitleme ile neredeyse aynı sonucu vermiştir. Bu araştırmada en düşük sistematik hata Kernel eşitleme yönteminde bulunmuştur. Kernel lineer eşitleme yöntemi en düşük rastgele hatayı üretirken bu hatanın en yüksek seviyesi geleneksel eşit yüzdelikli eşitleme yöntemi tarafından üretilmiştir. Kernel eşitleme yöntemlerinin geleneksel eşitleme yöntemlerine göre nispeten daha uygun eşitleme yöntemleri olduğu sonucuna varılmıştır.

Akın Arıkan (2019) yapmış olduğu çalışmada TIMSS 2019 fen bilimleri testinde kullanılan iki kitapçığın puanlarını (Kitapçık 1 ve 14) Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli, Kernel son tabakalama lineer ve Kernel

zincirleme lineer yöntemlerini kullanarak eşitlemiş ve elde edilen yüzde görelî hata (percent-relative error – PRE), SEE, SEED ve RMSD değerlerini karşılaştırmıştır. Araştırma sonucunda PRE değerlerinin eşit yüzdelikli eşitleme yöntemlerinde lineer eşitleme yöntemlerine göre daha düşük değerler sergilediği görülmüştür. Dağılımlar incelendiğinde, eşitleme fonksiyonları arasındaki farkın tüm puan aralığında ± 2 SEED dahilinde olduğu belirlenmiştir. Başka bir deyişle eşit yüzdelikli ve lineer fonksiyonlar birbirinden çok farklı değildir. Eşitleme yöntemleri arasındaki farklılıklar incelendiğinde Kernel son tabakalama ve Kernel zincirleme yöntemleri arasındaki farklar dışında Kernel eşit yüzdelikli eşitlemenin Kernel lineer eşitlemeye çok benzer sonuçlar ürettiği görülmüştür. Kernel son tabakalama ve Kernel zincirleme eşitleme yöntemlerine dayalı RMSD katsayıları karşılaştırıldığında, son tabakalama eşitleme yönteminin en az rastgele hatayı ve zincirleme lineer eşitlemenin maksimum rastgele hatayı verdiği belirlenmiştir. Kernel eşitleme yöntemleri için lineer eşitleme yöntemlerinin eşit yüzdelikli yöntemlere göre biraz daha yüksek SEE içerdiği görülmüştür.

Wang, Lee, Brennan ve Kolen (2008) yapmış oldukları çalışmada NEAT deseni altında frekans tahmini ve zincirleme eşit yüzdelikli eşitleme yöntemlerini karşılaştırmıştır. Bu çalışmada 60 maddelik bir matematik testinin dört farklı formundan elde edilen veriler kullanılmıştır. Her form için yaklaşık 3000 katılımcı testi almıştır. Araştırma sonucunda, frekans tahmini yönteminin zincirleme eşit yüzdelikli yöntemden sürekli olarak daha düşük standart hata ürettiği belirlenmiştir. Yanlılık incelendiğinde, frekans tahmini yönteminin zincirleme eşit yüzdelikli yöntemden daha fazla yanlılık ürettiği ve grup farklılıkları arttıkça yanlılıktaki farkın da arttığı görülmüştür. Bu sonuçlara dayanarak grup farklılıkları küçük olduğunda frekans tahmini yöntemi, grup farklılıkları büyük olduğunda ise zincirleme eşit yüzdelikli yöntem önerilmiştir.

Livingston (1993) yapmış olduğu çalışmada loglineer düzgünleştirmenin küçük örneklemelerde zincirleme eşit yüzdelikli eşitlemenin kesinliğini ne ölçüde iyileştirebileceğini araştırmıştır. Yüz soruluk bir teste verilen cevaplardan iki test formu ve bir ortak test oluşturulmuştur. Düzenlenen her bir koşulda 50 tekrar yapılmıştır. Çalışma 25, 50, 100 ve 200 kişilik örneklemelerde gerçekleştirilmiştir. Her örnek ile ilki düzgünleştirilmemiş dağılımlara dayalı ve üçü değişen düzeylerde düzgünleştirmeye dayalı dört eşitleme gerçekleştirilmiştir. İlk düzgünleştirmede yalnızca ilk iki tek değişkenli moment (ortalamalar ve standart sapmalar) ve ilk iki değişkenli moment (alt form ile ortak test arasındaki korelasyon), ikinci düzgünleştirmede bunlara ek olarak üçüncü tek değişkenli moment (çarpıklık) ve üçüncü düzgünleştirmede bunlara ek olarak

dördüncü tek değişkenli moment (basıklık) korunmuştur. Daha sonra sonuçlar kriter eşitleme yöntemi olan eşit yüzdelikli eşitleme ile karşılaştırılmıştır. Örneklem ne kadar küçük olursa, popülasyondan farklı olma olasılığı o kadar fazladır. Bu nedenle, düzgünleştirmenin en fazla küçük örneklerde fayda sağladığı belirlenmiştir. Aynı nedenle daha küçük örnekler, gözlenen verilerin daha az özelliğini koruyan düzgünleştirme yöntemleri gerektirmiştir ve düşük uç değerlerde eşitleme hatası (RMSD) daha büyük değerler almıştır.

İlgili araştırmalar özetlendiğinde; test eşitleme ile ilgili çalışmalarda NEAT deseninin sıklıkla kullanıldığı görülmüştür. Fakat bu desende ortak test, testin bütününe yeterince temsil etmediğinde güvenilir sonuçlar ortaya çıkmayacağı gözlemlenmiştir. Ayrıca her testte ortak maddenin olmadığı belirtilmiştir. Daha sonra Holland ve diğerleri (2007) sınava giren kişiler hakkında ek bilgi kullanmanın mantıklı olup olmadığı sorusunu gündeme getirmiş ve bunun ileri araştırmalar için önemli bir alan olduğunu öne sürmüştür. Bundan sonra literatürde ortak değişkenlerin ortak maddelerin yerine kullanılabileceğiyle ilgili çalışmalar yapılmış ve bu yeni araştırma deseni NEC olarak isimlendirilmiştir. Fakat hangi ortak değişkenlerin kullanılabileceğini ve bunların düzenli olarak nasıl toplanacağını incelemek araştırmaların önemli bir bölümüdür. Test puanlarıyla yüksek oranda korelasyon gösteren ortak değişkenler eşitleme durumunda iyi çalışma eğiliminde olduğundan, ortak değişkenleri seçerken ortak değişkenler ve test puanları arasındaki korelasyonu incelemek gerektiği belirtilmiştir. Literatürde ortak değişkenlerin Kernel eşitleme yöntemine dahil edildiği ve elde edilen sonuçların geleneksel eşitleme yöntemleriyle karşılaştırıldığı sonuçta daha doğru kestirimler yapıldığı çalışmalara rastlanmıştır. Literatürde eşitleme sürecinde ortak değişken kullanımına izin veren ve daha kesin kestirimler yaptığı belirtilen esnek bir yaklaşım olan PoB modeli önerilmiştir. Bu modelin daha küçük örneklerde avantajlı olduğu ve esnek yapısı sebebiyle eşitleme varsayımlarının sağlanmasına gerek duymadığı vurgulanmıştır.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, verilerin toplanması ve verilerin çözümlenmesine ilişkin bilgiler yer almaktadır.

Araştırmanın Modeli

Araştırmada PoB yöntemine göre elde edilen eşitlenmiş puanlar eşit yüzdelikli eşitleme ve Kernel eşitleme yöntemleriyle elde edilen eşitlenmiş puanların dağılımlarıyla karşılaştırılmıştır. Bu araştırma tarama modellerinden betimsel araştırma niteliği taşımaktadır. Tarama modellerinde üzerinde çalışma yapılan grup veya sürece müdahale olmadan veriler toplanır (Deniz ve Çelebi, 2022). Bu araştırmada kullanılan eşitleme yöntemlerinin performansları inceleneceğinden betimsel araştırma niteliği taşımaktadır. Betimsel araştırmalar, belirli bir durumu mümkün olduğunca eksiksiz ve dikkatli bir şekilde tanımlar (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2009).

Çalışma Grubu

Araştırmanın çalışma grubunu 2022-2023 eğitim-öğretim yılında Antalya ilinde öğrenim gören 15-16 yaş grubundaki dokuzuncu sınıf öğrencileri oluşturmaktadır. Antalya il sınırları içerisinde bulunan bütün dokuzuncu sınıf öğrencilerine ulaşılması zor olduğundan, araştırma verilerini toplamak için uygulanacak ölçme aracının örnekleme tipik durum örnekleme yöntemi kullanılarak oluşturulmuştur. Tipik durum örnekleme yöntemi, araştırma probleminin içinde bulunduğu evrendeki birçok durumdan tipik bir durumu belirleyerek bu örneklemden bilgi toplamayı amaçlar. Tipik durum örnekleme kullanılırken, “hangi durumların tipik olduğu ve tipikliği tanımlamak için hangi kriterlerin kullanıldığı konusunda geniş bir fikir birliği sağlamaya çalışmak” çok önemlidir (Patton, 2002). Burada temel prensip, olağan dışı olmayan, tipik bir durumun seçilmesidir. Verilerin analizinde kullanılacak olan ortak değişkenlerin (cinsiyet-LGS

puan düzeyi) alt grupları arasında oluşacak büyük farklar analizi etkileyebilir. Bu sebeple tipik durum örnekleme ile bu durumu kontrol altına almak amaçlanmıştır.

A ve B olmak üzere iki paralel formu olan yetenek testi deneme uygulaması için Antalya İl Milli Eğitim Müdürlüğüne bağlı beş resmi ortaöğretim kurumunda dokuzuncu sınıf düzeyindeki öğrencilere uygulanmıştır. Deneme uygulaması sonucunda toplam 706 öğrenci araştırmaya dahil edilmiştir. Optik formlar spirallemeye uygun olarak dağıtılmıştır. Bu durumda öğrencilerin 353'ü A formunu ve 353'ü B formunu almıştır. Nihai uygulama için ise beş merkez ilçede, toplam 26 resmi ortaöğretim kurumunda dokuzuncu sınıf düzeyindeki öğrencilere uygulanmıştır. Nihai uygulama sonucunda 4201 öğrenci araştırmaya dahil edilmiştir. Yine optik formlar spirallemeye uygun olarak dağıtılmıştır. Bu durumda öğrencilerin 2100'ü A formunu ve 2101'i B formunu almıştır.

Verilerin Toplanması

Araştırmada ölçme aracı olarak geliştirilen yetenek testinin geliştirme ve uygulama sürecinde izlenen basamaklar şu şekildedir (Crocker ve Algina, 2008):

1. Test puanının kullanım amaçlarının belirlenmesi
2. Yapıyı temsil eden davranışların belirlenmesi
3. Kapsamın belirlenmesi
4. Madde yazımı
5. Maddelerin gözden geçirilmesi
6. Deneme formunun hazırlanması
7. Deneme formunun hedeflenen grubu temsil edecek örnekleme uygulanması
8. Madde analizi
9. Nihai formların oluşturulması

1. Test Puanı Kullanım Amaçlarının Belirlenmesi

Bireyler farklı yetenek, ilgi ve tutumlara sahiptirler. Yeteneklerine uygun bir alanda eğitim gören ve sonrasında buna uygun bir alanda meslek edinen bireylerin, hem eğitim hem de meslek hayatlarında başarılı ve mutlu olma ihtimalleri daha yüksektir. Bireylerin başarılı olabilecekleri alanda eğitim almaları ya da çalışmaları yeteneklerinin bilinmesi ile mümkün olur.

Eğitim ve psikolojide testler insan özelliklerinin ölçülmesinde kullanılan ölçme araçlarıdır. İnsan özelliklerinin bir kısmı eğitimle ilgili olan öğrenme ürünleri iken, diğer kısmı daha çok psikolojik kökenli olan zekâ, yetenek ve kabiliyet gibi özelliklerdir. Bu açıdan ölçmenin dolayısıyla testin kullanılış amaçlarını biri eğitim diğeri de psikoloji yönünden olmak üzere iki kategoride değerlendirmek mümkündür. Bu bakımdan bazı yazarlar başarı testleri ve psikolojik testler olmak üzere iki gruba ayırırlar. Ancak testlerin öğrenme ürünlerinin ölçülmesindeki amaçlarını psikolojideki kullanılış amaçlarından ayırmak mümkün değildir (Baykul, 2015).

Eğitim sistemindeki sıkıntı kaynakları arasında girdiler ve süreçteki yetersizlikler önemli bir yer tutar. Girdilerden eğitim programlarının ve süreçte öğretim hizmetlerinin niteliğinin değerlendirilmesi ve değerlendirmede testlerden yararlanılması yaygın bir uygulama alanına sahiptir. Gerek öğrenme ortamının düzenlenmesinde gerekse öğretim faaliyetlerinin devam ettirilmesinde ön-şart oluş ilişkilerine dikkat edilmesi gerekir (Baykul, 2015).

Araştırmada geliştirilen yetenek testi ile lise eğitimi düzeyine gelmiş öğrencilerin hazırbulunuşluklarının incelenmesi amaçlanmaktadır.

2. Yapıyı Temsil Eden Davranışların Belirlenmesi

Yeteneklerin ölçülebilmesi için öncelikle yapının tanımlanması gerekmektedir. Bilişsel yeteneklerin yapısıyla ilgili teorilere yapılan atıflar zekâ testlerinde yaygındır. Testleri anlamak için teorilere biraz aşına olunması gerekmektedir (Hogan, 2019). Bu araştırmada Thurstone'un Çoklu Faktörler Modeli temel alınarak bir yetenek testi geliştirildiği için bu teori kısaca anlatılmıştır. Thurstone birçok testin hem geliştirilmesine hem de sonuçlarına önemli ölçüde katkıda bulunan faktör analizi tekniğini kullanarak, çoklu faktörler modeli olarak adlandırılan zekâ kuramını geliştirmiştir. Bu kurama göre zihinsel etkinlik gerektiren işler gruplanabilir ve her grupta bulunan işler diğerlerinden farklı bir zihinsel güç gerektirir. Thurstone yaptığı kapsamlı çalışmalara dayanarak “birincil zihinsel yetenekler” olarak adlandırdığı yaklaşık 12 faktör önermiştir. Thurstone ve diğer araştırmacıların (French, 1951; Thurstone, 1938; Thurstone ve Thurstone, 1941; Akt. Anastasi, 1976) çalışmalarında ise yedi faktör sıklıkla bahsedilmiştir. Bu kuramda yeteneklere ilişkin belirlenen yedi temel faktör Tablo 6'da verilmiştir.

Tablo 6

Thurstone Çoklu Faktörler Modeli

Faktörler
Sözel Yetenek
Kelime Akıcılığı
Sayısal Yetenek
Muhakeme Yeteneği
Yer ve Mekâna Uyma Faktörü
Bellek Faktörü
Algısal Faktörler

Kaynak. Anastasi, 1976

Thurstone'un ifade ettiği bu faktörler aşağıda özet olarak verilmiştir (Özgüven, 2004):

- **Sözel Yetenek:** Bu faktör, kelimelerin anlamlarını ve yazılı metinleri anlama yeteneğini belirtir. Yetenek testinin birçoğunda kelime dağarcığı, cümle düzenleme, cümle tamamlama, sözcüklerin eş anlamını ve zıt anlamını bulma gibi farklı şekilleri bulunmaktadır.
- **Kelime Akıcılığı:** Bu faktörde bir türden kelimenin hızlıca hatırlanma yeteneği büyük öneme sahiptir. Thurstone, “kelimeleri anlama” ve “kelime çokluğu” arasında .42’lik bir korelasyon tespit etmiştir. Fakat iki faktör de farklı zihinsel yetenekleri temsil eder. Kelimeleri anlama, kelimelerin zorluğu ve kolaylığıyla yakından ilişkilidir. Kelime akıcılığı faktöründe ise kelimenin zorluğu veya kolaylığı sonuçları pek etkilemez, asıl önemli olan hızlı bir şekilde çok sayıda kelime üretebilmektir.
- **Sayısal Yetenek:** Bu faktör, temel aritmetik işlemleri hızlı ve doğru bir şekilde hesaplama becerisini içerir.
- **Muhakeme Yeteneği:** Muhakeme faktörü karmaşık bir konudur. Burada tümevarımsal ve tümdengelimsel düşünme kastedilmekle beraber tümevarımsal düşünmenin daha önemli olduğu vurgulanmaktadır. Muhakeme, bir bireyin geçmiş deneyimlerini kullanarak iki veya daha fazla unsuru bir araya getirerek bir problemi çözmeye yönelik düşünme sürecidir.
 - **Genel Muhakeme Faktörü:** Bazı muhakeme faktörleri sayısal ilişkilerle bağlantılıdır. Genel muhakemeyi değerlendiren problem çözme alanı, muhakeme faktörlerinin en fazla var olduğu bilinen alandır.

- **Analoji:** Muhakemenin bir diğer türü analogi, bir konuda kanıtlara dayalı olarak sonuç çıkarmaktır. Özellikle keşfedilmesi gereken belirli bir unsura veya çözüme ihtiyaç yoktur; temel olarak, var olan verilerden sonuç çıkarmak önemlidir.
- **Sebep-Sonuç İlişkisi:** Bu muhakeme faktörü, birçok grup zekâ testinde gözlenmektedir. Bu faktörü ölçen test soruları genellikle iki fikir veya nesne arasındaki ilişkileri tanımlama (kuralı tespit etme) yeteneğine dayanır ve bu ilişkiyi diğer şeyler arasındaki ilişkileri bulma sürecinde kullanma gerekliliğini vurgular. Bu faktör, özellikle sözcük ve şekil ilişkilerini ölçme durumunda en iyi şekilde değerlendirilmektedir.
- **Yer ve Mekâna Uyuma Faktörü:** Bu faktör bir nesnenin görünmeyen yönleri dahil olmak üzere farklı uzamsal durumlarını tasarlama, hayal etme, zihinde canlandırma ve mekânsal ilişkileri algılama yeteneğini içerir.
- **Bellek Faktörü:** Bu faktör anlamsız unsurları, sözel, sayısal ve şekilsel sıralamaları ve materyalleri detaylarıyla hafızada tutma yeteneğini içerir.
- **Algısal Faktörler:** Bu faktör görsel olarak farklı şekiller arasındaki benzerlikleri, farklılıkları ve ayrıntıları hızlı bir şekilde algılayabilme yeteneğini ifade eder.

Yetenek testine ilişkin çerçevenin belirlenmesinde bu faktörler bir program geliştirme, bir ölçme ve değerlendirme, iki Türkçe ve dört matematik alan uzmanıyla birlikte incelenmiştir. Zekâ ve yetenek testlerine ait madde örnekleri incelenerek alan uzmanlarıyla birlikte program geliştirme uzmanının da görüşü alınarak belirlenen yaş düzeyine uygun yeterli ifadelerinin belirlenmesi için çalışma yapılmıştır. Belirlenen yeterli ifadelerinin kelime akıcılığı faktöründe kısa cevaplı maddeyle, diğer faktörlerde çoktan seçmeli maddelerle ölçülebilir olmasına ve madde yazmaya elverişli olacak şekilde ifade edilmesine dikkat edilmiştir. Araştırma kapsamında, konu alanı uzmanları ve program geliştirme ile ölçme ve değerlendirme uzmanlarının görüşleri alınarak öncelikle geliştirilen yetenek testi için uygun yeterli ifadeleri hazırlanmıştır. Yapılan çalışmanın sonucunda, yetenek testi ile ilişkilendirilen bu yeterli ifadeleri Tablo 7’de sıralanmıştır.

Tablo 7

Faktörlere Ait Yeterlik İfadeleri

Faktörler	Yeterlikler
Sözel Yetenek	*Verilen sözcüğün eş anlamlısını seçer. *Kelimeler gruplandığında dışarıda kalan kelimeyi bulur. *Verilen sözcük çifti ile benzer ilişki taşıyan sözcüğü veya sözcük çiftini tespit eder.
Kelime Akıcılığı	*Belirli bir zaman diliminde belirli bir kategoriye ait çok sayıda kelime bulur.
Sayısal Yetenek	*Toplama, çıkarma, çarpma ve bölme gibi çeşitli aritmetik işlemleri yapar.
Muhakeme Yeteneği	*Verilen problemin çözümünü, toplama, çıkarma, çarpma ve bölme işlemlini kullanarak yapar. *Verilen bilgilerden (uzunluk-kısalık, büyüklük-küçüklük, hafiflik-ağırlık vb.) sonuçlar çıkarır.
	*Okuduğu metin veya cümleyi uygun kelimelerle tamamlar. *Verilen metinden elde edilebilecek kesin sonuçları tespit eder ve çıkarımlarda bulunur. *Karmaşık bir şekilde verilen kelimeleri anlamlı bir cümleye dönüştürür. *Okuduğu metnin anlam bütünlüğü kazanması için hangi cümlelerin yer değiştirmesi gerektiğini bulur. *Okuduğu bir metinde düşüncenin akışını bozan cümleyi bulur. *Paragraftan çıkarılabilecek anlamı bulur.
	*Verilen sayı örüntüsünde eksik olan terimi tespit eder. *Karmaşık bir harf dizisi içerisinde belirli bir kurala göre gizlenmiş kelimeyi tespit eder. *Belirli bir kurallara uygun olarak verilen şekil örüntüsünde eksik olan ögeyi tespit eder. *Verilen iki şekil arasındaki bağıntıya benzer ilişki gösteren şekli bulur.
Yer ve Mekâna Uyuma Faktörü	*Nesne veya şekillerin durumları değiştirildiği zaman nasıl görüneceğini bulur. *Bir perspektiften verilen nesne veya şekillerin başka bir perspektiften görünümünü belirler.
Bellek Faktörü	*Birbiri ile ilişkili sözcük çiftleri arasındaki ilişkileri yakalar ve hatırd tutar. *Resim, şekil ve şema gibi görsel materyallerdeki içeriği hatırd tutar.
Algısal Faktörler	*Verilen şekli oluşturmak için parçaları seçer. *Verilen şekillerin içinde farklı olanı bulur.

3. Kapsamın Belirlenmesi

Thurstone'un Çoklu Faktörler Modeli çerçevesinde yeterlik ifadeleri belirlendikten sonra alan uzmanlarının da görüşleri alınarak her yeterlik ifadesi ile ilgili belirlenen yaş düzeyine uygun yazılacak madde sayısı belirlenmiştir. Bu maddelerin

bilişsel düzeyi belirlenerek belirtke tablosu oluşturulmuştur. Maddelerin bilişsel düzeyinin belirlenmesinde Yenilenmiş Bloom Taksonomisi kullanılmıştır. Her yeterlik ifadesi için en az anlama düzeyinde maddelerin yazılmasına dikkat edilmiştir. Belirtke tablosunun hazırlanmasından sonra madde yazımına başlamadan önce program geliştirme ve konu alanı uzmanlarının da görüşleri alınmıştır. Her bir paralel formda yeterlikleri ölçmek için belirlenen madde sayıları ve bilişsel düzeyleri Tablo 8’de Belirtke Tablosunda verilmiştir.

Tablo 8

Belirtke Tablosu

Bilişsel Düzey Yeterlikler	Hatırlama	Anlama	Uygulama	Çözümleme
*Verilen sözcüğün eş anlamlısını seçer. *Kelimeler gruplandığında dışarıda kalan kelimeyi bulur. *Verilen sözcük çifti ile benzer ilişki taşıyan sözcüğü veya sözcük çiftini tespit eder.		3 1	3	
*Belirli bir zaman diliminde belirli bir kategoriye ait çok sayıda kelime bulur.	2			
*Toplama, çıkarma, çarpma ve bölme gibi çeşitli aritmetik işlemleri yapar.		2	2	
*Verilen problemin çözümünü, toplama, çıkarma, çarpma ve bölme işlemlini kullanarak yapar. *Verilen bilgilerden (uzunluk-kısalık, büyüklük-küçüklük, hafiflik-ağırlık vb.) sonuçlar çıkarır.			3 1	5
*Okuduğu metin veya cümleyi uygun kelimelerle tamamlar. *Verilen metinden elde edilebilecek kesin sonuçları tespit eder ve çıkarımlarda bulunur. *Karmaşık bir şekilde verilen kelimeleri anlamlı bir cümleye dönüştürür. *Okuduğu metnin anlam bütünlüğü kazanması için hangi cümlelerin yer değiştirmesi gerektiğini bulur. *Okuduğu bir metinde düşüncenin akışını bozan cümleyi bulur. *Paragraftan çıkarılabilecek anlamı bulur.			1 1 1 1	1 1 1 1

(Devam ediyor)

Tablo 8 (Devam)

Belirtke Tablosu

Bilişsel Düzey Yeterlikler	Hatırlama	Anlama	Uygulama	Çözümleme
*Verilen sayı örüntüsünde eksik olan terimi tespit eder.			1	1
*Karmaşık bir harf dizisi içerisinde belirli bir kurala göre gizlenmiş kelimeyi tespit eder.			1	
*Belirli bir kurallara uygun olarak verilen şekil örüntüsünde eksik olan ögeyi tespit eder.			2	1
*Verilen iki şekil arasındaki bağıntıya benzer ilişki gösteren şekli bulur.			1	1
*Nesne veya şekillerin durumları değiştirildiği zaman nasıl görüneceğini bulur.				1
*Bir perspektiften verilen nesne veya şekillerin başka bir perspektiften görünümünü belirler.			2	
*Birbiri ile ilişkili sözcük çiftleri arasındaki ilişkileri yakalar ve hatırd tutar.	3			
*Resim, şekil ve şema gibi görsel materyallerdeki içeriği hatırd tutar.	4			
*Verilen şekli oluşturmak için parçaları seçer.			1	
*Verilen şekillerin içinde farklı olanı bulur.			1	

4. Madde Yazımı

Test geliştirme sürecinde madde yazımında uyulması gereken bazı ilkeler de bulunmaktadır. Bu ilkeler şu şekildedir (Thorndike ve Thorndike-Christ, 2017):

1. Test maddelerinin sözcük seviyesinin ve okuma güçlüğüne olabildiğince basit düzeyde olması gerekmektedir. Öğrencilerin, maddenin değerlendirmeyi amaçladığı içerik hakkında bildiklerini göstermelerini engelleyebilecek karmaşık cümle yapısından veya gereksiz zor sözcüklerden kaçınılmalıdır.
2. Her maddenin uzmanların üzerinde anlaşacağı en iyi veya tek bir doğru yanıtının olması gerekmektedir. Normalde, tartışmalı konularda farklı bakış açılarına ilişkin bilgilerin önemli olabileceği durumlar olmasına rağmen, tartışmalı nitelikteki ifadeler iyi maddeler oluşturmaz.
3. Her maddenin önemsiz şeylerle değil içerik alanının önemli bir yönü ile ilgili olması gerekmektedir.

4. Her madde birbirinden bağımsız olmalıdır. Bir maddenin yanıtı bir sonraki maddeyi çözmek için şart olarak aranmamalıdır. Her öğrenci her bir madde için kendi başına adil bir şansa sahip olmalıdır.
5. Çeldirici maddeler kullanmaktan kaçınmak gerekmektedir.

Maddelerin belirlenen yeterlik düzeylerini ölçecek nitelikte olmasına ve yeterlik ifadelerini yoklayacak sınırlılıkta olmasına dikkat edilmiştir. Madde türü olarak kelime akıcılığı faktöründe kısa cevaplı maddeye; diğer faktörlerde ise çoktan seçmeli (beş seçenekli) maddelere yer verilmesine karar verilmiştir. Madde yazarlarına öncelikle araştırmacı tarafından yetenek testleri ile ilgili bilgi verilmiş ve başarı testleri ile aralarındaki farklılıklar örnek maddeler üzerinde incelenmiştir.

Araştırmacı konu alanı uzmanlarıyla birlikte maddeleri inceleyip son halini vermiş, maddelerin tahmini güçlük düzeylerini belirlemiş ve madde havuzunu oluşturmuştur. Madde yazımı için yazarlara iki aylık süre verilmiş ve toplam 250 madde ile havuz oluşturulmuştur.

5. Maddelerin Gözden Geçirilmesi

Madde havuzu tamamlandıktan sonra maddelerin ölçülmek istenen yeterliliği ölçüp ölçmediği, bilimsel olarak hatalı olup olmadığı ve dil uzmanları tarafından anlaşılabilirliği incelenmiştir. Madde incelemesini yapan uzmanların verdiği düzeltme önerileri doğrultusunda maddeler gözden geçirilmiş ve son hali verilmiştir.

6. Deneme Formunun Hazırlanması

Deneme formu iki paralel form oluşacak şekilde düzenlenmiştir. Her bir formda yeterlikleri ölçecek 48 madde bulunması planlanmıştır. Havuzda bulunan diğer maddeler test üzerinde ilerde yapılacak çalışmalar için saklanmıştır.

7. Hedef Grubun Temsil Edildiği Örneklemde Deneme Uygulanması

Deneme formları hazırlandıktan sonra, gerekli incelemelerin yapılabilmesi için hedef grubu temsil eden bir örneklem uygulanmıştır. Deneme formlarının uygulaması 2022 yılı Kasım ve Aralık ayı içerisinde kağıt-kalem testi olarak yapılmıştır.

Uygulamadan önce, Antalya İl Milli Eğitim Müdürlüğünden gerekli izinler alınmıştır. Ayrıca deneme uygulamasından önce, uygulamanın yapılacağı okul yönetimine ve öğretmenlere çalışmanın amacıyla ilgili gerekli bilgiler sunulmuştur. Planlanan tarihlerde uygulama gerçekleştirilmiştir.

Deneme uygulamasına başlamadan önce araştırmacı öğrencilere testin amacını açıklamış ve onların sorularını cevaplamıştır. Deneme uygulaması sürecinde öğrencilerin sınav kurallarına uyması sağlanmıştır. Deneme uygulaması için öğrencilere verilen toplam süre yaklaşık 55 dakika olmuştur. Araştırmacı tarafından öğrencilerin maddeler ve test hakkındaki izlenimleri ve sınavı bitirme süreleri kaydedilmiştir. Test için verilen sürenin yeterli olup olmadığı belirlenmiştir.

8. Madde Analizi

Deneme uygulaması sonucunda çoktan seçmeli maddeler 0-1 şeklinde puanlanırken kısa cevaplı maddeler 0 ile 3 arasında kısmi olarak puanlanıp madde analizi ve geçerlik ve güvenirlik analizleri yapılmıştır. Bu süreçle ilişkin bilgiler verilerin çözümlenmesi başlığında ayrıntılı olarak sunulmuştur.

9. Nihai Formların Oluşturulması ve Uygulanması

Verilerin analizi sonucunda maddeler üzerinde gerekli düzeltmeler yapıp uygun görülen maddeler seçildikten sonra nihai formlar oluşturulmuştur. Oluşturulan formlar 2023 yılı Mart, Nisan ve Mayıs ayı içerisinde Antalya İl Milli Eğitim Müdürlüğünden gerekli izinler alınarak kağıt-kalem testi olarak uygulanmıştır. Nihai uygulama için öğrencilere verilen toplam süre yaklaşık 50 dakika olmuştur. Nihai formların oluşturulması ve nihai uygulama sonucunda test üzerinde yapılan analizler verilerin çözümlenmesi başlığında ayrıntılı olarak sunulmuştur.

Verilerin Çözümlenmesi

Araştırmanın bu bölümünde öncelikle geliştirilen yetenek testi için deneme uygulaması ve nihai uygulamaya ait madde ve test istatistiklerine ve geçerlik ve güvenirlik analizlerine yer verilmiştir. Daha sonra nihai veri için eşitleme varsayımları

test edilmiş ve veri yapılacak analizler için hazırlanmıştır. Verilerin çözümlenmesinde kullanılacak paketler tanıtılmış ve karşılaştırma kriteri belirlenmiştir.

Deneme uygulaması sonucunda elde edilen veriler, maddelerin ve ölçeğin psikometrik özellikleri açısından incelenmiştir. Bu amaçla test ve madde istatistikleri belirlenmiş ve güvenilirlik ve geçerlik analizleri yapılmıştır. Güvenirlik, ölçme sonuçlarının tesadüfî hatalardan ne kadar arındığını ölçen bir kavramdır (Turgut ve Baykul, 2019). Paralel formların güvenilirliğinin belirlenmesi için McDonald'ın Omega (ω) güvenilirlik katsayısı hesaplanmıştır. Geçerlik ise bir testin sadece hedeflenen değişkeni ölçmesi, bu değişkeni ölçerken başka değişkenlerle karıştırmamasıdır (Özçelik, 2016). Paralel formların yapı geçerliğinin belirlenmesi için doğrulayıcı faktör analizi (DFA) yapılmıştır ve ölçüt geçerliğinin belirlenmesi için öğrencilerin 2022 LGS yerleştirme puanları ölçüt puan olarak kullanılmıştır.

Deneme Uygulaması

Deneme uygulamasında test formlarında bulunan madde sayıları ve puanlama şekli Tablo 9'da verilmiştir.

Tablo 9

Test Formlarında Bulunan Maddeler

Form	A Formu		B Formu	
Madde Türü	Çoktan Seçmeli	Kısa Cevaplı	Çoktan Seçmeli	Kısa Cevaplı
Madde Sayısı	46 Madde (34 Madde + 12 Söзде Ortak Madde)	2 Madde	46 Madde (34 Madde + 12 Söзде Ortak Madde)	2 Madde
Puanlama	0-1	0,1,2,3 (Kısmi Puanlama)	0-1	0,1,2,3 (Kısmi Puanlama)

Her bir test formunda 12 madde söзде ortak madde olmak üzere toplam 46 çoktan seçmeli madde ve 2 kısa cevaplı madde bulunmaktadır. Çoktan seçmeli maddeler 0-1 şeklinde puanlanırken kısa cevaplı maddeler 0 ile 3 arasında kısmi olarak puanlanmıştır. Bu durumda bir test formundan alınabilecek en yüksek puan 52'dir.

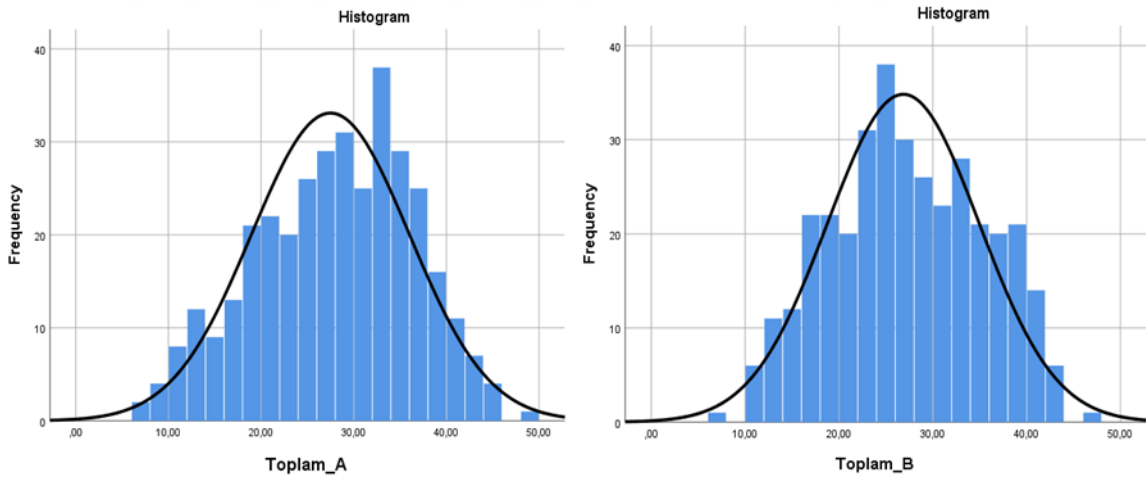
Yetenek Testi Deneme Uygulamasına Ait Geçerlik ve Güvenirlik Analizleri.

Deneme uygulaması sonucunda A ve B formlarına ait test istatistikleri ve histogram grafikleri Tablo 10 ve Şekil 4’te verilmiştir.

Tablo 10

Deneme Uygulamasına Ait Test İstatistikleri

Form	A	B
Değerlendirilen öğrenci sayısı	353	353
Testten alınabilecek en yüksek puan	52	52
Testten alınan en düşük puan	6	9
Testten alınan en yüksek puan	51	48
Mod (X_{mod})	33.00	24.00
Medyan (X_{ort})	28.00	26.00
Aritmetik ortalama ($\bar{X}$)	27.46	26.86
Standart sapma (S_s)	8.67	8.09
Çarpıklık (K_y)	-.275	.013
Basıklık (B_s)	-.534	-.791
Ortalama güçlük (p)	.54	.52
Ortalama ayırt edicilik (Çift serili korelasyon katsayısı – r_{jx})	.51	.49



Şekil 4. Formlara Ait Histogram Grafikleri

Formlarda çok yönlü uç değerleri belirlemek için Mahalanobis uzaklıkları incelenmiş ve her iki formda da uç değer tespit edilmemiştir. Formlara ait merkezi eğilim ölçüleri incelendiğinde her bir form için bu değerlerin birbirine yakın olduğu; çarpıklık ve basıklık katsayılarının ± 1 sınırları arasında olduğu belirlenmiştir. Şekil 4’te verilen

histogram grafikleri de incelendiğinde dağılımların normale yakın olduğu söylenebilir. A formunun orta güçlükte ($p = .54$) ve çok iyi ayırt edici olduğu ($r_{jx} = .51$); B formunun da orta güçlükte ($p = .52$) ve çok iyi ayırt edici olduğu ($r_{jx} = .49$) belirlenmiştir.

Deneme uygulaması sonucunda A ve B formlarına ait madde analizleri incelenmiştir. Elde edilen sonuçlar Tablo 11’de verilmiştir.

