



**AÇIK UÇLU SORULARIN PUANLANMASINDA PUANLAYICI
EĞİTİMİNİN PUANLAYICI DAVRANIŞLARINA ETKİSİNİN
İNCELENMESİ**

Umur Öç

DOKTORA TEZİ
EĞİTİM BİLİMLERİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

AĞUSTOS, 2024

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren on iki (12) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Umur
Soyadı : Öç
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim Tarihi :

TEZİN

Türkçe Adı : Açık Uçlu Soruların Puanlanmasında Puanlayıcı Eğitiminin
Puanlayıcı Davranışlarına Etkisinin İncelenmesi

İngilizce Adı : Examining the Effect of Rater Training on Rater Behaviors in
Scoring Open-Ended Questions

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğim ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı : Umur Öç

İmza :

JÜRİ ONAY SAYFASI

Umur ÖÇ tarafından hazırlanan “Açık uçlu soruların puanlanmasında puanlayıcı eğitiminin puanlayıcı davranışlarına etkisinin incelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Ana Bilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. İsmail KARAKAYA

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

.....

Başkan: Prof. Dr. Nuri DOĞAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

.....

Üye: Prof. Dr. Nilüfer KAHRAMAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

.....

Üye: Doç. Dr. Melek Gülşah ŞAHİN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

.....

Üye: Doç. Dr. Celal Deha DOĞAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

.....

Tez Savunma Tarihi: 30/07/2024

Bu tezin Eğitim Bilimleri Anabilim Dalı’nda Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban ÇETİN

.....

Eğitim Bilimleri Enstitüsü Müdürü



*Yaşadığı zulme dur
diyemediğim nice dev
yüreklerle ve eşime...*

TEŞEKKÜR

Doktora eğitimime başladığım andan itibaren her anlamda bana desteğini esirgemeyen, her zaman bilgi ve tecrübelerini benimle paylaşan, kişiliğiyle, duruşuyla kendime model olarak gördüğüm, öğrencisi olmaktan gurur duyduğum çok değerli danışmanım Sayın Prof. Dr. İsmail KARAKAYA'ya, tez izleme komitemde yer alan tezimin planlanması ve yürütülmesi sürecinde görüş ve önerileri ile tezime katkı sağlayan kıymetli hocalarım Prof. Dr. Nilüfer KAHRAMAN ve Prof. Dr. Nuri DOĞAN'a, Tez savunmamda yer alarak enerjileriyle, güler yüzüyle değerli eleştirilerini ve fikirlerini sunarak tezime katkı sağlayan saygıdeğer hocalarım Doç. Dr. Melek Gülşah ŞAHİN'e ve Doç. Dr. Celal Deha DOĞAN'a teşekkür ederim.

Araştırmamın veri toplama sürecinde desteklerini esirgemeyen Cumhuriyet Üniversitesi akademisyen ve öğrencilerine teşekkür ederim.

Doktora eğitimim sürecinde kazandığım ve iyi ki tanımışım dediğim, emeklerini hiçbir zaman esirgemeyen kıymetli kardeşlerim Tayfun EROĞLU ve Nazira TURSUNBAYEVA'ya teşekkür ederim.

Araştırma sürecinde ihtiyaç duyduğumda bana desteğini esirgemeyen Nurcan CEYLAN'a, Ensar YILDIZ'a ve Özgür ÇANKAYA'ya teşekkür ederim.

Eğitim sürecinde bana destek olan anneme, babama ve kardeşlerime teşekkür ederim.

Kısa bir teşekkürün yetersiz kalacağı, doktora eğitimime başlamamda ve tamamlamamda bana hep destek olan, onca sıkıntıya rağmen hep yanımda olan kıymetli eşime, can yoldaşıma teşekkür ederim. Ben çalışsam da babam kesin ders çalışıyordur diye odama hep sessizce gelen oğlum İlteriş Eymen'e ve dünyaya gelmek üzere olan kızıma teşekkür ederim.

AÇIK UÇLU SORULARIN PUANLANMASINDA PUANLAYICI EĞİTİMİNİN PUANLAYICI DAVRANIŞLARINA ETKİSİNİN İNCELENMESİ

(Doktora Tezi)

Umur Öç

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Ağustos 2024

ÖZ

Yapılan araştırma, açık uçlu soruların puanlanmasında puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisini incelemeyi amaçlamaktadır. Bu amaç doğrultusunda araştırmada denk olmayan kontrol gruplu desen kullanılmıştır. Araştırmanın çalışma grubu, 14 Türkçe öğretmenliği, 20 matematik öğretmenliği bölümünde öğrenimine devam eden puanlayıcılardan oluşmaktadır. Türkçe öğretmenliği bölümünde öğrenimine devam eden puanlayıcılardan yedi tanesi deney, yedi tanesi kontrol grubunda yer alırken matematik öğretmenliği bölümünde öğrenimine devam eden puanlayıcılardan on tanesi deney, on tanesi de kontrol grubunda yer almaktadır. Araştırma kapsamında hem Türkçe hem de matematik puanlayıcı grubundaki puanlayıcıların puanlayıcı katılığı, puanlayıcı cömertliği, merkeze eğilim, halo etkisi ve farklılaşan puanlayıcı katılığı ve farklılaşan puanlayıcı cömertliği olmak üzere altı çeşit puanlayıcı davranışı incelenmiştir. Puanlayıcı davranışları hem grup düzeyinde hem de bireysel düzeyde incelenmiştir. Araştırmada veri toplama aracı olarak araştırmacı tarafından geliştirilen ortaokul Türkçe ve Matematik dersleri kazanımlarını kapsayan beş Türkçe ve altı Matematik olmak üzere toplam 11 açık uçlu maddeden oluşan başarı testi kullanılmıştır. Ayrıca araştırmacı tarafından geliştirilen bütünsel dereceli puanlama anahtarı ve puanlayıcı değerlendirme formu kullanılmıştır. Araştırmada elde edilen veriler, çok yüzeyli rasch modeli, Mann-Whitney U testi ve Wilcoxon işaretli sıralar testi kullanılarak analiz edilmiştir. Çok yüzeyli rasch modeli kapsamında yapılan analizlerde tümüyle çaprazlanmış desen kullanılmıştır. Araştırmada deneysel işlemde hem Türkçe hem de matematik puanlayıcı grubuna puanlayıcı eğitimi verilmiştir. Puanlayıcılara puanlayıcı eğitimi verilirken puanlayıcı eğitim stratejilerinden puanlayıcı hatası eğitim stratejisi ve

referans çerçevesi eğitim stratejileri bir arada sunulmuştur. Sadece deney grubuna uygulanan deneysel işlem haftada bir saat olmak üzere dört hafta sürmüştür. Puanlayıcı eğitimi öncesinde hem Türkçe hem de matematik deney ve kontrol gruplarının puanlayıcı davranışları açısından benzer olduğu belirlenmiştir. Deneysel işlem öncesinde Türkçe puanlayıcı deney ve kontrol grubunda puanlayıcı katılımı ve puanlayıcı cömertliği davranışlarının hem grup hem de bireysel düzeyde ortaya çıktığı, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılımı ve farklılaşan puanlayıcı cömertliği davranışlarının ise sadece bireysel düzeyde ortaya çıktığı belirlenmiştir. Deneysel işlem öncesi Matematik puanlayıcı deney ve kontrol grubunda da puanlayıcı katılımı ve puanlayıcı cömertliği davranışlarının hem grup hem de bireysel düzeyde ortaya çıktığı, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılımı ve farklılaşan puanlayıcı cömertliği davranışlarının ise sadece bireysel düzeyde ortaya çıktığı belirlenmiştir. Puanlayıcı eğitiminden sonra Türkçe ve matematik puanlayıcı gruplarının davranışları incelendiğinde deney grubunda yer alan puanlayıcıların puanlayıcı davranışları tamamen ortadan kalkmasa da puanlayıcı davranışlarında azalma olduğu görülmüştür. Hem Türkçe hem de matematik puanlayıcı deney gruplarının puanlayıcı eğitimi öncesinde ve sonrasında yapılan karşılaştırmalarında istatistiksel olarak fark olduğu bulunmuştur. Türkçe puanlayıcı grubundaki puanlayıcıların katılım ve cömertlik durumlarında meydana gelen değişimin %60,2'sinin, matematik puanlayıcı grubundaki puanlayıcıların katılım ve cömertlik durumlarında meydana gelen değişimin de %57,8'inin puanlayıcı eğitimi ile açıklanabildiği bulunmuştur. Ayrıca puanlayıcı eğitiminin matematik puanlayıcı grubunda yer alan puanlayıcıların davranışlarında daha fazla azalmayı sağladığı ve dolayısıyla daha etkili olduğu belirlenmiştir. Araştırma sonuçlarından hareketle açık uçlu soruların puanlanmasında puanlama işlemi öncesinde puanlayıcılara puanlayıcı eğitimi verilmesi önerilmektedir.

Anahtar Kelimeler : Açık uçlu sorular, puanlayıcı eğitimi, puanlayıcı davranışları, geçerlik, güvenilirlik, çok yüzeyli rasch modeli
Sayfa Adedi : xxi+ 260
Danışman : Prof. Dr. İsmail KARAKAYA

EXAMINING THE EFFECT OF RATER TRAINING ON RATER BEHAVIORS IN SCORING OPEN-ENDED QUESTIONS

(PH. D THESIS)

Umur Öc

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

August 2024

ABSTRACT

The study aims to examine the effect of scorer training on scorer behaviors in the scoring of open-ended questions. To this end, a non-equivalent control group design was employed in the research. The study group consisted of scorers currently enrolled in 14 Turkish language teaching and 20 mathematics teaching programs. Seven scorers from the Turkish language teaching program were assigned to the experimental group and seven to the control group, while in the mathematics teaching program, ten scorers were in the experimental group and ten in the control group. In the research, six types of scorer behaviors were examined across both Turkish and mathematics scorer groups: scorer severity, scorer leniency, central tendency, halo effect, and differential scorer severity and leniency, analyzed at both group and individual levels. The research employed an achievement test consisting of 11 open-ended items, five in Turkish and six in mathematics, covering middle school Turkish and Mathematics curriculum outcomes, developed by the researcher as the data collection instrument. Additionally, a holistic graded scoring rubric and scorer evaluation form developed by the researcher were utilized. The data obtained in the research were analyzed using the Many-Facets Rasch Model, Mann-Whitney U test and Wilcoxon test. The Many-Facets Rasch Model analysis utilized a fully crossed design. In the experimental procedure of the study, scorer training was provided to both Turkish and mathematics scorer groups. Scorer training included scorer error training and reference frame training strategies presented together. The experimental procedure applied only to the experimental group, lasting one hour per week for four weeks. Prior to scorer training, it was determined that Turkish and mathematics experimental and control groups were similar in terms of scorer behaviors. Before the experimental procedure, scorer severity and leniency behaviors emerged at both group and individual levels in the Turkish scorer experimental and control

groups, while central tendency, halo effect, differential scorer severity, and leniency behaviors emerged only at the individual level. Similarly, in the mathematics scorer groups before the experimental procedure, scorer severity and leniency behaviors emerged at both group and individual levels, while central tendency, halo effect, differential scorer severity, and leniency behaviors emerged only at the individual level. After scorer training, when examining scorer behaviors of Turkish and mathematics scorer groups, a reduction in scorer behaviors was observed in the experimental group, although scorer behaviors did not completely disappear. Statistical differences were found in comparisons before and after scorer training in both Turkish and mathematics scorer experimental groups. It was found that 60.2% of the change in severity and leniency behaviors of Turkish scorers and 57.8% of the change in severity and leniency behaviors of mathematics scorers could be explained by scorer training. Additionally, it was determined that scorer training led to a greater reduction in behaviors among mathematics scorers, indicating its greater effectiveness. Based on the research findings, it is recommended that scorer training be provided to scorers before the scoring process of open-ended questions.