Tablo 11

A ve B Formlarına Ait Madde Analizleri

Faktör	Madde				Madde	B Formu			
	No	p	r_{jx}	Maddenin İncelenmesi		p	r_{jx}	Maddenin İncelenmesi	
Bellek	1	.80	.37	Yeterli	1	.85	.32	Yeterli	
Bellek	2	.81	.37	Yeterli	2	.80	.35	Yeterli	
Bellek	3	.87	.44	Yeterli	3	.86	.32	Yeterli	
Bellek	4	.65	.19	Çok zayıf	4	.77	.23	Düzeltilmeli	
Bellek	5	.61	.23	Düzeltilmeli	5	.79	.34	Yeterli	
Bellek	6	.52	.25	Düzeltilmeli	6	.57	.24	Düzeltilmeli	
Bellek	7	.56	.31	Yeterli	7	.50	.17	Çok zayıf	
Sözel	8	.69	.52	Yeterli	8	.65	.19	Çok zayıf	
Sözel	9	.72	.17	Çok zayıf	9	.48	.44	Yeterli	
Sözel	10*	.61	.30	Yeterli	10*	.56	.34	Yeterli	
Sözel	11*	.78	.48	Yeterli	11*	.74	.41	Yeterli	
Sözel	12	.50	.46	Yeterli	12	.48	.18	Çok zayıf	
Sözel	13	.48	.19	Çok zayıf	13	.75	.51	Yeterli	
Sözel	14	.77	.70	Yeterli	14	.86	.57	Yeterli	
Muhakeme/A	15	.72	.67	Yeterli	15	.43	.37	Yeterli	
Muhakeme/A	16	.68	.70	Yeterli	16	.50	.55	Yeterli	
Muhakeme/A	17	.72	.66	Yeterli	17	.83	.60	Yeterli	
Muhakeme/A	18	.68	.48	Yeterli	18	.33	.31	Yeterli	
Muhakeme/A	19	.67	.47	Yeterli	19	.47	.30	Yeterli	
Muhakeme/A	20*	.54	.49	Yeterli	20*	.61	.47	Yeterli	
Sayısal	21*	.46	.78	Yeterli	21*	.45	.73	Yeterli	
Sayısal	22	.69	.75	Yeterli	22	.67	.71	Yeterli	
Sayısal	23	.33	.70	Yeterli	23	.46	.79	Yeterli	
Sayısal	24*	.43	.79	Yeterli	24*	.51	.76	Yeterli	
Muhakeme/GM	25	.29	.61	Yeterli	25	.61	.70	Yeterli	
Muhakeme/GM	26	.15	.46	Yeterli	26	.25	.59	Yeterli	
Muhakeme/GM	27	.26	.64	Yeterli	27	.43	.71	Yeterli	
Muhakeme/GM	28	.35	.60	Yeterli	28*	.40	.72	Yeterli	
Muhakeme/GM	29*	.36	.67	Yeterli	29*	.16	.58	Yeterli	
Muhakeme/GM	30*	.23	.62	Yeterli	30	.32	.76	Yeterli	
Muhakeme/GM	31	.46	.44	Yeterli	31	.46	.63	Yeterli	
Muhakeme/GM	32	.64	.58	Yeterli	32	.48	.59	Yeterli	
Muhakeme/GM	33*	.19	.54	Yeterli	33*	.26	.55	Yeterli	
Muhakeme/SS	34	.28	.47	Yeterli	34	.50	.64	Yeterli	
Muhakeme/SS	35*	.72	.65	Yeterli	35*	.66	.67	Yeterli	
Muhakeme/SS	36	.59	.84	Yeterli	36	.28	.54	Yeterli	
Muhakeme/SS	37	.63	.40	Yeterli	37	.18	.32	Yeterli	

* Söзде Ortak Madde

(Devam ediyor)

Tablo 11 (Devam)

A ve B Formlarına Ait Madde Analizleri

Faktör	Madde No	A Formu			Madde No	B Formu		
		p	r_{jx}	Maddenin İncelenmesi		p	r_{jx}	Maddenin İncelenmesi
Yer-Mekân U.	38	.57	.49	Yeterli	38	.58	.56	Yeterli
Muhakeme/SS	39	.68	.46	Yeterli	39	.43	.58	Yeterli
Muhakeme/SS	40	.61	.60	Yeterli	40	.46	.63	Yeterli
Muhakeme/SS	41	.37	.52	Yeterli	41	.63	.59	Yeterli
Muhakeme/SS	42	.54	.75	Yeterli	42	.29	.46	Yeterli
Yer-Mekân U	43	.51	.48	Yeterli	43	.34	.48	Yeterli
Yer-Mekân U	44*	.43	.47	Yeterli	44*	.49	.51	Yeterli
Algısal	45	.39	.48	Yeterli	45	.39	.31	Yeterli
Algısal	46	.45	.50	Yeterli	46	.36	.51	Yeterli
Kelime A.	47	.49	.52	Yeterli	47	.54	.43	Yeterli
Kelime A.	48	.66	.44	Yeterli	48	.60	.45	Yeterli

* Söзде Ortak Madde

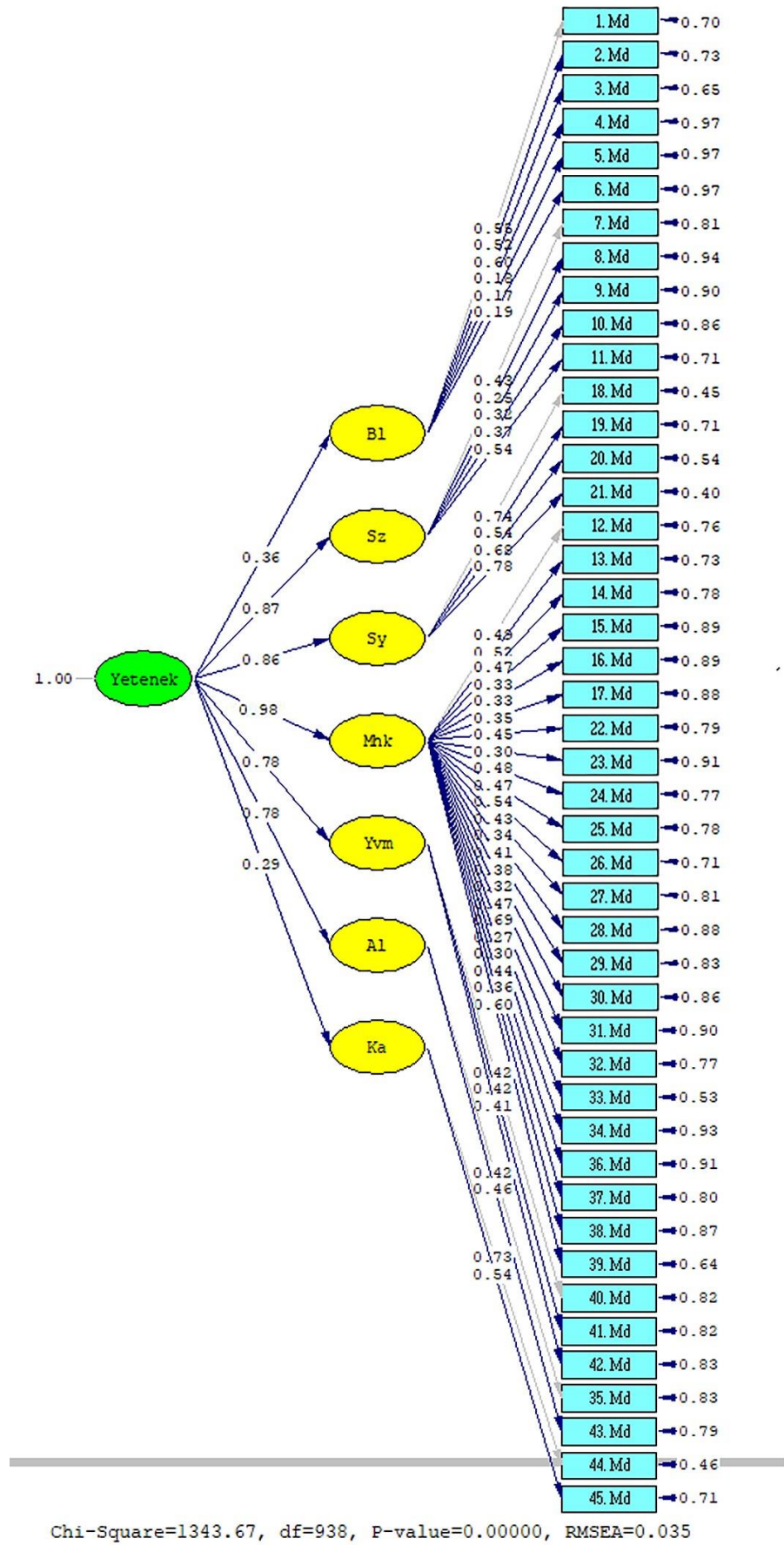
A ve B formlarında madde güçlük ve ayırt edicilik indeksleri (çift serili korelasyon katsayısı) hesaplanmıştır.

- A formunda madde güçlük indekslerinin .15 ile .87 aralığında değiştiği, genel olarak kolay ve orta güçlükte olan maddelerin ağırlıkta olduğu görülmektedir. Maddelerin ayırt edicilik indeksi olarak çift serili korelasyon katsayısı incelendiğinde ise .17 ile .84 aralığında değiştiği ve maddelerin ayırt ediciliğinin genel olarak yeterli olduğu belirlenmiştir. A formunda bellek faktöründen 5. ve 6. maddelerin düzeltilip teste alınmasına; bellek faktöründen 4. maddenin ve sözel faktörden 9. ve 13. maddelerin testten çıkarılmasına karar verilmiştir. Diğer maddelerin ise teste alınabileceği görülmektedir.
- B formunda madde güçlük indekslerinin .16 ile .86 aralığında değiştiği ve genel olarak orta güçlükte olan maddelerin ağırlıkta olduğu görülmektedir. Maddelerin ayırt edicilik indeksi olarak çift serili korelasyon katsayısı incelendiğinde ise .17 ile .79 aralığında değiştiği ve maddelerin ayırt ediciliğinin genel olarak yeterli olduğu belirlenmiştir. B formunda bellek faktöründen 4. ve 6. maddelerin düzeltilip teste alınmasına; bellek faktöründen 7. maddenin ve sözel faktörden 8. ve 12. maddelerin testten çıkarılmasına karar verilmiştir. Diğer maddelerin ise teste alınabileceği görülmektedir. Her iki formda da maddeler çıkarıldıktan sonra 45 madde ile testlere nihai halinin verilmesine karar verilmiştir.

Thurstone'un Çoklu Faktörler modelindeki 7 faktörlü yapı temel alınarak geliştirilen bu formların yapı geçerliğini belirlemek amacıyla İkinci Düzey DFA yapılmıştır. Elde edilen bulgular her form için sırasıyla aşağıda verilmiştir.

A formunda yeteneğin; bellek, sözel, sayısal, muhakeme, yer ve mekâna uyma, algısal ve kelime akıcılığı olmak üzere yedi gizil değişkeni ne ölçüde tanımladığını belirlemek amacıyla İkinci Düzey DFA yapılmıştır. Elde edilen model Şekil 5'te verilmiştir.





Şekil 5. A Formu İçin İkinci Düzey DFA

A formunda İkinci Düzey DFA sonucunda elde edilen uyum indeksleri Tablo 12’de verilmiştir.

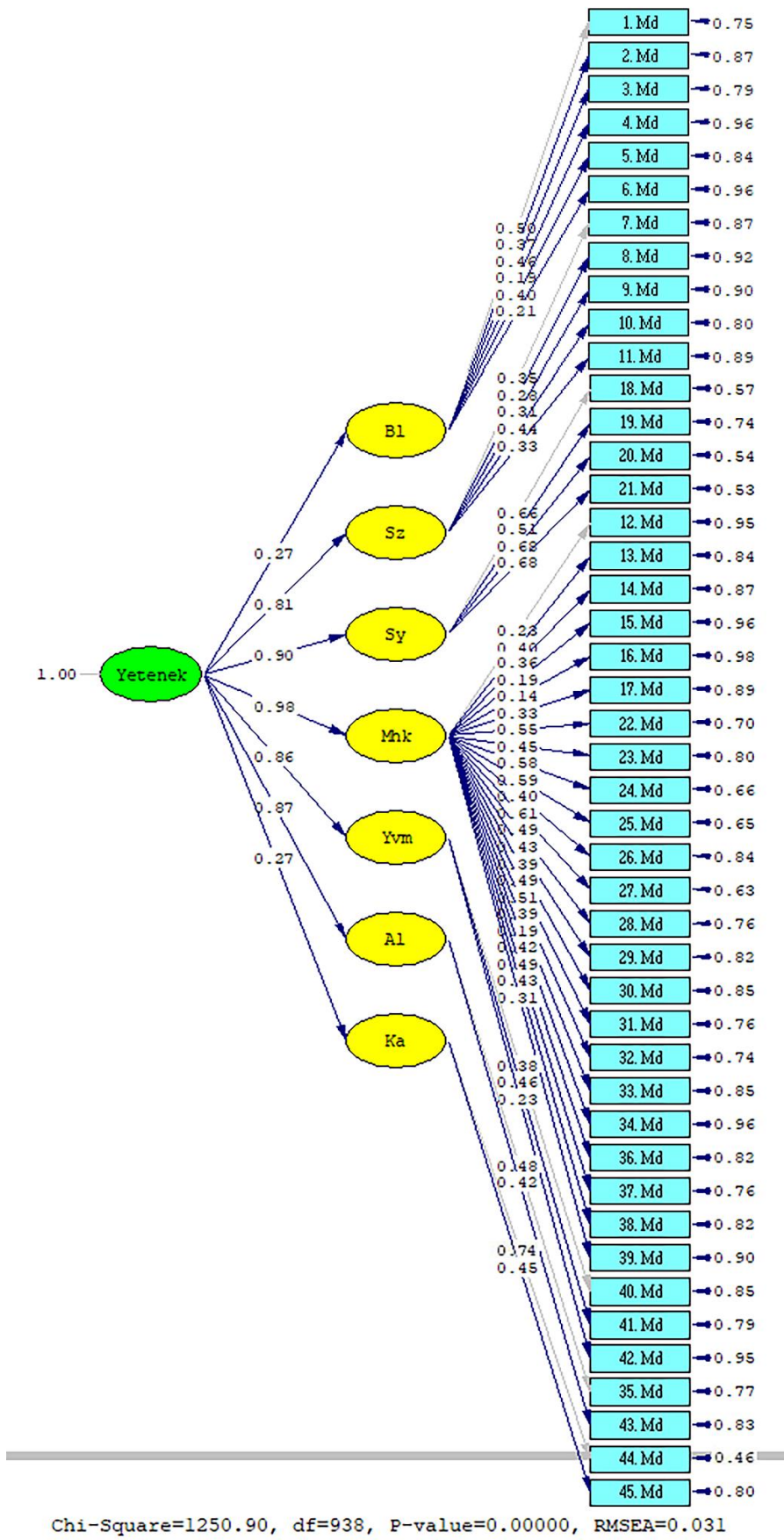
Tablo 12

A Formunda İkinci Düzey DFA Uyum İndeksleri

Uyum İndeksleri	İyi Uyum	Kabul Edilebilir Uyum	Model Uyum Değerleri	
			LISREL	R
χ^2			1343.67	1301.72
p			.000	.000
sd			938	938
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 < \chi^2/sd \leq 3$	1.436	1.387
RMSEA	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$	.035	.033
SRMR	$0 \leq SRMR \leq .05$	$.05 < SRMR \leq .08$	.053	.053
GFI	$.95 \leq GFI \leq 1$	$.90 \leq GFI < .95$	.85	
NNFI (TLI)	$.95 \leq NNFI \leq 1$	$.90 \leq NNFI < .95$	.95 (NNFI)	.91 (TLI)
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI < .95$	.96	.91

Modelde İkinci Düzey DFA sonucunda maddelerin t değerlerinin (2.61 – 19.29) anlamlı olduğu ve hata varyans değerlerinin (.40 – .97) ise bellek faktöründe 4., 5. ve 6. maddelerde .95’in üzerinde olduğu tespit edilmiştir. Bu maddelerin t değerleri anlamlı olduğundan modelde kalmaları yönünde karar verilmiştir. Elde edilen uyum indeksleri incelendiğinde χ^2/sd oranının 2’nin altında olması model veri uyumunun iyi olduğunun göstergesidir (Tabachnick ve Fidell, 2020). RMSEA değerinin .08’den küçük olması iyi uyuma işaret eder (Jöreskog ve Sörbom, 1993). SRMR değerinin de .08’den küçük olması iyi uyumun göstergesidir (Brown, 2006). NNFI (TLI) ve CFI değerlerinin 0.90 ve üzerinde olması kabul edilebilir uyuma; .95 ve üzerinde olması iyi uyuma işaret etmektedir (Sümer, 2000). GFI değerinin .90’ın altında olduğu görülmektedir. Modele ait uyum indeksleri bütün olarak incelendiğinde, GFI haricindeki uyum indekslerinin kabul edilebilir sınırlar arasında olduğu görülmektedir. GFI uyum indeksi faktör yüklerine ve örneklem büyüklüğüne daha hassas olduğu için düşük çıkma eğiliminde olabileceği belirtilmektedir (Shevlin ve Miles, 1998). Diğer uyum indekslerinin kabul edilebilir düzeyde olmasına dayanarak modelin doğrulandığı ifade edilebilir.

B formunda yeteneğin; bellek, sözel, sayısal, muhakeme, yer ve mekâna uyma, algısal ve kelime akıcılığı olmak üzere yedi gizil değişkeni ne ölçüde tanımladığını belirlemek amacıyla İkinci Düzey DFA yapılmıştır. Elde edilen model Şekil 6’da verilmiştir.



Şekil 6. B Formu İçin İkinci Düzey DFA

B formunda İkinci Düzey DFA sonucunda elde edilen uyum indeksleri Tablo 13'te verilmiştir.

Tablo 13

B Formunda İkinci Düzey DFA Uyum İndeksleri

Uyum İndeksleri	İyi Uyum	Kabul Edilebilir Uyum	Model Uyum Değerleri	
			LISREL	R
χ^2			1250.90	1258.59
p			.000	.000
sd			938	938
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 < \chi^2/sd \leq 3$	1.333	1.341
RMSEA	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$	.031	.031
SRMR	$0 \leq SRMR \leq .05$	$.05 < SRMR \leq .08$	.051	.051
GFI	$.95 \leq GFI \leq 1$	$.90 \leq GFI < .95$	.86	
NNFI (TLI)	$.95 \leq NNFI \leq 1$	$.90 \leq NNFI < .95$	.96 (NNFI)	.91 (TLI)
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI < .95$	.96	.92

Modelde İkinci Düzey DFA sonucunda maddelerin t değerlerinin (2.79 – 12.38) anlamlı olduğu ve hata varyans değerlerinin (.46 – .98) ise bellek faktöründe 4. ve 6. maddelerde, muhakeme faktöründe 15., 16. ve 34. maddelerde .95'in üzerinde olduğu tespit edilmiştir. Bu maddelerin t değerleri anlamlı olduğundan modelde kalmaları yönünde karar verilmiştir. Elde edilen uyum indeksleri bütün olarak incelendiğinde, GFI haricindeki uyum indekslerinin kabul edilebilir sınırlar arasında olduğu görülmektedir. Diğer uyum indekslerinin kabul edilebilir düzeyde olmasına dayanarak modelin doğrulandığı ifade edilebilir.

A ve B formlarında faktör yapısı doğrulandıktan sonra güvenilirlik analizleri yapılmış ve ölçüt geçerliği belirlenmiştir. Test güvenilirliğine yönelik McDonald'ın ω katsayısı, iç tutarlılık anlamında bilgi sağlamaktadır. Kestirilen ω katsayısı A formunda .90 ve B formunda .89 olduğundan formların iç tutarlılık anlamında güvenilirliğinin yeterli olduğu söylenebilir. Formların ölçüt geçerliğini belirlemek amacıyla öğrencilerin yetenek testinden aldıkları puanlar ile LGS yerleştirme puanları arasındaki Pearson Momentler Çarpımı korelasyon katsayısı (r_{xy}) hesaplanmıştır. A formunda $r_{xy} = .85$ ve B formunda $r_{xy} = .86$ olarak belirlenmiştir. Elde edilen sonuçlar incelendiğinde formların ölçüt geçerliğinin olduğu söylenebilir.

Deneme uygulaması sonucunda güvenilirlik ve geçerlik analizleri yapılan test formları gerekli düzeltmeler yapıp nihai uygulama için hazırlanmıştır.

Nihai Uygulama

Nihai uygulama sonucunda test formlarında bulunan madde sayıları ve puanlama şekli Tablo 14’te verilmiştir.

Tablo 14

Test Formlarında Bulunan Maddeler

Form	A Formu		B Formu	
Madde Türü	Çoktan Seçmeli	Kısa Cevaplı	Çoktan Seçmeli	Kısa Cevaplı
Madde Sayısı	43 Madde (31 Madde + 12 Sözde Ortak Madde)	2 Madde	43 Madde (31 Madde + 12 Sözde Ortak Madde)	2 Madde
Puanlama	0-1	0,1,2,3 (Kısmi Puanlama)	0-1	0,1,2,3 (Kısmi Puanlama)

Her bir test formunda 12 madde sözde ortak madde olmak üzere toplam 43 çoktan seçmeli madde ve 2 kısa cevaplı madde bulunmaktadır. Çoktan seçmeli maddeler 0-1 şeklinde puanlanırken kısa cevaplı maddeler 0 ile 3 arasında kısmi olarak puanlanmıştır. Bu durumda bir test formundan alınabilecek en yüksek puan 49’dur.

Nihai uygulama sonucunda her iki formda da çok yönlü uç değerleri belirlemek için Mahalanobis uzaklıkları incelenmiştir. A formunda uç değer tespit edilmezken B formunda bir veri uç değer olarak belirlenmiştir. B formunda bu veri örneklemden çıkarılmış ve her iki formda da 2100 kişilik örneklem ile madde ve test istatistikleri incelenmiştir.

Nihai uygulama sonucunda A ve B formları için KTK ve MTK’ye dayalı madde ve test istatistikleri hesaplanmıştır. MTK’ye dayalı madde istatistiklerinde, 0-1 şeklinde puanlanan çoktan seçmeli maddeler için iki kategorili 1 parametrelili lojistik model (1-PL), 2 parametrelili lojistik model (2-PL) ve 3 parametrelili lojistik model (3-PL) kestirimleri; 0 ile 3 arasında kısmi olarak puanlanan kısa cevaplı maddeler için çok kategorili Kısmi Puan Modeli (KPM), Genelleştirilmiş Kısmi Puan Modeli (GKPM) ve Aşamalı Tepki Modeli (ATM) kestirimleri yapılmıştır. Elde edilen sonuçlar Tablo 15 ve Tablo 16’da verilmiştir.

Tablo 15

A ve B Formlarında İki Kategorili Modeller Arasında Yapılan ANOVA Sonuçları

		Model	AIC	BIC	log.Lik	LRT	df	p
A Formu	1-PL ve	1-PL	101674.9	101923.5	-50793.48			
	2-PL	2-PL	100144.5	100630.3	-49986.23	1614.49	42	< .001
	2-PL ve	2-PL	100144.5	100630.3	-49986.23			
	3-PL	3-PL	99815.1	100543.9	-49778.55	415.36	43	< .001
B Formu	1-PL ve	1-PL	101155.81	101404.40	-50533.91			
	2-PL	2-PL	99407.13	99893.01	-49617.57	1832.68	42	< .001
	2-PL ve	2-PL	99407.13	99893.01	-49617.57			
	3-PL	3-PL	99095.13	99823.94	-49418.57	398	43	< .001

İki kategorili maddeler için ANOVA sonuçları incelenmiştir. Her iki formda da 1-PL ve 2-PL modeller incelendiğinde aralarında anlamlı bir fark bulunmuş ve 2-PL modelin AIC ve log.Lik uyum indekslerinin daha düşük olduğu görülmüştür. Bu durum 2-PL modelin veriye daha iyi uyum sağladığını göstermektedir. Daha sonra 2-PL ve 3-PL modeller incelendiğinde aralarında anlamlı bir fark bulunmuş ve 3-PL modelin AIC ve log.Lik uyum indekslerinin daha düşük olduğu görülmüştür. Bu durum 3-PL modelin veriye daha iyi uyum sağladığını göstermektedir. Sonuç olarak çoktan seçmeli maddeler için 3-PL model kullanılmıştır.

Tablo 16

A ve B Formlarında Çok Kategorili Modellerin Uyum İndeksleri

	Model	AIC	BIC	log.Lik
A Formu	KPM	-4703.153	9422.306	9467.503
	GKPM	-4693.812	9401.624	9441.172
	ATM	-4687.758	9391.515	9436.713
B Formu	KPM	-4458.238	8928.475	8971.373
	GKPM	-4451.072	8916.144	8962.341
	ATM	-4433.238	8882.476	8927.341

Çok kategorili maddeler için uyum indeksleri incelenmiştir. Her iki formda da ATM'de AIC ve log.Lik uyum indekslerinin diğer modeller içerisinde en düşük olduğu görülmüştür ve bu durum daha iyi uyuma işaretir. Sonuç olarak kısmi olarak puanlanan kısa cevaplı maddeler için ATM kullanılmıştır.

A Formunda Nihai Uygulama Verilerine Ait Madde ve Test İstatistikleri. A

formunda nihai uygulama verilerine ait madde istatistikleri Tablo 17 ve Tablo 18’de verilmiştir.

Tablo 17

A Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri

Faktörler	Madde No	KTK		MTK (3PL)		
		<i>p</i>	<i>r_{jx}</i>	<i>b</i> parametresi	<i>a</i> parametresi	<i>c</i> parametresi
Bellek	1	.79	.33	-2.055	.715	.002
Bellek	2	.81	.39	-2.145	.638	.027
Bellek	3	.86	.48	-2.225	.841	.020
Bellek	4	.61	.31	-1.152	.520	.003
Bellek	5	.54	.30	-.451	.495	.025
Bellek	6	.51	.32	-.004	.535	.023
Sözel	7	.70	.51	-1.087	.918	.004
Sözel	8*	.64	.34	-.190	.571	.237
Sözel	9*	.80	.49	-1.803	.885	.001
Sözel	10	.50	.40	.526	.747	.150
Sözel	11	.78	.61	-1.294	1.268	.002
Muhakeme/A	12	.72	.66	-.787	1.492	.051
Muhakeme/A	13	.63	.67	-.290	1.676	.100
Muhakeme/A	14	.72	.66	-.833	1.459	.026
Muhakeme/A	15	.67	.54	-.855	.996	.000
Muhakeme/A	16	.64	.52	-.470	1.003	.109
Muhakeme/A	17*	.57	.50	.070	1.067	.157
Sayısal	18*	.42	.73	.418	2.714	.073
Sayısal	19	.68	.69	-.378	1.938	.160
Sayısal	20	.33	.69	.758	2.772	.083
Sayısal	21*	.44	.75	.356	2.863	.085
Muhakeme/GM	22	.28	.61	1.046	2.823	.116
Muhakeme/GM	23	.15	.57	1.567	2.439	.055
Muhakeme/GM	24*	.27	.61	1.091	1.814	.062
Muhakeme/GM	25	.36	.59	.647	1.184	.009
Muhakeme/GM	26*	.33	.69	.742	2.230	.063
Muhakeme/GM	27*	.20	.59	1.327	2.910	.082
Muhakeme/GM	28	.49	.50	.298	1.009	.091
Muhakeme/GM	29	.60	.56	-.485	.971	.000
Muhakeme/GM	30*	.20	.65	1.206	3.564	.077
Muhakeme/SS	31	.29	.56	1.150	1.470	.080
Muhakeme/SS	32*	.70	.63	-.703	1.410	.072
Muhakeme/SS	33	.57	.77	-.085	2.319	.060
Muhakeme/SS	34	.63	.54	-.243	1.175	.155
Yer-Mekân U.	35*	.56	.54	-.292	.931	.002
Muhakeme/SS	36	.66	.55	-.812	.992	.000
Muhakeme/SS	37	.58	.56	-.114	1.125	.110
Muhakeme/SS	38	.38	.51	.814	1.034	.071
Muhakeme/SS	39	.51	.71	.136	1.862	.070
Yer-Mekân U	40	.49	.53	.560	1.421	.192
Yer-Mekân U	41*	.43	.57	.792	2.013	.183

* Söзде Ortak Madde

(Devam ediyor)

Tablo 17 (Devam)

A Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri

Faktörler	Madde No	KTK		MTK (3PL)		
		p	r_{jx}	b parametresi	a parametresi	c parametresi
Algısal	42	.43	.42	.537	.663	.016
Algısal	43	.40	.55	.513	1.007	.010
Kelime	44	.49	.51			
Kelime	45	.62	.52			

* Söзде Ortak Madde

A formunda maddelerin KTK'ye dayalı madde analizleri incelendiğinde madde güçlük indekslerinin .15 ile .86 aralığında değiştiği; en kolay maddenin 3. madde ve en zor maddenin 23. madde olduğu görülmektedir. Genel olarak orta güçlükte olan maddelerin ağırlıkta olduğu, iki maddenin çok kolay ve üç maddenin çok zor olduğu belirlenmiştir. KTK'ye dayalı çift serili korelasyonla ayırt edicilik indeksleri incelendiğinde .30 ile .77 aralığında değiştiği; en az ayırt edici maddenin 5. madde ve en çok ayırt edici maddenin 33. madde olduğu görülmektedir. Maddelerin ayırt ediciliklerinin genel olarak yeterli olduğu belirlenmiştir.

Çoktan seçmeli maddeler için MTK'ye dayalı madde analizleri incelenmiştir. MTK'ye dayalı madde güçlüklerini ifade eden b parametreleri incelendiğinde -2.225 ile 1.567 aralığında değiştiği; en kolay maddenin 3. madde ve en zor maddenin 23. madde olduğu belirlenmiştir. Maddelerin çok zor ya da çok kolay olması istenmediğinden b parametresinin pratikte -2.00 ve +2.00 aralığında olması istenir (DeMars, 2016). Bu durumda b parametreleri incelendiğinde üç maddenin çok kolay olduğu belirlenmiştir. MTK'ye dayalı madde ayırt ediciliklerini ifade eden a parametreleri incelendiğinde .495 ile 3.564 aralığında değiştiği; en az ayırt edici maddenin 5. madde ve en çok ayırt edici maddenin 30. madde olduğu görülmektedir. Baker (2016), maddelerin ayırt edicilik parametre değerlerine göre çok düşük (.01 – .34), düşük (.35 – .64), orta (.65 – 1.34), yüksek (1.35 – 1.69) ve çok yüksek (>1.70) ayırtıcılığa sahip maddeler şeklinde sınıflandırılabilceğini belirtmiştir. Buna göre MTK'ye dayalı yapılan ayırt edicilik analizinde genel olarak orta düzeyde ayırt edici maddelerin ağırlıkta olduğu belirlenmiştir. MTK'ye dayalı şans parametresi olan c parametresinin .002 ile .192 aralığında değiştiği görülmektedir. Bu değer .30'un altında olması, öğrencilerin maddeyi şans ile cevaplama olasılıklarının düşük olduğunu göstermektedir (DeMars, 2016). Şans parametresi sıfıra yaklaştığında madde ölçülmek istenen özellikle ilgili daha fazla bilgi vermektedir.

Tablo 18

A Formunda Kısa Cevaplı Maddelerin MTK'ye Dayalı Madde İstatistikleri

Faktör	Madde No	ATM			
		<i>a</i>	<i>b1</i>	<i>b2</i>	<i>b3</i>
Kelime	44	1.018	-3.685	-1.092	1.053
Kelime	45	3.672	-1.970	-.810	.012

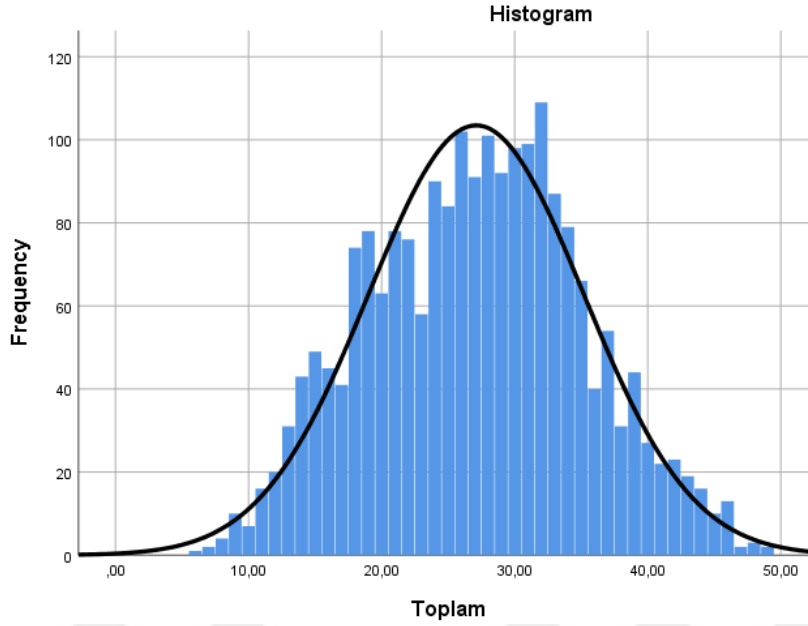
A formunda kısa cevaplı olan 44. ve 45. maddeler için ATM altında madde güçlük veya eşik parametreleri incelendiğinde beklendiği üzere eşik parametre değerleri en düşükten en yükseğe doğru sıralanmıştır. Bir başka ifadeyle her madde için *b1* en düşük eşik parametresine, *b3* en yüksek eşik parametresine sahiptir. Buna göre 44. maddenin orta ve 45. maddenin çok yüksek ayırt ediciliğe sahip olduğu söylenebilir.

A formunda nihai uygulama verilerine ait test istatistikleri Tablo 19 ve Şekil 7’de verilmiştir.

Tablo 19

A Formunda Nihai Uygulama Verilerine Ait Test İstatistikleri

A Formu Nihai Uygulama KTK'ye Dayalı Test İstatistikleri	
Değerlendirilen öğrenci sayısı	2100
Testten alınabilecek en yüksek puan	49
Testten alınan en düşük puan	6
Testten alınan en yüksek puan	49
Mod (X_{mod})	32.00
Medyan (X_{ort})	27.00
Aritmetik ortalama ($\bar{X}$)	27.10
Standart sapma (S_s)	8.10
Çarpıklık (K_y)	.031
Basıklık (B_s)	-.485
Ortalama güçlük (p)	.53
Ortalama ayırt edicilik (Çift serili korelasyon katsayısı – r_{jv})	.55
Güvenirlik (ω)	.91
A Formu Nihai Uygulama MTK'ye Dayalı Test İstatistikleri	
Madde ayırt edicilik parametresi ortalaması (a)	1.514
Çoktan seçmeli maddeler için madde güçlük parametresi ortalaması (b)	-.098
Çoktan seçmeli maddeler için şans parametresi ortalaması (c)	.068
Lord'un güvenilirlik katsayısı	.90



Şekil 7. A formuna Ait Histogram Grafiği

A formunda nihai uygulama sonucunda 45 madde için KTK ve MTK'ye dayalı test istatistikleri hesaplanmıştır. Nihai uygulama sonucunda 2100 öğrenci analize dahil edilmiştir. A formundan alınan en düşük puanın 6 ve en yüksek puanın 49 olduğu belirlenmiştir. A formuna ait merkezi eğilim ölçüleri incelendiğinde bu değerlerin birbirine yakın olduğu; çarpıklık ve basıklık katsayılarının ± 1 sınırları arasında olduğu görülmektedir. Şekil 7'de verilen histogram grafiği de incelendiğinde dağılımın normale yakın olduğu söylenebilir. KTK'ye dayalı test istatistiklerine göre A formunun orta güçlükte ($p = .53$) ve çok iyi ayırt edici olduğu ($r_{jx} = .55$) belirlenmiştir. Kestirilen ω katsayısı .91 olduğundan iç tutarlılık anlamında güvenilirliğinin yeterli olduğu söylenebilir.

A formunda çoktan seçmeli maddeler için hesaplanan güçlük parametresi (b) ortalamasının -.098 olduğu belirlenmiştir. Bu değer, 0'ın altında olduğu için A formunun biraz daha kolay olduğu ifade edilebilir. A formunun MTK'ye dayalı madde ayırtıcılık parametresi (a) ortalamasının 1.514 olduğu belirlenmiştir. Bu değer 2.00'a yakın olması testin ayırt ediciliğinin bir göstergesi olarak ifade edilebilir. A formunda çoktan seçmeli maddeler için hesaplanan şans parametresi (c) ortalaması .068 olup bu değer .30'dan düşük olduğu belirlenmiştir. Madde ve yetenek parametreleri arasında ilişki kuran Lord'un güvenilirlik katsayısı .90 olarak hesaplandığından testin güvenilirliğinin yeterli olduğu ifade edilebilir.

Sonuç olarak A formu için yapılan çalışmalar bütün olarak değerlendirilmiştir. A formunda 48 madde ile deneme uygulaması yapılmıştır. Bu uygulama sonucunda madde ve test istatistikleri incelenmiş; 3 maddenin testten çıkarılmasına ve 2 maddenin düzeltilip test alınmasına karar verilmiştir. Deneme uygulaması sonucunda yapı geçerliğini belirlemek için ikinci düzey DFA yapılmış, yapı doğrulanmış ve test nihai uygulama için hazırlanmıştır. Nihai uygulama sonucunda 45 madde ile madde ve test istatistikleri değerlendirilen A formunun geçerli ve güvenilir bir yetenek testi olduğu belirlenmiştir.