Key Words : Open-ended questions, rater training, rater behaviors, validity, reliability, many facet rasch model
Page Number : xxi+260
Supervisor : Prof. Dr. İsmail KARAKAYA

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI	ii
JÜRİ ONAY SAYFASI.....	iii
TEŞEKKÜR.....	v
ÖZ	vi
ABSTRACT	viii
İÇİNDEKİLER.....	x
TABLolar LİSTESİ.....	xvi
ŞEKİLLER LİSTESİ.....	xx
SİMGELER VE KISALTMALAR LİSTESİ.....	xxi
BÖLÜM I	1
GİRİŞ.....	1
Problem Durumu	1
Problem İfadesi	10
Araştırma Soruları.....	10
Araştırmanın Amacı	11
Araştırmanın Önemi.....	11
Sınırlılıklar.....	12

Tanımlar	13
BÖLÜM II.....	14
KAVRAMSAL ÇERÇEVE	14
Üst Düzey Zihinsel Beceriler	14
Üst Düzey Zihinsel Becerilerin Ölçümünde Açık Uçlu Maddeler	15
Açık Uçlu Maddelerin Puanlanması	17
Puanlayıcı Davranışları	18
Puanlayıcı Katılığı ve Puanlayıcı Cömertliği.....	21
Merkeze Eğilim.....	22
Halo Etkisi.....	24
Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Puanlayıcı Yanlılığı)....	26
Puanlayıcı Güvenirliği	27
Klasik Test Kuramı Temelli Puanlayıcı Güvenirliği Belirleme Yöntemleri.....	28
Genellenebilirlik Kuramı Temelli Puanlayıcı Güvenirliği Yöntemleri.....	29
Madde Tepki Kuramı Temelli Puanlayıcı Güvenirliği Yöntemleri	30
Çok Yüzeyle Rasch Modeli	31
<i>Çok Yüzeyle Rasch Modeli Puanlama Desenleri.....</i>	<i>42</i>
<i>Çaprazlanmış Desen.....</i>	<i>42</i>
<i>Yuvalanmış Desen.....</i>	<i>43</i>
Puanlayıcı Davranışlarını Benzer Hale Getirmeye Yönelik Uygulamalar	44
Birden Fazla Puanlayıcı Kullanılması.....	44
Dereceli Puanlama Anahtarı	45
<i>Bütünsel dereceli puanlama anahtarı</i>	<i>47</i>
<i>Analitik dereceli puanlama anahtarı.....</i>	<i>47</i>
Puanlayıcı Eğitimi	48

<i>Öz Liderlik Eğitimi</i>	<i>49</i>
<i>Davranış Gözlem Eğitimi</i>	<i>50</i>
<i>Performans Boyutu Eğitimi</i>	<i>51</i>
<i>Puanlayıcı Hatası Eğitimi</i>	<i>52</i>
<i>Referans Çerçevesi Eğitimi</i>	<i>53</i>
<i>Puanlayıcı Değişkenliği Eğitimi</i>	<i>54</i>
BÖLÜM III	56
İLGİLİ ARAŞTIRMALAR.....	56
Puanlayıcı Davranışları ile İlgili Araştırmalar	56
BÖLÜM IV	61
YÖNTEM.....	61
Araştırmanın Deseni	61
Çalışma Grubu	61
Veri Toplama Araçları	62
Başarı Testlerinin Hazırlanması	63
Dereceli Puanlama Anahtarının Hazırlanması	63
Açık Uçlu Maddelerin Geçerlik Çalışmaları	66
Bütünsel Dereceli Puanlama Anahtarının (BDPA) Geçerlik Çalışmaları.....	67
Kapsam geçerliği	67
Puanlayıcı Değerlendirme Formu.....	69
Verilerin Eldesi ve Deneysel İşlem	70
Puanlayıcı Eğitimi	71
Verilerin Analizi.....	74
Tek boyutluluk.....	74

Yerel bağımsızlık	74
Model veri uyumu	75
BÖLÜM V	77
BULGULAR VE YORUM	77
Birinci araştırma sorusuna ait bulgular	77
Türkçe puanlayıcı grubu uygunluk değerleri ve model ölçümleri	78
Matematik puanlayıcı grubu uygunluk değerleri ve model ölçümleri ..	87
İkinci araştırma sorusuna ait bulgular	98
Puanlayıcı Eğitimi Öncesinde Türkçe Puanlayıcı Grubu Puanlayıcı Davranışları	99
<i>Katılık ve Cömertlik</i>	99
<i>Merkeze Eğilim</i>	101
<i>Halo Etkisi</i>	104
<i>Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)</i>	106
Puanlayıcı Eğitimi Öncesinde Matematik Puanlayıcı Grubu Puanlayıcı Davranışları	107
<i>Katılık ve Cömertlik</i>	108
<i>Merkeze Eğilim</i>	110
<i>Halo Etkisi</i>	113
<i>Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)</i>	115
Araştırma sonuçlarını destekleyen çalışmalar	116
<i>Katılık ve Cömertlik</i>	117
<i>Merkeze Eğilim</i>	118
<i>Halo Etkisi</i>	119
<i>Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)</i>	120
Üçüncü araştırma sorusuna ait bulgular	120

Puanlayıcı Eğitimi Sonrasında Türkçe Puanlayıcı Grubu Puanlayıcı Davranışları	121
<i>Katılık ve Cömertlik</i>	121
<i>Merkeze Eğilim</i>	125
<i>Halo Etkisi</i>	128
<i>Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)</i>	129
Puanlayıcı Eğitimi Sonrasında Matematik Puanlayıcı Grubu Puanlayıcı Davranışları	132
<i>Katılık ve Cömertlik</i>	132
<i>Merkeze Eğilim</i>	136
<i>Halo Etkisi</i>	138
<i>Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)</i>	140
Araştırma sonuçlarını destekleyen çalışmalar	143
BÖLÜM VI	148
SONUÇLAR	148
Öneriler	153
Araştırmacılara Öneriler	153
Uygulamacılara Öneriler	155
KAYNAKLAR	156
EKLER	185
EK 1. Türkçe Dersi Açık Uçlu Sorular	186
EK 2. Türkçe Dersi Dereceli Puanlama Anahtarı	192
EK 3. Türkçe Dersi Puanlayıcı Değerlendirme Formu	197
EK 4. Matematik Dersi Açık Uçlu Sorular	198
EK 5. Matematik Dersi Dereceli Puanlama Anahtarı	202

EK 6. Matematik Dersi Puanlayıcı Değerlendirme Formu	208
EK 7. Ön Test Türkçe Puanlayıcı Grubu <i>Puanlayıcı X Birey</i> Etkileşimleri	209
EK 8. Son Test Türkçe Puanlayıcı Grubu <i>Puanlayıcı X Birey</i> Etkileşimleri	220
EK 9. Ön Test Matematik Puanlayıcı Grubu <i>Puanlayıcı X Birey</i> Etkileşimleri.....	231
EK 10. Son Test Matematik Puanlayıcı Grubu <i>Puanlayıcı X Birey</i> Etkileşimleri.....	246



TABLÖLAR LİSTESİ

Tablo 1 <i>Puanlayıcı Davranışları</i>	20
Tablo 2 <i>KTK Temelli Puanlayıcı Güvenirliği Belirleme Yöntemleri</i>	28
Tablo 3 <i>Klasik Test Kuramı ve Genellenebilirlik Kuramlarının Karşılaştırılması</i>	30
Tablo 4 <i>Tümüyle Çaprazlanmış Desen</i>	43
Tablo 5 <i>Yuvalanmış Desen</i>	44
Tablo 6 <i>Kapsam Geçerlik Oranı (KGO) Minimum Değerleri</i>	65
Tablo 7 <i>Türkçe Dersi Açık Uçlu Maddelere Ait KGO Değerleri</i>	66
Tablo 8 <i>Matematik Dersi Açık Uçlu Maddelere Ait KGO Değerleri</i>	67
Tablo 9 <i>Türkçe BDPA Taslak Formu Kriterlere Ait KGO Değerleri</i>	68
Tablo 10 <i>Matematik BDPA Taslak Formu Kriterlere Ait KGO Değerleri</i>	69
Tablo 11 <i>Türkçe Öğretmen Adayları Standartlaştırılmış Artık Oranları</i>	75
Tablo 12 <i>Matematik Öğretmen Adayları Standartlaştırılmış Artık Oranları</i>	75
Tablo 13 <i>Türkçe Puanlayıcı Grubu Madde Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	78
Tablo 14 <i>Türkçe Puanlayıcı Grubu Birey Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	80
Tablo 15 <i>Türkçe Puanlayıcı Grubu Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	82
Tablo 16 <i>Türkçe Puanlayıcı Grubu Madde Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	84

Tablo 17 <i>Türkçe Puanlayıcı Grubu Birey Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	85
Tablo 18 <i>Türkçe Puanlayıcı Grubu Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	86
Tablo 19 <i>Matematik Puanlayıcı Grubu Madde Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	88
Tablo 20 <i>Matematik Puanlayıcı Grubu Birey Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	90
Tablo 21 <i>Matematik Puanlayıcı Grubu Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	92
Tablo 22 <i>Matematik Puanlayıcı Grubu Madde Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	93
Tablo 23 <i>Matematik Puanlayıcı Grubu Birey Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	95
Tablo 24 <i>Matematik Puanlayıcı Grubu Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	97
Tablo 25 <i>Deney ve Kontrol Gruplarının Ön Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları</i>	101
Tablo 26 <i>Deney ve Kontrol Gruplarının Ön Test Sonuçlarına Ait Kategori İstatistikleri</i>	102
Tablo 27 <i>Deney ve Kontrol Grubu Ön Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri</i>	103
Tablo 28 <i>Türkçe Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu</i>	105
Tablo 29 <i>Türkçe Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri</i>	107
Tablo 30 <i>Deney ve Kontrol Gruplarının Ön Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları</i>	110
Tablo 31 <i>Matematik Puanlayıcı Grubuna Ait Ön Test Sonuçlarına Ait Kategori İstatistikleri</i>	111

Tablo 32 Deney ve Kontrol Grubu Ön Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri	112
Tablo 33 Matematik Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu	114
Tablo 34 Matematik Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri	116
Tablo 35 Deney ve Kontrol Gruplarının Son Test t İstatistikleri Farklarına Ait Mann-Whitney U Testi Sonuçları	123
Tablo 36 Deney Grubu Ön Test ve Son Test t Değerlerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları	124
Tablo 37 Kontrol Grubu Birinci ve İkinci Uygulama t Değerlerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları	124
Tablo 38 Türkçe Puanlayıcı Grubuna Ait Son Test Sonuçlarına Ait Kategori İstatistikleri	125
Tablo 39 Deney ve Kontrol Grubu Son Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri	127
Tablo 40 Türkçe Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu	129
Tablo 41 Türkçe Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri	131
Tablo 42 Deney ve Kontrol Gruplarının Son Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları	134
Tablo 43 Deney Grubu Ön Test ve Son Test t İstatistiklerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları	135
Tablo 44 Kontrol Grubu Ön Test ve Son Test t İstatistiklerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları	136
Tablo 45 Matematik Puanlayıcı Grubuna Ait Son Test Sonuçlarına Ait Kategori İstatistikleri	136
Tablo 46 Deney ve Kontrol Grubu Son Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri	138

Tablo 47 <i>Matematik Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu</i>	140
Tablo 48 <i>Matematik Puanlayıcı Grubu Son Test Sonuçlarına Ait Anlamli Puanlayıcı x Birey Etkileşimleri</i>	142



ŞEKİLLER LİSTESİ

Şekil 1. Puanlayıcı Eğitimi Genel Yapısı.....	72
Şekil 2. Türkçe puanlayıcı deney grubu ön test verilerinden elde edilen t değerleri.	100
Şekil 3. Türkçe puanlayıcı kontrol grubu ön test verilerinden elde edilen t değerleri.	100
Şekil 4. Matematik puanlayıcı deney grubu ön test verilerinden elde edilen t değerleri. ..	109
Şekil 5. Matematik puanlayıcı kontrol grubu ön test verilerinden elde edilen t değerleri.	109
Şekil 6. Türkçe puanlayıcı deney grubu son test verilerinden elde edilen t değerleri.	122
Şekil 7. Türkçe puanlayıcı kontrol grubu son test verilerinden elde edilen t değerleri.....	122
Şekil 8. Matematik puanlayıcı deney grubu son test verilerinden elde edilen t değerleri..	133
Şekil 9. Matematik puanlayıcı kontrol grubu son test verilerinden elde edilen t değerleri.	133

SİMGELER VE KISALTMALAR LİSTESİ

ADPA	: Analitik Dereceli Puanlama Anahtarı
AFA	: Açımlayıcı Faktör Analizi
ÇYRM	: Çok Yüzeyle Rasch Modeli
IELTS	: International English Language Testing System
KTK	: Klasik Test Kuramı
LGS	: Liselere Giriş Sınavı
MEB	: Millî Eğitim Bakanlığı
MTK	: Madde Tepki Kuramı
ÖSYM	: Ölçme Seçme ve Yerleştirme Merkezi
PISA	: Programme for International Student Assessment
RMSE	: Root Mean Square Error
TOEFL	: Test of English as a Foreign Language
χ^2	: Ki-kare İstatistiği

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu ortaya konmuş, araştırmanın amacı, alt problemleri, önemi ifade edilmiş, sınırlılıklarına ve ilgili tanımlarına yer verilmiştir.

Problem Durumu

Dünya'nın gelişip değiştiği süreçte eğitim ve öğretim süreçlerinde de bazı farklılıklar meydana gelmektedir. Geçmişteki eğitim bileşenleri, öğretmeni merkeze alan ve bilgiyi aktaran olarak görürken günümüzde öğrenci merkezli bir hale evrilmiş ve öğrencilerin bilgiyi günlük yaşamda nasıl ve ne kadar kullandıklarına odaklanmıştır. Ayrıca günümüzdeki eğitim bileşenleri geçmiş yılların aksine daha üst düzey zihinsel becerilere odaklanmaktadır. Eğitim bileşenlerindeki bu değişim değerlendirme yaklaşımlarına da yansımıştır. Günümüzde sadece bilgiye değil bilginin nasıl yapılandırıldığına, nasıl kullanıldığına odaklanılmaktadır. Geçmişte bireylerle ilgili karar verirken ya da durum belirlerken doğru- yanlış, çoktan seçmeli testler, sözlü yoklamalar gibi klasik (geleneksel) ölçme yaklaşımları kullanılmaktayken günümüzde performans değerlendirme, dereceli puanlama anahtarı, portfolyo, öz değerlendirme, akran değerlendirme gibi tamamlayıcı (çağdaş, alternatif, süreç odaklı) ölçme ve değerlendirme yaklaşımları kullanılmaktadır. Klasik (geleneksel) ölçme ve değerlendirme yaklaşımlarında cevapların puanlayıcılara göre değişmemesi ve puanlama sürecinin objektif olması bu yaklaşımların avantajı olsa da bireylerin çoktan seçmeli bir testte doğru yanıtı seçmesi, bir ifadenin doğru ya da yanlışlığına karar vermesi ya da boşluğa uygun kelime yerleştirmesi bilginin değerlendirilmesi için yeterli değildir (Kutlu, Doğan, & Karakaya, 2017, s. 8). Bu yüzden üst düzey becerilerin ölçme ve değerlendirme sürecinde geleneksel (klasik) yöntemlere ek olarak tamamlayıcı

(çağdaş, alternatif, süreç odaklı) yöntemlerin kullanılmasının etkili olduğu bilinen bir gerçektir (Haladyna, 1997; Linn & Gronlund, 1995).

Üst düzey zihinsel beceriler bireylerin bilgiyi yapılandırmalarına bağlı olarak bireyden bireye değişen hem süreç hem de ürün odaklı becerilerdir. Üst düzey zihinsel beceriler bireylerin bilgi edinmesinin yanında karar verme, eleştirel düşünme, problem çözme gibi becerileri kapsamaktadır. Üst düzey zihinsel becerilerin ölçülmesi esnasında bireylere kendi cevaplarını yapılandırma fırsatı sunulur. Üst düzey zihinsel beceriler başarının göstergesi olduğu için geçerli ve güvenilir bir şekilde ölçülmeleri oldukça önemlidir (Haladyna, 1997).

Ulusal düzeyde yapılan sınavlar düşünüldüğünde eğitim sistemimizde yaygın olarak kullanılan ölçme araçlarından biri de çoktan seçmeli testlerdir. Çoktan seçmeli testlerde fazla soru sorularak daha fazla davranış ölçülür. Bu sayede yapılan ölçümlerin geçerli ve güvenilir sonuçlar üretmesi sağlanabilir. Ayrıca çoktan seçmeli testlerin puanlanmasını farklı puanlayıcılar yaptığında ölçüm sonuçları değişmez, yani objektiftir (Turgut & Baykul, 2019, s. 188). Aynı anda daha fazla becerinin ölçülmesine imkân tanınması, kalabalık gruplarda uygulaması ve değerlendirmesi kolay olması ve objektif puanlanması bakımından çoktan seçmeli testler diğer ölçme araçlarına göre daha ön plana çıkmaktadır.

Çoktan seçmeli testlerde tahminle doğru cevaplama ihtimalinin olması sınırlılıklarından bir tanesidir (Güler, 2019, s. 74). Bireylerin maddenin cevabını tahminle bulmasına şans başarısı adı verilir ve şans başarısında bireylerin soru hakkında hiç fikri yoktur ya da bireyler sorudaki ipucuna göre doğru cevabı bulmaktadır (Rogers'tan aktaran Güler, 2008, s. 3). Şans başarısı, ölçülecek özellik ile ilgisiz varyansın artmasına sebep olmaktadır. Varyansın artması da güvenilirliğin ve geçerliğin düşmesine neden olur. Şans başarısını azaltmanın yani öğrencilerin bilmedikleri soruları boş bırakmalarını sağlamanın bir yolu da düzeltme formülleridir. Düzeltme formülü belli sayıdaki yanlış sorunun (seçenek sayısının bir eksiği) doğru yanıtlardan birinin iptaline dayanır. Örneğin Liselere Giriş Sınavı'nda her test için yanlış yanıt sayısının üçte biri, doğru yanıt sayısından çıkarılır ve bireyin o testten aldığı ham puan hesaplanır.

Çoktan seçmeli testlerin bir diğer sınırlılığı kopya çekmeye daha elverişli olmasıdır. Kopya çekme davranışını önlemek ya da en aza indirmek için de paralel formların kullanımı ya da çeşitli kopya taramalarının yapılması uygundur (Turgut & Baykul, 2019, s. 240). Çoktan seçmeli testlerin diğer bir sınırlılığı ise test tekniği becerisini de içeriyor olmasıdır. Bireyler gerekli bilgiye sahip olmadan doğru cevap verme şanslarını artırıcı bilgiye sahiplerse doğru

cevabı bulabilirler. Bu durumda asıl ölçülmek istenen dışında farklı bir özelliğin ölçülmesi yapı geçerliğini olumsuz etkilemektedir.

Çoktan seçmeli testlerle ilgili yukarıda da bahsedildiği gibi bazı sınırlılıklar vardır ancak bu sınırlılıkları azaltmak ya da tamamen ortadan kaldırmak mümkündür. Ancak çoktan seçmeli testlerin önlenemeyen sınırlılığı üst düzey zihinsel becerileri yeterince ölçememesi, üst düzey zihinsel becerilerin ölçümünde yetersiz kalmasıdır. Birey performans değerlendirme sürecinde kullanılan çoktan seçmeli testler problem çözme, empati kurma, analitik düşünme, eleştirel düşünme, okuduğunu anlama, karar verme, yaratıcılık gibi üst düzey zihinsel becerilerin ölçülmesinde yetersizdir (Kutlu, 2006). Çoktan seçmeli madde tipi daha çok alt düzey zihinsel becerilerin ölçümüne uygundur (Ebel & Frisbie, 1991, s. 156; Kutlu vd., 2017, s. 27). Çünkü çoktan seçmeli testlerde bireyler cevabı kendileri yapılandırmaz, doğru cevabı seçenekler arasından bulurlar. Bireyler zamanlarının çoğunu maddeleri okumak ve doğru yanıtı bulmakla geçirdikleri için yazmaya ayrılan süre çok azdır.

Ülkemizde üst eğitim kurumuna geçen bireylerin belirlenmesi sürecinde sadece bilişsel açıdan değerlendirme imkânı sunan sınavlar yer almaktadır. Millî Eğitim Bakanlığı (MEB)'nın Liselere Geçiş Sınavı (LGS), Ölçme Seçme ve Yerleştirme (ÖSYM)'nin Temel Yeterlik, Alan Yeterlik ve Yabancı Dil Testi, Kamu Personeli Seçme Sınavı gibi sınavlarda sadece çoktan seçmeli maddeler yer almaktadır. Öğretim programlarında uygun öğrenme çıktılarına ulaşmak için uygun madde kullanımı oldukça önemlidir (Reiner, Bothell, Sudweeks, & Wood, 2002, s. 14). Çoktan seçmeli testlerin sahip olduğu sınırlılıklar geniş ölçekli yapılan sınavlarda farklı yöntemlerin kullanılması gereğini ortaya çıkarmıştır. Üst düzey zihinsel becerilerin ölçülmesi için klasik (geleneksel) ölçme yaklaşımlarından farklı olarak günlük hayattan bir kesite dayalı hem ürüne hem de sürece odaklanan, bireyin güçlü ve zayıf yönlerini fark etmesini sağlayan, daha fazla problem çözmeye odaklanan ölçme ve değerlendirme yöntemleri kullanılmalıdır (Brown & Hudson, 1998; Khattri, 1995; Moore, 2009, s. 1). Bu yöntemlerden birisi de açık uçlu maddelerin kullanımıdır. Açık uçlu maddeler öğrencilerin üst düzey zihinsel becerilerini kullanmalarına imkân tanır (Popham, 2001). Açık uçlu madde ile bireyler yanıtlarını yapılandırırken ön bilgilerini ve yorumlama yeteneklerini kullanmaktadırlar.