B Formunda Nihai Uygulama Verilerine Ait Madde ve Test İstatistikleri. B formunda nihai uygulama verilerine ait madde istatistikleri Tablo 20 ve Tablo 21’de verilmiştir.

Tablo 20

B Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri

Faktörler	Madde No	KTK		MTK (3PL)		
		<i>p</i>	<i>r_{jk}</i>	<i>b</i> parametresi	<i>a</i> parametresi	<i>c</i> parametresi
Bellek	1	.86	.33	-2.269	.556	.001
Bellek	2	.81	.36	-2.135	.568	.000
Bellek	3	.86	.36	-2.250	.561	.003
Bellek	4	.77	.37	-1.680	.574	.002
Bellek	5	.77	.32	-1.750	.330	.014
Bellek	6	.48	.30	1.507	.105	.071
Sözel	7	.50	.40	.011	.577	.000
Sözel	8*	.54	.43	-.270	.669	.001
Sözel	9*	.74	.34	-1.572	.511	.000
Sözel	10	.75	.47	-1.463	.834	.000
Sözel	11	.79	.61	-1.905	1.693	.140
Muhakeme/A	12	.41	.37	1.493	1.029	.240
Muhakeme/A	13	.49	.52	.062	.882	.000
Muhakeme/A	14	.79	.45	-1.206	1.348	.051
Muhakeme/A	15	.37	.39	1.226	.704	.083
Muhakeme/A	16	.48	.33	.227	.453	.000
Muhakeme/A	17*	.59	.52	-.479	.894	.000
Sayısal	18*	.38	.73	.563	2.561	.082
Sayısal	19	.64	.66	-.085	2.338	.235
Sayısal	20	.44	.73	.415	2.640	.117
Sayısal	21*	.50	.73	.249	2.590	.127
Muhakeme/GM	22	.60	.69	.023	2.447	.208
Muhakeme/GM	23*	.23	.63	1.188	3.376	.100
Muhakeme/GM	24	.39	.68	.485	1.554	.016
Muhakeme/GM	25*	.39	.67	.627	2.090	.108
Muhakeme/GM	26*	.16	.58	1.543	2.604	.063
Muhakeme/GM	27	.32	.70	.782	2.174	.059
Muhakeme/GM	28	.46	.68	.392	2.016	.111

* Söзде Ortak Madde

(Devam ediyor)

Tablo 20 (Devam)

B Formunda Nihai Uygulama Verilerine Ait Madde İstatistikleri

Faktörler	Madde No	KTK		MTK (3PL)		
		p	r_{jx}	b parametresi	a parametresi	c parametresi
Muhakeme/GM	29	.44	.57	.508	1.278	.096
Muhakeme/GM	30*	.27	.58	1.226	1.693	.085
Muhakeme/SS	31	.49	.64	.160	1.393	.043
Muhakeme/SS	32*	.65	.67	-.310	1.801	.125
Muhakeme/SS	33	.27	.54	1.315	1.736	.114
Muhakeme/SS	34	.19	.39	1.739	3.335	.137
Yer-Mekân U.	35*	.56	.51	-.313	.876	.000
Muhakeme/SS	36	.41	.54	.558	1.022	.039
Muhakeme/SS	37	.45	.69	.282	1.640	.047
Muhakeme/SS	38	.60	.63	-.391	1.307	.000
Muhakeme/SS	39	.31	.52	1.069	.913	.001
Yer-Mekân U	40	.35	.43	1.370	1.128	.160
Yer-Mekân U	41*	.44	.56	.780	2.352	.230
Algısal	42	.36	.39	1.509	1.166	.205
Algısal	43	.30	.52	1.271	1.839	.152
Kelime	44	.55	.51			
Kelime	45	.58	.52			

* Söзде Ortak Madde

B formunda çoktan seçmeli maddelerin KTK'ye dayalı madde analizleri incelendiğinde madde güçlük indekslerinin .16 ile .86 aralığında değiştiği; en kolay maddelerin 1. ve 3. madde ve en zor maddenin 26. madde olduğu görülmektedir. Genel olarak orta güçlükte olan maddelerin ağırlıkta olduğu, üç maddenin çok kolay ve iki maddenin çok zor olduğu belirlenmiştir. KTK'ye dayalı çift serili korelasyonla ayırt edicilik indeksleri incelendiğinde .30 ile .73 aralığında değiştiği; en az ayırt edici maddenin 6. madde ve en çok ayırt edici maddelerin 20. ve 21. madde olduğu görülmektedir. Maddelerin ayırt ediciliklerinin genel olarak yeterli olduğu belirlenmiştir.

Çoktan seçmeli maddeler için MTK'ye dayalı madde analizleri incelenmiştir. MTK'ye dayalı madde güçlüklerini ifade eden b parametreleri incelendiğinde -2.269 ile 1.739 aralığında değiştiği; en kolay maddenin 1. madde ve en zor maddenin 34. madde olduğu belirlenmiştir. Bu durumda b parametreleri genel olarak incelendiğinde üç maddenin çok kolay olduğu belirlenmiştir. MTK'ye dayalı madde ayırt ediciliklerini ifade eden a parametreleri incelendiğinde .105 ile 3.376 aralığında değiştiği; en az ayırt edici maddenin 6. madde ve en çok ayırt edici maddenin 23. madde olduğu görülmektedir. Baker'ın (2016) sınıflamasına göre çok yüksek düzeyde ayırt edici maddelerin ağırlıkta olduğu belirlenmiştir. MTK'ye dayalı şans parametresi olan c parametresinin .000 ile .240 aralığında değiştiği ve bu değer .30'dan düşük olduğu görülmektedir.

Tablo 21

B Formunda Kısa Cevaplı Maddelerin MTK'ye Dayalı Madde İstatistikleri

Faktör	Madde No	ATM			
		<i>a</i>	<i>b1</i>	<i>b2</i>	<i>b3</i>
Kelime	44	1.754	-2.768	-1.305	.658
Kelime	45	1.928	-2.647	-1.152	.321

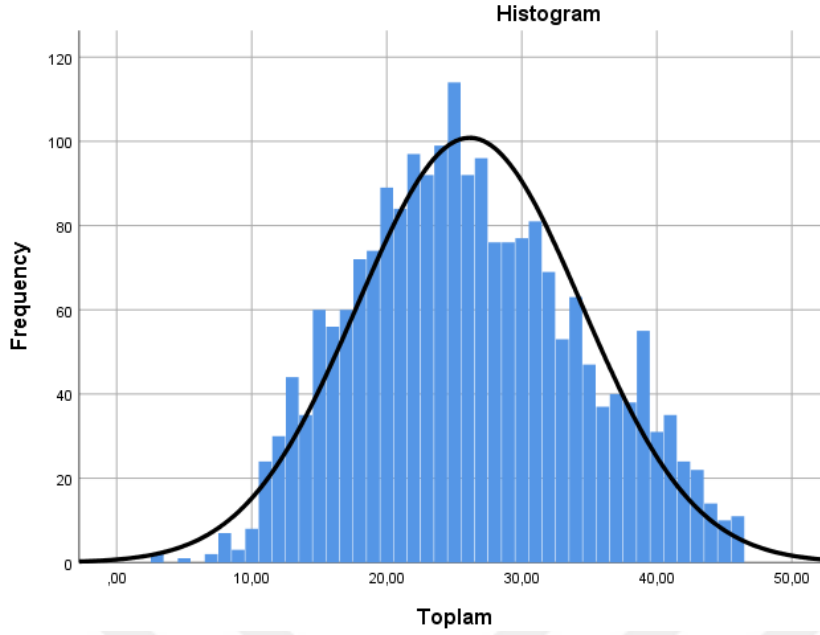
B formunda kısa cevaplı olan 44. ve 45. maddeler için ATM altında madde güçlük veya eşik parametreleri incelendiğinde beklendiği üzere eşik parametre değerleri en düşükten en yükseğe doğru sıralanmıştır. Bir başka ifadeyle her madde için *b1* en düşük eşik parametresine, *b3* en yüksek eşik parametresine sahiptir. Buna göre 44. ve 45. maddenin çok yüksek ayırt ediciliğe sahip olduğu söylenebilir.

B formunda nihai uygulama verilerine ait test istatistikleri Tablo 22 ve Şekil 8'de verilmiştir.

Tablo 22

B Formunda Nihai Uygulama Verilerine Ait Test İstatistikleri

B Formu Nihai Uygulama KTK'ye Dayalı Test İstatistikleri	
Değerlendirilen öğrenci sayısı	2100
Testten alınabilecek en yüksek puan	49
Testten alınan en düşük puan	3
Testten alınan en yüksek puan	46
Mod (X_{mod})	25.00
Medyan (X_{ort})	25.00
Aritmetik ortalama ($\bar{X}$)	26.14
Standart sapma (S_s)	8.31
Çarpıklık (K_y)	.203
Basıklık (B_s)	-.565
Ortalama güçlük (p)	.51
Ortalama ayırt edicilik (Çift serili korelasyon katsayısı - r_{jx})	.52
Güvenirlilik (ω)	.90
B Formu Nihai Uygulama MTK'ye Dayalı Test İstatistikleri	
Madde ayırt edicilik parametresi ortalaması (a)	1.463
Çoktan seçmeli maddeler için madde güçlük parametresi ortalaması (b)	.105
Çoktan seçmeli maddeler için şans parametresi ortalaması (c)	.076
Lord'un güvenirlilik katsayısı	.89



Şekil 8. B formuna Ait Histogram Grafiği

B formunda nihai uygulama sonucunda 45 madde için KTK ve MTK'ye dayalı test istatistikleri hesaplanmıştır. Nihai uygulama sonucunda 2100 öğrenci analize dahil edilmiştir. B formundan alınan en düşük puanın 3 ve en yüksek puanın 46 olduğu belirlenmiştir. B formuna ait merkezi eğilim ölçüleri incelendiğinde bu değerlerin birbirine yakın olduğu; çarpıklık ve basıklık katsayılarının ± 1 sınırları arasında olduğu görülmektedir. Şekil 8'de verilen histogram grafiği de incelendiğinde dağılımın normale yakın olduğu söylenebilir. KTK'ye dayalı test istatistiklerine göre B formunun orta güçlükte ($p = .51$) ve çok iyi ayırt edici olduğu ($r_{jx} = .52$) belirlenmiştir. Kestirilen ω katsayısı .90 olduğundan iç tutarlılık anlamında güvenilirliğinin yeterli olduğu söylenebilir.

B formunda çoktan seçmeli maddeler için hesaplanan güçlük parametresi (b) ortalamasının .105 olduğu belirlenmiştir. Bu değer, 0'ın üzerinde olduğu için B formunun biraz daha zor olduğu ifade edilebilir. B formunun MTK'ye dayalı madde ayıricılık parametresi (a) ortalamasının 1.463 olduğu belirlenmiştir. Bu değer 2.00'a yakın olması testin ayırt ediciliğinin bir göstergesi olarak ifade edilebilir. B formunda çoktan seçmeli maddeler için hesaplanan şans parametresi (c) ortalaması .076 olup bu değer .30'dan düşük olduğu belirlenmiştir. Madde ve yetenek parametreleri arasında ilişki kuran Lord'un güvenilirlik katsayısı .89 olarak hesaplandığından testin güvenilirliğinin yeterli olduğu söylenebilir.

Sonuç olarak B formu için yapılan çalışmalar bütün olarak değerlendirilmiştir. B formunda 48 madde ile deneme uygulaması yapılmıştır. Bu uygulama sonucunda madde ve test istatistikleri incelenmiş; 3 maddenin testten çıkarılmasına ve 2 maddenin düzeltilip test alınmasına karar verilmiştir. Deneme uygulaması sonucunda yapı geçerliğini belirlemek için ikinci düzey DFA yapılmış yapı doğrulanmış ve test nihai uygulama için hazırlanmıştır. Nihai uygulama sonucunda 45 madde ile madde ve test istatistikleri değerlendirilen B formunun geçerli ve güvenilir bir yetenek testi olduğu belirlenmiştir.

A ve B formlarının geçerlik ve güvenirlik analizlerinden sonra formlar test eşitleme için hazırlanmıştır. Bu formlarda çoktan seçmeli ve kısa cevaplı maddeler bulunmaktadır. Çoktan seçmeli maddeler 0-1 şeklinde puanlanırken kısa cevaplı maddeler 0 ile 3 arasında kısmi olarak puanlanmıştır. Maddelerin değerlendirilmesinde modeli etkilemesini önlemek amacıyla bütün maddeler aynı kategori sayısında alınmıştır. Kısmi puanlanan maddelerde tamamen doğru yanıtlardan oluşan 3 kategorisi 1 olarak, kısmen doğru ve yanlış yanıtlardan oluşan 2, 1 ve 0 kategorileri 0 olacak şekilde yeniden kodlanarak iki kategoriye dönüştürülmüştür.

Test Eşitleme Varsayımlarının Test Edilmesi

Bu araştırmada, geliştirilen yetenek testinin A ve B formlarının gerçekte ortak maddeleri yoktur. Testlerde her faktörün altında madde yazılırken bu maddelerin paralel olmasına dikkat edilmiştir. Eşitlenecek her iki testte de aynı faktör altında bulunan maddelerden güçlükleri, ayırt edicilikleri ve güvenirlikleri benzer olacak şekilde maddeler seçilmiş ve paralel iki sözde ortak test oluşturulmaya çalışılmıştır. Bu alt testlerin madde sayısı testin toplam madde sayısının en az %20'si olacak şekilde düzenlenmiştir.

Bu araştırmada geliştirilen yetenek testinin farklı formları üzerinde eşitleme çalışmaları gerçekleştirilmiştir. Öncelikle eşitleme varsayımlarının incelenmesine yer verilmiştir. Test eşitlemede test formlarının istatistiksel özellikler açısından benzer olmaları önemlidir (Kolen ve Brennan, 2014). Test geliştirme sürecinde yapılan analizler sonucunda A ve B formlarına ait test istatistiklerinin birbirine yakın olduğu belirlenmiştir. Tüm maddeleri iki kategorili olacak şekilde düzenlenen A ve B formları ve bunların sözde ortak testleri ile ilgili test istatistikleri bütün olarak Tablo 23'te verilmiştir. Eşitleme varsayımlarını kontrol etmek amacıyla KTK ve MTK'ye dayalı test istatistiklerinde anlamlı farklılık olup olmadığı incelenmiştir.

Tablo 23

A ve B Formları ve Sözde Ortak Testler İçin Test İstatistikleri

İstatistik	Hedef Test		Temel Test	
	A Formu	A-Sözde Ortak	B Formu	B-Sözde Ortak
N	2100	2100	2100	2100
Madde Sayısı	45	12	45	12
Testten alınan en düşük puan	2	0	1	0
Testten alınan en yüksek puan	45	12	42	12
Aritmetik ortalama ($\bar{X}$)	23.66	5.54	22.56	5.46
Medyan (X_{ort})	24.00	5.00	22.00	5.00
Mod (X_{mod})	27.00	5.00	20.00	4.00
Standart sapma (S_s)	8.59	2.81	8.09	2.80
Çarpıklık (K_y)	.033	.421	.244	.382
Basıklık (B_s)	-.574	-.437	-.600	-.570
Ortalama güçlük (p)	.50	.47	.50	.46
Ortalama ayırt edicilik (Çift serili korelasyon katsayısı – r_{jx})	.54	.59	.51	.58
Güvenirlik (α)	.89	.82	.88	.81
b parametresi ortalaması	-.019	.251	.169	.270
a parametresi ortalaması	1.415	1.914	1.400	1.834
c parametresi ortalaması	.066	.091	.073	.077
Lord'un güvenilirlik katsayısı	.90	.81	.89	.81

Eşitleme için ortak maddeler testin bütünü temsil edecek şekilde oluşturulmalıdır. Bu sebeple her iki formda da sözde ortak testler ile testin bütünü arasındaki korelasyon katsayısı (Pearson r) incelenmiştir. A formunda $r_{xy} = .91$ ve B formunda $r_{xy} = .92$ olarak hesaplanmıştır. Sözde ortak testlerden alınan puanlar ile testin bütünü arasında pozitif yönlü oldukça yüksek bir korelasyon bulunmaktadır. Bu sonuca göre sözde ortak testlerin formları güçlü bir şekilde temsil ettiği söylenebilir.

Yapılan analizler sonucunda A ve B formları ile sözde ortak testlere ait güvenilirlik katsayılarının .80'den büyük olduğu ve testlerin güvenilirliklerinin yüksek olduğu görülmüştür. Güvenirlikler arasındaki farklılığın manidarlığını belirlemek için Fisher Z_r dönüşümü yapılmıştır (Akhun, 1984; Karadavut, 2021). Elde edilen Fisher Z değerleri A ve B testleri için sırasıyla 1.422 ve 1.376'dır. A ve B testlerinde Z değerleri arasındaki fark standart hataya bölündüğünde $Z = .211$ olarak bulunmuştur. Bu değer Z değeri ile karşılaştırıldığında H_0 hipotezinin bölgesinde bulunduğu görülmüştür. Dolayısıyla A ve B testlerinin güvenilirlikleri arasında manidar farklılığın olmadığı belirlenmiştir ($p > .05$). Her iki formda da güvenilirliklerin benzer olduğu söylenebilir. Benzer durum sözde ortak testlerin ($Z = .063$) karşılaştırılmasında da geçerlidir ve bu testlerin de güvenilirliklerinin benzer olduğu söylenebilir.

Bununla birlikte A ve B testleri ile sözde ortak testlere ait ortalama madde güçlükleri .40 ile .60 arasında olduğundan testler orta güçlüktedir. Veri seti normal dağıldığından madde güçlükleri açısından testler arasındaki farklılık bağımsız örneklemeler t testi ile incelenmiştir. Sonuç olarak madde güçlüklerine göre A ve B formları arasında anlamlı farklılık bulunmamıştır ($t_{(88)} = .769, p = .444$). Benzer şekilde A sözde ortak testi ve B sözde ortak testi ($t_{(22)} = .121, p = .905$); A formu ve A sözde ortak testi ($t_{(55)} = 1.050, p = .298$); B formu ve B sözde ortak testi ($t_{(55)} = .701, p = .486$) arasında madde güçlüklerine göre anlamlı farklılık olmadığı belirlenmiştir.

A ve B testleri ile sözde ortak testlere ait madde ayırt edicilikleri .40'ın üzerinde olduğundan testler yeterli ayırt ediciliktedir. Benzer şekilde madde ayırt edicilikleri açısından testler arasındaki farklılık incelenmiştir. Sonuç olarak madde ayırt ediciliklerine göre A ve B formları arasında anlamlı farklılık bulunmamıştır ($t_{(88)} = .932, p = .354$). Benzer şekilde A sözde ortak testi ve B sözde ortak testi ($t_{(22)} = .246, p = .808$); A formu ve A sözde ortak testi ($t_{(55)} = -1.209, p = .232$); B formu ve B sözde ortak testi ($t_{(55)} = -1.450, p = .153$) arasında madde ayırt ediciliklerine göre anlamlı farklılık olmadığı belirlenmiştir.

A ve B testleri ile sözde ortak testlere ait b parametreleri ortalaması incelendiğinde B formunun A formundan; B sözde ortak testinin A sözde ortak testinden ve sözde ortak testlerin test formlarından biraz daha zor olduğu belirlenmiştir. Daha sonra b parametreleri açısından testler arasındaki farklılık incelenmiştir. Sonuç olarak b parametrelerine göre A ve B formları arasında anlamlı farklılık bulunmamıştır ($t_{(88)} = -.812, p = .419$). Benzer şekilde A sözde ortak testi ve B sözde ortak testi ($t_{(22)} = -.050, p = .960$); A formu ve A sözde ortak testi ($t_{(55)} = -.840, p = .405$); B formu ve B sözde ortak testi ($t_{(55)} = -.274, p = .785$) arasında b parametrelerine göre anlamlı farklılık olmadığı belirlenmiştir.

A ve B testleri ile sözde ortak testlere ait a parametreleri ortalaması incelendiğinde A formunun B formundan; A sözde ortak testinin B sözde ortak testinden ve sözde ortak testlerin test formlarından biraz daha ayırt edici olduğu belirlenmiştir. Sonuç olarak a parametreleri açısından testler arasındaki farklılık incelenmiştir. Buna göre a parametrelerine göre A ve B formları arasında anlamlı farklılık bulunmamıştır ($t_{(88)} = -.073, p = .942$). Benzer şekilde A sözde ortak testi ve B sözde ortak testi ($t_{(22)} = -.208, p = .837$); A formu ve A sözde ortak testi ($t_{(55)} = -1.867, p = .067$); B formu ve B sözde ortak testi ($t_{(55)} = -1.556, p = .125$) arasında a parametrelerine göre anlamlı farklılık olmadığı belirlenmiştir.

A ve B testleri ile sözde ortak testlere ait c parametreleri ortalaması incelendiğinde bu değerin bütün testlerde .30'un altında olduğu belirlenmiştir. Daha sonra c parametreleri açısından testler arasındaki farklılık incelenmiştir. Sonuç olarak c parametrelerine göre A ve B formları arasında anlamlı farklılık bulunmamıştır ($t_{(88)} = -.502, p = .617$). Benzer şekilde A sözde ortak testi ve B sözde ortak testi ($t_{(22)} = .501, p = .882$); A formu ve A sözde ortak testi ($t_{(55)} = -1.264, p = .212$); B formu ve B sözde ortak testi ($t_{(55)} = -.174, p = .862$) arasında c parametrelerine göre anlamlı farklılık olmadığı belirlenmiştir.

Sonuç olarak test ve madde istatistikleri ve güvenirlikleri incelendiğinde A ve B formlarının benzer testler olduğu belirlenmiştir. Aynı durum sözde ortak testler için de geçerlidir. Bu sonuçlar, eşitlenecek olan iki testin istatistiksel özellikler açısından benzer olduğunu ortaya koymaktadır.

Bu araştırmadaki yetenek testi istatistiksel olarak EG ve eş değer olmayan gruplar (NEAT ve NEC) desenlerine ait denklemler kullanılarak eşitlenmiştir. Bu sebeple eş değer ve eş değer olmayan iki grup oluşturulmuştur. Her okulda spirallemeye uygun olarak dağıtılan ve geniş bir örnekleme ulaşılan verinin tamamı EG deseni için kullanılmıştır. Daha sonra okulların 2022 LGS yerleştirme puanları dikkate alınmış ve bu okullar iki gruba ayrılmıştır. Bir grupta A formunu alan öğrenciler ve diğer grupta B formunu alan öğrenciler seçilerek yeni bir grup oluşturulmuştur ve bu grup eş değer olmayan gruplar deseni için kullanılmıştır. Sonuç olarak oluşturulan bu gruplarda A ve B formlarını alan öğrencilerin sayısı ve bu öğrencilerin ortalama 2022 LGS yerleştirme puanları Tablo 24'te verilmiştir.

Tablo 24

Gruplarda 2022 LGS Yerleştirme Puan Ortalaması

	EG		NEAT VE NEC	
	A Formu	B Formu	A Formu	B Formu
N	2100	2100	1000	1000
LGS Ortalaması	337.544	336.878	334.125	351.414

Araştırmada cinsiyet; LGS puanları; cinsiyet ve LGS puanları birlikte ortak değişken olarak kullanılmıştır. Sürekli olan LGS puanları kategorik hale getirilmiştir. Puanlar kategorik hale getirilirken öğrencilerin yüzdelik dilimleri dikkate alınmıştır. LGS puanları “500 – 370” arasında olan yüzdelik dilimi yaklaşık olarak “0 – 15” arasında olan

öğrenciler yüksek; “370 – 300” arasında olan yüzdelik dilimi “15 – 36” arasında olan öğrenciler orta; “300 puanın altında” olan yüzdelik dilimi “36’nın üzerinde” olan öğrenciler düşük düzey olarak kategorilendirilmiştir. Ortak değişkenlerin alt kategorilerinde bulunan öğrenci sayıları Tablo 25’te verilmiştir.

Tablo 25

Ortak Değişkenlerin Alt Kategorilerinde Bulunan Öğrenci Sayıları

Ortak Değişken	Kategori	A Formu	B Formu
Cinsiyet	Kız	595	496
	Erkek	405	504
LGS Puan Düzeyi	Yüksek	327	316
	Orta	334	300
	Düşük	339	384
Cinsiyet ve LGS Puan Düzeyi	Yüksek-Kız	177	146
	Orta-Kız	203	135
	Düşük-Kız	215	215
	Yüksek-Erkek	150	170
	Orta-Erkek	131	165
	Düşük-Erkek	124	169

Bu araştırmadaki eşitleme süreci hedef testin temel teste eşitlenmesi şeklinde yürütülmüştür. A formu hedef test ve B formu temel test olarak belirlenmiştir. Bu doğrultuda A formunun puanları B formuna eşit yüzdelikli eşitleme, Kernel eşitleme ve PoB yöntemleri kullanılarak eşitlenmiştir. EG ve NEAT desenlerinde eşit yüzdelikli ve Kernel eşitleme; NEC deseninde Kernel eşitleme ve PoB yöntemleri kullanılarak formlar birbirine eşitlenmiştir. Bu şekilde farklı eşitleme yöntemlerinin performansları incelenmiştir. PoB yöntemiyle yapılan eşitlemenin diğer yöntemlerle yapılan eşitlemelerle karşılaştırılması sağlanmıştır. Verilerin analizi R (for Windows 4.2.2) programlama dilinde RStudio (version 2022.07.2) ara yüzünde yapılmıştır. Eşit yüzdelikli eşitlemeler “equate” paketi (Albano, 2016) ve Kernel eşitlemeler “kequate” paketi (Andersson, Bränberg ve Wiberg, 2013) kullanılarak yapılmıştır. PoB yöntemiyle eşitleme için Yurtçu’nun (2018), Gonzalez ve diğerlerinin (2015a; 2015b) çalışmalarında kullanmış olduğu formüllerden yararlanarak R programlama dilinde yazmış olduğu kodlar kullanılmıştır.

Eşit yüzdelikli eşitleme yönteminde hem düzgünleştirilmiş hem de düzgünleştirilmemiş puan dağılımları üzerinde eşitlenmiş puanlar elde edilmiştir. Eşit yüzdelikli eşitlemede örneklem büyüklüğünün oldukça geniş olması gerektiği

belirtilmektedir (Rapp, 1999). Alanyazında eşit yüzdelikli eşitlemenin doğru sonuçlar vermesi için örneklem büyüklüğünün 1500 ve üzerinde olması önerilmektedir (Kolen ve Brennan, 2014; Zhu, 1998). Hanson, Zeng ve Colton (1994) ise daha küçük örneklemelerde düzgünleştirme işlemi yapılmasının, eşit yüzdelikli eşitlemeden elde edilen sonuçların doğruluğunu artırdığını belirtmiştir. Başka bir çalışmada Cui ve Kolen (2009) düzgünleştirmenin random ve sistematik hatayı azaltarak eşitleme kestiriminin doğruluğunu artırdığını belirtmiştir. Düzgünleştirme yöntemleri arasında ön düzgünleştirmenin her zaman son düzgünleştirmeye göre daha düşük standart hata ürettiği belirtilmiştir (Hagge vd., 2011). Polinomial Loglineer ön düzgünleştirme, Holland ve Thayer (1987) tarafından tanımlanan ve en sık kullanılan ön düzgünleştirme yöntemidir. Bu yöntem esnek olduğu için daha çok tercih edilmektedir. Ayrıca bu yöntem diğerlerine göre daha kolay uygulanmaktadır (Holland ve Thayer, 2000; Moses ve Holland, 2010). Bu sebeplerle çalışmada Polinomial Loglineer ön düzgünleştirme yöntemi seçilmiş ve düzgünleştirme yapılan ve yapılmayan eşitleme sonuçları karşılaştırılmıştır. Kernel eşitlemede ise literatürde ortak bir seçim olan Gauss Kernel düzgünleştirme tercih edilmiştir.

Eşit yüzdelikli ve Kernel eşitlemede “Bootstrap” yöntemi kullanılmıştır. Bu yöntem Efron ve Tibshirani (1993) tarafından önerilmiştir ve “yeniden örnekleme” yöntemi olarak da adlandırılabilir. Bu yöntemde başlangıçta evrenden seçkisiz olarak bir örneklem seçilir. Seçilen örneklemde verilerin yerleri değiştirilip, seçkisiz bir şekilde birçok farklı örneklem (bootstrap) çekilir. Efron ve Tibshirani (1993) replikasyon sayısının 25 ile 200 arasında yeterli olacağını belirtmiştir. Bu çalışmada 100 replikasyon yapılarak örneklem çekilmiştir. Her replikasyon için ilgili istatistikler hesaplanmış ve eşitlenmiş puanlar elde edilmiştir. Eşitleme işleminden sonra kullanılan her bir yöntem için hata değerleri hesaplanmıştır.

PoB yönteminde kullanılan ortak değişkenler bilgi vermeyen önseller olarak modele eklenmiştir. van de Schoot ve Depaoli (2014) bilgi vermeyen ve zayıf bilgi veren önsellerin seçiminden, sonsal dağılımın çok etkilenmeyeceğini belirtmiştir. Bu şekilde daha objektif değerlendirme yapılması amaçlanmıştır. PoB yönteminde Gonzalez ve diğerlerinin (2015a; 2015b) çalışmalarında uygun sonuç verdiği belirtilen sabitler kullanılmış ve analiz adımları başlatılmıştır. PoB yöntemi kullanılarak yapılan eşitlemede ortak değişkenlerin analize dahil edilmesi ile, hedef ve temel test puanları arasında kurulacak ilişki şekillenmektedir. Bu modelde, parametre kestiriminde uygun sonuçlara ulaşmak için MCMC örnekleme süreci kullanılmıştır. MCMC örnekleme yönteminde

amaç, evreni temsil edecek örneklem elde ederek doğru istatistiksel kestirimler yapabilmektir. Ortak değişkenlerden yararlanılarak örneklem ile ilgili genel bilgiler elde edilmektedir. MCMC örnekleme süreci ile oluşturulan dosyalarda DBPP modeli kullanılarak veriye uygun parametreler ve ortak değişkenler birleştirilmiştir. MCMC süreçleri A ve B formları için ayrı ayrı yürütülmüştür. Sonsal çıkarımlarda her bir veri seti için 150000 MCMC örneği alınmış ve iterasyonlar en uygun değerlere ulaşıncaya kadar sürdürülmüştür. Sonrasında ise eşitlenmiş puanlar elde etmek için eşitleme dönüşümleri yapılmıştır. Ortak değişkenlerin kategorilerine göre elde edilen ortak profillerin dağılımları, aynı yüzdelik dilime yerleştirilerek eşitleme süreci tamamlanmıştır. Puan dağılımları kestirilken %95 güven aralıkları (highest density interval) kullanılmış ve grafiklerde bu aralıklar gri renk tonu ile vurgulanmıştır. Güven aralıkları oluşturulurken çeyrek dilimler (.025 – .5 ve .975) kullanılmıştır. Çalışmada kestirilen puan dağılımlarına ilaveten güven aralıkları da incelenmiş ve puan dağılımları güven aralıklarına bağlı olarak yorumlanmıştır.

Karşılaştırma Kriteri

Geleneksel eşitleme yöntemlerinde SEE, BIAS ve RMSE gibi ölçütler kullanılmaktadır. Bu çalışmada da eşit yüzdelikli ve Kernel eşitlemeler için bu hatalar elde edilmiştir ve bu iki yöntem arasında karşılaştırma yapılabilmektedir. Fakat PoB yöntemiyle yapılan eşitlemede hata üretimine ihtiyaç duyulmamaktadır. Bu sebeple PoB ile farklı temelleri olan yöntemleri karşılaştırmak son derece zordur. Bu çalışmada diğer yöntemler ile PoB yönteminden elde edilen sonuçları karşılaştırmak için, eşitlenmiş puanların dağılımının hedef teste uzaklıklarını inceleyen Hellinger Uzaklığı kullanılmıştır. Bayes analizi genellikle araştırmacının MCMC tekniklerini kullanarak bir sonsal dağılımdan örnekler oluşturmasını gerektirir ve bu da çıkarım yapmak için kullanılır. Günümüzde zincirin yakınsamasını ve sonuçta elde edilen çıkarımların hassasiyetlerini belirlemek için çeşitli yaklaşımlar geliştirilmiştir. Bu yaklaşımlardan biri de MCMC tanımlaması için Hellinger uzaklığıdır. Bu uzaklık hesaplanırken dağılımların birbirlerine olan noktasal uzaklıkları toplanmaktadır. Birçok farklı formu bulunan bu uzaklık için, araştırmaya uygun olarak; f ve g herhangi iki dağılım olmak üzere bunlar arasındaki Hellinger Uzaklığı formülü aşağıda verilmiştir (Boone, Merrick ve Krachey, 2012).

$$\begin{aligned}
\hat{H}(f, g) &= \left[\frac{1}{2} \int \left(\sqrt{\hat{f}(x)} - \sqrt{\hat{g}(x)} \right)^2 dx \right]^{1/2} \\
&\approx \left[\frac{1}{2} \sum_{l=1}^k \left(\sqrt{\hat{f}(x)} - \sqrt{\hat{g}(x)} \right)^2 (x_l - x_{l-1}) \right]^{1/2}.
\end{aligned} \tag{70}$$

Yapılan her eşitlemede eşitlenmiş puanların dağılımının hedef testteki dağılıma olan uzaklığını gösteren grafikler sunulmuştur.



BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde her bir araştırma sorusuna yanıt aranacak şekilde bulgular ve bulgulara dayalı yapılan yorumlar yer almaktadır. Araştırma soruları ayrı başlıklar altında sunulmuştur.

EG ve NEAT Desenlerinde Eşit Yüzdelikli Eşitleme Sonuçlarının İncelenmesi

Yetenek testinin paralel formlarından elde edilen puanlar EG ve NEAT desenlerinde hem düzgünleştirilen hem de düzgünleştirilmeyen puan dağılımları üzerinde eşit yüzdelikli eşitleme yöntemi ile eşitlenmiş ve elde edilen bulgular yorumlanmıştır. EG deseninde eşit yüzdelikli eşitleme yöntemi ile; NEAT deseninde eşit yüzdelikli frekans tahmini ve eşit yüzdelikli zincirleme eşitleme yöntemleri ile eşitlenmiş puanlar elde edilmiştir. Her eşitleme işlemi sonucunda eşitlenmiş puan ortalamalarının (eqYX), SEE, BIAS ve RMSE değerlerinin ne düzeyde olduğu ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunduğu incelenmiştir.

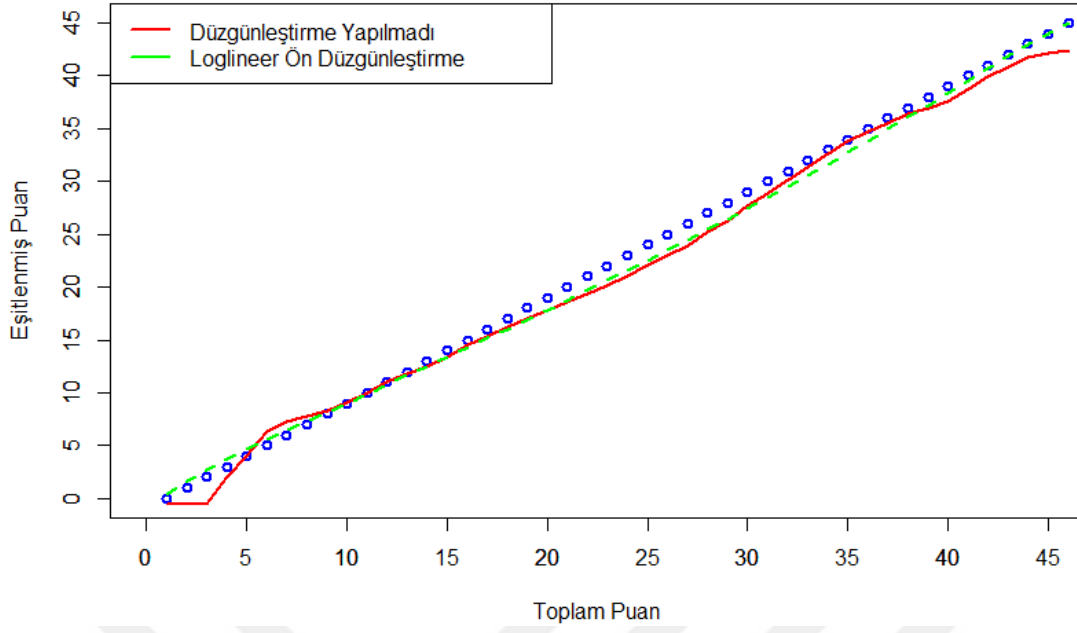
- a) EG deseninde puanların eşit yüzdelikli eşitleme yöntemi kullanılarak eşitlenmesiyle elde edilen eşitleme sonuçları Tablo 26’da verilmiştir.

Tablo 26

EG Deseninde Eşit Yüzdelikli Eşitleme Sonuçları

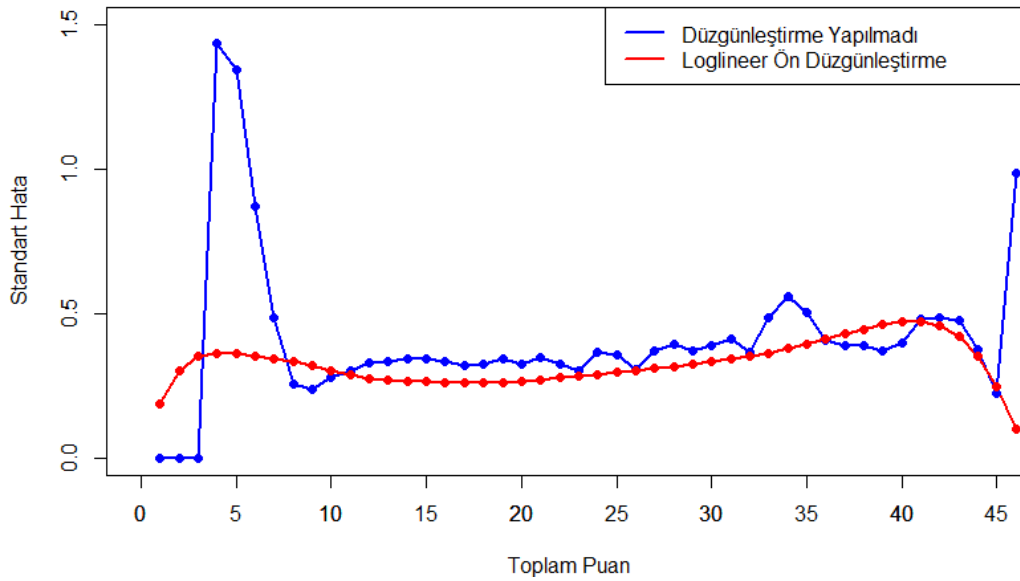
	Düğünlleştirme Yapılmadı	Loglineer Ön Düğünlleştirme
eqYx	21.612	21.870
SEE	.413	.325
BIAS	-.887	-.629
RMSE	.978	.708

Tablo 26 incelendiğinde ön düğünlleştirme sonrası puan ortalamalarının arttığı belirlenmiştir. Ön düğünlleştirme sonrası SEE, BIAS ve RMSE değerlerinin düştüğü hatanın azaldığı görülmektedir. Eşitlenmiş puanların ve SEE değerlerinin dağılımı incelenmiş, bu dağılımlar Şekil 9 ve Şekil 10’da verilmiştir.



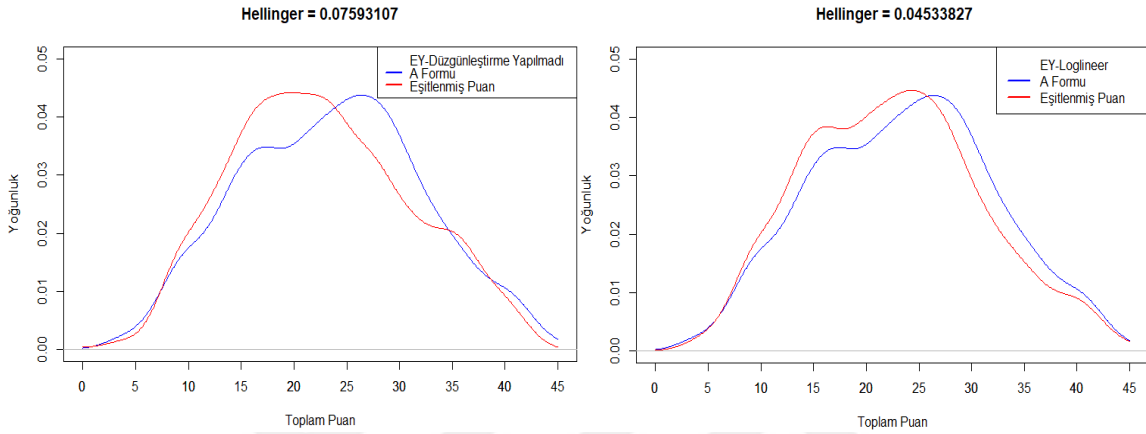
Şekil 9. EG Deseninde Eşit Yüzdelikli Eşitleme Puan Dağılımları

Şekil 9 incelendiğinde eşitleme sonucunda düzgünleştirme yapılan ve yapılmayan puan dağılımlarının 7 puan ile 30 puan arasında birbirine benzer olduğu; 5 – 7 puan ile 31 – 36 puan arasında birbirinden az da olsa farklılaştığı ve uçlardaki puanlarda bu farklılaşmanın daha da arttığı belirlenmiştir.



Şekil 10. EG Deseninde Eşit Yüzdelikli Eşitleme SEE Dağılımları

Şekil 10 incelendiğinde eşitleme sonucunda düzgünleştirme yapılan ve yapılmayan puan dağılımlarının standart hataları incelendiğinde 7 ve 44 puan aralığında farklılıklar daha az iken uçlardaki puanlarda daha belirgin farklılıklar görülmüştür. Her iki yöntemde eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 11’de verilmiştir.



Şekil 11. EG Deseninde Eşit Yüzdelikli Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

Olasılık yoğunluk dağılımları incelendiğinde eşitleme sonucunda hata değerleri daha düşük olan Loglineer ön düzgünleştirme sonucunda elde edilen puan kestirimlerinin hedef teste daha yakın olduğu belirlenmiştir.

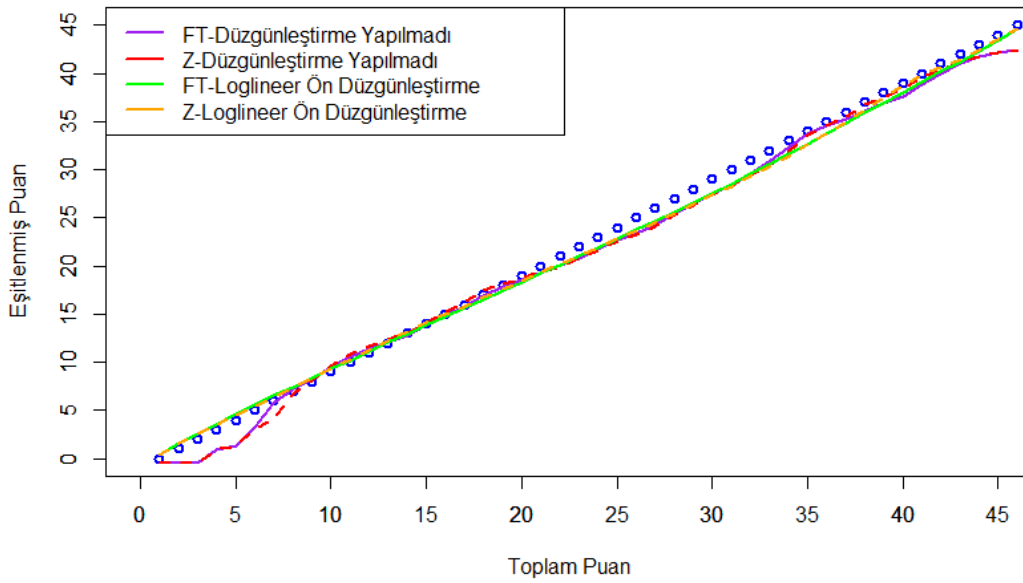
- b) NEAT deseninde puanların eşit yüzdelikli frekans tahmini ve eşit yüzdelikli zincirleme eşitleme yöntemleri kullanılarak eşitlenmesiyle elde edilen eşitleme sonuçları Tablo 27’de verilmiştir.

Tablo 27

NEAT Deseninde Eşit Yüzdelikli Eşitleme Sonuçları

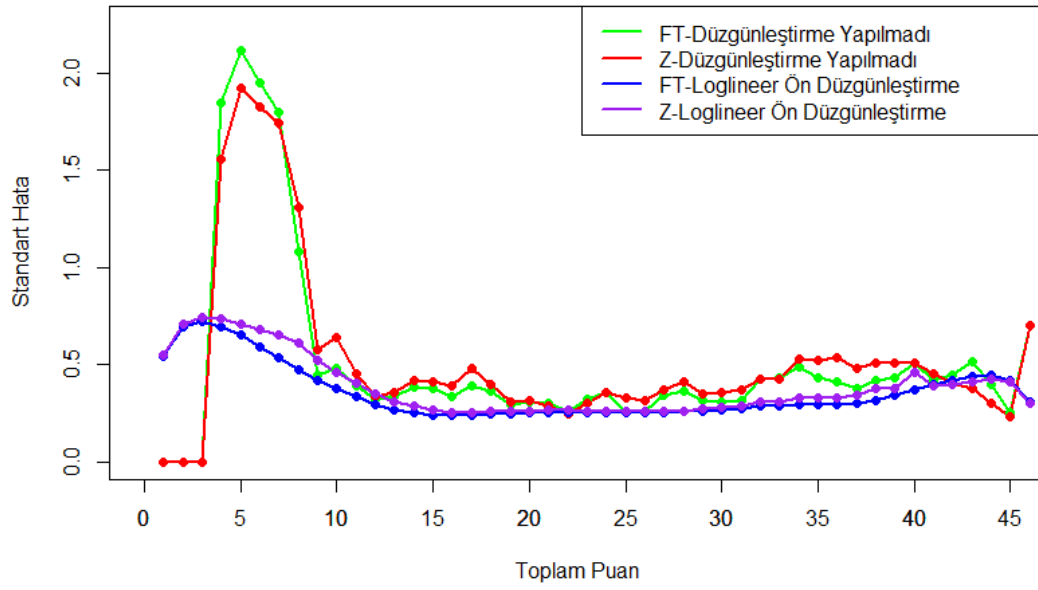
	Frekans Tahmini		Zincirleme	
	Düzgünleştirme Yapılmadı	Loglineer Ön Düzgünleştirme	Düzgünleştirme Yapılmadı	Loglineer Ön Düzgünleştirme
eqYx	21.56	21.93	21.61	22.00
SEE	.504	.354	.521	.383
BIAS	-.941	-.574	-.893	-.499
RMSE	1.067	.674	1.033	.629

Tablo 27 incelendiğinde her iki yöntemde de düzgünleştirme sonrası puan ortalamalarının arttığı; SEE, BIAS ve RMSE değerlerinin düştüğü hatanın azaldığı görülmektedir. Eşit yüzdelikli frekans tahmini ve eşit yüzdelikli zincirleme eşitleme yöntemleri karşılaştırıldığında ise frekans tahmininde SEE değerinin, zincirleme eşitlemede BIAS ve RMSE değerlerinin daha düşük olduğu hatanın azaldığı belirlenmiştir. Eşitlenmiş puanların ve SEE değerlerinin dağılımı incelenmiş, bu dağılımlar Şekil 12 ve Şekil 13’te verilmiştir.



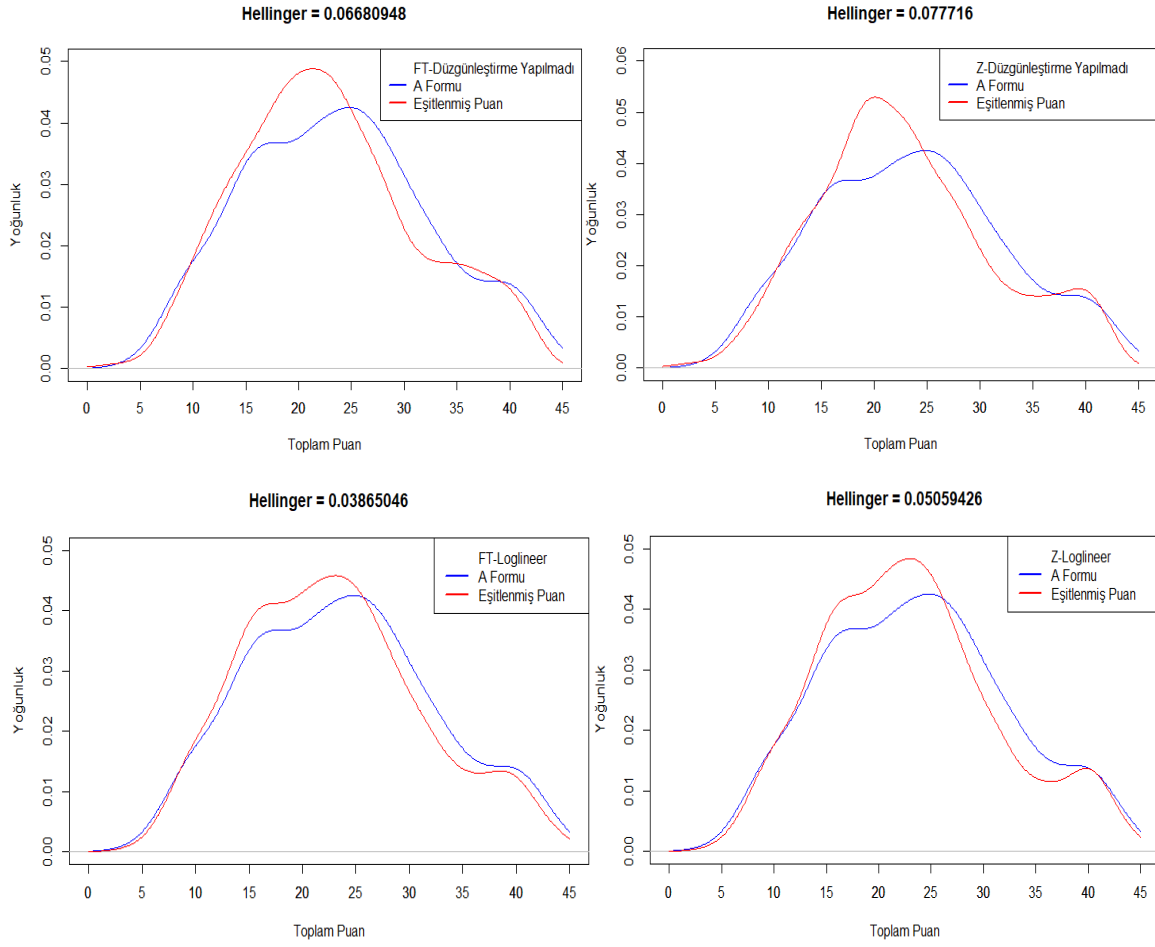
Şekil 12. NEAT Deseninde Eşit Yüzdelikli Eşitleme Puan Dağılımları

Şekil 12 incelendiğinde eşitleme sonucunda düzgünleştirme yapılan ve yapılmayan puan dağılımlarının 7 puan ile 43 puan arasında birbirine benzer olduğu; uçlardaki puanlarda düzgünleştirilmemiş puanların hedef testten uzaklaştığı belirlenmiştir.



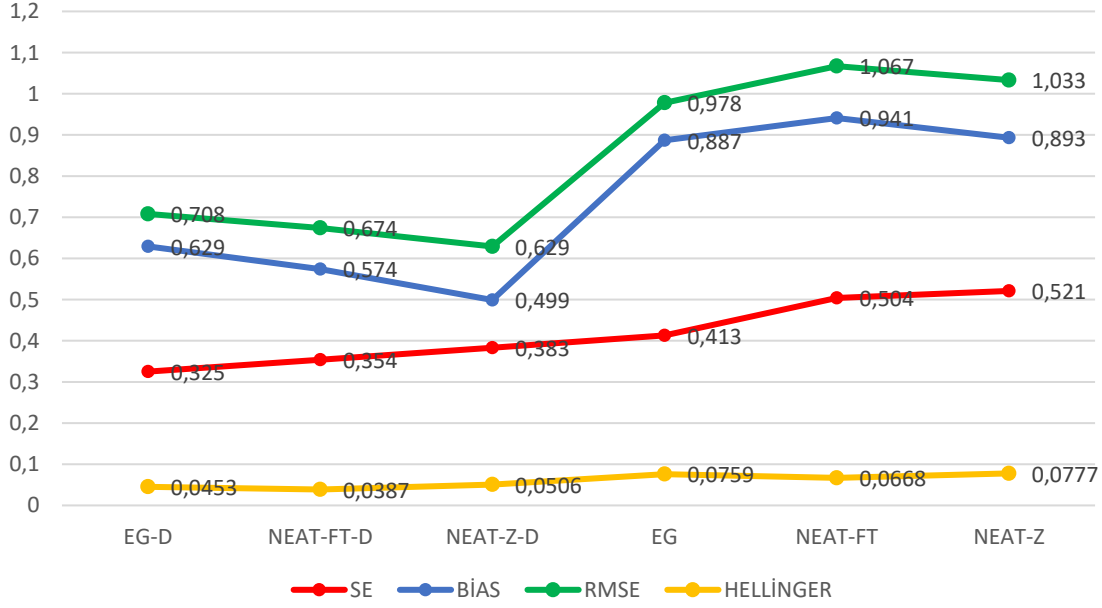
Şekil 13. NEAT Deseninde Eşit Yüzdelikli Eşitleme SEE Dağılımları

Şekil 13 incelendiğinde puan dağılımlarının standart hatalarının 7 ve 44 puan aralığında farklılıkları daha az iken uçlardaki puanlarda düzgünleştirme yapılan ve yapılmayan dağılımlarda belirgin farklılıklar görülmüştür. Bu yöntemlerde eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 14’te verilmiştir.



Şekil 14. NEAT Desende Eşit Yüzdelikli Eşitleme Sonucunda Puanların Olasılık Yoęunluk Daęılımları ve Hedef Teste Uzaklıkları

Olasılık yoęunluk daęılımları incelendięinde düzgünleştirme yapılmayan puan daęılımlarında eşitlenmiş puanların daęılımının hedef testin daęılımından daha sivri olduęu görülmektedir. Eşitleme sonucunda hata deęerleri daha düşük olan Loglineer ön düzgünleştirmede elde edilen puan kestirimlerinin hedef teste daha yakın olduęu belirlenmiştir. Loglineer ön düzgünleştirme yapılan puanlarda ise frekans tahmini sonucunda elde edilen eşitlenmiş puan kestirimlerinin zincirleme eşitlemeye göre hedef teste daha yakın olduęu belirlenmiştir.



Şekil 15. Eşit Yüzdelikli Eşitleme Sonuçlarının Değerlendirilmesi

Eşit yüzdelikli eşitleme sonucunda elde edilen bütün bulgular Şekil 15'te verilmiştir. Eşit yüzdelikli eşitleme sonucunda her iki desende de düzgünleştirme yapılan puan dağılımlarında SEE, BIAS ve RMSE hata değerlerinin azaldığı belirlenmiştir. Bu bulgu literatürde yapılan çalışmaların bulgularıyla örtüşmektedir (Hagge vd., 2011; Livingston, 1993; Livingston ve Kim, 2010; Skaggs, 2005). Eşitleme sonuçları incelendiğinde düzgünleştirme yapılmayan puan dağılımında uçlardaki puanlar dışında önemli bir sapma meydana gelmediği belirlenmiştir. Hem düzgünleştirilen hem de düzgünleştirilmeyen puan dağılımlarında en düşük SEE değeri EG desende elde edilmiştir. NEAT desende eşit yüzdelikli frekans tahmini ve eşit yüzdelikli zincirleme eşitleme karşılaştırıldığında yöntemler arasında küçük de olsa farklılıklar görülmüştür. Frekans tahmininin zincirleme eşitlemeye göre daha düşük SEE ürettiği belirlenmiştir ve bu bulgu literatürde yapılan çalışmaların bulgularıyla örtüşmektedir (Hagge vd., 2011; Liu ve Kolen, 2011; Wang vd., 2008). Ayrıca frekans tahmininin zincirleme eşitlemeye göre daha büyük BIAS içerdiği belirlenmiştir ve bu bulgu literatürde yapılan çalışmaların bulgularıyla benzerdir (Liu ve Kolen, 2011; Wang vd., 2008). Literatürde zincirleme eşitlemenin eşitlenecek formları alan sınav grupları arasında yetenek farkı olduğunda frekans tahmini yönteminden daha az yanlı sonuçlar ürettiği belirtilmiştir (Holland, Sinharay, von Davier ve Han, 2008; Sinharay ve Holland, 2007; Wang vd., 2008). Ayrıca grup farklılıkları küçük olduğunda frekans tahmininin daha az RMSE gösterdiği ve grup

farklılıkları arttıkça zincirleme eşit yüzdelikli yöntemin daha az RMSE gösterme eğiliminde olduğu belirlenmiştir (Wang vd., 2008).

Olasılık yoğunluk dağılımları incelendiğinde ise düzgünleştirme yapılmayan dağılımlarda eşitlenmiş puanların biraz daha sivri olduğu diğerlerinde hedef teste daha yakın olduğu belirlenmiştir. Her bir puan için noktasal olarak Hellinger uzaklığı hesaplanmıştır. Hem düzgünleştirilen hem de düzgünleştirilmeyen puan dağılımlarında frekans tahmini yönteminde elde edilen puan kestirimlerinin hedef teste en yakın olduğu belirlenmiştir. Frekans tahmini yöntemi EG desenine göre daha yüksek SEE üretmiş olsa da hedef teste daha yakın kestirimler yaptığı şeklinde yorumlanabilir.

EG ve NEAT Desenlerinde Kernel Eşitleme Sonuçlarının İncelenmesi

Yetenek testinin paralel formlarından elde edilen puanlar EG ve NEAT desenlerinde Kernel eşitleme yöntemi ile eşitlenmiş ve elde edilen bulgular yorumlanmıştır. EG deseninde Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri ile; NEAT deseninde Kernel eşitleme yöntemlerinden son tabakalama eşit yüzdelikli, zincirleme eşit yüzdelikli, son tabakalama lineer ve zincirleme lineer yöntemleri ile eşitlenmiş puanlar elde edilmiştir. Her eşitleme işlemi sonucunda eşitlenmiş puan ortalamalarının (eqYX), SEE, BIAS ve RMSE değerlerinin ne düzeyde olduğu ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunduğu incelenmiştir.

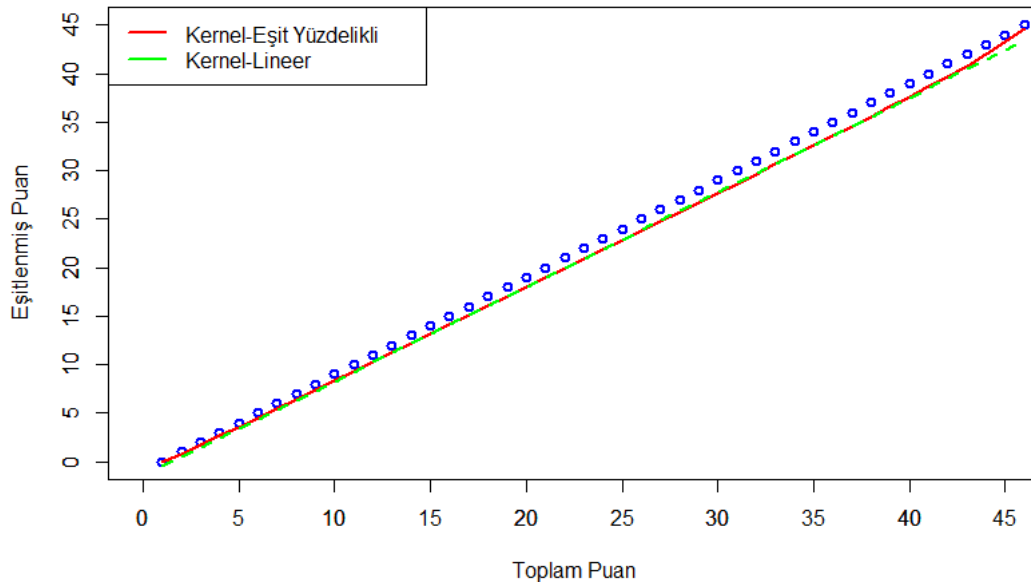
- a) EG deseninde puanların Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri kullanılarak eşitlenmesiyle elde edilen eşitleme sonuçları Tablo 28’de verilmiştir.

Tablo 28

EG Deseninde Kernel Eşitleme Sonuçları

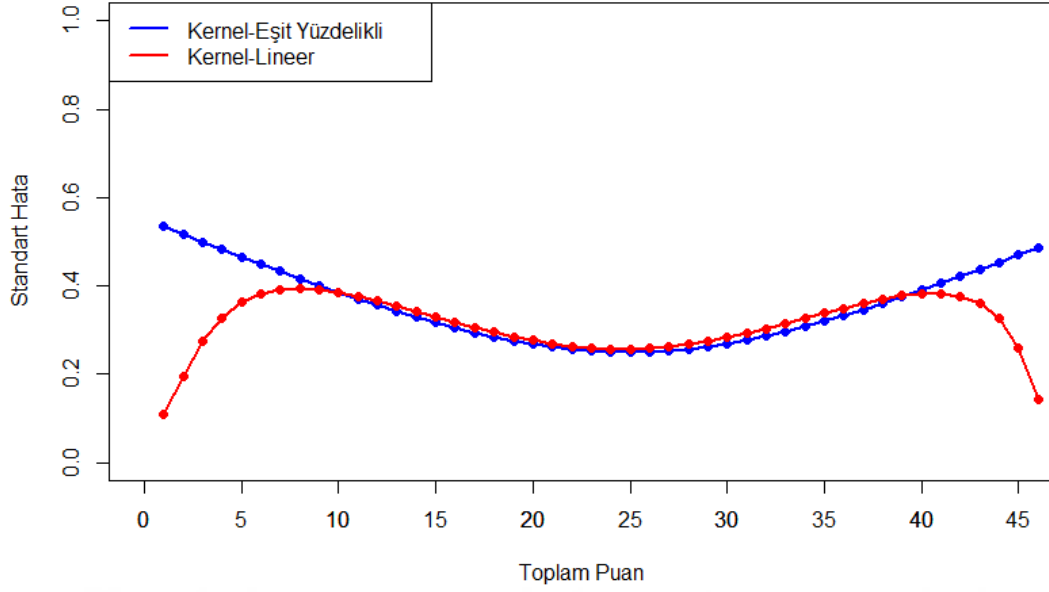
	Eşit Yüzdelikli	Lineer
eqYx	21.423	21.717
10.PREYx	-.117	-1.115
SEE	.351	.314
BIAS	-.642	-.573
RMSE	.731	.653

Tablo 28 incelendiğinde eşitlenmiş puan ortalamalarının Kernel lineer eşitlemede daha yüksek olduğu görülmüştür. Kernel eşit yüzdelikli eşitlemede puanların Kernel lineer eşitlemeye göre daha yüksek SEE, BIAS ve RMSE ile kestirildiği belirlenmiştir. Onuncu momentteki PRE değerleri incelendiğinde Kernel eşit yüzdelikli eşitlemede %1.12 ve Kernel lineer eşitlemede %1.12 yüzde görelî hata bulunmuştur. Eşitlenmiş puanların ve SEE değerlerinin dağılımı incelenmiş, bu dağılımlar Şekil 16 ve Şekil 17’de verilmiştir.



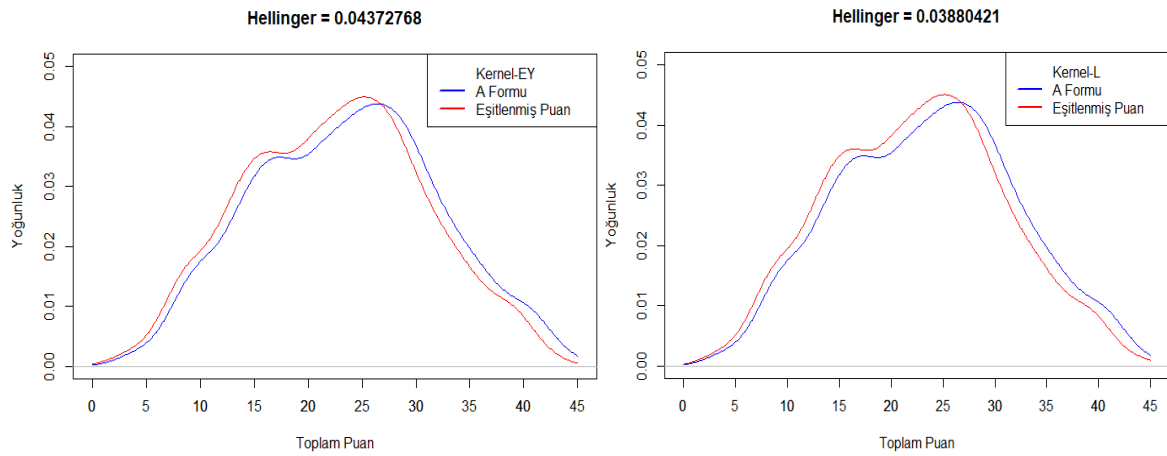
Şekil 16. EG Deseninde Kernel Eşitleme Puan Dağılımları

Şekil 16 incelendiğinde eşitlenmiş puan dağılımlarının doğrusal olduğu ve genel olarak birbirine benzediği fakat 43 – 45 arasındaki puanlarda dağılımların az da olsa farklılaştığı belirlenmiştir. Ayrıca 25 – 43 arasındaki puanlarda eşitlenmiş puanların hedef testten az da olsa düşük olduğu görülmektedir.



Şekil 17. EG Deseninde Kernel Eşitleme SEE Dağılımları

Şekil 17 incelendiğinde puan dağılımlarının standart hatalarının 10 ve 40 puan aralığında birbirine genel olarak benzer olduğu söylenebilirken uçlardaki puanlarda Kernel eşit yüzdelikli eşitlemede hata artarken, Kernel lineer eşitlemede azaldığı görülmüştür. Bu yöntemlerde eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 18’de verilmiştir.



Şekil 18. EG Deseninde Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

Olasılık yoğunluk dağılımları incelendiğinde Kernel eşit yüzdellikli ve Kernel lineer eşitleme sonucunda oluşan dağılımların birbirine çok benzer olduğu görülmüştür. Her bir puan için noktasal olarak Hellinger uzaklığı hesaplanmıştır. Eşitleme sonucunda hata değerleri daha düşük olan Kernel lineer eşitlemede elde edilen puan kestirimlerinin hedef teste daha yakın olduğu belirlenmiştir.

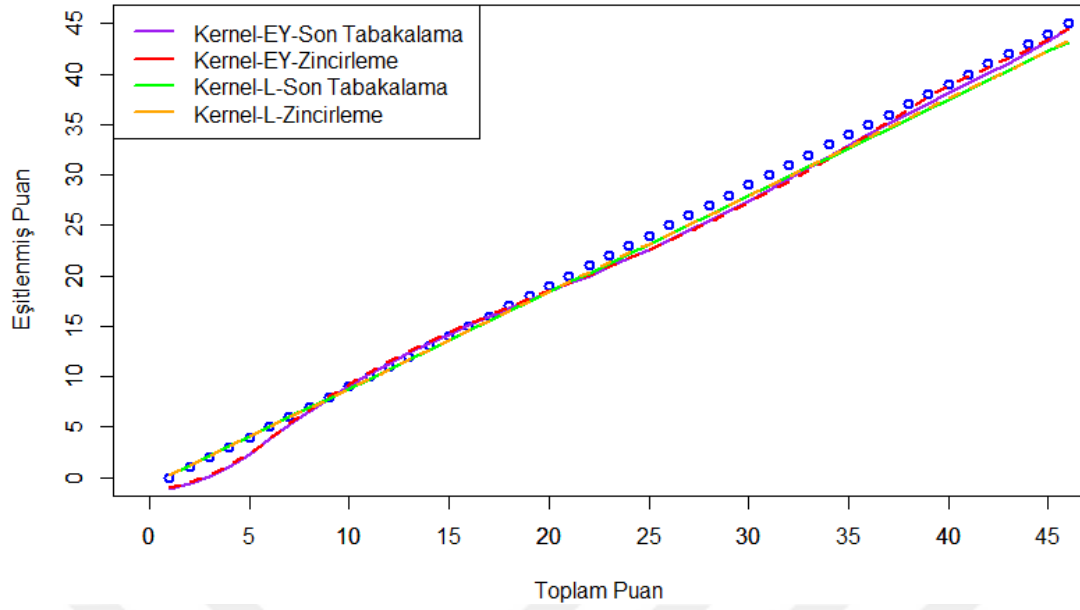
- b) NEAT deseninde puanların Kernel eşitleme yöntemlerinden son tabakalama eşit yüzdellikli, zincirleme eşit yüzdellikli, son tabakalama lineer ve zincirleme lineer yöntemleri kullanılarak eşitlenmesiyle elde edilen eşitleme sonuçları Tablo 29’da verilmiştir.

Tablo 29

NEAT Deseninde Kernel Eşitleme Sonuçları

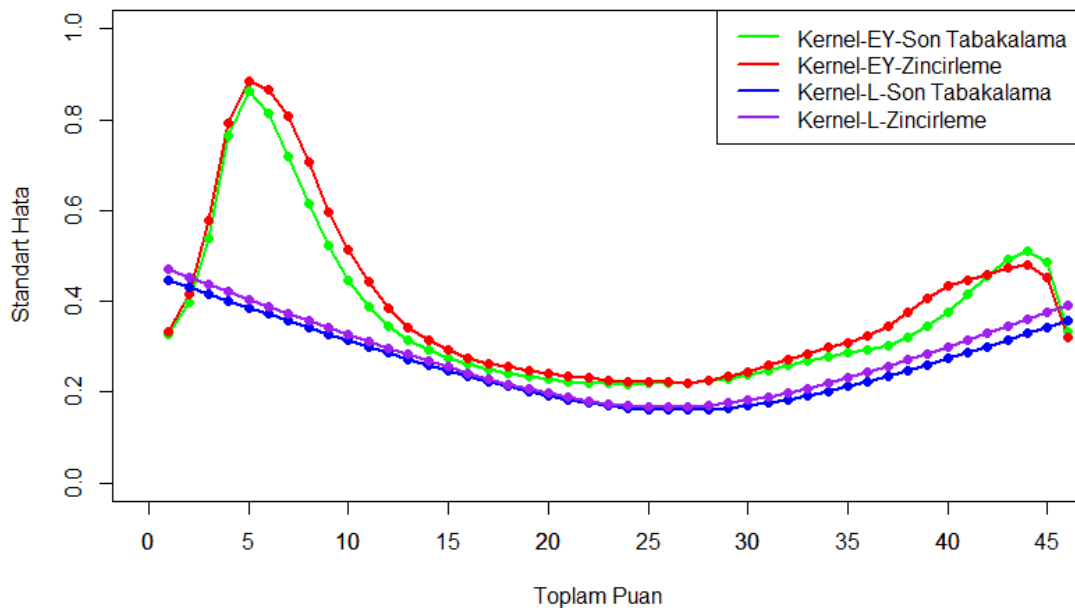
	Eşit Yüzdellikli		Lineer	
	Son Tabakalama	Zincirleme	Son Tabakalama	Zincirleme
eqY _x	21.617	21.753	21.700	21.736
10.PREY _x	-.277		-1.522	
10.PREAx		.786		-2.730
10.PREYa		-2.576		5.031
SEE	.321	.350	.261	.277
BIAS	-.542	-.651	-.457	-.523
RMSE	.629	.739	.526	.591

Tablo 29 incelendiğinde eşitlenmiş puan ortalamalarının Kernel zincirleme eşit yüzdellikli eşitlemede diğerlerine göre daha yüksek olduğu görülmüştür. Kernel lineer eşitlemede puanların Kernel eşit yüzdellikli eşitlemeye göre daha düşük SEE, BIAS ve RMSE ile kestirildiği belirlenmiştir. Kernel son tabakalama ve Kernel zincirleme eşitlemeler karşılaştırıldığında ise Kernel son tabakalamada puanların Kernel zincirleme eşitlemeye göre daha düşük SEE, BIAS ve RMSE ile kestirildiği sonucuna ulaşılmıştır. Onuncu momentteki PRE değerleri incelendiğinde %.79 ile %5.03 arasında yüzde göreceli hata bulunduğu belirlenmiştir. Eşitlenmiş puanların ve SEE değerlerinin dağılımı incelenmiş, bu dağılımlar Şekil 19 ve Şekil 20’de verilmiştir.



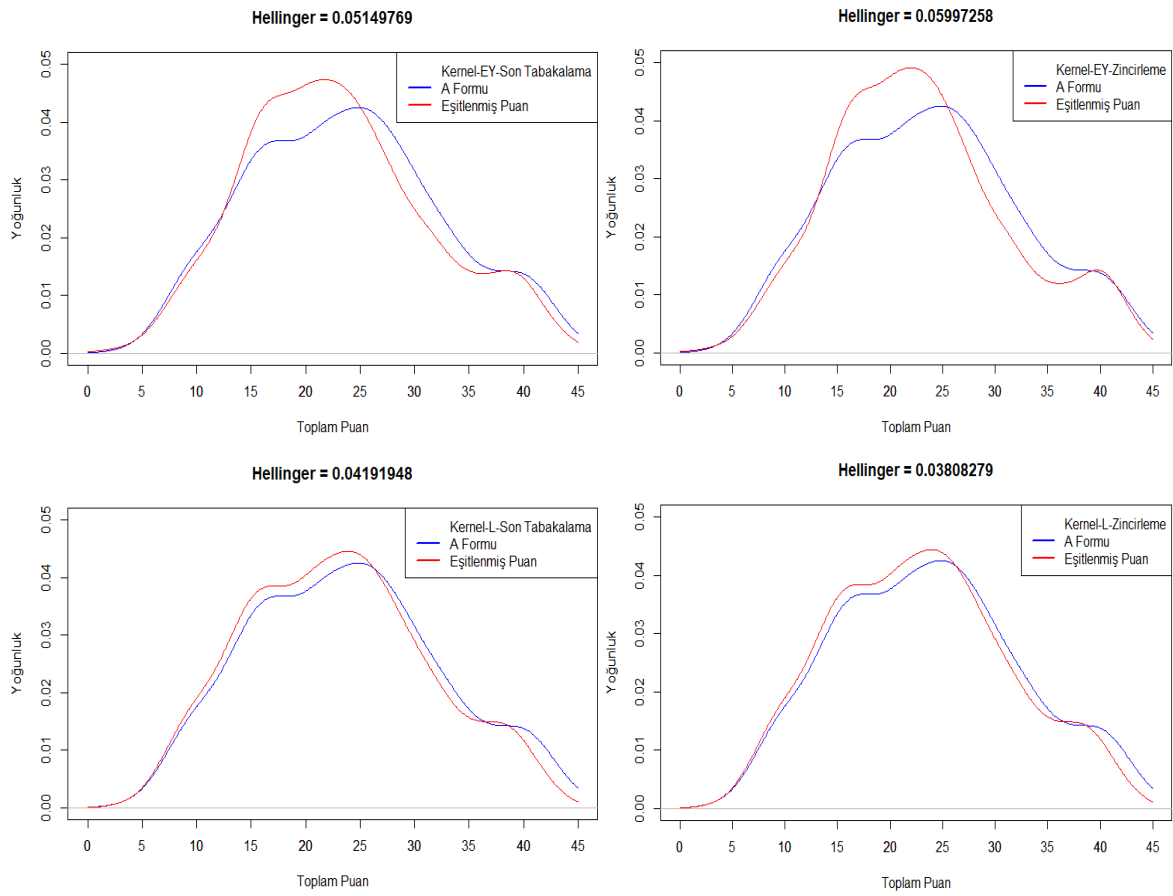
Şekil 19. NEAT Deseninde Kernel Eşitleme Puan Dağılımları

Şekil 19 incelendiğinde eşitlenmiş puan dağılımlarının 7 puan ile 37 puan arasında birbirine benzer olduğu, uçlardaki puanlarda ise farklılıklar olduğu görülmüştür. Özellikle 0 – 6 puan arasında Kernel eşit yüzdelikli eşitlemelerde puanların hedef testten belirgin şekilde uzaklaştığı; 42 – 45 puan arasında ise Kernel lineer eşitlemelerde puanların hedef testten biraz uzaklaştığı belirlenmiştir.