Açık uçlu maddeler bireyin özellikle üst düzey zihinsel düşünme becerilerini ölçmekte kullanılabilir (Turgut & Baykul, 2019, s. 147). Yenilenmiş Bloom sınıflandırmasına göre analiz etme, değerlendirme ve yaratma basamağındaki; Haladyna sınıflandırmasına göre ise problem çözme, eleştirel düşünme, yaratıcı düşünme becerilerinin ölçümüne yönelik madde

hazırlanması halinde açık uçlu maddelerin kullanımı uygundur. Açık uçlu maddelerin net bit tanımı olmamakla birlikte tek bir doğru yanıt olmayan veya birden fazla çözüm yolu içeren maddeler olarak ifade etmek mümkündür (Brookhart, 2014, s. 52).

Açık uçlu maddeler bireylerin kendi fikirlerini yapılandırmasını ve formülize etmelerini gerektirir. Bu sayede açık uçlu maddeler bireylerin alternatif düşünme becerilerini geliştirir. Açık uçlu maddeler öğrencileri daha fazla düşünmeye ve birden fazla yanıt aramaya yönlendirir. Böylece açık uçlu maddeler bireyleri detaylı öğrenmeye teşvik eder. Açık uçlu maddeler bireylerin yazma becerilerini geliştirir. Bireyler cevaplarını yapılandırırken sadece üst zihinsel becerileri değil yazma becerileri de gelişir (Karakaya & Şata, 2022, s. 32). Ayrıca açık uçlu maddeler çoktan seçmeli maddelerde sıklıkla karşımıza çıkan şans başarısının da ortadan kalmasını sağlamaktadır. Açık uçlu maddelerde şans başarısı olmadığı için çoktan seçmeli maddelere göre ölçme hatası daha azdır (Haladyna & Rodriguez, 2013, s. 191).

Açık uçlu maddeler uluslararası sınavlarda sıklıkla kullanılmaktadır. ICILS, PISA, PIRLS, TIMSS gibi birçok değerlendirme çalışmalarında açık uçlu sorular yer almaktadır. Ülkemiz de öğrencilerin başarı seviyelerini yükseltmek, eğitim sistemini daha güncel ve güçlü hale getirmek, diğer eğitim sistemlerine göre kıyaslama yapabilmek için uluslararası başarı belirleme faaliyetlerine katılmaktadır. Ancak ulusal geniş ölçekli sınavlarda, okullarda açık uçlu maddelerin yeterince yer almaması bireylerin zorlanmalarına sebep olmaktadır.

MEB ve ÖSYM üst düzey zihinsel becerilerin ölçülmesi için sınavlarda açık uçlu maddelerin yer alması gerektiğini belirtmiştir. Bu kapsamda MEB (2015), 8. sınıf öğrencilerine yönelik üst düzey zihinsel becerilerini ölçmek için farklı madde türlerini içeren beceri testlerini geliştirmek ve bireylerin beceri seviyelerini belirlemek için Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) projesini faaliyete geçirmiştir. Ayrıca açık öğretim sınavlarında yapılan yönetmelik değişikliği ile bazı derslerin sınavlarında çoktan seçmeli maddelere ek olarak kısa cevaplı, eşleştirme, doğru yanlış, açık uçlu madde türlerinin de kullanılmasına imkân tanınmıştır (Anadolu Üniversitesi, 2017). Tüm bu çalışmalara ek olarak ÖSYM de İç İşleri Bakanlığı Kaymakam Adaylığı Girişi Sınavı'nda da açık uçlu sınavlara yer verileceğini ifade etmiştir (ÖSYM, 2015).

Yapılan yenilikler açık uçlu maddelerden oluşan testlerin puanlanması sürecindeki işlemlerin incelenmesi ihtiyacını ortaya çıkarmıştır. Açık uçlu maddeler yanıtlanırken bireyler kendi bakış açılarıyla değerlendirme yaptıkları için farklı yanıtlarla karşılaşmak oldukça muhtemeldir. Dolayısıyla açık uçlu maddelerden oluşan testlerin puanlama yöntemleri ve puanlama güvenirliği oldukça önemli bir hal almaktadır (Alkan, 2013, s. 4).

Açık uçlu maddeleri katılımcı bireylerin tamamının aynı şeyi anlayacağı şekilde oluşturmak zordur (Moelands, 2014'ten aktaran Karakaya & Şata, 2022, s. 33). Ayrıca açık uçlu maddenin geçerli ve güvenilir bir şekilde puanlanabilmesi için nitelikli bir dereceli puanlama anahtarı hazırlamak gerekmektedir. Dereceli puanlama anahtarı hazırlamak ise uzmanlık gerektiren bir işlemdir. Açık uçlu maddenin niteliğinin yanında puanlamayı yapan kişinin de alan uzmanı ve eğitilmiş olması gereklidir. Çünkü açık uçlu maddeler puanlanırken puanlayıcı hatası daha çok karışmaktadır (Karakaya & Şata, 2022, s. 32)

Açık uçlu maddelerin değerlendirilmesinde en önemli sınırlılığın güvenilirlik sorunu olduğu söylenebilir. Puanlama sürecinde objektifliğin sağlanması oldukça zordur (Romagnano, 2001). Puanlayıcıların objektif olmalarını sağlamak için temel iki şarttan bahsedilebilir. Bunlar iyi hazırlanmış bir dereceli puanlama anahtarı ve puanlayıcı davranışlarıdır. Objektif puanlama için iyi hazırlanmış dereceli puanlama anahtarı ve hatalardan bağımsız puanlama yapabilen puanlayıcılar gerekmektedir (Şata, 2022, s. 82).

Puanlayıcı davranışları ile ilgili literatür incelendiğinde, puanlayıcı etkisi, puanlayıcı hatası, puanlayıcı yanlılığı, problemli puanlayıcı davranışları gibi farklı ifadelerle tanımlandığı görülmektedir (Abu Kassim, 2011; Brennan, Gao, & Colton, 1995; Farrokhi, Esfandiari, & Dalili, 2011; Şata, 2019, s. 5). Bu çalışmada puanlayıcı davranışları ifadesi kullanılacaktır.

Puanlayıcı görüşüne dayalı olarak yapılan değerlendirmelerde en büyük varyans kaynağı puanlayıcılarıdır (Doğan & Anadol, 2017; Russell & Airasian, 2011, s. 220). Açık uçlu maddelerin değerlendirilmesi sürecinde puanlayıcı özellikleri, puanlayıcı sayısı, puanlayıcıların bakış açısı, çevre etkisi gibi çeşitli hatalarla karşılaşılabilir. Puanlama işlemi esnasında ortaya çıkan puanlayıcı davranışları bireylerin beceri ölçümlerinin hatalı kestirilmesine sebep olur. Hatalı puan kestirimi de yapılan ölçümlerin güvenilirliğini olumsuz etkilemektedir. Ayrıca birden fazla puanlayıcının yer aldığı süreçte en temel geçerlik kanıtı puanlayıcılar arasındaki uyumdur (Downing, 2005). Yapılan puanlama işleminde amaçtan farklı değişkenler de değerlendirmeye alınacağından geçerlik de olumsuz etkilenecektir (Abu Kassim, 2011; Congdon & MeQueen, 2000). Bu yüzden farklı puanlayıcı davranışlarının ortadan kaldırılması ya da en aza indirilmesi oldukça önemlidir (YoungKoung Kim, 2009, s. 2; Linacre, 1994, s. 11).

Literatür incelendiğinde puanlama sürecinde objektifliği sağlamak için bir dizi yöntem önerilmiştir. Bu yöntemler; birden fazla puanlayıcının kullanılması, dereceli puanlama anahtarının kullanılması, otomatik puanlama (Aktaş & Alıcı, 2017; Attali, Bridgeman, & Trapani, 2010; Bernardin & Buckley, 1981; Burstein vd., 1998; Dunbar, Brooks, & Kubicka-

Miller, 2006; Ebel & Frisbie, 1991, s. 194; Gronlund, 1977, s. 85; Haladyna, 1997, s. 143; İlhan & Çetin, 2014; Kubiszyn & Borich, 2013, s. 170; Kutlu, Doğan, & Karakaya, 2014, s. 51; Landauer, Laham, & Foltz, 2003, s. 100; Lumley & McNamara, 1995; Oosterhof, 2003, s. 81) ve puanlayıcı eğitimidir (Bernardin & Buckley, 1981; Feldman, Lazzara, Vanderbilt, & DiazGranados, 2012; Haladyna, 1997, s. 143; Hauenstein & McCusker, 2017; Stamoulis & Hauenstein, 1993; Weigle, 1998; Zedeck & Cascio, 1982). Önerilen yöntemlerin her birinin kendine has üstünlükleri veya zayıf noktaları, avantaj ve dezavantajları vardır. Ancak kullanılan yöntem ne olursa olsun puanlayıcılar arasında uyumun sağlanması zordur (Haladyna, 1997, s. 137). Bu çalışmada bu önerilerden birden fazla puanlayıcı, dereceli puanlama anahtarı ve puanlayıcı eğitimi kullanılmıştır.

Puanlayıcı eğitimi, ölçümlere karışan puanlayıcı kaynaklı varyansı azaltmak için sıklıkla kullanılmaktadır. Eğitimin temel amacı puanlayıcılardan kaynaklanan ve geçerliği önemli ölçüde etkileyen; hedef yapıyla bağlantısız varyansı azaltmak, tamamen ortadan kaldırmak ya da kontrol altına almaktır (Brijmohan, 2016, s. 11). Bir başka deyişle puanlayıcı eğitiminin amacı birey performanslarının değerlendirilme sürecinde daha geçerli ve güvenilir ölçümler yapılmasını sağlamaktır (Moser, Kemter, Wachsmann, Köver, & Soucek, 2018). Bu kapsamda örnek uygulamalarla puanlayıcılar arasında ortak bir anlayış oluşturulmaya çalışılmaktadır (Cellar, Curtis, Kohlepp, Poczapski, & Mohiuddin, 1989; Fahim & Bijani, 2011).

Puanlayıcı eğitimi aşırı uçlardaki puanları azaltarak ölçümlerin hem geçerliğine hem de güvenilirliğine katkıda bulunmaktadır. Ancak puanlayıcı eğitimi her ne kadar puanlayıcılardan kaynaklanan varyansı azaltsa da bu varyansı tamamen ortadan kaldıramamaktadır (McIntyre, 1993). Çünkü puanlama sürecinde varyans kaynağı sadece puanlayıcıların puanlama davranışları değildir. Puanlayıcının dikkat seviyesi, motivasyonu, puanlama anahtarının anlaşılma düzeyi gibi varyans kaynakları da yer almaktadır. Ancak yine de puanlayıcı eğitiminin puanlayıcıların puanlama davranışlarındaki aşırı farklılaşmanın önüne geçtiği, puanlayıcıların ön yargılarını azalttığı ve bu sayede puanlayıcı iç tutarlılığını artırdığı belirtilmiştir (Weigle, 1994).

Puanlayıcı eğitiminin işlevselliği ile ilgili tartışmalar olsa da eğitim içeriği ile ilgili çok az şey bilinmektedir (Fahim & Bijani, 2011). Puanlayıcı eğitimi programları genel olarak üç öge içermektedir (Hauenstein & McCusker, 2017). Bu ögeler performansın boyutlarıyla ilgili eğitim, eğitim içeriği ve uygulamanın değerlendirilmesine dair dönüt olarak ifade edilebilir. Puanlayıcı eğitimi çalışmalarında ilk olarak performansın sahip olduğu boyutlar

ve bu boyutların göstergeleri iyi belirlenmelidir. Bu sayede puanlayıcıların tutarlılığı artırılabilir (Şata, 2022, s. 89). Eğer performansla ilgili doğru tanımlamalar yapılamazsa eğitimin puanlama davranışlarının geçerlik ve güvenilirlikleri üzerindeki etkisini kestirmek zor olacaktır (Hauenstein & McCusker, 2017). Bu yüzden puanlayıcı eğitimlerinde performans boyutları eğitimi sabit olup eğitim içeriği yenilenmektedir. Bu eğitim içerikleri farklı yaklaşımlara sahip olsalar da asıl amacın puanlayıcıların yapmış oldukları puanlamaların doğru ve güvenilir olmasını sağlamaktır.

Literatür incelendiğinde birçok puanlayıcı eğitim stratejisinin yer aldığı görülmektedir. En sık kullanılan puanlayıcı eğitimi stratejilerinin öz liderlik eğitimi, davranış gözlem eğitimi, puanlayıcı değişkenliği eğitimi, performans boyutları eğitimi, puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi olduğu görülmektedir (Woehr & Huffcutt, 1994; Zedeck & Cascio, 1982).

Öz liderlik eğitimi stratejisinde öncelikle puanlayıcıların öz güvenlerini artırmak amaçlanmaktadır. Öz liderlik, bireyin kendini gerçekleştirmek için gerekli olan öz yönetim ve öz motivasyonunu meydana getirmek amacıyla bireyin kendini ikna etme sürecidir (Neck, 1996). Neck, Stewart, & Manz (1995) tarafından geliştirilen modelde puanlayıcının sahip olduğu bilişsel yapıların doğruluğu ya da yanlışlığı puanlama davranışını etkileyeceği için puanlayıcıya özel eğitimler verilmesi gerektiği ifade edilmektedir.

Davranış gözlem eğitimi stratejisinde puanlayıcıların gözlem yapma davranışlarına odaklanılmaktadır (Woehr & Huffcutt, 1994). Bu stratejide puanlayıcıların gözlem yapma yeteneklerini geliştirmek için gözlem kayıtları kullanılmaktadır (Antonioni & Woehr, 2001). Puanlayıcı performansı değerlendirilen bireyi ne kadar çok hatırlarsa değerlendirmenin de o kadar objektif olacağı vurgulanmaktadır.

Performans boyutları eğitimi stratejisinde odak noktası ölçülen performans ve bu performansın boyutlarıdır. Puanlayıcılara ilgili beceri ve beceriye sahip olan bireylerin davranışları hakkında bilgi verilir. Örneğin futbolda penaltı atışı değerlendirilirken penaltı atışının tanımı, hangi durumlarda meydana geldiği, nereden kullanıldığı, nasıl kullanılması gerektiği gibi belirli davranışlardan oluşan eğitim puanlayıcılara verilebilir. Bu eğitim stratejisinin temel amacının puanlayıcıların bilişsel olarak gelişmesi sağlamak olduğu söylenebilir (DeNisi, Robbins, & Williams, 1989; Feldman, 1981).

Puanlayıcı hatası eğitimi stratejisinin temel odak noktası, yapılan puanlama işleminde öznel değerlendirmeleri ortadan kaldırmaktır. Bu eğitim stratejisi Woehr & Huffcutt (1994)

tarafından birden fazla puanlayıcının yer aldığı puanlamalarda ölçüm sonuçlarının geçerliğini ve güvenilirliğini artırmak için ortaya atılmıştır. Ancak bu eğitim stratejisi diğer eğitim stratejilerine göre daha teoriktir ve uygulama boyutu sınırlıdır (Şata, 2022, s. 93). Literatürdeki araştırmalar da incelendiğinde puanlayıcı hatası eğitimi stratejisinin tek başına kullanılmasının yeterli olmadığı, puanlayıcılar arası tutarlılık ve değerlendirme doğruluğu üzerinde pozitif anlamlı bir etkisi olmadığı ifade edilmiştir (Bernardin & Pence, 1980; Borman, 1975).

Referans çerçevesi eğitimi stratejisi en fazla tercih edilen eğitim stratejisidir (Roch, Woehr, Mishra, & Kieszczyńska, 2012). Bu durumun sebebi referans çerçevesi eğitim stratejisinin ilgili performansa ilişkin ortak bir anlayış geliştirme ve kavramsallaştırma oluşturulmasının hedeflenmesidir (Aguinis, Mazurkiewicz, & Heggstad, 2009; Athey & McIntyre, 1987; Woehr & Huffcutt, 1994). Bu eğitim stratejisinde puanlayıcılara örnek davranışlarla ilgili örnekler sunulur ve puanlama esnasında puanlayıcılar örneklerden faydalanabilirler. Bu sayede puanlayıcıların vermiş oldukları puanlar arasında eşleştirme yapma fırsatı olacaktır (Woehr & Huffcutt, 1994). Yapılan çalışmalar referans çerçevesi eğitim stratejisinin değerlendirme doğruluğunu artırdığını, en etkili yöntemin referans çerçevesi eğitim stratejisi olduğunu göstermektedir (Aguinis, 2012, s. 180; Roch vd., 2012; Stamoulis & Hauenstein, 1993; Woehr, 1994).

Puanlayıcı değişkenliği eğitimi stratejisi diğer eğitim stratejilerinden farklı olarak harmanlanmış bir stratejidir (Şata, 2019, s. 34). Bu eğitim stratejisinde puanlayıcı hatası eğitimi ve referans çerçevesi eğitiminin güçlü yönleri kullanılarak daha objektif değerlendirme yapılması hedeflenmektedir (Stamoulis & Hauenstein, 1993). Puanlayıcı hatası eğitiminde puanlayıcılar hatalı puanlamalar hakkında bilgi sahibi olurken değerlendirme konusunda eksik kalmaktadırlar. Referans çerçevesi eğitiminde ise puanlayıcılar hatalı puanlamalar hakkında eksik kalmaktadırlar. Bu strateji her iki yöntemin güçlü yanları ile oluşturulmuştur. Kısaca puanlayıcı değişkenliği eğitimi araç olarak referans çerçevesi eğitimi kullanırken hedefine puanlayıcı hatası eğitimi yer almaktadır (Hauenstein, 1998). Bu sayede farklı puanlayıcı davranışlarının ortadan kaldırılması/en aza indirilmesi amaçlanmaktadır.

Ele alınan puanlayıcı eğitim stratejilerinin farklı yapılara sahip olduğu ancak hepsinin temel amacının puanlayıcı güvenilirliğinin ve puanlama doğruluğunun artırılması, puanlayıcı hatalarının azaltılması olduğu görülmektedir. Yapılan çalışmalar da puanlayıcı eğitimlerinin puanlayıcı güvenilirliğini ve puanlama doğruluğunu artırdığı, puanlayıcı hatalarını azalttığını

gösterse de bazen puanlayıcı eğitiminin, puanlayıcı hatalarını tamamen ortadan kaldırmadığı bireysel farklılıkların ortadan kalmasında yeterli olmadığını göstermektedir (Cook, Dupras, Beckman, Thomas, & Pankratz, 2009; Eckes, 2011; Galati, 2007, s. 7; Gingerich, Kogan, Yeates, Govaerts, & Holmboe, 2014; Gingerich, Regehr, & Eva, 2011; Holmboe, Hawkins, & Huot, 2004; Sebok & Syer, 2015). Ayrıca puanlayıcı eğitimlerinin etkili olabilmesi için düzenli olarak belirli aralıklarla yapılması gerektiği belirtilmiştir (Wang, 2010). IELTS, TOEFL gibi sınavlarda da puanlayıcı güvenilirliğini ve tutarlığını artırmak için puanlayıcılara belirli aralıklarla eğitim verildiği bilinmektedir.

Özetle literatür incelendiğinde birden fazla puanlayıcının yer aldığı değerlendirme sürecinde performans ile alakalı olmayan, puanlayıcıdan kaynaklanan varyansın ölçümlere karıştığı ve bu durumun ölçümlerin geçerlik ve güvenilirliğini olumsuz etkilediği görülmektedir. Puanlayıcılarda yaygın olarak görülen puanlayıcı davranışlarının katılık, cömertlik, merkeze eğilim, puanlayıcı yanlılığı olduğu görülmektedir (Myford & Wolfe, 2003, 2004).

Puanlayıcı davranışlarının belirlenmesi sürecinde farklı yöntemler kullanılmaktadır. Bu yöntemler klasik test kuramı (KTK)'na dayalı yöntemler, genellenebilirlik kuramı (GK)'na dayalı yöntemler ve madde tepki kuramı (MTK)'na dayalı yöntemler olarak ifade edilebilir. Bu yöntemler içerisinde genellenebilirlik kuramı ve çok yüzeyli rasch modeli en fazla tercih edilen yöntemlerdir (Royal & Hecker, 2016). KTK, varsayımlarının karşılanmasının kolay olmasından dolayı uzun yıllardır alanda kullanılan yöntem olmasına rağmen madde ve yetenek parametrelerinin gruba bağlı kestirilmesi, tüm bireyler için bir hata kestirimi yapması gibi sınırlılıkları mevcuttur (Köse, Usta, & Yandı, 2016). KTK'nin sınırlılıklarını varsayımlarla ortadan kaldıran ölçme kuramı ise MTK'dir. MTK, bireyin her bir maddeye vermiş olduğu yanıtları matematiksel modeller kullanarak ölçülen özelliğe sahip olma durumunu ortaya koyan ölçme kuramıdır (Babcock, 2009, s. 1). Temelinde MTK olan, madde güçlüğüne örneklemden bağımsız tahmin eden ve tek parametrelili olan modele Rasch modeli adı verilmektedir (Rasch, 1960'dan aktaran Linacre, 1994). Rasch modeline puanlayıcı davranışlarının (katılık, cömertlik) eklenmesiyle çok yüzeyli rasch modeli (ÇYRM) geliştirilmiştir (Linacre, 1993a). Çok yüzeyli rasch modeli farklı değişkenlik kaynaklarının da dikkate alınmasını sağlayan bir modeldir (Lynch & McNamara, 1998).

Açık uçlu soruların puanlanmasında verilen puanlayıcı eğitiminin puanlayıcı davranışlarına etkisinin ne olduğu bu araştırmanın problemi oluşturmaktadır.

Problem İfadesi

Açık uçlu soruların puanlanmasında puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisi nasıldır?

Araştırma Soruları

Problem ifadesi doğrultusunda aşağıdaki sorulara cevap aranmıştır;

1. Puanlayıcı eğitimi öncesinde açık uçlu soruların puanlanması sürecinde;
 - a. Türkçe öğretmen adaylarından oluşan puanlayıcı grubunun ön testte ve son testte birey madde ve puanlayıcı yüzeylerinden elde edilen verilerin uygunluk değerleri ve model ölçümleri nasıldır?
 - b. Matematik öğretmen adaylarından oluşan puanlayıcı grubunun ön testte ve son testte birey madde ve puanlayıcı yüzeylerinden elde edilen verilerin uygunluk değerleri ve model ölçümleri nasıldır?
2. Puanlayıcı eğitimi öncesinde açık uçlu soruların puanlanması sürecinde;
 - a. Türkçe öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların puanlayıcı davranışları nelerdir?
 - b. Türkçe öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların t değerleri değerleri arasında istatistiksel olarak fark var mıdır?
 - c. Matematik öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların puanlayıcı davranışları nelerdir?
 - d. Matematik öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların t değerleri değerleri arasında istatistiksel olarak fark var mıdır?
3. Puanlayıcı eğitimi sonrasında açık uçlu soruların puanlanması sürecinde;
 - a. Türkçe öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların puanlayıcı davranışları nelerdir?
 - b. Türkçe öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların t değerleri arasında istatistiksel olarak fark var mıdır?
 - c. Matematik öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların puanlayıcı davranışları nelerdir?
 - d. Matematik öğretmen adaylarından oluşan deney ve kontrol grubunda yer alan puanlayıcıların t değerleri arasında istatistiksel olarak fark var mıdır?