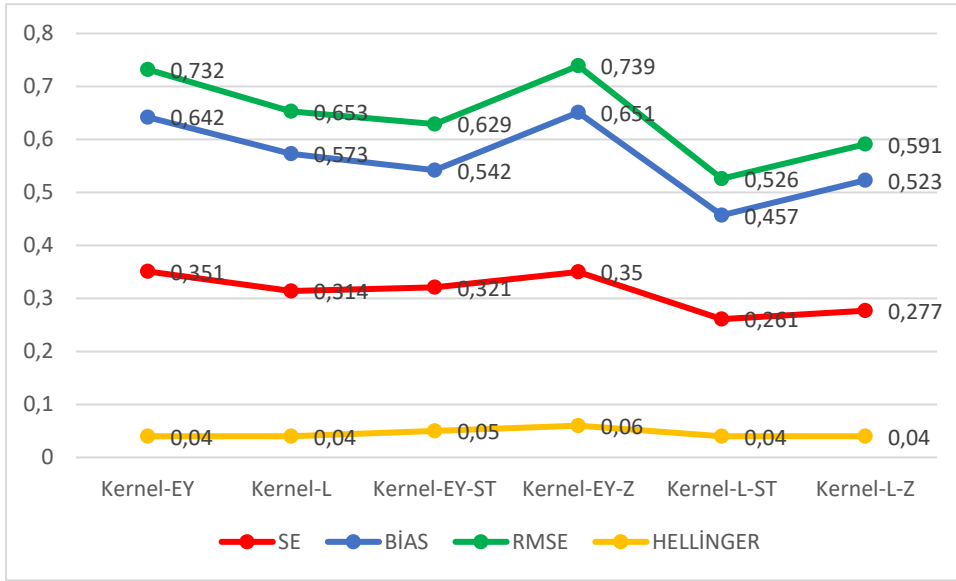
Şekil 20. NEAT Deseninde Kernel Eşitleme SEE Dağılımları

Şekil 20 incelendiğinde puan dağılımlarının standart hatalarının Kernel lineer eşitlemelerde daha düşük olduğu görülmüştür. Kernel eşit yüzdelikli eşitlemelerde ise uçlardaki puanların standart hatasında artış meydana geldiği özellikle alt uçlardaki puanlarda bu artışın daha belirgin olduğu belirlenmiştir. Bu yöntemlerde eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 21’de verilmiştir.



Şekil 21. NEAT Desende Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

Olasılık yoğunluk dağılımları incelendiğinde eşitleme sonucunda hata değerleri daha düşük olan Kernel lineer eşitlemelerde puan kestirimlerinin hedef teste daha yakın olduğu belirlenmiştir. Hedef teste en yakın puanların Kernel lineer zincirleme eşitlemede kestirildiği görülmüştür.



Şekil 22. EG ve NEAT Desenlerinde Kernel Eşitleme Sonuçlarının Değerlendirilmesi

EG ve NEAT desenlerinde Kernel eşitleme sonucunda elde edilen bütün bulgular Şekil 22’de verilmiştir. Kernel lineer eşitlemeden elde edilen SEE’lerin Kernel eşit yüzdelikli eşitlemeden daha düşük olduğu belirlenmiştir. Bu bulgu literatürde yapılan çalışmaların bulgularıyla örtüşmektedir (Akın Arıkan, 2019; Akın Arıkan ve Gelbal, 2018; Choi, 2009; Liou, Cheng ve Johnson, 1997). Kernel eşit yüzdelikli ve lineer eşitlemelerde son tabakalama eşitleme yönteminin zincirleme eşitlemeye göre daha az eşitleme hatasına sahip olduğu belirlenmiştir. Bu sonuç literatürde yapılan çalışmaların bulgularıyla tutarlıdır (Akın Arıkan, 2020; Akın Arıkan ve Gelbal, 2021; Liu ve Low, 2008; Sezer Başaran, 2023; Wiberg, 2021). Holland, von Davier, Sinharay ve Han (2006) gruplar arasındaki yetenek dağılımı bezer olduğunda Kernel son tabakalama ve Kernel zincirleme eşitleme yöntemlerinin birbirine yakın, farklı olduğu durumlarda ise zincirleme eşitleme yönteminin daha iyi sonuçlar verdiğini belirtmiştir. Liu ve Low (2008) popülasyonlar ortak testte aynı dağılıma sahip olduğunda Kernel son tabakalama ve Kernel zincirleme eşitleme yöntemlerinin benzer sonuçlar üretebileceğini belirtmiştir. Choi (2009) ise örneklem büyüklüğü arttıkça Kernel son tabakalama ve Kernel zincirleme eşitleme arasındaki farkın biraz daha azaldığı sonucuna ulaşmıştır. Bu çalışmada da yöntemler arasındaki farkın çok yüksek olmamasının örneklem büyüklüğünden kaynaklandığı söylenebilir. Bu çalışmada Kernel eşit yüzdelikli eşitlemelerde uç noktalarda puanların daha yüksek hata ile kestirildiği görülmüştür. Mao (2006) Kernel eşitleme yöntemlerinde uç noktalardaki hataların daha yüksek olmasının Gauss Kernel

süreklileştirme yöntemindeki ölçeğin $-\infty$ ve $+\infty$ aralığında olmasından kaynaklandığını belirtmiştir. Bu çalışmada loglineer ön düzgünleştirme yapılan eşit yüzdelikli eşitlemenin Kernel eşitlemeye yakın sonuçlar verdiği belirlenmiştir.

Choi (2009) Kernel eşitlemenin bazı içsel sorunları olduğunu belirtmiştir. İlk olarak süreklileştirme adımı bant genişliği seçimine çok fazla ağırlık verilmektedir. Fakat eşitlemeyi etkileyen pek çok başka faktör olabilir: test uzunluğu, örneklem büyüklüğü, düzgünleştirme derecesi, ortak test sayısı, popülasyonlar arasındaki farklılıklar bunlardan sadece birkaçıdır. Örneğin, NEAT deseni altında eşitlemede popülasyon farkının eşitleme için seçilen bant genişliğinin üzerinde büyük etkileri olabilir. von Davier ve diğerleri (2004) süreklileştirme adımı en az altı momentin korunmasını önermektedir. Ancak Gauss çekirdeği kullanılarak yapılan süreklileştirme adımı Kernel eşitlemede yalnızca ortalama ve varyans korunur. van der Linden (2006) süreklileştirme adımı daha yüksek momentlerin korunmasının düşünülebileceğini belirtmiştir. Ayrıca Kernel eşitlemenin geliştiricileri sadece Gauss çekirdeğini dikkate almıştır.

Olasılık yoğunluk dağılımları incelendiğinde ise Kernel eşit yüzdelikli eşitlemelerde eşitlenmiş puanların biraz daha sivri olduğu Kernel lineer eşitlemelerde ise hedef teste daha yakın olduğu belirlenmiştir. Her bir puan için noktasal olarak Hellinger uzaklığı hesaplanmıştır. Kernel lineer zincirleme eşitlemede elde edilen puan kestirimlerinin hedef teste en yakın olduğu belirlenmiştir.

NEC Deseninde Kernel Eşitleme Sonuçlarının İncelenmesi

Yetenek testinin paralel formlarından elde edilen puanlar NEC deseninde cinsiyet; LGS puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri ile eşitlenmiş ve elde edilen bulgular yorumlanmıştır. Her eşitleme işlemi sonucunda eşitlenmiş puan ortalamalarının (eqYX), SEE, BIAS ve RMSE değerlerinin ne düzeyde olduğu ve eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki bulunduğu incelenmiştir.

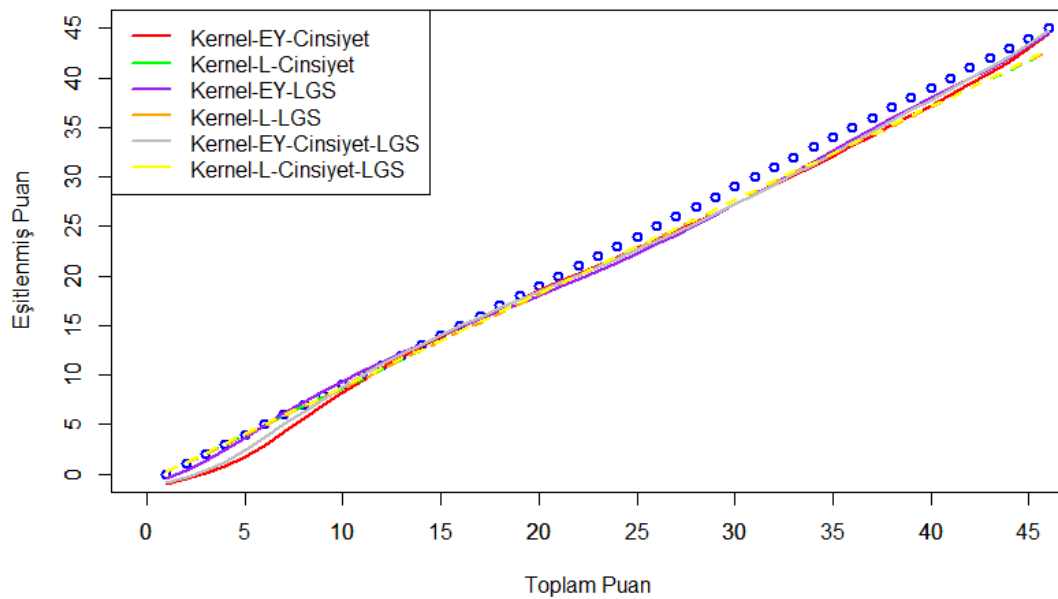
NEC deseninde puanların Kernel eşit yüzdelikli ve Kernel lineer eşitleme yöntemleri kullanılarak eşitlenmesiyle elde edilen eşitleme sonuçları Tablo 30’da verilmiştir.

Tablo 30

NEC Deseninde Kernel Eşitleme Sonuçları

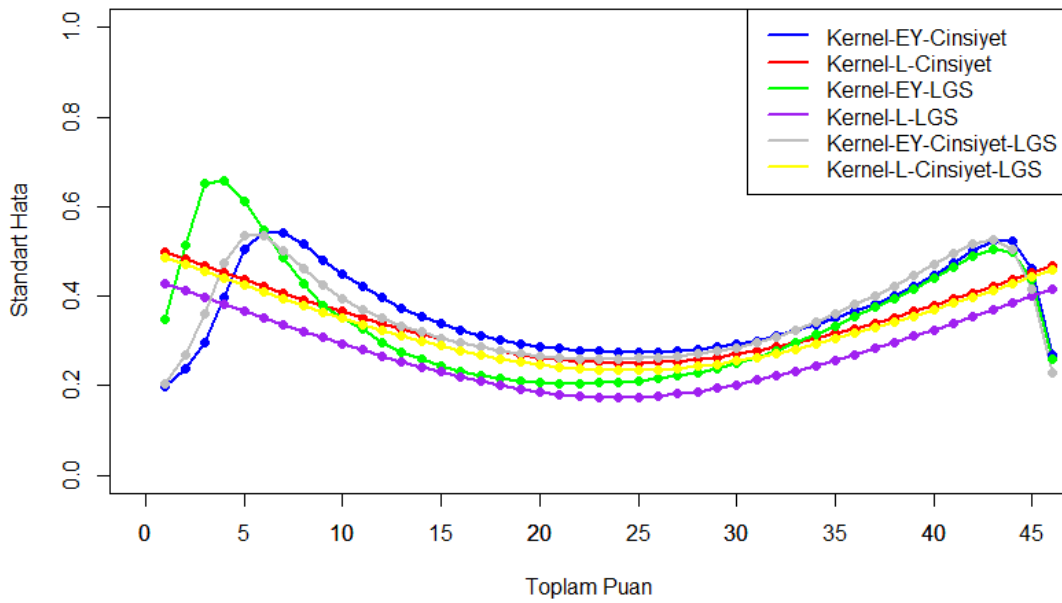
	Cinsiyet		LGS Puan Düzeyi		Cinsiyet-LGS Puan Düzeyi	
	Eşit Yüzdelikli	Lineer	Eşit Yüzdelikli	Lineer	Eşit Yüzdelikli	Lineer
eqYx	21.246	21.451	21.649	21.453	21.493	21.522
SEE	.352	.301	.327	.275	.339	.294
BIAS	-.735	-.714	-.665	-.612	-.703	-.695
RMSE	.814	.774	.741	.671	.780	.754

Tablo 30 incelendiğinde eşitlenmiş puan ortalamalarının LGS puan düzeylerinin ortak değişken olarak kullanıldığı Kernel eşit yüzdelikli eşitlemede en yüksek olduğu görülmüştür. Kernel lineer eşitlemede puanların Kernel eşit yüzdelikli eşitlemeye göre daha düşük SEE, BIAS ve RMSE ile kestirildiği belirlenmiştir. Kernel eşit yüzdelikli ve lineer eşitlemede en düşük hata kestirimleri LGS puan düzeylerinin ortak değişken olarak kullanıldığı eşitlemede, en yüksek hata kestirimleri ise cinsiyetin ortak değişken olarak kullanıldığı eşitlemede yapılmıştır. Eşitlenmiş puanların ve SEE değerlerinin dağılımı incelenmiş, bu dağılımlar Şekil 23 ve Şekil 24’te verilmiştir.



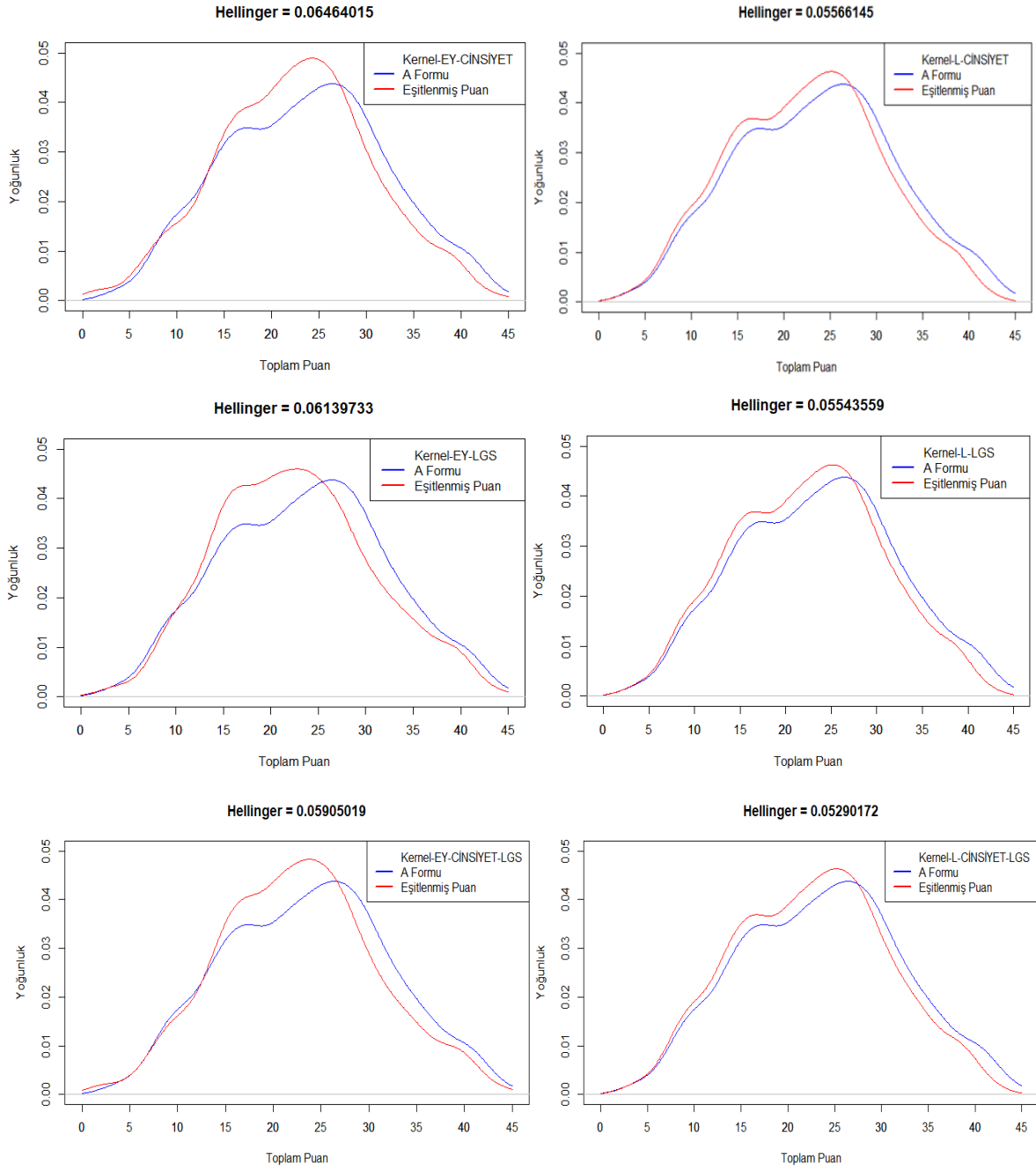
Şekil 23. NEC Deseninde Kernel Eşitleme Puan Dağılımları

Şekil 23 incelendiğinde eşitlenmiş puan dağılımlarının 9 puan ile 41 puan arasında birbirine benzer olduğu, uçlardaki puanlarda ise farklılıklar olduğu görülmüştür. Özellikle 0 – 8 puan arasında cinsiyetin ortak değişken olduğu Kernel eşit yüzdelikli eşitlemede ve cinsiyet ve LGS puan düzeylerinin birlikte ortak değişken olduğu Kernel eşit yüzdelikli eşitlemede puanların hedef testten uzaklaştığı belirlenmiştir. Diğer uçta 42 – 45 puan arasında ise Kernel lineer eşitlemenin üçünde de puanların hedef testten az da olsa uzaklaştığı görülmüştür.



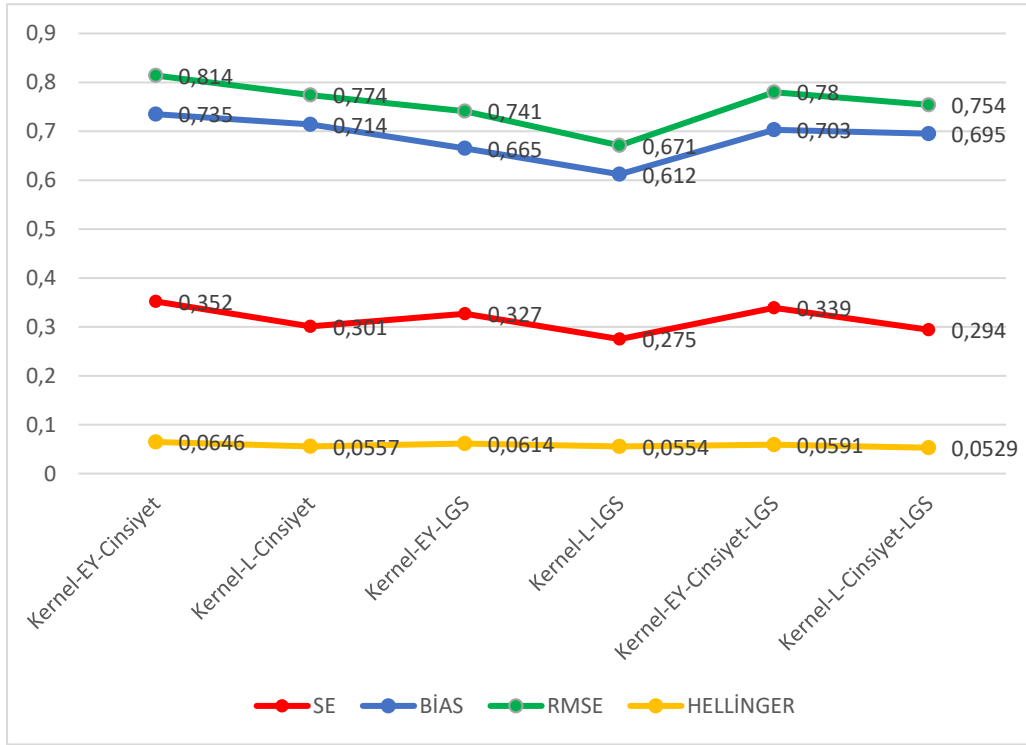
Şekil 24. NEC Deseninde Kernel Eşitleme SEE Dağılımları

Şekil 24 incelendiğinde puan dağılımlarının standart hatalarının LGS puan düzeylerinin ortak değişken olarak kullanıldığı Kernel lineer eşitlemede uçlardaki puanlar hariç en düşük olduğu; cinsiyetin ortak değişken olarak kullanıldığı ve cinsiyet ve LGS puan düzeylerinin birlikte ortak değişken olarak kullanıldığı Kernel lineer eşitlemelerde ise dağılımların benzer olduğu belirlenmiştir. Kernel lineer eşitlemelerde standart hataların uçlardaki puanlara doğru artış gösterdiği, Kernel eşit yüzdelikli eşitlemelerde ise uçlardaki puanlara doğru önce artış gösterdiği sonra azaldığı görülmüştür. Bu yöntemlerde eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 25'te verilmiştir.



Şekil 25. NEC Deseninde Kernel Eşitleme Sonucunda Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

Olasılık yoğunluk dağılımları incelendiğinde Kernel lineer eşitlemelerde puan kestirimlerinin hedef teste daha yakın olduğu belirlenmiştir. Cinsiyet ve LGS puan düzeylerinin birlikte ortak değişken olarak kullanıldığı Kernel lineer eşitlemede puan kestirimlerinin hedef teste en yakın olduğu görülmüştür.



Şekil 26. NEC Deseninde Kernel Eşitleme Sonuçlarının Değerlendirilmesi

NEC deseninde Kernel eşitleme sonucunda elde edilen bütün bulgular Şekil 26’da verilmiştir. Daha önce de belirtildiği gibi Kernel lineer eşitlemeden elde edilen SEE’lerin Kernel eşit yüzdelikli eşitlemeden daha düşük olduğu belirlenmiştir. Bu çalışmada NEC deseninden elde edilen sonuçlar diğer yöntemlerden elde edilen hata değerleriyle kıyaslandığında SEE değerlerinin yakın olduğu fakat BIAS ve RMSE değerlerinin biraz yükseldiği belirlenmiştir. Bu çalışmada yöntemler arasındaki fark genel olarak çok yüksek değildir. Literatürde yapılan çalışmalar incelendiğinde NEC deseni ile NEAT desenindeki son tabakalama yöntemlerinden elde edilen SEE değerlerinin birbirine yakın olduğuna (Akın Arıkan, 2020; Sezer Başaran, 2023; Wallin ve Wiberg, 2017; Wiberg ve Branberg, 2015); bu SEE değerlerinin EG deseninden daha düşük olduğuna (Sezer Başaran, 2023; Wiberg ve Branberg, 2015) dair bulgular mevcuttur. Akın Arıkan (2020) NEC deseninde de son tabakalama eşitleme yöntemlerinin kullanılmasının, NEAT son tabakalama ve NEC desenleri sonucunda elde edilen SEE değerlerinin benzerliğine sebep olduğunu belirtmiştir.

Branberg ve Wiberg (2011) test puanlarıyla ilişkili değişkenler gibi daha fazla bilgi kullanmanın standart hatayı azaltıp eşitleme parametre tahminlerinin verimliliğini artırabileceğini belirtmiştir. Bu bilgileri toplamak çok zor değildir, bazı test

programlarında rutin olarak toplanmaktadır. Branberg ve Wiberg (2011) bir ortak test kullanmanın aslında ortak değişken kullanmaktan daha verimli olduğunu belirtse de ortak testin kullanılmadığı ya da ortak testin çok az katılımcıya uygulandığı durumlarda sistematik farklılıkları kontrol etmek için diğer olasılıkların da araştırılması gerektiğini belirtmiştir.

Bu çalışma sonucunda NEC desenindeki yöntemler karşılaştırıldığında LGS puan düzeylerinin ortak değişken olarak kullanılması en düşük hata değerlerine sebep olurken, cinsiyet değişkeninin ortak değişken olarak kullanılması en yüksek hata değerlerine sebep olmuştur. Cinsiyet değişkeni ile test arasındaki korelasyon değeri istatistiksel olarak anlamlı değilken, LGS puan düzeyleri ile yüksek düzeyde ve anlamlı bir korelasyon bulunmaktadır. Cinsiyet değişkenine LGS puan düzeylerinin de eklenip birlikte ortak değişken olarak kullanılması ise hata değerlerini düşürmüştür. Akın Arıkan'ın (2020) cinsiyet, sosyo ekonomik düzey ve bunları birlikte ortak değişken olarak kullandığı çalışmada da benzer bir sonuç elde edilmiştir. Bu çalışmada da cinsiyet değişkeni ile test arasında anlamlı olmayan bir korelasyon mevcutken, sosyoekonomik düzey değişkeni ile düşük düzeyde anlamlı bir korelasyon mevcuttur. Sosyo ekonomik düzey değişkeninin ortak değişken olarak kullanılması en düşük hata değerine sebep olurken, cinsiyet değişkeni ise en yüksek hata değerine sebep olmuştur. Cinsiyet değişkenine sosyo ekonomik düzey değişkeninin eklenmesi ise hata değerini düşürmüştür. Literatürde öğrencinin cinsiyetini ortak değişken olarak kullanarak testlerin eşitlendiği diğer çalışmalarda da genellikle farklı ortak değişkenlerle daha doğru eşitleme sonuçlarının elde edildiği belirtilmiştir (Altıntaş ve Wallin, 2021; Branberg ve Wiberg, 2011; Lu ve Guo, 2018; Özsoy, 2021; Yurtçu vd., 2021).

Wiberg ve Branberg (2015) bir ortak test olsa bile geniş bir test puanı aralığında daha düşük SEE'ye yol açabileceğinden ortak değişkenlerin yine de kullanılabileceğini belirtmiştir. Eşitlemedeki farkın boyutu ortak değişkenlerin test puanlarını ne kadar iyi tahmin ettiğine ve arka plan değişkenlerinin test puanlarındaki farklılıkları ne kadar iyi "açıkladığına" bağlıdır. "Doğru" ortak değişkenlerin (gruplar arasındaki farklılıkları açıklayan değişkenler) seçilmesi, gelecekteki araştırmalar için önemli bir alan olabilir. Uygun ortak değişkenleri seçerken vurgulanması gereken iki sorun ortak değişkenler içerisindeki kategori sayısı ve ortak değişkenler sürekli olduğunda kategorilerin nasıl oluşturulacağıdır. Tek bir ortak değişken kullanıldığında ve bu değişkenin test ile korelasyonu yüksek olduğunda, ortak test kullanımına benzer bir sonuç ortaya çıkar. Daha fazla ortak değişken eklendikçe kategori sayısında hızlı bir artış elde edilir. Bu da

bazı hücrelerde seyrek veriye sahip olma riskinin arttığı anlamına gelmektedir. Sonuç olarak daha iyi ortak değişkenler bulunursa muhtemelen eşitleme daha iyi hale gelecek ve NEAT deseniyle elde edilen sonuçlara daha çok benzeyecektir. Düşük korelasyonlu ortak değişkenlerin kullanımında son tabakalama yönteminin kullanımının sorgulanabileceği unutulmamalıdır. Bu nedenle daha ileri araştırmalar için göze çarpan bir alan, korelasyondan daha az etkilendiği için zincirleme eşitlemenin kullanımını incelemek olacaktır (Wiberg ve Branberg, 2015).

Puanların olasılık yoğunluk dağılımları ve hedef teste uzaklıkları incelendiğinde ise birbirine yakın olduğu görülmekle birlikte cinsiyet değişkeni ve LGS puan düzeyleri birlikte ortak değişken olarak kullanıldığında hedef teste en yakın kestirimler yapılmıştır. Yurtçu ve diğerleri (2021) yaptıkları çalışmada benzer bir sonuç elde etmişlerdir. Bu çalışmalarında cinsiyet, matematik öz yeterliği ve ortak testle elde edilen eşitlenmiş puanların hedef teste uzaklıklarının birbirine yakın olduğu ayrıca cinsiyet ve matematik özyeterliğinin birlikte ortak değişken olarak kullanıldığı NEC desenindeki eşitleme sonuçlarının hedef teste en yakın dağılıma sahip olduğu bulunmuştur.

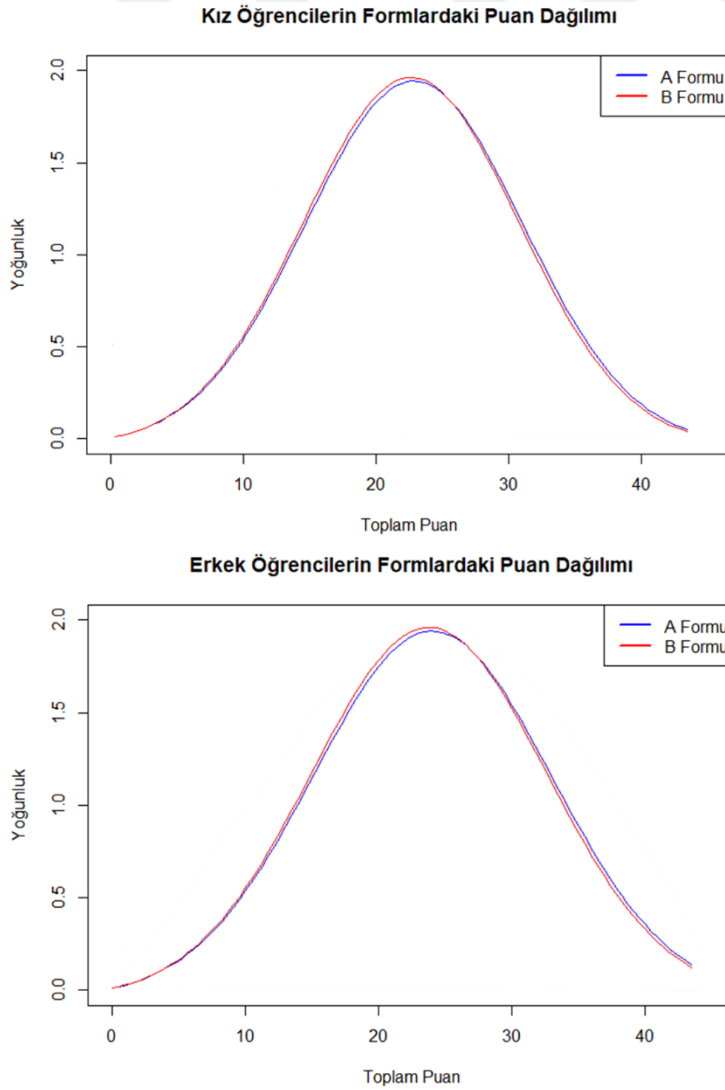
NEC Deseninde PoB Eşitleme Sonuçlarının İncelenmesi

Yetenek testinin paralel formlarından elde edilen puanlar NEC deseninde cinsiyet; LGS puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak PoB modeli ile eşitlenmiş ve elde edilen bulgular yorumlanmıştır. Her bir değişkenin alt kategorilerinde eşitlenmiş puanlar elde edilmiş; eşitlenmiş puanlar arasında, eşitlenmiş puanlar ile hedef testin dağılımları arasında nasıl bir ilişki olduğu incelenmiştir.

Eşitlemede PoB modelinin kullanılması hem değişken hem de değişkenin alt kategorileri ile ilgili ayrıntılı bilgi elde edilmesini ve her bir profile ait dağılımlar aracılığıyla farklı karşılaştırmaların yapılabilmesini sağlamaktadır. Bu model ile yapılan eşitlemelerde grup değişmezliği varsayımı kontrol altında tutulmaktadır. Bu sebeple ortak profillerden yararlanılarak formlar için eşitlenmiş puanlar elde edilebilir, ayrıca aynı test formunu alan kız ve erkek öğrenciler arasında veya LGS puan düzeyleri arasında da eşitleme yapılabilir. Sonuçta, LGS puan düzeyi düşük olan öğrenciler ile yüksek olan öğrencilerin cinsiyetleri, göstermiş oldukları performans ve aldıkları formlar ile ilgili yorum yapmaya ve bunlar arasındaki ilişkiyi incelemeye imkân sağlamaktadır.

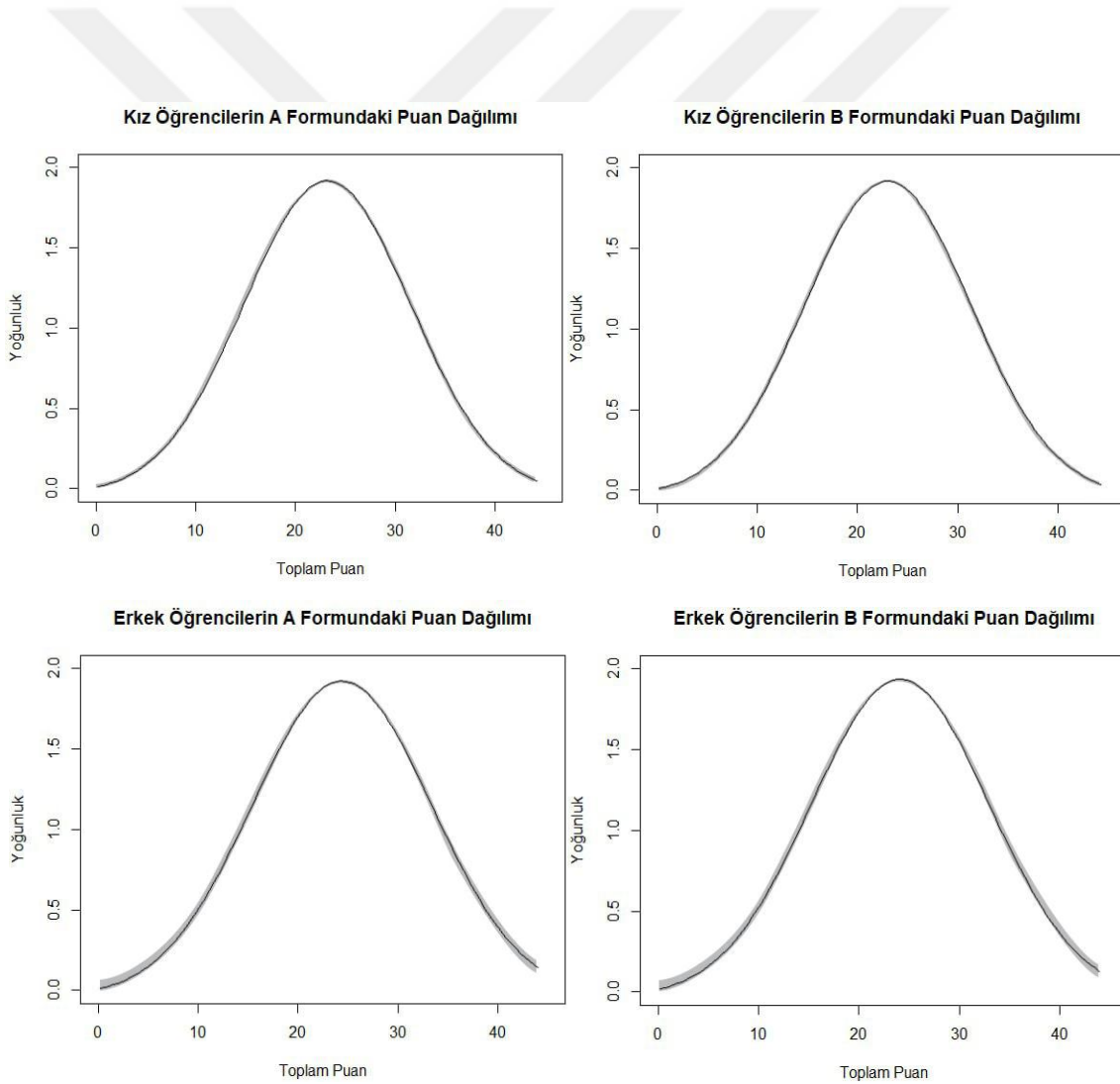
- a) NEC deseninde cinsiyet ortak deęiřken olarak alınmıř ve PoB yntemi ile ğrencilerin puanları birbirine baęlanmıřtır. Cinsiyet deęiřkenine gre puan daęılımları oluřturulmuř ve bu daęılımlar aracılıęıyla PoB modeli kullanılarak eřitlenmiř puanlar elde edilmiřtir.