Araştırmanın Amacı

Yapılan araştırma, açık uçlu soruların puanlanmasında puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisini incelemeyi amaçlamaktadır. Bu amaç doğrultusunda puanlayıcılara literatürde yer alan puanlayıcı eğitimi desenlerinden referans çerçevesi eğitimi ve puanlayıcı hatası eğitimi harmanlanarak oluşturulan eğitim verilmiştir.

Araştırmanın alt amaçlarından biri de öğrencilerin açık uçlu sorulara vermiş oldukları cevapları puanlamak olduğu için bütünsel (holistik) dereceli puanlama anahtarı geliştirilmiştir.

Araştırmanın Önemi

Ulusal literatür incelendiğinde puanlayıcı davranışları ile ilgili birçok çalışmanın yer aldığı (Akın & Baştürk, 2012; Atılgan, 2005b; Baştürk, 2010; Karakaya, 2015; Ok, Bilgiç, & Bilgiç, 2013; Şahan, 2018; Tobaş, 2020) görülmektedir. Ancak puanlayıcı davranışlarına yönelik eğitim ve puanlayıcı eğitiminin etkinliğine ile ilgili az sayıda çalışma (Çetin & İlhan, 2017; Şata, 2019) olduğu görülmektedir. Performans değerlendirme sürecinde puanlayıcı davranışlarının hem geçerliği hem de güvenilirliği olumsuz etkilediği göz önünde bulundurulduğunda, bu olumsuz etkilerin azaltılmasına yönelik çalışmalar önem arz etmektedir.

Uluslararası literatürde eğitim bilimleri alanında performans değerlendirme sürecinde puanlayıcı davranışlarından kaynaklanan varyansı ortadan kaldırmak ya da azaltmak için birçok puanlayıcı eğitimin modelinin yer aldığı görülmektedir. Ulusal literatürde puanlayıcı eğitimi ile ilgili İlhan & Çetin (2014) ve Şata'ya (2019) ait çalışmaların yer aldığı görülmektedir. Ayrıca yapılan çalışmalarda puanlayıcı eğitim stratejilerinden performansa yönelik 360 derece performans geri bildirim yönteminin ve referans çerçevesi eğitim stratejisinin kullanıldığı görülmektedir. Bu çalışma ile ulusal literatüre puanlayıcı eğitiminin hem kuramsal açıdan hem de uygulamalı olarak katkı sağlamasının önemli olacağı düşünülmektedir.

Bilginin yanında bilginin kullanımının da ön plana çıktığı, önem kazandığı çağımızda bireylerin kendine ve topluma yönelik görevlerini yerine getirebilmesi için üst düzey zihinsel becerilerle donatılmış olması gerekmektedir. Üst düzey zihinsel becerilere sahip ve görevlerini yerine getirebilen bireyler daha üretken, verimli, eleştirel düşünebilen, problem çözebilen bireylerdir. Bu noktada bireylerin performanslarının doğru değerlendirilmesi

önemlidir. Doğru bir değerlendirme ile bireylere en az hata ile dönüt verilebilir ve bireyin gelişimine katkıda bulunulabilir. Literatürde üst düzey zihinsel beceriler ölçülürken klasik (geleneksel) değerlendirmeden (çoktan seçmeli testler, doğru- yanlış testleri, sözlü yoklamalar, vb.) ziyade performans değerlendirme tekniklerinin (portfolyo, öz ve akran değerlendirme, dereceli puanlama anahtarı vb.) kullanımının daha uygun olduğu belirtilmiştir (Kutlu, Doğan, & Karakaya, 2010, s. 4). Ayrıca performans değerlendirme tekniklerinin en büyük sınırlılığının da puanlayıcıların objektifliğini sağlamanın zorluğu göz önüne alındığında performans değerlendirmede puanlayıcı objektifliğini artıracak yöntemlerin incelenmesi, puanlama sonuçlarının güvenilirliği ve geçerliği açısından oldukça önemlidir.

Ulusal sınavlarda açık uçlu sorular kullanılırken puanlanma sürecinde bazı risklerin olduğu ve performans görevlerinin geleneksel yöntemlerden daha avantajlı olduğu düşünüldüğünde (Brown & Hudson, 1998; Khattri, 1995; Moore, 2009, s. 69) bireylerin güçlü ve zayıf yönlerinin belirlenmesi, bireylere daha detaylı dönüt verilebilmesi, üst zihinsel becerilerin gelişimine, takibine imkân vermesi açısından yapılan puanlama işlemi oldukça önemlidir. Bu kapsamda yapılan çalışmanın literatüre katkı sağlayacağı düşünülmektedir.

Farklı disiplinlerde (tıp, mühendislik, pazarlama vb.) performans değerlendirme sürecinde birden fazla puanlayıcı kullanılmaktadır. Birden fazla puanlayıcının yer aldığı sınavların en tartışmalı bölümünün değerlendirme aşaması olduğu sıklıkla dile getirilmektedir. Bu durumda alınan kararlara ilişkin geçerlik ve güvenilirlik kaygıları ortaya çıkmaktadır. Bu kaygıyı ortadan kaldırmak ya da azaltmak için alınacak önlemler arasında puanlayıcı eğitimi de yer almaktadır. Bu çalışmanın birden fazla puanlayıcının yer aldığı alanlara da katkı sağlayacağı düşünülmektedir.

Sınırlılıklar

Araştırmanın sınırlılıkları aşağıda verilmiştir;

- Literatürde çeşitli puanlayıcı eğitim stratejileri yer alırken mevcut çalışmada referans çerçevesi eğitimi ve puanlayıcı hatası eğitimi olmak üzere iki puanlayıcı eğitimi stratejisi kullanılmıştır
- Araştırmada sayısal ders olarak Matematik, sözel ders olarak da Türkçe dersinden sorular yer almaktadır.
- Puanlama desenlerinden tümüyle çaprazlanmış desen kullanılmıştır.

Tanımlar

Puanlayıcı Davranışları: Performans değerlendirme sürecinde ölçülecek özellikle ilgisiz değişkenlerin ölçümleri etkileme ihtimali her zaman bulunmaktadır. Performans değerlendirme sürecinde meydana gelen bu davranışlar puanlayıcı davranışları olarak ifade edilmektedir.

Açık uçlu madde: Tek bir doğru cevabı olmayan, yanıtı bireyler tarafından oluşturulan, bilginin kullanılmasına imkân veren madde türüne açık uçlu madde adı verilmektedir.

Ayrırma oranı: İstatistiksel olarak birbirinden farklı, ayırt edilebilen performans düzeylerinin sayısıdır.

Ayrırma indeksi: Örneklem ile örneklemle aynı standart sapmaya sahip normal dağılımda istatistiksel olarak farklı performans düzeylerinin sayısıdır.

RMSE: Ölçümlerin standart hatalarının ortalamasıdır.

BÖLÜM II

KAVRAMSAL ÇERÇEVE

Bu bölümde araştırmaya ait temel kavramsal çerçeve ile ilgili belirlemeler ve değerlendirmeler yer almaktadır. Bu bölümde üst düzey zihinsel beceriler, açık uçlu maddeler, açık uçlu maddelerin puanlanması, puanlayıcı davranışları, puanlayıcı davranışlarının belirlenmesi, azaltılması ya da kontrol altına alınmasına yönelik uygulamalar ve puanlayıcı güvenilirliği ile ilgili bilgiler yer almaktadır.

Üst Düzey Zihinsel Beceriler

Bireyin bazı davranışları sergilerken kullandığı bilişsel, duyuşsal ve psikomotor özelliklerin kendi özellikleriyle harmanlayarak kullanmasına üst düzey zihinsel beceriler adı verilir. Othman & Mohamad (2014), üst düzey zihinsel becerileri, bilgi toplama, bilgiyi organize etme, yönetme, bilginin kalitesini değerlendirme ve doğru bilgiyi üretme yeteneği olarak tanımlamıştır. Bu beceriler kolay ulaşılamayan, birden fazla çözüm yolu, farklı bakış açıları gerektiren beceriler olarak tanımlanabilir (Wellman, 1997'den aktaran Erman Aslanoğlu, 2022, s. 4). Alt düzey zihinsel beceriler ezberleme, hatırlama, doğrudan çıkarım yapma, tanımlama, olayları anlatma gibi davranışları kapsar ve tek cevaplı yapıya sahiptir. Örneğin çoktan seçmeli test çözümü alt düzey zihinsel becerileri geliştirmektedir. Üst düzey zihinsel beceriler ezberleme, hatırlama gibi davranışlardan daha çok sahip olunan bilgiyi farklı durumlara transfer etmeyi gerektirir (Anderson & Krathwohl, 2001, s. 63; Brookhart, 2010, s. 5; Choorapanthiyil, 2007, s. 53). Üst düzey zihinsel becerilerin kullanımını gerektiren durumlar, geçmişte var olan bilginin yeni bilgilerle harmanlanıp başka bir duruma uygulanmasını gerektirir. Örneğin bireyin belli bir konu ile ilgili özgün bir kompozisyon yazması üst düzey zihinsel becerilerin kullanımını gerektirmektedir.

Üst düzey zihinsel becerilerin ölçülmesi 1950'lerin ortalarına doğru Bloom ve arkadaşları tarafından geliştirilen eğitim hedeflerinin taksonomik hale gelmesiyle daha da önem kazanmıştır. Ancak bu becerilerin ölçülmesi günümüzde daha da önem kazanmıştır. Bu durumun temel sebebi iş gücü tanımının değişmesidir (Levy & Murnane, 2004, s. 1). Geçmişte rutin olarak yapılan işleri artık bilgisayarlar yapmaktayken bireylerden, bilgileri elde edip problem çözme ve karar verme sürecinde kullanmaları beklenmektedir. Bu beklentiyi dikkate alan eğitim sistemleri bireyleri üretken, eleştirel düşünebilen, problem çözebilen ve hayat boyu öğrenmeye açık bireyler olarak yetiştirmeyi hedeflemektedir. Bu yüzden bireylere üst düzey zihinsel becerileri öğretmek ve bireylerin bu becerilerini geliştirmek için uygun öğretim programları ile ölçme ve değerlendirme modelleri uygulamak gerekmektedir.

Üst düzey zihinsel becerilerin ölçülmesi sürecinde bireyin bilgiyi zihninde nasıl yapılandığı, sahip olduğu bilgiyi nasıl kullandığı ve karşılaştığı problemi nasıl çözdüğü hakkında bilgi elde edilir. Ancak üst düzey zihinsel becerilerin ölçülmesinde klasik ölçme araçlarının (çoktan seçmeli, doğru yanlış, kısa yanıtlı vb.) kullanılması bu becerileri ölçmede yetersiz kalmaktadır. Çünkü klasik ölçme araçları genellikle alt düzey zihinsel becerileri temel almaktadır (Erman Aslanoğlu, 2022, s. 3). Bu durum üst düzey zihinsel becerilerin ölçümünde performans değerlendirme ve açık uçlu soruların kullanımının ön plana çıkmasına sebep olmaktadır.

Üst Düzey Zihinsel Becerilerin Ölçümünde Açık Uçlu Maddeler

Açık uçlu maddeler, birden fazla çözüm yolu içeren ve dolayısıyla doğru yanıtı birden fazla olan maddeler olarak ifade edilebilir (Brookhart, 2014; Stalnaker, 1951). Açık uçlu maddeler, bireyin düşüncelerini yapılandırmasını ve düşüncelerin formüle edilmesini gerektirir. Bu sayede problemlere farklı bakış açısıyla bakma fırsatı doğacaktır (Hancock, 1995).

Açık uçlu maddeler sadece “ne” sorusunun cevabını değil “ne zaman, nerede, niçin” gibi soruların cevabını da aramaktadır (Dochy, 2001). Bu durum açık uçlu soruların üst düzey zihinsel becerilere odaklanmasını sağlar. Birey yanıtını kendi yapılandırır ve açıklama fırsatı bulur. Ayrıca bireye özgürce kendini ifade etme imkânı verir.

Maddenin açık uçlu madde olarak adlandırılması için sahip olması gereken özellikler vardır (Reiner vd., 2002);

1. Yanıtlar bireyler tarafından oluşturulmalı ve en az bir cümle içermelidir.
2. Yanıtların değerlendirilmesi sürecinde uzmanlara ihtiyaç vardır.
3. Birden fazla bilginin kullanılmasına fırsat vermelidir.
4. Birden fazla çözüm yolu ve doğru yanıt olmalıdır.

Bu özellikleri taşıyan maddeler açık uçlu madde olarak adlandırılabilir. Açık uçlu maddeler yanıtı sınırlandırılmış ve yanıtı sınırlandırılmamış açık uçlu maddeler olmak üzere ikiye ayrılmaktadır (Başol, 2019, s. 45; Çetin, 2019, s. 128-131; Reiner vd., 2002).

Yanıtı sınırlandırılmış açık uçlu maddelerde çözüm yolu ya da doğru cevap birden fazladır ancak sonsuz değildir. Bu tür maddeler puanlanırken bütünsel dereceli puanlama anahtarı kullanılması önerilmektedir. Ayrıca ulusal ve uluslararası geniş ölçekli sınavlarda, sınıf içi uygulamalarda yanıtı sınırlandırılmış açık uçlu maddelerin kullanımı daha uygundur (Karakaya & Şata, 2022, s. 31).

Yanıtı sınırlandırılmamış açık uçlu maddelerde tek bir doğru yanıt yoktur. Doğru yanıt her bir birey kendince farklı bir şekilde oluşturup gerekçelendirebilir. Bu tür maddeler daha çok kompozisyon yazma tipindeki sınavlarında kullanılmaktadır. Bu maddelerin puanlanmasında analitik dereceli puanlama anahtarı kullanılmaktadır. Analitik dereceli puanlama kullanılmasının sebepleri ölçülen özelliğin karmaşık olması, ölçütün fazla olması, detaylı dönüt verme isteği olabilir (Karakaya & Şata, 2022, s. 31)

Açık uçlu maddelerin kullanımının avantajları olduğu gibi dezavantajları da vardır. Açık uçlu maddelerin avantajları (Bağcan Büyükturan & Çıkrıkçı Demirtaşlı, 2013; Başol, 2019, s. 45; Birenbaum & Tatsuoka, 1987; Coffman, 1971; Güler, 2019, s. 67; Haladyna & Rodriguez, 2013; Karakaya & Şata, 2022, s. 31; Reiner vd., 2002; Ward, Frederiksen, & Carlson, 1980);

- Kopya gibi olumsuz davranışlar daha az gözlemlenir.
- Şans başarısını ortadan kaldırır. Şans başarısını ortadan kaldırdığı için ölçme hatası daha az olmaktadır. Bu durum ölçme aracının hem geçerliğini hem de güvenilirliğini artırmaktadır.
- Bireyin kendi fikirlerini yapılandırmasına ve formüle etmesine imkân tanırken bireyin üst düzey düşünme becerilerini geliştirir. Ayrıca günlük hayat problemlerine benzer problemlere yer verildiğinde bireylerin problem çözme becerilerinin gelişimine de katkı sağlar.

- Bireyleri daha fazla düşünmeye ve yanıt aramaya yönlendirirken bireylere ayrıntılı ve kalıcı öğrenme imkânı sunar.
- Bireylerin yazma becerilerini gelişimine katkı sağlar. Bireyler cevaplarını yapılandırırken sadece üst düzey zihinsel becerileri değil aynı zamanda yazma becerileri de gelişir.
- Açık uçlu maddeler kısmi puanlamaya uygundur ve çoktan seçmeli testlere göre daha kolay hazırlanabilir.
- Objektif madde türleri ile kıyaslandığında, günlük yaşamla ilgili madde oluşturmaya daha uygundur.

Açık uçlu maddelerin dezavantajları (Abedi, 2006, s. 379; Bennett, Ward, Rock, & LaHart, 1990; Downing & Yudkowsky, 2009, s. 40; Karakaya & Şata, 2022, s. 32; Livingston, 2009);

- Madde yazımı ve puanlama anahtarının geliştirilmesi zordur. Hazırlanan maddenin sınavı alan tüm bireyler tarafından aynı şekilde yorumlaması oldukça zordur. Buna ek olarak puanlama anahtarının geliştirilmesi de zordur ve uzmanlık gerektirir.
- Hazırlanması, uygulanması ve puanlanması çoktan seçmeli testlere göre daha uzun zaman gerektirir. Yani çoktan seçmeli maddelere göre yüksek maliyetlidir.
- Sınavı alan bireyler için de maddelerin cevaplandırılması uzun zaman alabilmektedir.
- Konu ile ilgili yazılacak madde sayısının az olması durumunda kapsam geçerliği düşük olabilmektedir.
- En büyük sınırlılığı diğer madde türlerine göre (çoktan seçmeli, doğru yanlış, eşleştirme vb.) puanlayıcı hatası daha çok karışır. Puanlayıcıların uzman ve puanlayıcı eğitimi almış olması gerekmektedir.

Açık Uçlu Maddelerin Puanlanması

Açık uçlu maddelerin yer aldığı testlerde maddenin kalitesi kadar maddenin puanlanması da önemlidir. Puanlanmanın kalitesini belirlemek amacıyla puanlamalar arasındaki tutarlılığa bakılır. Aynı maddeyi farklı zamanlarda puanlayan puanlayıcı ya da aynı maddeyi puanlayan farklı puanlayıcıların puanlamaları arasında tutarlılık bulunması puanlama güvenilirliğinin sağlandığına kanıttır (Çetin, 2019, s. 138). Açık uçlu maddelerin sınırlılıklarından bir tanesi puanlama sürecinde objektifliği sağlamanın zorluğudur. Yani puanlayıcılar bireyleri puanlarken amacın dışındaki farklı değişkenlerin de ölçümlere karışması ya da karışma

ihtimali puanlama sonuçlarını olumsuz etkilemektedir. Bu hataları azaltmak, en aza indirmek için iki şartı yerine getirmek gerekmektedir. Bunlardan ilki iyi hazırlanmış bir dereceli puanlama anahtarı iken ikincisi puanlayıcı eğitimidir.

Dereceli puanlama anahtarının hazırlanırken bazı kaynaklara göre iki temel yaklaşım (Kutlu vd., 2017, s. 54) yer almaktayken bazı kaynaklara göre de üç yaklaşım yer almaktadır (Morrow, Mood, Disch, & Kang, 2015). Bu yaklaşımlardan hangisinin uygun olacağı testin amacına göre belirlenmektedir. Bireylerden gelen cevapların doğruluk dereceleri ile ilgileniliyorsa analitik puanlama, ikna edici kanıtlar ve anlatım biçimi ile ilgileniliyorsa bütünsel puanlama yaklaşımının kullanılması daha uygundur (Cunningham, 1998). Kutlu vd. (2017, s. 54) açık uçlu maddenin ölçtüğü yapı birden fazla alt boyuta sahipse analitik puanlama yaklaşımı, yapı tek boyuttan oluşuyor ya da alt boyutlara ayrılamıyorsa da bütünsel puanlama yaklaşımının kullanılması gerektiğini belirtmektedir. Morrow vd. (2015) tarafından önerilen ve bağıl (göreceli) puanlama adı verilen yaklaşımda bireylerin cevapları yeterlik durumuna göre sıralanır (iyi, orta, kötü vb.) ve puanlama yapılır. Bu sayede puanlamanın tutarlılığı artırılmış olur. Bağıl (göreceli) puanlama yöntemi daha çok karmaşık zihinsel becerilerin ölçülmesinde etkilidir (Morrow vd., 2015). Dereceli puanlama anahtarının hazırlanması sürecinde kullanılan yöntemlerin birbirlerine üstünlükleri ya da zayıflıkları olmakla birlikte puanlayıcı tutarlılığı üzerinde olumlu etkisi vardır. Hazırlanan açık uçlu maddenin değerlendirilmesinde geçerli ve güvenilir sonuçlar elde etmek için madde amacına uygun dereceli puanlama anahtarı hazırlama yaklaşımının kullanılması önemlidir (Karakaya, 2022, s. 55).

Açık uçlu maddelerin puanlanmasında puanlayıcı eğitimi verilmesi önemlidir. Puanlayıcı eğitiminin temel amacı puanlayıcıların sebep olduğu ya da olabileceği ve geçerliği olumsuz etkileyen varyansı azaltmak, en aza indirmek ya da kontrol altına almaktır (Brijmohan, 2016, s. 11). Bu amaca yönelik olarak kullanılacak puanlama ölçekleri puanlayıcılara tanıtılır (Fahim & Bijani, 2011), örnek uygulamalar yaptırılır (Cellar vd., 1989) ve bu sayede puanlayıcıların ortak bir anlayış geliştirmeleri amaçlanır (Alderson, Clapham, & Wall, 2005, s. 105).

Puanlayıcı Davranışları

Bireylerin puanlanması sürecinde bazı puanlayıcılar genel puanlama seyrine göre farklı puanlama davranışı sergileyebilirler. Bu durum ölçme sonuçlarının geçerliğini ve

güvenirliğini olumsuz etkilemektedir. Geçerliğı ve güvenilirliğı olumsuz yönde etkileyen bu davranışlar literatürde, puanlayıcı davranışları, problemli puanlayıcı davranışları, puanlayıcı etkisi, puanlayıcı hatası ve puanlayıcı yanlılığı olarak adlandırılmaktadır (Farrokhi vd., 2011; Haladyna, 1997, s. 139; Myford & Wolfe, 2004; Royal & Hecker, 2016). Bazı araştırmacılar puanlayıcı davranışları türlerini halo etkisi, merkeze eğilim, puanlayıcı katılığı, puanlayıcı cömertliğı, ranj sınırlaması, tutarsızlık ve yanlılık olarak ifade etmektedir (Myford & Wolfe, 2004). Bazı araştırmacılar da puanlayıcı davranışları türlerini cömertlik, halo, ilgi eksikliği, katılık, merkeze eğilim, yakınlık, yanıt hatası, kişiye özgü, mantıksal hata gibi isimlerle ifade etmektedir (Haladyna, 1997, s. 139). Literatürde sıklıkla gözlenen 30 puanlayıcı davranışından bahsedilmektedir (Royal & Hecker, 2016). Tablo 1’de literatürde yer alan puanlayıcı davranışları yer almaktadır.