Ortak deęiřken olarak cinsiyet modele dahil edildięinde, A ve B formlarını alan kız ve erkek ğrencilerin daęılımları karřılařtırılmıř, ayrıca ortak deęiřkenin alt kategorileri iin ayrı ayrı eřitleme iřlemi yapılmıřtır. Ortak deęiřkenin alt kategorilerinin gruplarda benzer daęılım gsterdięi varsayılarak test formları arasındaki farklılıklar da incelenmiřtir. ğrencilerin cinsiyetlerine gre formlardaki puan daęılımları řekil 27’de verilmiřtir.



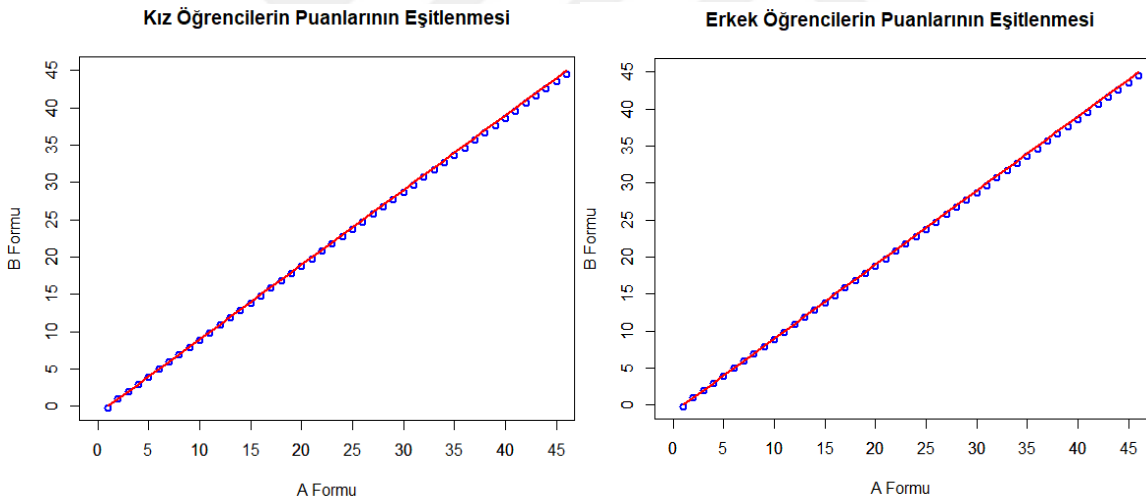
řekil 27. Cinsiyet Deęiřkenine Gre Formlardaki Puan Daęılımları

Kız öğrencilerin ve erkek öğrencilerin formlardaki puan dağılımları benzer olmakla birlikte kız öğrencilerin puan dağılımının erkek öğrencilere göre biraz daha sivri olduğu söylenebilir. Kız ve erkek öğrencilerin kendi içerisinde de A ve B formlarındaki puan dağılımları benzer olmakla birlikte çok az da olsa farklılıklar görülmektedir. Kız ve erkek öğrencilerin her ikisinin dağılımında da yaklaşık olarak 15 ile 25 puan arasında B formunda yoğunluğun biraz daha fazla; 30 ile 45 puan arasında A formunda yoğunluğun biraz daha fazla olduğu belirlenmiştir. NEC deseninde en önemli adım olan ortak kategorilere ait dağılımların gruplar arasında benzer olduğu varsayımı ile hem kız hem de erkek öğrenciler için A formunun daha kolay sorular içerdiği söylenebilir. Kız ve erkek öğrenciler için kestirilen dağılımların doğruluğunu belirlemek için güven aralıkları analiz edilmiştir. Elde edilen sonuçlar Şekil 28’de verilmiştir.



Şekil 28. Cinsiyet Değişkenine Göre Formlardaki Puan Dağılımları İçin Güven Aralıkları

Kız öğrenciler için formlarda kestirilen puanların dağılımlarına ait güven aralıkları benzerlik göstermektedir. Ortadaki puanlarda güven aralıklarının nispeten daraldığı görülmektedir. Erkek öğrenciler için de formlarda kestirilen puanların dağılımlarına ait güven aralıkları benzerlik göstermektedir. Alt ve üst uçtaki puanlarda güven aralıklarının nispeten genişlediği görülmektedir. Kız ve erkek öğrenciler karşılaştırıldığında ise erkek öğrenciler için formlarda kestirilen güven aralıklarının nispeten daha geniş olduğu söylenebilir. Dağılımlarda geniş güven aralıkları, puan kestirimlerinde kesinliğin azaldığını veya bu noktalardaki kestirimlerde belirsizlikler olduğunu göstermektedir. Bu durumun, puanların kestirilmesi için kullanılan veride yeterli örneklemin bulunmamasından veya örneklem içerisinde bu profildeki bireylerin geniş bir aralıkta puanlar almasından kaynaklandığı söylenebilir. Cinsiyet değişkenine göre oluşturulan profiller temel alındığında A formunun B formuna eşitlenmesi ile elde edilen eşitlenmiş puanlar Şekil 29’da verilmiştir.

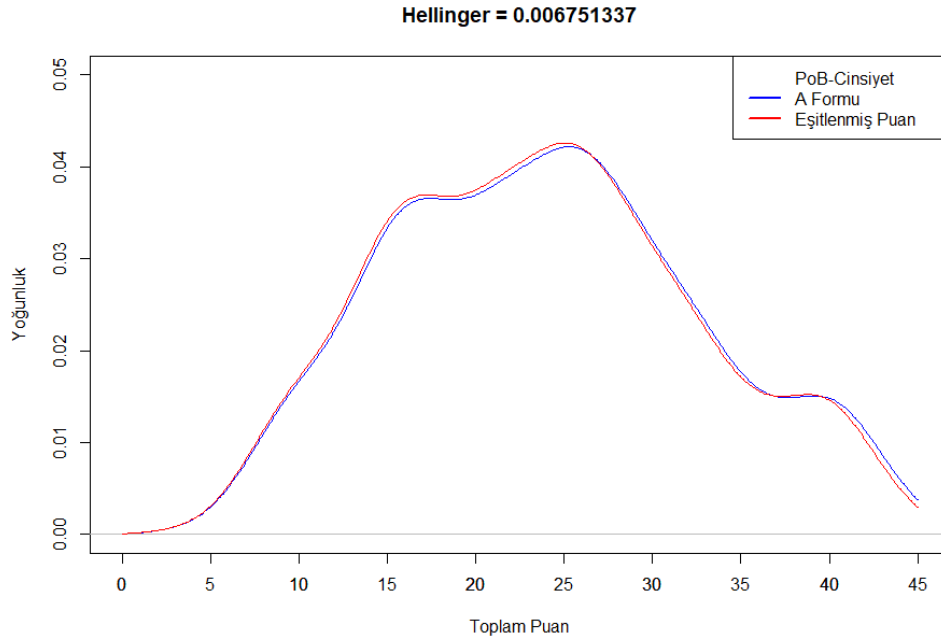


Şekil 29. Kız ve Erkek Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi

Kız ve erkek öğrenciler için formlardan elde edilen puanlar PoB yöntemine göre eşitlendiğinde her iki grupta da formlar arasında doğrusala yakın bir ilişki görülmüştür. Bu durum cinsiyet değişkeninin formlar arasındaki farkı açıklamada yeterli olmadığı veya formların birbiri ile benzerlik gösterdiği şeklinde ifade edilebilir. Fakat A formunun nispeten daha kolay sorular içerdiği söylenebilir. Dolayısıyla formlar arasındaki farkı açıklamada cinsiyet değişkeninin yeterince etkili olmadığı görülmektedir.

Literatürde PoB yönteminde, cinsiyetin ortak değişken olarak kullanıldığı çalışmalarda da eşitleme fonksiyonunun cinsiyete bağlı olmadığı belirlenmiştir (Gonzalez vd., 2015a, 2015b; Gonzalez ve Wiberg, 2017; Yurtçu vd., 2021).

Cinsiyet ortak değişkeni kullanılarak her bir öğrencinin puanı PoB modeline göre eşitlenmiştir. Eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 30’da verilmiştir.

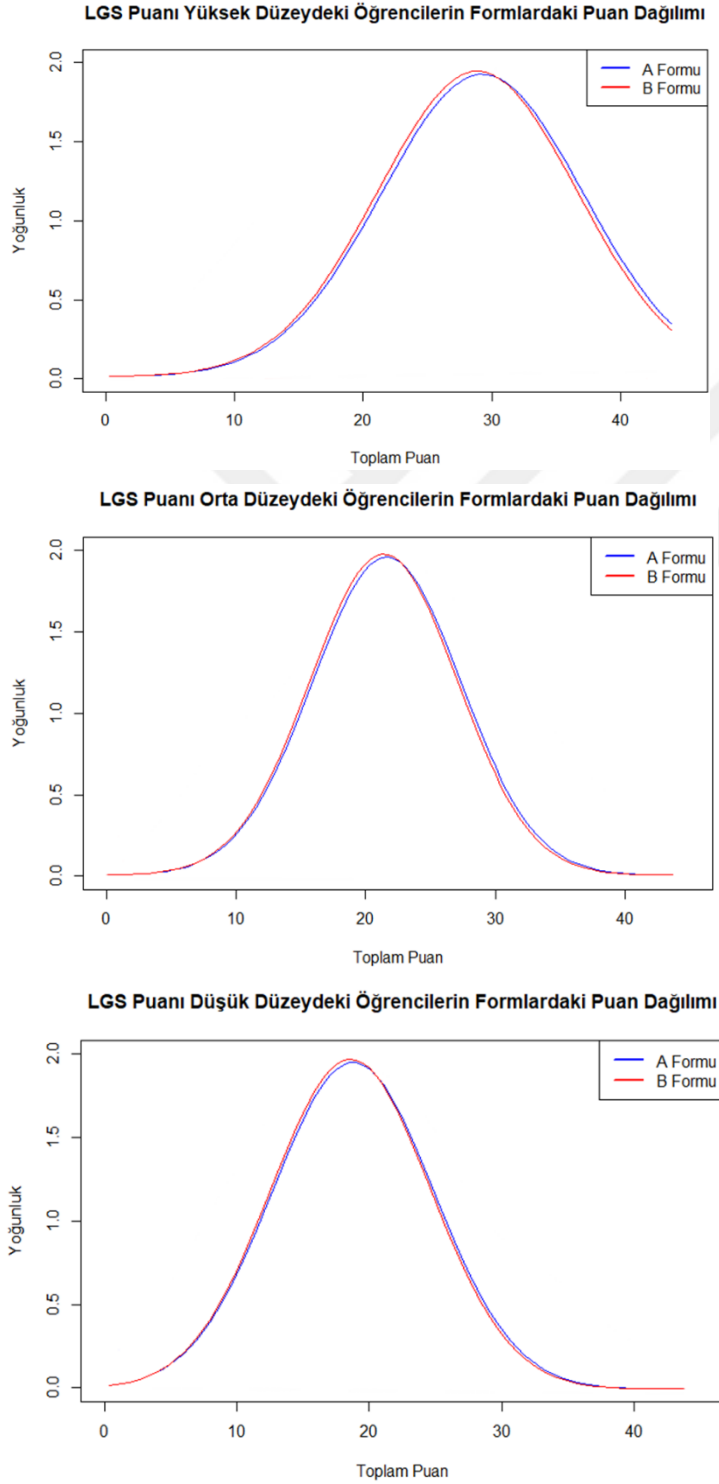


Şekil 30. Cinsiyet Ortak Değişkenine Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

Her bir öğrenci için cinsiyet değişkeni ve test puanları birleştirilerek PoB yönteminde puanlar eşitlenmiştir. Puanlara ait olasılık yoğunluk dağılımları incelendiğinde eşitlenmiş puanların dağılımı hedef teste ait puanların dağılımına çok yakındır. Bu uzaklık eşit yüzdelikli eşitleme ve Kernel eşitleme ile elde edilen uzaklığın yaklaşık onda biridir.

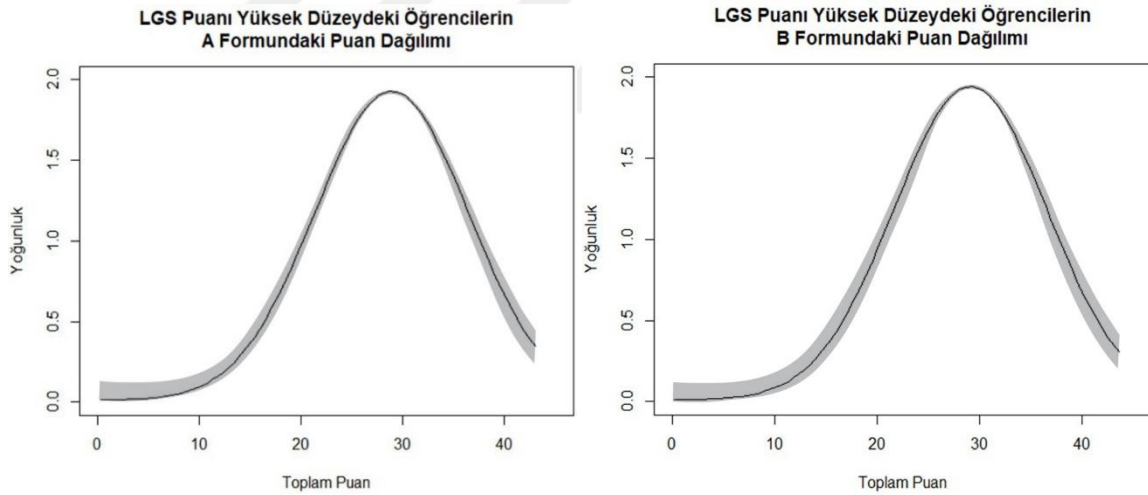
- b) NEC deseninde yüksek, orta ve düşük olarak kategorilendiren LGS puan düzeyleri ortak değişken olarak alınmış ve PoB yöntemi ile öğrencilerin puanları birbirine bağlanmıştır. LGS puan düzeylerine göre puan dağılımları oluşturulmuş ve bu dağılımlar aracılığıyla PoB modeli kullanılarak eşitlenmiş puanlar elde edilmiştir.

Yüksek, orta ve düşük düzeyi temsil edecek şekilde üç kategoriye ayrılan LGS puanlarının alt kategorileri için puan dağılımları oluşturulmuş ve eşitlenmiş puanlar elde edilmiştir. LGS puan düzeylerine göre öğrencilerin formlardan aldıkları puan dağılımları Şekil 31’de verilmiştir.



Şekil 31. LGS Puan Düzeylerine Göre Formlardaki Puan Dağılımları

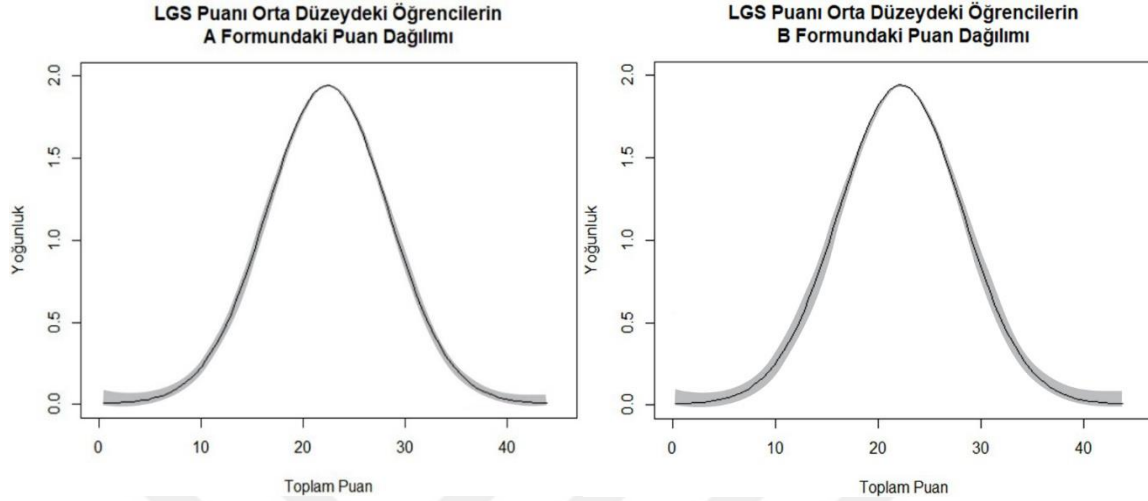
Dağılımların LGS puan düzeylerine göre farklılaştığı belirlenmiştir. Yüksek, orta ve düşük olarak kategorilendirilen LGS puan düzeylerinin kendi içerisinde, formlardan alınan puan dağılımları genel olarak benzer olmakla birlikte çok az da olsa farklılıklar görülmektedir. Bu farklılıklar LGS puanı yüksek düzeydeki öğrencilerin puan dağılımlarında biraz daha belirgindir. LGS puanı yüksek düzeydeki öğrencilerin puanlarının ortanın üzerinde; düşük düzeydeki öğrencilerin ortanın altında ve orta düzeydeki öğrencilerin ortadaki puanlarda yoğunlaştığı görülmektedir. LGS puanı yüksek düzeyde olan öğrencilerin düşük puan alma olasılığının azaldığı; düşük düzeyde olan öğrencilerin yüksek puan alma olasılığının azaldığı belirlenmiştir. Her düzey için yüksek puan alma olasılığının A formunda nispeten daha fazla olduğu görülmektedir. Dolayısıyla bu grafikler incelendiğinde de A formunda daha kolay soruların bulunduğu söylenebilir. Formlara göre elde edilen dağılımlara ait kestirimlerin doğruluğunu belirlemek için güven aralıkları analiz edilmiştir. Elde edilen sonuçlar Şekil 32, Şekil 33 ve Şekil 34’te verilmiştir.



Şekil 32. LGS Puanı Yüksek Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları

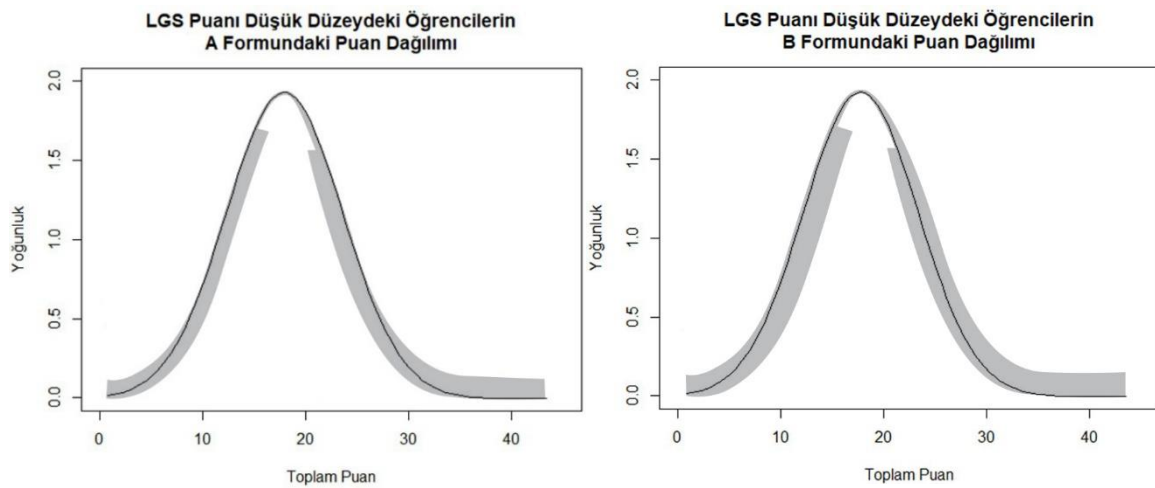
LGS puanı yüksek düzeydeki öğrencilerin formlardaki dağılımlarına ait güven aralıkları ayrı ayrı ele alınmıştır. Formlardaki dağılımlarda büyük bir farklılığın olmadığı ancak B formundaki güven aralığının nispeten daha geniş olduğu söylenebilir. Her iki formda da güven aralıkları uçlardaki puanlara doğru genişlemiştir. Güven aralıklarının genişleyip belirsizliğin artması, uçlardaki puanlarda kestirimler için önsel bilginin

yetersiz olmasından ya da aynı profildeki bireylerin farklı puanlar almasından kaynaklanabilir.



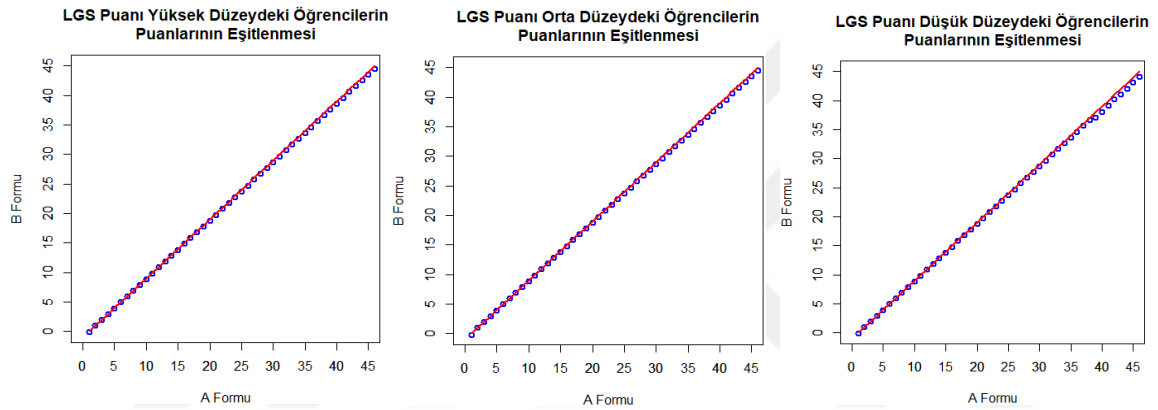
Şekil 33. LGS Puanı Orta Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları

LGS puanı orta düzeydeki öğrenciler için formlardaki dağılımlar benzerlik gösterirken güven aralıklarındaki farklılığın biraz daha belirgin olduğu görülmüştür. Güven aralığının B formunda nispeten daha geniş olduğu söylenebilir.



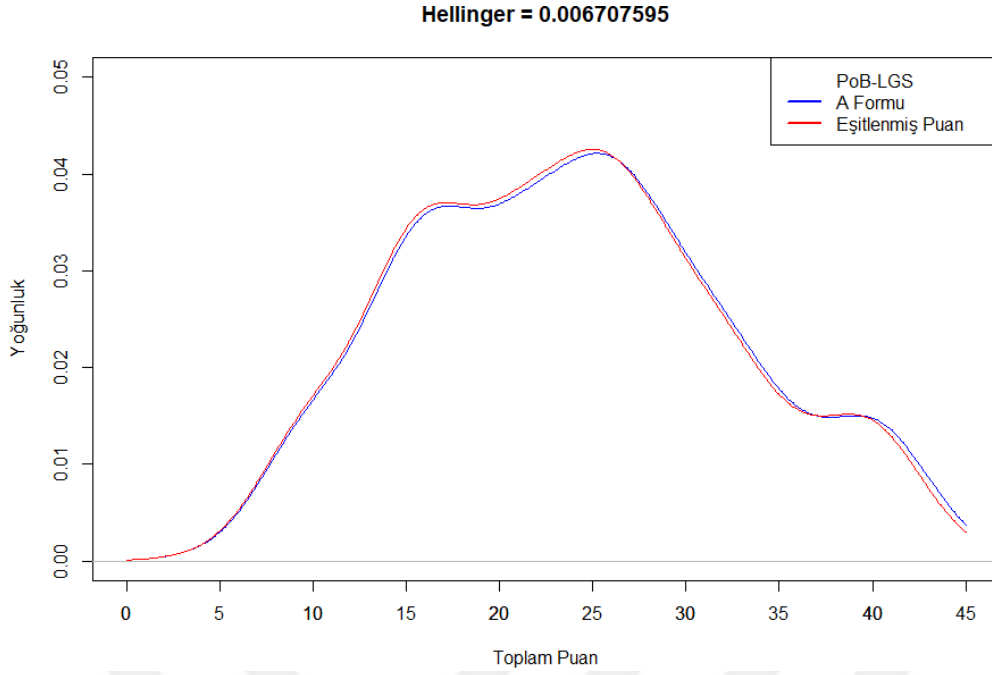
Şekil 34. LGS Puanı Düşük Düzeydeki Öğrencilerin Formlardaki Puan Dağılımları İçin Güven Aralıkları

LGS puanı düşük düzeydeki öğrenciler için formlardaki dağılımlar benzerlik göstermektedir. Güven aralıklarının sadece uçlarda değil ortadaki puanlarda da belirgin olduğu, tepe noktasına doğru daraldığı görülmektedir. Dağılımlar incelendiğinde B formunda güven aralığının nispeten daha geniş olduğu belirlenmiştir. B formunda güven aralığının geniş olması, kestirimlerde A formuna göre nispeten daha fazla hata yapıldığını göstermektedir. Güven aralıklarının genişlemesi, puan kestirimleri için önsel bilginin yetersiz olmasından ya da aynı profildeki bireylerin farklı puanlar almasından kaynaklanabilir. LGS puan düzeylerine göre oluşturulan profiller temel alındığında A formunun B formuna eşitlenmesi ile elde edilen eşitlenmiş puanlar Şekil 35’te verilmiştir.



Şekil 35. LGS Puan Düzeylerine Göre Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi

LGS puan düzeylerine göre formlardan elde edilen puanlar PoB yöntemine göre eşitlendiğinde formlar arasında doğrusala yakın bir ilişki görülmüştür. LGS puanı yüksek ve orta düzeydeki öğrencilerin puan dağılımlarının benzer olduğu; düşük düzeydeki öğrencilerde ise üst uçta biraz farklılaştığı belirlenmiştir. Dağılımlardan ve eşitlenmiş puanlardan görüldüğü üzere LGS puanı kategorik olarak alındığında, bu değişkenin formlar arasındaki farkı açıklamada yeterli olmadığı veya formların birbiri ile benzerlik gösterdiği söylenebilir. Fakat yine de A formunun nispeten daha kolay sorular içerdiği ifade edilebilir. Özellikle LGS puanı düşük düzeydeki öğrenciler için formların eşitlenmesinde bu sonucu görmek nispeten mümkündür. Eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 36’da verilmiştir.

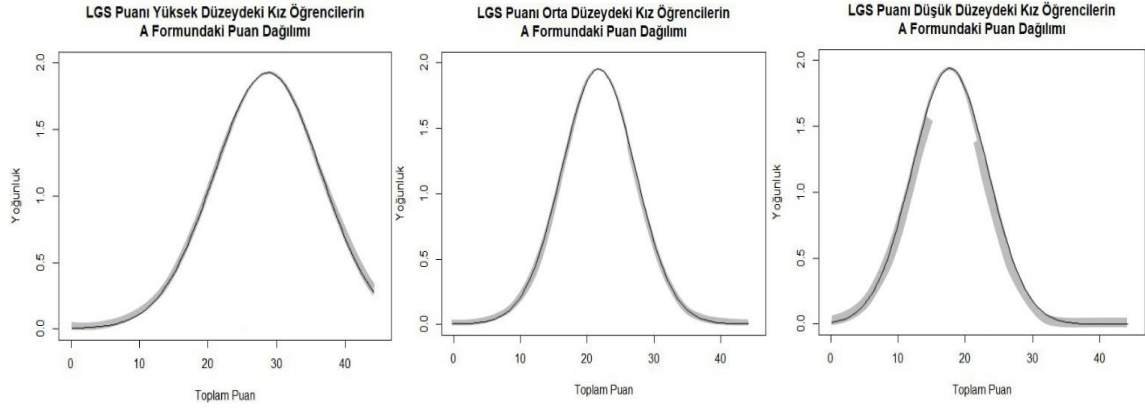


Şekil 36. LGS Puan Düzeyleri Ortak Değişkenine Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

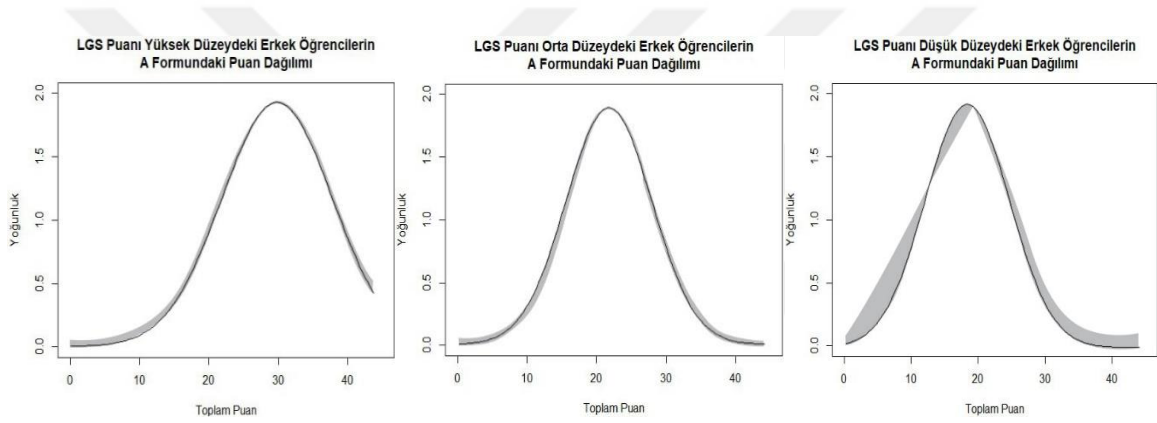
Her bir öğrenci için, kategorilere ayrılan LGS puanı değişkeni ve test puanları birleştirilerek PoB yönteminde puanlar eşitlenmiştir. Puanlara ait olasılık yoğunluk dağılımları incelendiğinde eşitlenmiş puanların dağılımı hedef teste ait puanların dağılımına çok yakındır. PoB yönteminde LGS puanı ortak değişken olarak kullanıldığında, cinsiyet ortak değişkenine göre yapılan kestirimden daha yakın puanlar elde edilmiştir.

- c) NEC deseninde cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınmış ve PoB yöntemi ile öğrencilerin puanları birbirine bağlanmıştır. Cinsiyet ve LGS puan düzeylerine göre puan dağılımları ve bu dağılımlar aracılığıyla PoB modeli kullanılarak eşitlenmiş puanlar elde edilmiştir.

Cinsiyet ve LGS puanları birlikte incelenmiş, kız ve erkek öğrencilerin LGS puan düzeylerine göre puan dağılımları oluşturulmuş ve güven aralıkları elde edilmiştir. Her iki form için öğrencilerin tüm kategorilere ait grafikleri aşağıda verilmiştir.



Şekil 37. LGS Puan Düzeylerine Göre A Formunda Kız Öğrencilerin Puan Dağılımları İçin Güven Aralıkları

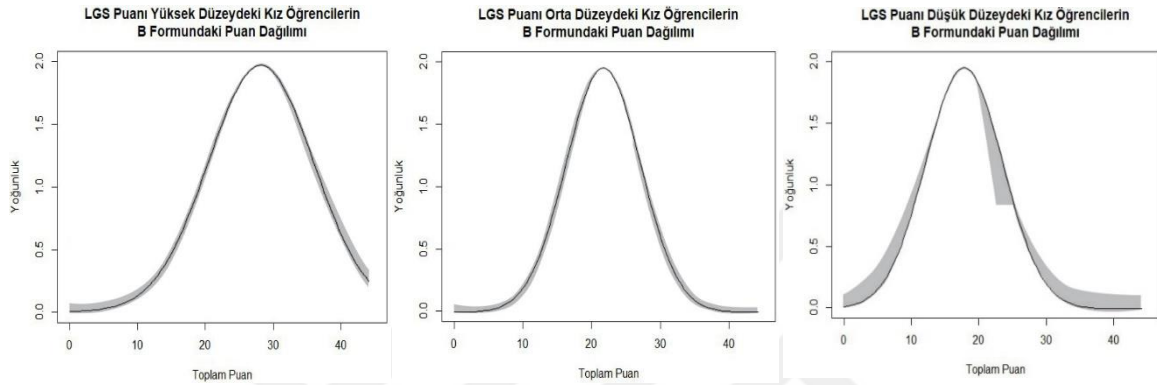


Şekil 38. LGS Puan Düzeylerine Göre A Formunda Erkek Öğrencilerin Puan Dağılımları İçin Güven Aralıkları

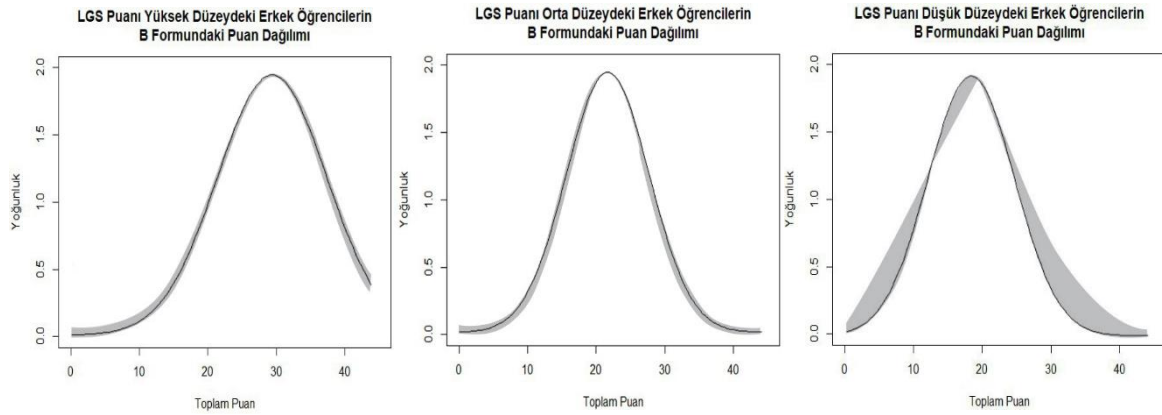
Şekil 37 ve Şekil 38’de A formunda kız ve erkek öğrencilerin puanlarının değişkenin kategorilerine göre dağılımları ve güven aralıkları verilmiştir. Grafiklerde puan dağılımlarının benzerlik gösterdiği fakat puan dağılımlarının kestirimindeki kesinliği gösteren güven aralıklarında cinsiyet değişkenine göre farklılık olduğu görülmüştür.

Kız öğrenciler için grafikler incelendiğinde LGS puan düzeylerine göre dağılımların farklılık gösterdiği belirlenmiştir. LGS puanı düşük düzeyde olan kız öğrencilerin güven aralıkları diğer düzeylerden farklılaşmaktadır. LGS puan düzeyi yüksek ve orta olması durumunda kız öğrencilerde güven aralığının dar ve kestirimlerin daha kesin olduğu; LGS puanı düşük düzeydeki kız öğrencilerde ise puan dağılımlarının kestirimindeki hatanın nispeten daha fazla olduğu belirlenmiştir.

A formunda erkek öğrencilerin LGS puan düzeylerine göre dağılımlarının kız öğrencilerle benzer olduğu görülmüştür. Kız öğrencilerdeki gibi, güven aralıkları da LGS puanı düşük düzeydeki erkeklerde genişlemiştir. Ayrıca bu düzeyde ortadaki puanlarda güven aralığının kızlara göre nispeten daha geniş olduğu belirlenmiştir. Güven aralıklarının genişlemesi, puan kestirimleri için önsel bilginin yetersiz olmasından ya da aynı profildeki bireylerin farklı puanlar almasından kaynaklanabilir.



Şekil 39. LGS Puan Düzeylerine Göre B Formunda Kız Öğrencilerin Puan Dağılımları İçin Güven Aralıkları

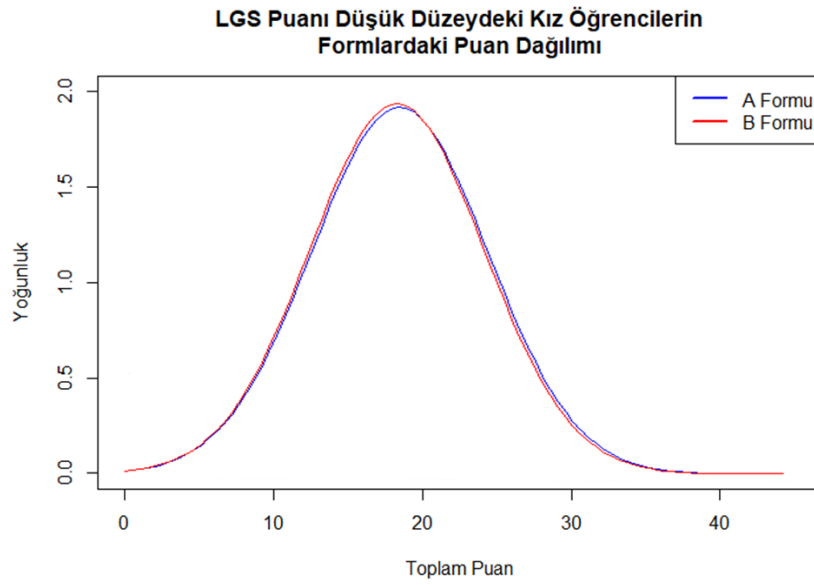
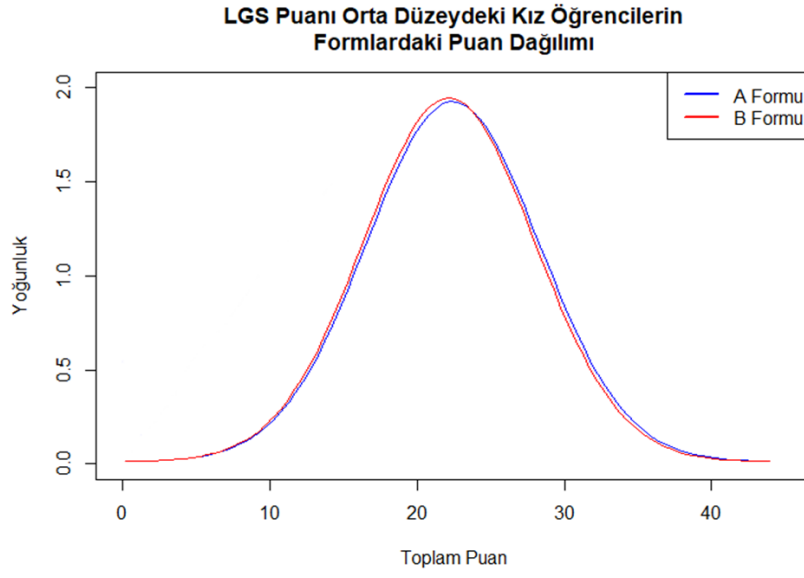
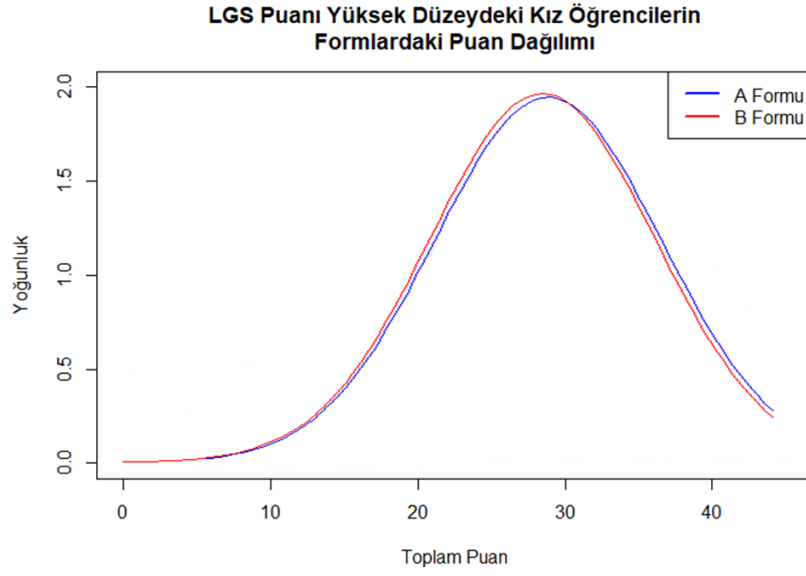


Şekil 40. LGS Puan Düzeylerine Göre B Formunda Erkek Öğrencilerin Puan Dağılımları İçin Güven Aralıkları

Şekil 39 ve Şekil 40'ta B formunda kız ve erkek öğrencilerin puanlarının değişkenin kategorilerine göre dağılımları ve güven aralıkları verilmiştir. B formundan elde edilen sonuçlar A formu ile benzerlik göstermektedir.

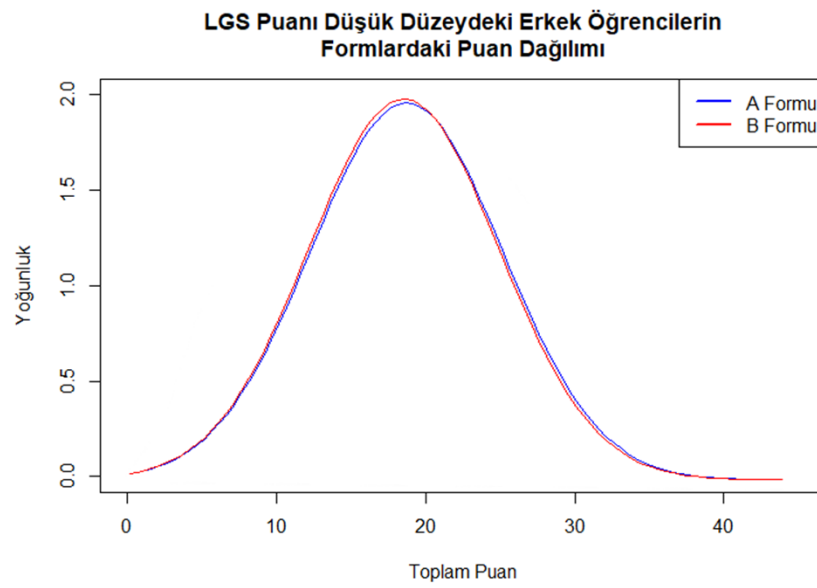
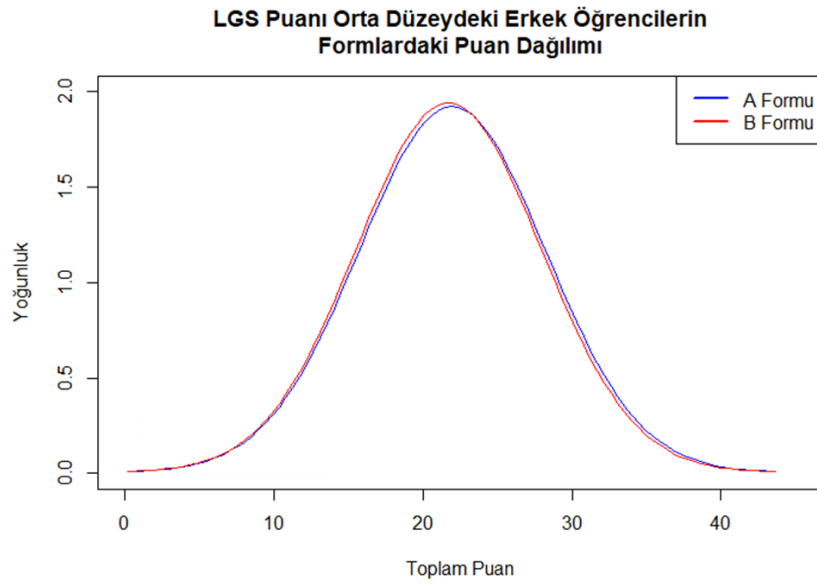
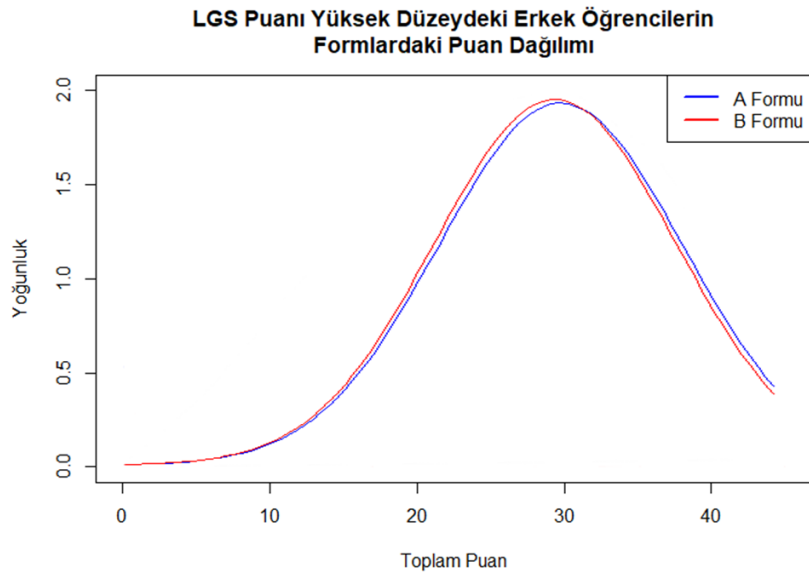
B formunda da LGS puanı düşük düzeyde olan kız öğrencilerin LGS puan düzeylerine göre dağılımları farklılık göstermektedir. LGS puanı düşük düzeyde olan kız öğrencilerin güven aralıklarının diğer düzeylerden daha geniş olduğu ve puan dağılımlarının kestirimindeki hatanın nispeten daha fazla olduğu belirlenmiştir. LGS puan düzeyleri yüksek ve orta olduğunda güven aralıklarının daha dar ve kestirimlerin daha kesin olduğu görülmüştür.

B formunda da erkek öğrencilerin LGS puan düzeylerine göre dağılımlarının kız öğrencilerle benzer olduğu görülmüştür. Kız öğrencilerdeki gibi, güven aralıkları da LGS puanı düşük düzeydeki erkekler için genişlemiştir. Ayrıca bu düzeyde ortadaki puanlarda güven aralığının kızlara göre nispeten daha geniş olduğu görülmektedir. Güven aralıklarının genişlemesi, puan kestirimleri için önsel bilginin yetersiz olmasından ya da aynı profildeki bireylerin farklı puanlar almasından kaynaklanabilir.



Şekil 41. LGS Puan Düzeylerine Göre Kız Öğrencilerin Formlardaki Puan Dağılımı

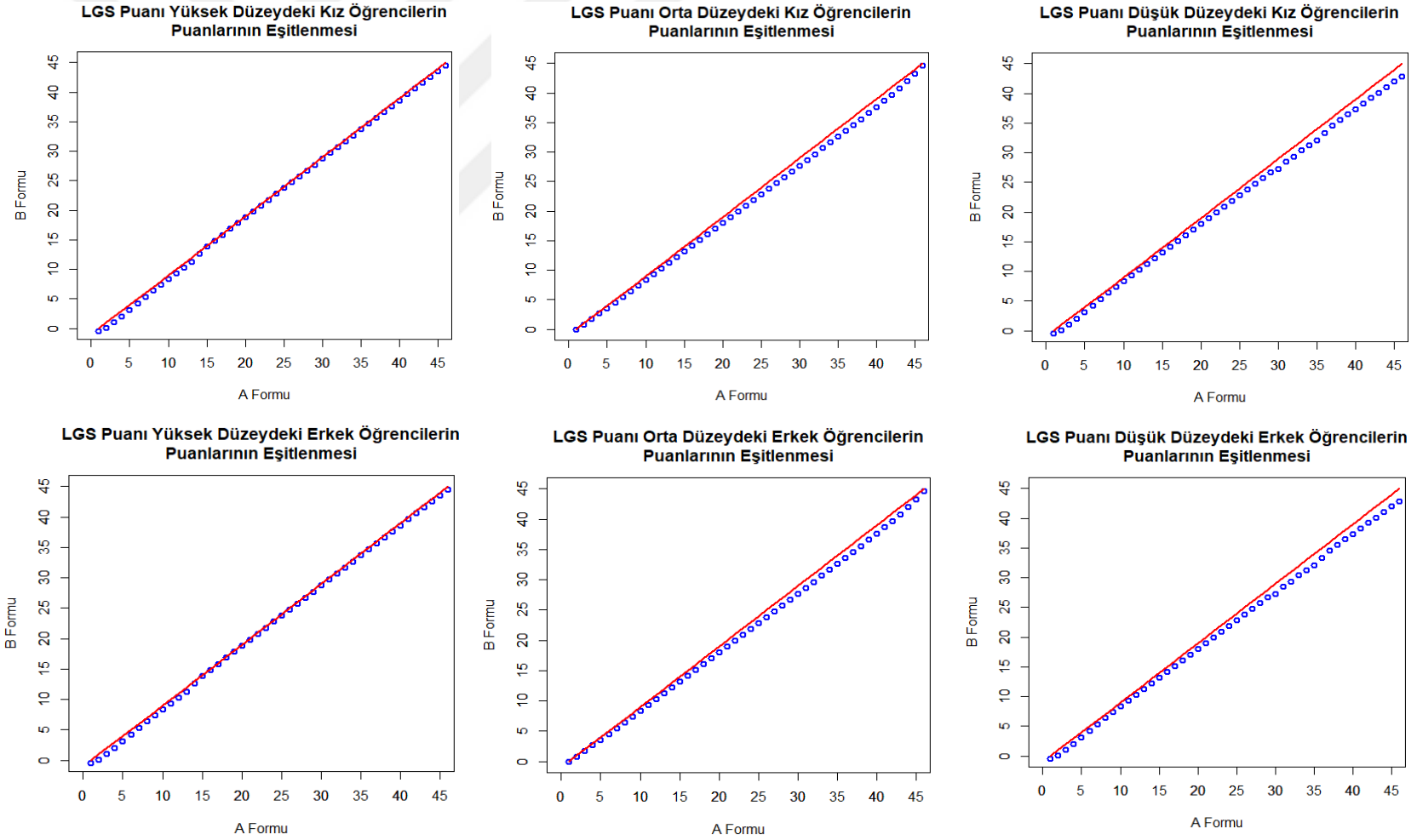
Kız öğrencilerin LGS puan düzeylerine göre formlardaki dağılımı Şekil 41’de verilmiştir. Bu dağılımlarda her bir LGS puan düzeyi için kız öğrencilerin puanları ortak profildir. Aynı profile sahip kız öğrencilerin farklı formlardan aldıkları puanların dağılımı analiz edilmektedir. LGS puan düzeylerine göre öğrencilerin puan dağılımlarının farklılaştığı görülmektedir. LGS puan düzeylerinin kendi içerisinde, formlardan alınan puan dağılımları genel olarak benzer olmakla birlikte çok az da olsa farklılıklar görülmektedir. Bu farklılıklar LGS puanı yüksek düzeydeki kız öğrencilerin puan dağılımlarında biraz daha belirgindir. LGS puanı yüksek düzeyde olan öğrencilerin düşük puan alma yoğunluğunun azaldığı ve aynı şekilde düşük puan alan öğrencilerin ise yüksek puan alma yoğunluğunun azaldığı görülmektedir. Her düzeydeki dağılımlar için yüksek puan alan öğrencilerin sayısı A formunda biraz daha fazladır. Bu dağılımların formlara göre karşılaştırılması ele alındığında A formunun B formundan daha kolay olduğu söylenebilir. Benzer şekilde erkek öğrenciler için de puan dağılımları incelenmiştir.



Şekil 42. LGS Puan Düzeylerine Göre Erkek Öğrencilerin Formlardaki Puan Dağılımı

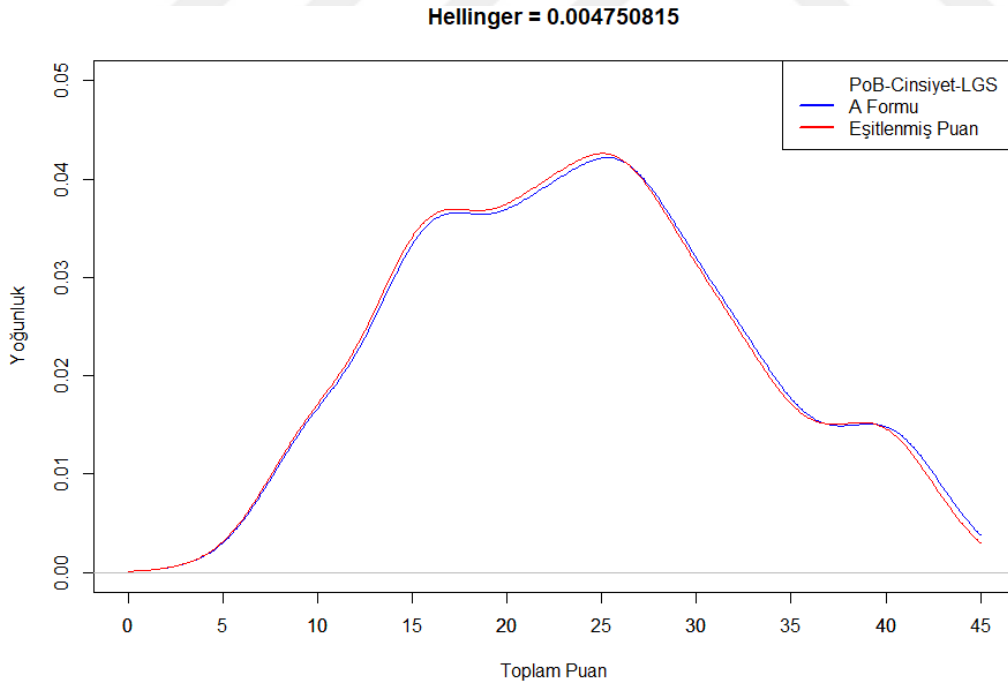
Erkek öğrencilerin LGS puan düzeylerine göre formlardaki dağılımı Şekil 42’de verilmiştir. LGS puan düzeylerine göre erkek öğrenciler için oluşturulan formlardaki dağılımlar kız öğrencilerin dağılımı ile benzerlik göstermektedir. Kız öğrenciler için oluşturulan puan dağılımlarının erkek öğrencilere göre nispeten daha sivri olduğu belirlenmiştir. Erkek öğrenciler için de LGS puan düzeylerine göre dağılımların farklılaştığı görülmektedir. LGS puan düzeylerinin kendi içerisinde, formlardan alınan puan dağılımları genel olarak benzer olmakla birlikte çok az da olsa farklılıklar görülmektedir. Bu farklılıklar LGS puanı yüksek düzeydeki erkek öğrencilerin puan dağılımlarında biraz daha belirgindir. LGS puanı yüksek düzeyde olan öğrencilerin düşük puan alma yoğunluğunun azaldığı ve aynı şekilde düşük puan alan öğrencilerin ise yüksek puan alma yoğunluğunun azaldığı görülmektedir. Erkek öğrencilerde de her düzeydeki dağılımlar için yüksek puan alan öğrencilerin sayısı A formunda biraz daha fazladır. Kız ve erkek öğrencilere ait puan dağılımları genellikle örtüşmektedir.

Kız ve erkek öğrenciler için, LGS puan düzeylerine göre oluşan profillerde A ve B formları arasında yapılan eşitlemeler sonucunda elde edilen grafikler Şekil 43’te verilmiştir.



Şekil 43. LGS Puan Düzeylerine Göre Kız ve Erkek Öğrencilerin Puanlarının PoB Yöntemiyle Eşitlenmesi

LGS puan düzeyi ve cinsiyet ortak değişken olarak alındığında, oluşan kategorilerdeki puan dağılımları arasında eşitleme yapılmıştır. Her bir kategorideki puan dağılımlarında gözlenen farklılık gibi eşitlenmiş puan dağılımlarında da nispeten farklılıklar görülmektedir. Kız ve erkek öğrenciler için eşitleme sonucunda oluşan dağılımlar birbirine benzerlik göstermektedir. LGS puanı yüksek düzeydeki öğrencilerde alt uçlardaki puanlarda, LGS puanı orta düzeydeki öğrencilerde orta ve üzerindeki puanlarda, LGS puanı düşük düzeydeki öğrencilerde üst uçtaki puanlarda doğrusaldan nispeten uzaklaşma görülmüştür. Puan dağılımlarından ve eşitlenmiş puanlardan da görüldüğü üzere, değişkenin formlar arasındaki farkı açıklamada yeterli olmadığı veya formların birbiri ile benzerlik gösterdiği söylenebilir. Eşitlenmiş puanlar karşılaştırıldığında öğrencilerin A formundan daha yüksek puanlar aldığı görülmektedir. Örneğin LGS puan düzeyi düşük olan ve A formundan 45 puan alan bir öğrencinin B formundan alacağı puan 42'dir. Eşitlenmiş puanlar ve hedef test için oluşan olasılık yoğunluk dağılımları Hellinger uzaklığı ile incelenmiş ve elde edilen sonuçlar Şekil 44'te verilmiştir.



Şekil 44. İki Ortak Değişkene Göre PoB Yöntemiyle Eşitlenen Puanların Olasılık Yoğunluk Dağılımları ve Hedef Teste Uzaklıkları

İki ortak değişken birlikte modele alındığında, eşitlenmiş puanlara ait dağılımın hedef teste çok yakın olduğu görülmüştür. Puanlara ait olasılık yoğunluk dağılımları PoB modeli ile elde edilen diğer modellere benzerdir. Ayrıca iki değişken kullanıldığında dağılımlar arasındaki yakınlığı temsil eden Hellinger uzaklığının tek değişken kullanılan durumlara kıyasla oldukça küçük olduğu belirlenmiştir.

PoB yöntemi ile yapılan eşitlemeler diğer yöntemlerle kıyaslandığında, noktasal olarak mesafeyi hesaplayan Hellinger uzaklıklarının bu yöntemde hedef teste en yakın sonuçları ürettiği belirlenmiştir. Klasik yöntemlere göre Bayes yönteminin daha iyi kestirim yaptığını ve daha yararlı çıkarımlar için kullanılabileceğini ifade eden çalışmalar elde edilen sonuçları desteklemektedir. Karabatsos ve Walker (2009) test eşitleme için PoB modelini tanıtmıştır. Bayesçi eşitleme modelinin Kernel, lineer, ortalama ve eşit yüzdelikli eşitleme modellerinden daha fazla hesaplama çabası ve matematiksel uzmanlık gerektirdiğini fakat modelin temel avantajları göz önüne alındığında bu çabanın göz ardı edilebileceğini belirtmiştir. Önceki dört modelin veriler hakkında oldukça gerçekçi olmayan varsayımlarda bulunduğu ve asimetri, ayırık puanlar gibi diğer teknik sorunları taşıdığı iddia edilmiştir. Bayes modeli, bu tür sorunları önlemesi açısından diğer modellerden daha iyi performans göstermiştir. Ayrıca eşit yüzdelikli, lineer ve ortalama eşitlemenin önsel bilgiler içeren PoB modelinin özel durumları olduğu kanıtlanmıştır. Gonzalez ve diğerleri (2015a) yapmış oldukları çalışmada eşitleme dönüşümüne yol açan puan dağılım fonksiyonlarının tahmininde ortak değişkenlerin kullanımına izin veren, test eşitleme için esnek bir yaklaşım olan PoB modelini kullanmıştır. Bu yöntemde eşitleme dönüşümleri ortak değişken değerlerine göre oluşturulur. Bu modelle yapılan eşitleme; Kernel ve eşit yüzdelikli eşitleme yöntemlerine göre ya daha iyi ya da onlarla eşit performans göstermiştir. Bu yöntem uç değerlerde ve sıfır frekanslı değerlerde eşitleme fonksiyonunun kestirimini kolaylaştırmıştır. PoB yönteminin esnek bir yöntem olduğu, eşitleme için gerekli olan birçok varsayımın karşılanmasına gerek olmadığı ve ortak maddenin olmadığı durumlarda da eşitlemenin yapılabileceği belirtilmiştir. Yurtçu ve diğerleri (2021) yapmış oldukları çalışmada PoB yöntemiyle yapılan test eşitlemede klasik yöntemlere göre daha fazla bilgi elde edildiği, ortak maddelerin yerine ortak değişkenlerin kullanılabileceği hatta daha iyi sonuçlar elde edilebileceği sonucuna ulaşmıştır. Ayrıca PoB ile elde edilen eşitlenmiş puanların hedef teste daha yakın sonuçlar verebileceği görülmüştür.

Testte ortak madde bulunmadığında, ortak maddeler eşitleme için gereken özellikleri taşımadığında veya test eşitleme varsayımlarının ihlal edildiği durumlarda

ortak deęiřken kullanılmasının daha doęru eřitleme sonuları vereceęini destekleyen alıřmalar bulunmaktadır (Branberg ve Wiberg, 2011; Dorans ve Holland,2000; Liou, Cheng ve Li, 2001).

Bayesi bir yaklařımı benimsemenin birok faydası ve avantajı olduęuna inanılmasına raęmen sıklıkla analiz ile ilgili bazı potansiyel zorluklar olduęu kabul edilmektedir. Bayes yntemlerini benimsemek isteyen arařtırmacıların ęrenme srecinde gerekli zaman ve abayı harcamaları gerekmektedir. Bununla birlikte yeni bir yaklařımın faydalarını anlamanın ve benimsemenin en iyi yolu denemektir (Kruschke vd., 2012). Bayesi yntemler model karřılařtırması iin mkemmeldir nk parametre sayısı ve yapısal karmařıklıktaki farklılıkları otomatik olarak dikkate alırlar. Bu nedenle Bayesi yntemler, gl ıkarımlar yapabilmek iin resmi bir mekanizma saęlar (Gray ve Cooper, 2010; Leavitt, Mitchell ve Peterson, 2010). Son olarak, Bayes analizinin neredeyse sihirli bir ıkarımsal sre olduęu ileri srlmřtr ve “gerekten zor problemler iin ok amalı bir istatistiksel zm olarak ařırı bir řekilde tanıtılmıřtır” (Gelman, 2008).

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

Bu bölümde araştırma sonucunda elde edilen bulgulara yönelik sonuçlar ve öneriler yer almaktadır.

Sonuçlar

Bu araştırmada Thurstone'un Çoklu Faktörler Modeli temel alınarak dokuzuncu sınıf öğrencileri için iki paralel formu olan bir yetenek testi geliştirilmiştir. Bu test Thurstone'un Çoklu Faktörler Modeli için kabul gören yedi temel faktörü içermektedir. Bu faktörler: bellek, sözel, sayısal, muhakeme, yer ve mekâna uyma, algısal ve kelime akıcılığıdır. Bu faktörler için çerçeve ve yeterlik ifadeleri belirlenmiştir. Belirlenen yeterlik ifadelerinin kelime akıcılığı faktöründe kısa cevaplı maddeyle, diğer faktörlerde çoktan seçmeli maddelerle ölçülebilir olmasına ve madde yazmaya elverişli olacak şekilde ifade edilmesine dikkat edilmiştir. Maddelerin bilişsel düzeyi belirlenerek belirtke tablosu oluşturulmuştur. Maddeler yazılmış, gözden geçirilmiş ve her bir formda 48 madde olacak şekilde deneme formları hazırlanmıştır. Deneme formları Antalya ili beş merkez ilçede hedef grubu temsil eden örneklem üzerinde uygulanmıştır. Deneme uygulaması sonucunda elde edilen veriler üzerinde test ve madde istatistikleri belirlenmiş, geçerlik ve güvenirlik analizleri yapılmıştır. Paralel formların güvenirliğinin belirlenmesi için ω güvenirlik katsayısı hesaplanmıştır. Yapı geçerliğinin belirlenmesi için doğrulayıcı faktör analizi yapılmış ve ölçüt geçerliğinin belirlenmesi için öğrencilerin 2022 LGS yerleştirme puanları ölçüt puan olarak kullanılmıştır. Deneme uygulaması sonucunda gerekli incelemeleri yapılan formların geçerli ve güvenilir bir yetenek testi olduğu belirlenmiştir. Yapılan incelemeler sonucunda her bir formda 45 madde olacak şekilde oluşturulan nihai formların Antalya ili beş merkez ilçede nihai uygulaması gerçekleştirilmiştir. Nihai uygulama sonucunda elde edilen veri üzerinde eşitleme çalışmaları yapılmış ve formlar eşitlenmiştir.

Bu arařtırmada eřit y zdelikli, Kernel ve PoB y ntemleri ile formlar eřitlenmiřtir. Klasik y ntemlerle yapılan eřitlemeler test eřitlemede ortak deęiřken kullanımına izin veren ve esnek bir y ntem olan PoB y ntemi ile karřılařtırılmıřtır. Test formlarında s zde ortak maddeler bulunmaktadır. Ortak deęiřken olarak cinsiyet ve kategorik hale getirilen LGS puan d zeyleri kullanılmıřtır. EG ve NEAT desenlerinde eřit y zdelikli ve Kernel eřitleme; NEC deseninde Kernel eřitleme ve PoB y ntemleri kullanılarak formlar birbirine eřitlenmiřtir. Bu arařtırmada dięer y ntemler ile PoB y nteminden elde edilen sonu ları karřılařtırabilmek i in kriter olarak eřitlenmiř puanların daęılımının hedef teste uzaklıklarını inceleyen Hellinger Uzaklıęı kullanılmıřtır. Yapılan eřitlemeler sonucunda elde edilen bulgular ve yorumlara ait sonu lar ařaęıda verilmiřtir.

- Arařtırmanın birinci alt probleminde EG ve NEAT desenlerinde d zg nleřtirilen ve d zg nleřtirilmeyen puan daęılımları  zerinde eřit y zdelikli eřitleme yapılmıřtır. Eřit y zdelikli eřitleme sonucunda her iki desende de d zg nleřtirme yapılan puan daęılımlarında SEE, BIAS ve RMSE hata deęerlerinin daha d ř k olduęu belirlenmiřtir. Hem d zg nleřtirilen hem de d zg nleřtirilmeyen puan daęılımlarında en d ř k SEE deęeri EG deseninde elde edilmiřtir. NEAT deseninde eřit y zdelikli frekans tahmini ve eřit y zdelikli zincirleme eřitleme karřılařtırıldığında y ntemler arasında k   k de olsa farklılıklar g r lm řtir. Frekans tahmininin zincirleme eřitlemeye g re daha d ř k SEE fakat daha y ksek BIAS ve RMSE  rettięi belirlenmiřtir. Eřit y zdelikli eřitleme sonucunda elde edilen eřitlenmiř puanların hedef teste uzaklıkları incelendięinde hem d zg nleřtirilen hem de d zg nleřtirilmeyen puan daęılımlarında frekans tahmini y nteminde elde edilen puan kestirimlerinin hedef teste en yakın olduęu belirlenmiřtir. Frekans tahmini y ntemi EG desenine g re daha y ksek SEE  retmiř olsa da hedef teste daha yakın kestirimler yaptığ  g r lm řtir.
- Arařtırmanın ikinci alt probleminde EG ve NEAT desenlerinde Kernel eřitleme yapılmıřtır. Elde edilen bulgular sonucunda Kernel lineer eřitlemede SEE, BIAS ve RMSE deęerlerinin Kernel eřit y zdelikli eřitlemeden daha d ř k olduęu belirlenmiřtir. Kernel eřit y zdelikli eřitlemelerde u  noktalarda puanların daha y ksek hata ile kestirildięi g r lm řtir. Kernel eřit y zdelikli ve lineer eřitlemelerde son tabakalama eřitleme y nteminin zincirleme eřitlemeye g re daha d ř k hata deęerlerine sahip olduęu belirlenmiřtir. Bu  alıřmada y ntemler arasında elde edilen farklar  ok y ksek deęildir. Bu durumun  rneklem

büyükliğünden kaynaklandığı söylenebilir. Eşit yüzdelikli eşitleme ve Kernel eşitleme sonucunda elde edilen bulgular birlikte değerlendirildiğinde ise loglineer ön düzgünleştirme yapılan eşit yüzdelikli eşitlemenin Kernel eşitlemeye yakın sonuçlar verdiği belirlenmiştir. Kernel eşitleme sonucunda eşitlenmiş puanların hedef teste uzaklıkları incelendiğinde, Kernel eşit yüzdelikli eşitlemelerde eşitlenmiş puanların biraz daha sivri olduğu Kernel lineer eşitlemelerde ise hedef teste daha yakın olduğu belirlenmiştir. Hedef teste en yakın puan kestirimleri ise Kernel lineer zincirleme eşitlemede yapılmıştır.

- Araştırmanın üçüncü alt problemde NEC deseninde cinsiyet; LGS puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak Kernel eşitleme yapılmıştır. Elde edilen bulgular sonucunda Kernel lineer eşitlemeden elde edilen SEE, BIAS ve RMSE değerlerinin Kernel eşit yüzdelikli eşitlemeden daha düşük olduğu belirlenmiştir. Bu çalışmada NEC deseninde Kernel eşitleme ile elde edilen sonuçlar diğer yöntemlerden elde edilen hata değerleriyle kıyaslandığında SEE değerlerinin yakın olduğu fakat BIAS ve RMSE değerlerinin biraz yükseldiği sonucuna ulaşılmıştır. Ortak değişken olarak kullanılan cinsiyet değişkeni ile test arasındaki korelasyon değeri istatistiksel olarak anlamlı değilken, LGS puan düzeyleri ile yüksek düzeyde ve anlamlı bir korelasyon bulunmaktadır. Bu çalışma sonucunda NEC desenindeki yöntemler karşılaştırıldığında LGS puan düzeylerinin ortak değişken olarak kullanılması en düşük hata değerlerine sebep olurken, cinsiyet değişkeninin ortak değişken olarak kullanılması en yüksek hata değerlerine sebep olmuştur. Cinsiyet değişkenine LGS puan düzeylerinin de eklenip birlikte ortak değişken olarak kullanılması ise hata değerlerini düşürmüştür. Her ortak değişken için elde edilen eşitlenmiş puanların hedef teste uzaklıkları incelendiğinde birbirine yakın olduğu görülmekle birlikte cinsiyet değişkeni ve LGS puan düzeyleri birlikte ortak değişken olarak kullanıldığında hedef teste en yakın kestirimler yapılmıştır. Ortak değişken kullanılarak yapılan Kernel eşitleme sonucunda eşitlenmiş puanların hedef teste uzaklıkları diğer yöntemlerden elde edilen uzaklıklara yakındır.
- Araştırmanın dördüncü alt problemde NEC deseninde cinsiyet; LGS puan düzeyleri; cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alınarak PoB yönteminde eşitleme yapılmıştır. Modelde sadece cinsiyet ortak değişken olarak alındığında kız ve erkek öğrencilerin formlarda benzer dağılım gösterdiği görülmekle birlikte kız öğrencilerin erkek öğrencilerden biraz daha sivri bir

dağılıma sahip olduğu söylenebilir. Kız ve erkek öğrencilerin kendi içerisinde de A ve B formlarındaki puan dağılımları benzer olmakla birlikte çok az da olsa farklılıklar görülmüştür. Kız ve erkek öğrenciler için formlardan elde edilen puanlar PoB yöntemine göre eşitlendiğinde her iki grupta da formlar arasında doğrusala yakın bir ilişki belirlenmiştir. Dağılımlar ve eşitlenmiş puanlar incelendiğinde cinsiyetin yetenek testinden alınan puanlar üzerinde önemli bir etkisinin olmadığı sonucuna ulaşılmıştır. Cinsiyet değişkeninin formlar arasındaki farkı ortaya çıkarmada yetersiz olduğu görülse de A formunun B formundan biraz daha kolay sorular içerdiği belirlenmiştir. Her öğrencinin cinsiyeti ve testten aldığı puanlardan yararlanılarak oluşturulan eşitlenmiş puanların dağılımının hedef teste çok yakın olduğu görülmüştür. Kullanılan diğer yöntemlerden elde edilen dağılımlarla karşılaştırıldığında ise PoB yönteminde, eşitlenmiş puanların hedef teste oldukça yakın olduğu ve bu yöntem ile daha kesin puan kestirimleri yapıldığı sonucuna ulaşılmıştır.

- Dördüncü alt problemde modele sadece LGS puan düzeyleri ortak değişken olarak alındığında düzeylere göre öğrencilerin puan dağılımlarının farklılaştığı belirlenmiştir. Yüksek, orta ve düşük olarak kategorilendirilen LGS puan düzeylerinin kendi içerisinde, formlardan alınan puan dağılımları genel olarak benzer olmakla birlikte çok az da olsa farklılıklar görülmüştür. Bu farklılıklar LGS puanı yüksek düzeydeki öğrencilerin puan dağılımlarında biraz daha belirgindir. LGS puanı yüksek düzeyde olan öğrencilerin düşük puan alma olasılığının azaldığı; düşük düzeyde olan öğrencilerin yüksek puan alma olasılığının azaldığı belirlenmiştir. Her düzey için yüksek puan alma olasılığının A formunda nispeten daha fazla olduğu görülmüştür. Dolayısıyla bu grafikler incelendiğinde de A formunda daha kolay soruların bulunduğu sonucuna ulaşılmıştır. LGS puan düzeyleri için formlardan elde edilen puanlar PoB yöntemine göre eşitlendiğinde formlar arasında doğrusala yakın bir ilişki belirlenmiştir. Her iki formda da dağılımlar ve eşitlenmiş puanlar incelendiğinde değişkenin formlar arasındaki farkı ortaya çıkarmada yeterli olmadığı görülmüştür. LGS puan düzeyinin ortak değişken olarak kullanılması ile cinsiyet değişkenine göre hedef teste daha yakın dağılımlar elde edildiği sonucuna ulaşılmıştır.
- Dördüncü alt problemde modele cinsiyet ve LGS puan düzeyleri birlikte ortak değişken olarak alındığında tek değişkenin ortak değişken olduğu durumlara göre

daha detaylı bilgiler elde edilmiştir. Kız ve erkek öğrenciler için ayrı ayrı LGS puan düzeylerine göre dağılımlar incelenmiştir. Hem kız hem de erkek öğrenciler için LGS puan düzeylerine göre dağılımların farklılaştığı belirlenmiştir. Yine kız ve erkek öğrenciler için LGS puan düzeylerinin kendi içerisinde, formlardan alınan puan dağılımları genel olarak benzer olmakla birlikte çok az da olsa farklılıklar görülmüştür. Bu farklılıklar LGS puanı yüksek düzeydeki öğrencilerin puan dağılımlarında biraz daha belirgindir. Burada elde edilen puan dağılımlarında da LGS puanı yüksek düzeyde olan öğrencilerin düşük puan alma yoğunluğunun azaldığı ve aynı şekilde düşük puan alan öğrencilerin ise yüksek puan alma yoğunluğunun azaldığı görülmüştür. Eşitlenmiş puanlar karşılaştırıldığında öğrencilerin A formundan daha yüksek puanlar aldığı belirlenmiştir. Puan dağılımları ve eşitlenmiş puanlar incelendiğinde değişken formlar arasındaki farklılığı açıklamada yeterli olmamış veya formlar birbiri ile benzerlik göstermiştir. Eşitlenmiş puanların hedef teste uzaklığı incelendiğinde ise iki değişken ortak değişken olarak kullanıldığında tek değişken kullanılan durumlara göre hedef teste daha yakın kestirimler yapıldığı sonucuna ulaşılmıştır.

- PoB yönteminde ortak değişkenlere göre yapılan eşitleme sonucunda dağılımların formlara göre büyük farklılıklar göstermediği belirlenmiştir. Bütün modellerde farklı ortak değişken kullanılsa bile A formunun B formundan daha kolay olduğu görülmüştür. Eşitlemede PoB modelinin kullanılması hem değişken hem de değişkenin alt kategorileri ile ilgili ayrıntılı bilgi elde edilmesini ve her bir profile ait dağılımlar aracılığıyla farklı karşılaştırmaların yapılabilmesini sağlamaktadır. Bu model ile yapılan eşitlemelerde grup değişmezliği varsayımı kontrol altında tutulmaktadır. Bu sebeple ortak profillerden yararlanılarak formlar için eşitlenmiş puanlar elde edilmiş, ayrıca aynı test formunu alan kız ve erkek öğrenciler arasında veya LGS puan düzeyleri arasında da eşitleme yapılabilmiştir. Böylece LGS puan düzeyi düşük olan öğrenciler ile yüksek olan öğrencilerin cinsiyetleri, göstermiş oldukları performans ve aldıkları formlar hakkında yorum yapmak ve aralarındaki ilişkiyi incelemek mümkün olmuştur.
- PoB yönteminde eşitlenmiş puanların hedef teste uzaklıklarını belirlemek için Hellinger uzaklıkları incelendiğinde hedef teste çok yakın dağılımlar elde edildiği görülmüştür. Bu dağılımlar içerisinde en yakın uzaklık iki ortak değişken birlikte kullanıldığında elde edilmiştir. PoB yöntemiyle elde edilen uzaklıkların eşit yüzdelikli ve Kernel eşitleme yöntemleri sonucunda oluşanların yaklaşık onda biri

olduğu belirlenmiştir. Eşitlenmiş puanların dağılımları ve hedef teste uzaklıkları incelendiğinde diğer yöntemler ile PoB modelleri arasında karşılaştırma yapmak oldukça kolaydır. Fakat farklı ortak değişkenler kullanıldığında PoB modelinin kendi içerisinde bu karşılaştırmayı yapmak zorlaşmaktadır. Hellinger uzaklığıyla elde edilen bilgiler karar almada yetersiz olsa da dağılımların grafikleri ve eşitlenmiş puanlar da bu bilgileri destekler niteliktedir.

- PoB yönteminde eşitleme yaparken eşitlenen formları alan birey sayısının aynı olması gibi bir zorunluluk yoktur. Hatta alt grupların eşitlenmesinde de formları alan birey sayısının aynı olması gerekmez. Alt gruplar arasında eşitleme yaparken kişi sayısının azalması veya aynı profile sahip bireylerin geniş bir ranjda puan alması güven aralıklarının genişlemesine neden olmuştur. Fakat geniş güven aralıklarında bile puan kestirimleri klasik yöntemlere göre hedef teste aşırı yakınlık göstermiştir. Bu araştırmada modelin klasik yöntemlere göre daha kullanışlı olduğu ve yapılan puan kestirimlerinin hedef teste daha yakın olduğu sonucuna ulaşılmıştır.

Öneriler

Uygulayıcılara ve araştırmacılara yönelik olmak üzere öneriler iki başlık altında verilmiştir.

Uygulayıcılara Yönelik Öneriler

- Araştırmada ortak değişken seçilirken test ile korelasyonunun yüksek olmasına ve değişkenin alt kategorilerinde seyrekleşme olmamasına dikkat edilmiştir. Kernel eşitlemede ortak değişken seçiminde değişkenin alt kategorilerdeki dağılımına da korelasyon kadar önem verilmelidir.
- Araştırmada iki ortak değişken kullanılmıştır. Kernel eşitlemede ortak değişken sayısına ve ortak değişkenlerin kategorilendirmesine dikkat edilmelidir. Çünkü kategori sayısı arttıkça bu kategorilerdeki birey sayısı azalacak ve bu da puan kestirimlerinde hataya sebep olacaktır.

- Araştırmada sürekli değişken için üç kategori belirlenmiştir. Sürekli değişkenin kategorilerinin belirlenmesinde puan dağılımlarındaki değişim dikkate alınmalıdır.
- Araştırmada eşit yüzdelikli ve Kernel eşitleme için bootstrap kullanılmış ve yeniden örnekleme yapılmıştır. PoB yönteminde MCMC örnekleme 150000 olarak belirlenmiştir. Dağılımların evrene yaklaşması için örnekleme sürecine dikkat edilmelidir.

Araştırmacılara Yönelik Öneriler

- Formlarda Değişen Madde Fonksiyonu içeren madde incelemesi yapıp bu maddeler dikkate alınarak çalışma tekrarlanabilir.
- Araştırmada formlarda çoktan seçmeli ve kısa cevaplı maddeler bulunmaktadır. Çoktan seçmeli maddeler 0-1 şeklinde puanlanırken kısa cevaplı maddeler 0 ile 3 arasında kısmi olarak puanlanmıştır. Maddelerin değerlendirilmesinde modeli etkilemesini önlemek amacıyla bütün maddeler aynı kategori sayısında alınmıştır. Farklı kategori sayısı dikkate alınarak karma formatta eşitleme yapıp sonuçlar incelenebilir.
- Araştırmada kategorik ortak değişkenler kullanılmıştır. Sürekli ve kategorik farklı ortak değişkenler kullanılarak eşitleme yapıp sonuçlar incelenebilir.
- Formlarda ortak değişkenlerin alt kategorilerindeki birey sayıları birbirine yakındır. Ortak değişkenlerin alt kategorileri arasındaki fark yüksek olduğunda eşitleme yapıp güven aralıkları incelenebilir.
- Araştırmada model olarak DBPP kullanılmıştır. Bayesçi farklı yöntemler kullanılarak sonuçlar incelenebilir.
- Farklı koşullar altında test eşitlemede PoB yönteminin performansını inceleyen ve klasik yöntemlerle karşılaştıran simülasyon çalışmaları yapıp sonuçlar incelenebilir.

KAYNAKLAR

- Akhun, İ. (1984). İki Korelasyon Katsayısı Arasındaki Farkın Manidarlığının Test Edilmesi. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 17(1), 1-7. doi:10.1501/Egifak_0000001034
- Akın Arıkan, Ç. (2019). A comparison of Kernel equating methods based on NEAT design. *Eurasian Journal of Educational Research*, 19(82), 27-44. doi:10.14689/ejer.2019.82.2
- Akın Arıkan, Ç. (2020). The impact of covariate variables on Kernel equating under the non-equivalent groups. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 362-373. doi: 10.21031/epod.706835
- Akın Arıkan, Ç., & Gelbal, S. (2018). A comparison of traditional and Kernel equating methods. *International Journal of Assessment Tools in Education*, 5(3), 417-427. doi:10.21449/ijate.409826
- Akın Arıkan, Ç., & Gelbal, S. (2021). A comparison of Kernel equating and item response theory equating methods. *Eurasian Journal of Educational Research*, 93, 179-198. doi:10.14689/ejer.2021.93.9
- Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74(8), 1–36. doi:10.18637/jss.v074.i08
- Altıntaş, Ö., & Wallin, G. (2021). Equality of admission tests using Kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, 8(4), 729-743. doi:10.21449/ijate.976660
- Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan Publishing.
- Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the Kernel method of test equating with the package kequate. *Journal of Statistical Software*, 55(6), 1-25. doi:10.18637/jss.v055.i06
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.) *Educational measurement* (pp.508–600). Washington, DC: American Council on Education.
- Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, 28(2), 147–162. doi:10.1111/j.1745-3984.1991.tb00350.x
- Baker, N. (2016). *Madde Tepki Kuramının Temelleri* (M. İlhan, Çev.) N. Güler (Ed.). Ankara: Pegem Akademi.

- Barrientos, A. F., Jara, A., & Quintana, F. A. (2012). On the support of MacEachern's dependent dirichlet processes and extensions. *Bayesian Analysis*, 7(2), 277-310. doi:10.1214/12-BA709
- Barrientos, A. F., Jara, A., & Quintana, F. A. (2017). Fully nonparametric regression for bounded data using dependent Bernstein polynomials. *Journal of the American Statistical Association*, 112(518), 806-825. https://www.mat.uc.cl/~ajara/Publications_files/DependentBernstein.pdf
- Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem Akademi.
- Boehm, A. E., & White, M. A. (1982). *The parents' handbook on school testing*. New York, NY: Teacher's College Press.
- Boone, E. L., Merrick, J. R. W., & Krachey, M. J. (2012). A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation*, 84(4), 833-849. doi:10.1080/00949655.2012.729588
- Boone, E. L., Ye, K., & Smith, E. P. (2009). Using data augmentation via the Gibbs Sampler to incorporate missing covariate structure in linear models for ecological assessments. *Environmental and Ecological Statistics*, 16(1), 75-87. doi:10.1007/s10651-007-0050-z
- Branberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement*, 48(4), 419-440. doi:10.1111/j.1745-3984.2011.00153.x
- Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P.W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.
- Brossman, B. G., & Lee, W. C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement*, 37(6), 460-481. doi:10.1177/0146621613484083
- Brown, G. W. (1981). Bayes' formula: Conditional probability and clinical medicine. *American Journal of Diseases of Children*, 135(12), 1125-1129. doi:10.1001/archpedi.1981.02130360033012
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (1st ed.). New York, NY: Guilford publications.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2009). *Bilimsel Araştırma Yöntemleri*. Ankara: Pegem Akademi.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytics studies*. New York, NY: Cambridge University Press.

- Cashen, L. H., & Geiger, S. W. (2004). Statistical power and the testing of null hypotheses: A review of contemporary management research and recommendations for future studies. *Organizational Research Methods*, 7(2), 151-167. doi:10.1177/1094428104263676
- Choi, S. I. (2009). *A comparison of Kernel equating and traditional equipercentile equating methods and the parametric bootstrap methods for estimating standard errors in equipercentile equating* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Department of Educational Psychology, Urbana.
- Cohen, R. J., & Swerdlik, M. E. (2009). *Psychological testing and assessment: A introduction to tests and measurement* (7th ed.). McGraw-Hill Companies.
- Cook, L. L., & Eignor, D. R. (1991). IRT equating methods. *Educational measurement: Issues and practice*, 10(3), 37-45. doi:10.1111/j.1745-3992.1991.tb00207.x
- Cook, L. L., & Paterson, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11(3), 225-244. doi:10.1177/014662168701100302
- Cortina, J. M., & Landis, R. S. (2011). The earth is not round (p $\frac{1}{4}$.00). *Organizational Research Methods*, 14(2), 332-349. doi:10.1177/1094428110391542
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.
- Cui, Z., & Kolen, M. J. (2009). Evaluation of two new smoothing methods in equating: The cubic b-spline presmoothing method and the direct presmoothing method. *Journal of Educational Measurement*, 46(2), 135-158. doi:10.1111/j.1745-3984.2009.00074.x
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Publications.
- DeMars, C. (2016). *Madde tepki kuramı* (H. Kelecioğlu, Çev.). Ankara: Nobel Akademi.
- Deniz, K. Z. ve Çelebi, S. (2022). *Herkes İçin Bilimsel Araştırmakolay*. Ankara: Nobel Akademi.
- Dorans, N. J., & Holland, P. W. (2000). Population in variance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37(4), 281-306. doi:10.1111/j.1745-3984.2000.tb01088.x
- Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College and University*, 73(2), 24-34.
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2009). Equating test scores: Toward best practices. In A. A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 21-42). New York, NY: Springer Science.

- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, 2010(2), i-41. doi:10.1002/j.2333-8504.2010.tb02236.x
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Springer Science.
- Emaikwu, S. O. (2014). Aptitude testing in the 21st century in Nigeria: The basic issues to rivet. *Research Journal in Organizational Psychology & Educational Studies*, 3(3), 145-151.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferguson, T. S. (1973). A Bayesian analysis of some non parametric problems. *The Annals of Statistics*, 1(2), 209-230.
- Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In M. H. Rizvi, J. Rustagi & D. Siegmund (Eds.), *Recent advances in statistics: Papers in honor of Herman Chernoff on his sixtieth birthday*. Bibliohound (pp. 287-302). New York: Academic Press.
- Flanagan, J. C. (1951). Units, scores, and norms. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 695–763). Washington, DC: American Council on Education.
- Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, 3(3), 445-450. doi:10.1214/08-BA318
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton: CRC Press.
- Gonzalez, J. (2014). SNSequate: standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software*, 59(7), 1-30. doi:10.18637/jss.v059.i07
- Gonzalez, J., Barrientos, A. F., & Quintana, F. A. (2015a). Bayesian nonparametric estimation of test equating functions with covariates. *Computational Statistics & Data Analysis*, 89, 222-244. doi:10.1016/j.csda.2015.03.012
- Gonzalez, J., Barrientos, A. F., & Quintana, F. A. (2015b). A dependent Bayesian nonparametric model for test equating. In R. E. Milsap, D. M. Bolt, L. A. van der Ark, & WC Wang (Eds.), *Quantitative psychology research* (pp. 213-226). Springer International Publishing.
- Gonzalez, J., & von Davier, M. (2013). Statistical models and inference for the true equating transformation in the context of local equating. *Journal of Educational Measurement*, 50(3), 315-320. doi:10.1111/jedm.12018
- Gonzalez, J., & Wiberg, M. (2017). *Applying test equating methods*. Cham: Springer International Publishing.

- Grant, M. C., Zhang, L., & Damiano, M. (2009). An evaluation of Kernel equating: Parallel equating with classical methods in the SAT subject tests™ program. *ETS Research Report Series*, 2009(1), i-25. doi:10.1002/j.2333-8504.2009.tb02163.x
- Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods*, 13(4), 620-643. doi:10.1177/1094428109356114
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42(2), 111–132. doi:10.1177/0022219408326207
- Gündüz, T. (2015). *Test Eşitlemede Madde Tepki Kuramına Dayalı Yetenek Parametresine Yönelik Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Haberman, S. J. (2015). Pseudo-equivalent groups and linking. *Journal of Educational and Behavioral Statistics*, 40(3), 254–273. doi:10.3102/1076998615574772
- Hagge, S. L., Liu, C., He, Y., Powers, S. J., Wang, W., & Kolen, M. J. (2011). A comparison of IRT and traditional equipercentile methods in mixed-format equating. In M. J. Kolen & WC Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating* (pp. 19-50). Iowa: CASMA
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory, principles and applications*. New York: Springer Science.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3–24. doi:10.1177/0146621602026001001
- Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comparison of presmoothing and postsmoothing methods in equipercentile equating* (ACT Research Report 94-4). New York: American College Testing Program.
- Hogan, T. P. (2019). *Psychological testing: A practical introduction*. Hoboken, NJ: John Wiley & Sons.
- Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5-30). New York, NY: Springer New York.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (pp. 187-220). Westport, CT: Praeger.
- Holland, P. W., Dorans, N. J., & Petersen, N. S. (2007). 6 Equating test scores. *Handbook of statistics*, 26, 169-203. doi:10.1016/S0169-7161(06)26006-1

- Holland, P. W., Sinharay, S., von Davier, A. A., & Han, N. (2008). An approach to evaluating the missing data assumptions of the chain and post-stratification equating methods for the NEAT design. *Journal of Educational Measurement*, 45(1), 17-43. doi:10.1111/j.1745-3984.2007.00050.x
- Holland, P. W., & Thayer, D. T. (1987). Notes on the use of log-linear models for fitting discrete probability distributions. *ETS Research Report Series*, 1987(2), i-40. doi:10.1002/j.2330-8516.1987.tb00235.x
- Holland, P. W., & Thayer, D. T. (1989). The Kernel method of equating score distributions. *ETS Research Report Series*, 1989(1), i-45. doi:10.1002/j.2330-8516.1989.tb00333.x
- Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, 25(2), 133-183. doi:10.3102/10769986025002133
- Holland, P. W., von Davier, A., Sinharay, S., & Han, N. (2006). Testing the untestable assumptions of the chain and poststratification equating methods for the NEAT design. *ETS Research Report Series*, 2006(1), i-38. doi:10.1002/j.2333-8504.2006.tb02023.x
- Joint Committee on Standards for Educational Evaluation (2003). *The student evaluation standards: How to improve evaluations of students*. Thousand Oaks, CA: Author.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood: Scientific Software International.
- Kaplan D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673). New York, NY: The Guilford Press.
- Karabatsos, G., & Walker, S. (2009). A Bayesian nonparametric approach to test equating. *Psychometrika*, 74(2), 211-232. doi:10.1007 / S11336-008-9096-6
- Karabatsos, G., & Walker, S. (2011). A Bayesian nonparametric approach to test equating. In A. A. von Davier (Ed.) *Statistical Models for Test Equating, Scaling, and Linking* (pp. 175-184). New York, NY: Springer Sciences.
- Karadavut, T. (2021). Bağımsız Korelasyon Katsayıları için Hipotez Testleri: SPSS ve Microsoft Excel Uygulamaları. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), 375-389. doi:10.17240/aibuefd.2021.21.62826-810642
- Karmel, L. J., & Karmel, M. O. (1978). *Measurement and evaluation in the schools*. New York, NY: Macmillan Publishing Co., Inc.
- Katz, L. J., & Brown, F. C. (2019). Aptitude and achievement testing. In G. Goldstein, D. N. Allen & J. Deluca (Eds.), *Handbook of psychological assessment* (pp. 143-168). Academic Press.

- Kazanç, E. A. (2019). *Şans Başarısının Test Eşitlemeye Etkisinin Farklı Eşitleme Teknikleriyle Araştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kıran, B. (2020). Yetenek Testleri. C. Şahin (Ed.), *Bireyi Tanıma Teknikleri* (ss. 45-59). Ankara: Pegem Akademi.
- Kilmen, S. (2010). *Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinden Kestirilen Eşitleme Hatalarının Örneklem Büyüklüğü ve Yetenek Dağılımı Değişkenlerine Göre Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22(2), 131-143. doi:10.1177/01466216980222003
- Kim, S., Livingston, S. A., & Lewis, C. (2009, April). *Effectiveness of collateral information for improving equating in small samples*. American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), San Diego.
- Kim, S., Livingston, S. A., & Lewis, C. (2011). Collateral information for equating in small samples: A preliminary investigation. *Applied Measurement in Education*, 24(4), 302-323. doi:10.1080/08957347.2011.607057
- Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice*, 7(4), 29-37. doi:10.1111/j.1745-3992.1988.tb00843.x
- Kolen, M. J. (2004). Linking assessments: Concept and history. *Applied Psychological Measurement*, 28(4), 219-226. doi:10.1177/0146621604265030
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14(7), 293-300. doi:10.1016/j.tics.2010.05.001
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 14(2), 573-603. doi:10.1037/a0029146
- Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4), 722-752. doi:10.1177/1094428112457829
- Kuzgun, Y. (2017). Zeka ve Yetenekler. Y. Kuzgun ve D. Deryakulu (Edtörler), *Eğitimde Bireysel Farklılıklar* (ss. 13-70). Ankara: Nobel akademi.
- Lance, C. E. (2011). More statistical and methodological myths and urban legends. *Organizational Research Methods*, 14(2), 279-286. doi:10.1177/1094428110391814

- Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods*, 13(4), 644-667. doi:10.1177/1094428109345156
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83-102. doi:10.1207/s15324818ame0601_5
- Liou, M., Cheng, P. E., & Johnson, E. G. (1997). Standard errors of the Kernel equating methods under the common-item design. *Applied Psychological Measurement*, 21(4), 349-369. doi:10.1177/01466216970214005
- Liou, M., Cheng, P. E., & Li, M. (2001). Estimating comparable scores using surrogate variables. *Applied Psychological Measurement*, 25(2), 197-207. doi:10.1177/01466210122032000
- Lissitz, R. W., & Samuels, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437-448. doi:10.3102/0013189X07311286
- Liu, C., & Kolen, M. J. (2011). A comparison among IRT equating methods and traditional equating methods for mixed-format tests. In M. J. Kolen & WC Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating*, (pp. 75-94). Iowa: CASMA.
- Liu, J., & Low, A. C. (2008). A comparison of the Kernel equating method with traditional equating methods using SAT® data. *Journal of Educational Measurement*, 45(4), 309-323. doi:10.1111/j.1745-3984.2008.00067.x
- Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement*, 30(1), 23-39. doi:10.1111/j.1745-3984.1993.tb00420.x
- Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd ed.). Educational testing service.
- Livingston, S. A., & Kim, S. (2010). Random-groups equating with samples of 50 to 400 test takers. *Journal of Educational Measurement*, 47(2), 175-185. doi:10.1111/j.1745-3984.2010.00107.x
- Longford, N. T. (2015). Equating without an anchor for nonequivalent groups of examinees. *Journal of Educational and Behavioral Statistics*, 40(3), 227- 253. doi:10.3102/1076998615574773
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed score "equatings". *Applied Psychological Measurement*, 8(4), 453-461. doi:10.1177/014662168400800409
- Lorentz, G. G. (1986). *Bernstein Polynomials*. Chelsea Publishing Company.

- Lu, R., & Guo, H. (2018). A simulation study to compare nonequivalent groups with anchor test equating and pseudo-equivalent group linking. *ETS Research Report Series*, 2018(1), 1-16. doi:10.1002/ets2.12196
- MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science* (pp. 50-55).
- Mao, X. (2006). *An investigation of the accuracy of the estimates of standard errors for the Kernel equating functions* (Unpublished doctoral dissertation). University of Iowa, Department of Educational Measurement and Statistics, Iowa.
- Mao, X., von Davier, A. A., & Rupp, S. (2006). Comparisons of the Kernel equating method with the traditional equating methods on PRAXISTM data. *ETS Research Report Series*, 2006(2), i-31. doi:10.1002/j.2333-8504.2006.tb02036.x
- Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects* (Policy Information Rep.). Princeton, NJ: ETS.
- Moses, T., & Holland, P. (2007). Kernel and traditional equipercentile equating with degrees of presmoothing. *ETS Research Report Series*, 2007(1), 1-39. doi:10.1002/j.2333-8504.2007.tb02057.x
- Moses, T., & Holland, P. W. (2010). A comparison of statistical selection strategies for univariate and bivariate log-linear models. *British Journal of Mathematical and Statistical Psychology*, 63(3), 557-574. doi:10.1348/000711009X478580
- Muchinsky, P. M. (2009). Aptitude testing. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (Vol. 4). John Wiley & Sons. doi:10.1002/9780470479216.corpsy0077
- Müller, P., & Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19(1), 95-110. doi:10.1214/088342304000000017
- Oh, H. J., Guo, H., & Walker, M. E. (2009). Improved reliability estimates for small samples using empirical Bayes techniques. *ETS Research Report Series*, 2009(2), i-23. doi:10.1002/j.2333-8504.2009.tb02203.x
- Öner, N. (1997). *Türkiye’de Kullanılan Psikolojik Testler*. İstanbul: Boğaziçi Üniversitesi Yayınları.
- Özçelik, D. A. (2016). *Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Özdemir, B. (2017). Equating TIMSS mathematics subtests with nonlinear equating methods using neat design: Circle-arc equating approaches. *International Journal of Progressive Education*, 13(2), 116-132. <https://files.eric.ed.gov/fulltext/EJ1145605.pdf>
- Özgüven, İ. E. (2004). *Psikolojik Testler*. Ankara: Pdrem Yayınları.

- Özsoy, Ş. N. (2021). *Kernel Eşitleme Yöntemlerinin Karşılaştırılması: TIMSS 2019 Fen Testi Örneği*. (Yayımlanmamış Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü, Bolu.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage
- Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26(3), 373-393.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2006). Branded by a test. *Educational Leadership*, 63(7), 86-87.
- Press, S. (2014). Aptitude testing. In S. Press (Ed.), *Evaluating aptitude and progress* (pp. 52-58).
- Rapp, J. (1999). Linear and equipercentile methods for equating PET. *National Institute for Testing and Evaluation Report*, 1-22. <https://nite.org.il/files/reports/e266.pdf>
- Ricker, K. L., & von Davier, A. A. (2007). The impact of anchor test length on equating results in a nonequivalent groups design. *ETS Research Report Series*, 2007(2), i-19. doi:10.1002/j.2333-8504.2007.tb02086.x
- Schultz, D. P., & Schultz, S. E. (2007). Modern Psikoloji Tarihi (Y. Aslay, Çev.). *İstanbul: Kaknüs Yayınları*.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2), 639-650. <https://people.eecs.berkeley.edu/~jordan/sail/readings/archive/sethuraman.pdf>
- Sezer Başaran, E. (2023). Farklı Ortak Değişkenlerle Test Eşitlemenin Ortak Maddeli Test Eşitlemeyle Karşılaştırılması (Yayımlanmamış Doktora Tezi). Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Sezer Başaran, E., Mutluer, C. ve Çakan, M. (2023). A comparison of covariates, equating designs, and methods in equating TIMSS 2019 science tests. *Participatory Educational Research*, 10(5), 41-63. doi:10.17275/per.23.74.10.5
- Shevlin, M., & Miles, J. N. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, 25(1), 85-90. doi:10.1016/S0191-8869(98)00055-5
- Sinharay, S., & Holland, P. W. (2007). Is it necessary to make anchor tests mini-versions of the tests being equated or can some restrictions be relaxed? *Journal of Educational Measurement*, 44(3), 249-275. doi:10.1111/j.1745-3984.2007.00037.x

- Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement*, 42(4), 309-330. doi:10.1111/j.1745-3984.2005.00018.x
- Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational psychologist*, 27(1), 5-32. doi:10.1207/s15326985ep2701_3
- Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, 2(3-4), 536-560. doi:10.1037/1076-8971.2.3-4.536
- Stemler, S. E., & Sternberg, R. J. (2013). The assessment of aptitude. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology: Vol. 3. Testing and assessment in school psychology and education* (pp. 281-296). Washington, DC: American Psychological Association.
- Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment*, 15(1), 3–13. doi:10.1027/1015-5759.15.1.3
- Sternberg, R. J., & the Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34(4), 321–350. doi:10.1016/j.intell.2006.01.002
- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Sümer, N. (2000). Yapısal Eşitlik Modelleri. *Türk Psikoloji Yazıları*, 3(6), 49-74. http://www.nebisumer.com/wp-content/uploads/2015/03/SumerN.2000.YEM_TPY.pdf
- Tabachnick, B. G., & Fidell, L. S. (2020). *Çok Değişkenli İstatistiklerin Kullanımı* (M. Baloğlu, Çev.). Ankara: Nobel akademik Yayıncılık.
- Tan, Ş. (2015). Küçük Örneklemlerde Beta4 ve Polynomial Loglineer Ön Düzgünleştirme ve Kübik Eğri Son Düzgünleştirme Metotlarının Uygunluğu. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 35(1), 123-151. <https://dergipark.org.tr/en/download/article-file/77508>
- Thorndike, E. L. (1922). On finding equivalent scores in tests of intelligence. *Journal of Applied Psychology*, 6(1), 29–33. doi:10.1037/h0072880
- Thorndike, R. M. ve Thorndike-Christ, T. (2017). *Psikolojide ve Eğitimde Ölçme ve Değerlendirme* (M. Otrar, Çev.). Ankara: Nobel Yayın Dağıtım.
- Turgut, M. F. ve Baykul, Y. (2019). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.

- Uysal, İ. (2014). *Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinin Karma Modeller Üzerinde Karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü, Bolu.
- van der Linden, W. J. (2006). Book review of *The Kernel Method of Test Equating* by A. A. von Davier, P. W. Holland, & D. T. Thayer. *Journal of Educational Measurement*, 43(3), 291-294.
- van der Linden, W. J. (2010). Local observed-score equating. In A. A. von Davier (Ed.), *Statistical Models for Equating, Scaling, and Linking* (pp. 201-223). New York, NY: Springer.
- van der Linden, W. J., & Wiberg, M. (2010). Local observed-score equating with anchor-test designs. *Applied Psychological Measurement*, 34(8), 620-640. doi:10.1177/0146621609349803
- van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, 16(2), 75-84.
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken M. A. G. (2013). A gentle introduction to bayesian analysis: applications to developmental research. *Child Development*, 85(3), 842-860. doi:10.1111/cdev.12169
- Varas, I. M., Gonzalez, J., & Quintana, F. A. (2020). A Bayesian nonparametric latent approach for score distributions in test equating. *Journal of Educational and Behavioral Statistics*, 45(6), 639-666. doi:10.3102/1076998620907381
- von Davier, A. A. (Ed.). (2011). *Statistical models for test equating, scaling, and linking*. New York, NY: Springer Sciences.
- von Davier, A. A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). An evaluation of the Kernel equating method: A special study with pseudo tests constructed from real test data. *ETS Research Report Series*, 2006(1), i-31. doi:10.1002/j.2333-8504.2006.tb02008.x
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel method of test equating*. New York, NY: Springer.
- Wallin, G., & Wiberg, M. (2017). Non-equivalent groups with covariates design using propensity scores for Kernel equating. In L. A. van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas & WC Wang (Eds), *Quantitative Psychology – 81st annual meeting of the psychometric society, Asheville, North Carolina, 2016* (pp. 309-319). Cham: Springer International Publishing.
- Wang, T., Lee, W. C., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 32(8), 632-651. doi:10.1177/0146621608314943

- Wiberg, M. (2015). A note on equating test scores with covariates. In E.Fackle-Fornius (Ed.), *Festschrift in Honor of Hans Nyquist on the Occasion of his 65th Birthday* (s. 96-99). Stockholm: Stockholm University, Department of Statistics.
- Wiberg, M. (2021) On the use of different linkage plans with different observed-score equipercentile equating methods. *Practical Assessment, Research, and Evaluation*, 26(23), 1-16. doi:10.7275/21481778
- Wiberg, M., & Banberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, 39(5), 349–361. doi:10.1177/0146621614567939
- Wiberg, M., & von Davier, A. A. (2017). Examining the impact of covariates on anchor tests to ascertain quality over time in a college admissions test. *International Journal of Testing*, 17(2), 105-126. doi:10.1080/15305058.2016.1277357
- Woldbeck, T. (1998, April). *Basic concepts in modern methods of test equating*. Paper presented at the Annual Meeting of the Southwest Psychological Association, New Orleans. <https://files.eric.ed.gov/fulltext/ED417215.pdf>
- Yıldırım Seheryeli, M. (2022). *Ortak Değişkenlerin Özelliklerine Göre Kernel Eşitleme Sonuçlarının Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Yurtçu, M. (2018). *Parametrik Olmayan Bayes Yöntemiyle Ortak Değişkenlere Göre Yapılan Test Eşitlemelerinin Karşılaştırılması* (Tez No. 516810) [Doktora Tezi, Hacettepe Üniversitesi]. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Yurtçu, M., Kelecioğlu, H., & Boone, E. (2021). The comparison of the equated tests scores by using various covariates using Bayesian nonparametric model. *Journal of Measurement and Evaluation in Education and Psychology*, 12(2), 192-211. doi:10.21031/epod.864744
- Zhang, Z., Hamagami, F., Lijuan Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31(4), 374-383. doi:10.1177/0165025407077764
- Zhu, W. (1998). Test equating: what, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23. doi:10.1080/02701367.1998.10607662



EK 1. Etik Kurul Onayı

ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 27/06/2022
Toplantı Sayısı : 13
Karar Sayısı : 179

179-Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı doktora öğrencilerinden **Şerife Zeybekoğlu**'nun "Yarı Parametrik ve Parametrik Olmayan Eşitleme Yöntemleri ile Bir Yetenek Testinin Farklı Formlarının Eşitlenmesi" başlıklı tezi ile ilgili "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı doktora öğrencilerinden **Şerife Zeybekoğlu**'nun "Yarı Parametrik ve Parametrik Olmayan Eşitleme Yöntemleri ile Bir Yetenek Testinin Farklı Formlarının Eşitlenmesi" başlıklı tezinin araştırma protokolüne uyulması ve etik onay tarihinden itibaren geçerli olması koşuluyla uygulanmasının etik açıdan uygun olduğuna oy birliği ile karar verildi.

ASLININ AYNIDIR
27/06/2022

Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

EK 2. Antalya İl Milli Eğitim Müdürlüğü Araştırma İzni



T.C.
ANTALYA VALİLİĞİ
İl Milli Eğitim Müdürlüğü

Sayı : E-98057890-20-58202169
Konu : Anket Uygulaması

19.09.2022

İL MİLLİ EĞİTİM MÜDÜRLÜĞÜNE ANTALYA

İlgi : 21/01/2020 tarih ve 1563890 sayılı Millî Eğitim Bakanlığına Bağlı Okul ve Kurumlarda Yapılacak Araştırma, Yarışma ve Sosyal Etkinlik İzinlerine Yönelik İzin ve Uygulama Genelgesi.

Ankara Üniversitesi Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Doktora Programı öğrencisi Şerife ZEYBEKOĞLU Doç. Dr. Kaan Zülfiyar DENİZ danışmanlığında yürüttüğü "Yarı Parametrik ve Parametrik Olmayan Eşitleme Yöntemleri ile Bir Yetenek Testinin Farklı Formlarının Eşitlemesi" isimli araştırmasını, ilimiz merkez ilçelerindeki resmi ve özel ortaöğretim kurumlarında öğrenim gören 9. sınıf öğrencilerine yönelik uygulama isteği ile ilgili Ankara Üniversitesinin 01/08/2022 tarih ve 603475 sayılı yazısı Müdürlüğümüz Ar-Ge Birimi Değerlendirme ve İnceleme Komisyonunca incelenmiş olup;

Adı geçenin ilgi Genelge kapsamında 2022-2023 Eğitim Öğretim Yılı içerisinde olmak üzere, ilimiz merkez ilçelerindeki resmi ve özel ortaöğretim kurumlarında öğrenim gören 9. sınıf öğrencilerine yönelik araştırmasını, Okul Müdürlüklerinin sorumluluğunda Eğitim Öğretim faaliyetlerini aksatmaksızın yapması,

Söz konusu araştırmanın bitimine müteakip; sonuç raporunun bir örneğinin CD ortamında Müdürlüğümüz Ar-Ge bürosuna gönderilmesi kaydıyla uygulanması, Komisyonca uygun görülmüştür.

Makamlarınızca da uygun görüldüğü takdirde, Valilik Makamının 25/08/2020 tarih ve 24911 sayılı yetki devrine göre olurlarınıza arz ederim.

Fatma Zeynep ŞERAN
Müdür a.
Müdür Yardımcısı

OLUR

Emre ÇALIŞKAN
Vali a.
İl Milli Eğitim Müdürü

Bu belge güvenli elektronik imza ile imzalanmıştır.

Adres : Soğuksu Mah. Haridize Cad. No:59 Muratpaşa/Antalya

Belge Doğrulama Adresi : <https://www.turkiye.gov.tr/meh-alyys>

Tel/Fax No : 0 (242) 238 60 00

E-Posta : projeler07@meh.gov.tr

Kap Adresi : meh@hs01.kcp.tr

İnternet Adresi : <http://antalya.meh.gov.tr/>

Bilgi için: Etilal EK

Unvan : Teknisyen

Faksa:2422386111

Bu evrak güvenli elektronik imza ile imzalanmıştır. <https://evrakorgu.meh.gov.tr> adresinden: 8af2-b881-3fd7-a17b-8496 kodu ile teyit edilebilir.



BENZERLİK BİLDİRİMİ

“Yarı Parametrik ve Parametrik Olmayan Eşitleme Yöntemleri ile Bir Yetenek Testinin Farklı Formlarının Eşitlenmesi” başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihal Tespit Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) “Yedi sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan alıntılar” ve (4) “Lisansüstü tezim ile ilgili yapmış olduğum kendi yayınlarım ve yararlandığım mevzuat metinleri gibi kaynaklar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 20/07/2024
Gönderim Numarası	: 2370133291
Sayfa Sayısı	: 162
Sözcük Sayısı	: 34856
Karakter Sayısı	: 237591
Benzerlik Oranı	: %8
Savunma Tarihi	: 12/06/2024

Yukarıda belirtilen sonuçları gösteren Turnitin İntihal Tespit Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin, tek bir kaynakla eşleşme oranının ise %2’den fazla olduğunun belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Şerife ZEYBEKOĞLU

Tarih: 20/07/2024

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Şerife ZEYBEKOĞLU

E-Posta Adresi :

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Öğretmen	MEB – Alaylı Ortaokulu	2012-2015
Öğretmen	MEB – Taşlıca Ahmet Hamdi Akseki Ortaokulu	2015-2019
Öğretmen	MEB – Antalya Ölçme Değerlendirme Merkezi	2019-Devam ediyor

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Doktora	Eğitimde Ölçme ve Değerlendirme	Ankara Üniversitesi	2019-2024
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Akdeniz Üniversitesi	2015-2019
Lisans	İlköğretim Matematik Öğretmenliği	Abant İzzet Baysal Üniversitesi	2005-2009

Yayınlar:

Zeybekoğlu, Ş., & Koğar, H. (2022). Investigation of variables explaining science literacy in PISA 2015 Turkey sample. *Journal of Measurement and Evaluation in Education and Psychology*, 13(2), 145-163. doi: 10.21031/epod.798106

Zeybekoğlu, Ş., Bilicioğlu Güneş, A., & Yalçın, E. (2024). Examination of the differential item functioning in the PISA 2018 reading test items. *Eğitim ve Bilim*, 49(217), 201-223. doi: 10.15390/EB.2023.12123