Tablo 1
Puanlayıcı Davranışları

Davranış	Davranışın Özellikleri
Aktarma Etkisi	Bireyi puanlarken bir önceki bireyin performansından etkilenecek (olumlu/olumsuz) puan verme.
Amigo Etkisi	Performansı nasıl olursa olsun bireylere destekçi görünmek için tüm bireylere yüksek puan verme.
Ani Yıkım	Performansın, puanlayıcının olumsuz tepkisine neden olan yönlerinden dolayı daha düşük puan verme
Asimilasyon Etkisi	“Aşırı” görünmemek için diğer puanlayıcılarla benzer puanlama davranışı sergileme
Aşırılık	Dereceli puanlama anahtarının sadece uç noktalarını kullanma
Benzerlik Etkisi	Kendine benzeyen bireye daha yüksek puan verme
Bireysellik Etkisi	Bir özelliğe odaklanıp (güncellik, iletişim vb.) o özelliğe göre daha düşük/yüksek puan verme
Cömertlik	Sürekli yüksek/şişirme puan verme
Çelişen Standartlar	Kişisel standartları ile ölçme aracındaki standartlar arasındaki çelişkiler nedeniyle bireylere daha düşük puan verme
Çelişen Değerler	Performansta ifade edilen değerler ile puanlayıcının değerleri arasındaki çelişkiler nedeniyle daha düşük puan verme
Dış Bağlantı Etkisi	Bireyin başka kişilerle olan ilişkisine göre (düşük/yüksek) puan verme
Gestalt Olgusu	Birey performansı ayırt edilemediği için genel izlenime göre puan verme
Güvensizlik	Yeterli donanımına sahip olmadığını düşündüğü için yüksek puan verme eğiliminde olma
Hale Etkisi	Bireye yüksek puan verileceği algısı oluşturma
İlk İntiba/İlk İzlenim	Performansın genelini göz ardı edip ilk intibaya göre puan verme
İltimas	Çevrede çok sevilen bireyin bazı kusurlarını görmezden gelerek puan verme
Katılık	Sürekli düşük puan verme eğiliminde olma
Kişilik Çatışması	Bireyin bir özelliğini beğenmediği için düşük puan verme
Korkaklık	Bireyden farklı sebeplerden dolayı (meydan okuma, karşılık verme vs.) korkma ve bu yüzden yüksek puan verme
Merkeze Eğilim	Performanstan bağımsız olarak dereceli puanlama anahtarının orta kategorisini sık kullanma
Ölçeği Yanlış	Ölçeği ya da puanlamayı yanlış anladığı için uyumsuz puan verme
Yorumlama	
Öz Değerlendirme	Performans hakkında yeterli bilgiye sahip olmadığı için uyumsuz puan verme
Puanlayıcı Yetkinliği	Puanlayıcı yetkinliğinin az olması
Ranj Sınırlaması	Birey performanslarını ayırt edememe
Sapma	Zamanla daha az ya da daha çok puan verme eğiliminde olma
Sempati Etkisi	Performansın puanlayıcıyı duygusal olarak etkilemesinden dolayı ya da bireyin çabasının yüksek puanlanması
Süre Etkisi	Daha uzun performansla daha yüksek puan verme
Tekrarlama Etkisi	Aynı ya da benzer performansın sıklıkla gözlenmesi durumunda düşük puan verme
Tükenmişlik	Yorgunluktan performanstan bağımsız puanlama yapma
Yüzeysel Tarama	Birey performansının bir bölümüne göre puan verme

Johnson, R. L., Penny, J. A., & Gordon, B. (2009). *Assessing performance: designing, scoring, and validating performance tasks*. New York: Guilford Publications.

Wesolowski, B.C., Wind, S.A., Engelhard G. Rater fairness in music performance assessment: evaluating modeldata fit and differential rater functioning. *Music Science*, 19(2), 147–70.

Tablo 1 incelendiğinde bireylerin puanlanması sürecinde puanlayıcıların sergiledikleri birçok davranış olduğu görülmektedir. Bu davranışlardan bazıları kişilik özelliklerinden kaynaklanırken (hale etkisi, gestalt olgusu vb.) bazıları da ölçme aracından (ölçeği yanlış yorumlama, puanlayıcı yetkinliği vb.) kaynaklanmaktadır (Royal & Hecker, 2016). Katılık, cömertlik, farklılaşan katılık ve cömertlik, tutarsızlık gibi puanlayıcı davranışları yapılarından dolayı diğer puanlayıcı davranışlarına göre daha fazla ortaya çıkmaktadır

(Myford & Wolfe, 2003, 2004). Katılık ve cömertlik kişilik özelliklerinden bağımsız gerçekleşen bir eğilimdir (Guilford, 1954, s. 278).

Puanlayıcı davranışları, ölçülen yapı ile ilgisi olmayan varyansı artırabileceği için ölçümlerin geçerliği ve güvenilirliğinde etkisi çoktur. Bu yüzden, değerlendirme sürecinde puanlayıcı davranışlarının sebep olabileceği varyansın en aza indirilmesi gerekmektedir (Bachman, 2004, s. 149; Eckes, 2015, s. 189). Puanlayıcı davranışlarının belirlenmesi sürecinde birçok istatistiksel teknik yer almaktadır. Bu teknikler korelasyonel, grafiksel, regresyon, gizil sınıf modelleri ve çok yüzeyli modeller olarak ifade edilebilir. Ancak Madde Tepki Kuramı (Çok yüzeyli Rasch Modeli) ve Genellenebilirlik Kuramı yaklaşımları son yıllarda giderek popüler hale gelmiştir (Royal & Hecker, 2016).

En yaygın olarak görülen puanlayıcı davranışları puanlayıcı cömertliği, puanlayıcı katılığı, halo etkisi, merkeze eğilim ve ranj daralmasıdır (Engelhard, 2013, s. 212; Saal, Downey, & Lahey, 1980). Bu davranışlardan puanlayıcı katılığı, puanlayıcı cömertliği, merkeze eğilim, halo etkisi, farklılaşan katılık ve cömertlik (yanlılık) davranışları aşağıda açıklanmıştır.

Puanlayıcı Katılığı ve Puanlayıcı Cömertliği

Puanlayıcının diğer puanlayıcılara göre daha düşük puan verme ya da ölçütlerden düşük olanı seçme eğilimi puanlayıcı katılığı olarak ifade edilirken, puanlayıcının diğer puanlayıcılara göre daha yüksek puan verme ya da ölçütlerden yüksek olanı seçme eğilimi puanlayıcı cömertliği olarak ifade edilmektedir (Jackson, Schuler, & Werner, 2018, s. 317; Myford & Wolfe, 2004; Wu, 2017). Kısaca puanlayıcının ölçeğin orta noktasının altında puan verme eğilimini puanlayıcı katılığı, orta noktasının üstünde puan verme eğilimini puanlayıcı cömertliği şeklinde ifade etmek mümkündür (Saal vd., 1980). Puanlayıcının katılığı ya da cömertliği kişilik özelliğinden bağımsızdır ve birey, ölçüt, zaman gibi değişkenlerle değişebilmektedir (Guilford, 1954, s. 278; Lynch & McNamara, 1998; Wigglesworth, 1993). Puanlayıcının katı ya da cömert davrandığı durumlarda bireylerin sıralaması değişebilir, kesme noktasında olan bireyler için avantaj ya da dezavantaj durumu ortaya çıkabilir (McNamara, 1996, s. 10). Örneğin yetenek seviyeleri kesme noktasında olan iki bireyden birinin katı, diğerinin de cömert puanlayıcı tarafından puanlandığını düşünelim. Katı puanlayıcının puanladığı birey kalacakken, cömert puanlayıcı tarafından puanlanan birey geçecektir. Puanlayıcı davranışlarının karıştığı bu durum iki birey için verilen kararların geçerliğini büyük ölçüde olumsuz etkileyecektir. Bu yüzden birden fazla

puanlayıcının görev aldığı durumlarda puanlayıcı davranışlarının belirlenmesi ölçüm sonuçlarının geçerliğine ve güvenilirliğine katkı sağlayacaktır.

Puanlayıcı katılımı ve cömertliği davranışlarını azaltmak için bazı yöntemler önerilmektedir (Myford & Wolfe, 2003). Bu yöntemler ölçme aracı ve puanlayıcılar ile ilgilidir.

Ölçme aracı için yapılacaklar;

- Ölçülecek özellik detaylı bir şekilde açıklanmalı ve ölçekte yer alan her bir kategori için ortak tanımlamalar yapılmalıdır. Bu sayede puanlayıcı kategoriler arasındaki ayrımı kolay yapabilecek ve her bir kategorinin ne anlama geldiğini rahatlıkla anlayacaktır.
- Ölçme aracında yer alan kategorilerin yarısı olumlu yarısı olumsuz olacak şekilde tasarlanmalıdır.

Puanlayıcılar için yapılacaklar;

- Puanlayıcılar katılım ya da cömert davrandıklarında sonuçları ne kadar etkileyeceklerini anlayabilmeleri için eğitilmelidirler.
- Puanlayıcılardan ölçme aracının her kategorisine belli sayıda bireyi yerleştirmeleri istenmelidir.

Merkeze Eğilim

Puanlayıcılar, vermiş oldukları puanları sistemli olarak puanlama ölçeğinin alt kümesiyle sınırlandırır. Sınava giren bireylerin performansları doğru olarak ölçülemez. Bu durum birey performans değerlendirme sürecinde birey performanslarındaki değişikliğin az görünmesine sebep olur (Eckes, 2011; Myford & Wolfe, 2004). Merkeze eğilim davranışının ranj daralması özel bir durumdur (Myford & Wolfe, 2003; Saal vd., 1980). Merkeze eğilim davranışı, puanlayıcıların puanlama ölçeğinde genellikle orta kategoriyi ya da belirli kategorileri kullanma eğiliminde olmalarıdır (Engelhard, 1994; Wind, 2019; Wolfe, Jiao, & Song, 2015). Merkeze eğilim davranışı, Kingsbury (1922)'nin sınıflandırmış olduğu puanlayıcı davranışlarından biridir. Bu davranış, puanlayıcının puanlama ölçeğinin ortalama (mod/medyan) etrafındaki sınırlı bir alanda puan verme ve ölçeğin uç değerlerini kullanmaktan kaçınma eğilimi göstermesidir (Saal vd., 1980). Merkeze eğilim davranışı, puanlayıcının ölçeği kullanım şekli hakkında bilgi veren bir davranıştır (Engelhard, 1994). Bu davranışa sahip puanlayıcılar yüksek yeteneğe sahip bireyleri düşük puanlama

eğilimindeyken düşük yeteneğe sahip bireyleri de yüksek puanlama eğilimindedirler (Wind, 2019). Örneğin en düşük yetenek seviyesindeki bireyin 1, en yüksek yetenek seviyesinin de 5 puan aldığı ölçekte merkeze eğilim davranışına sahip puanlayıcı en yüksek yetenek seviyesindeki bireye 3 puan verme eğilimindeyken en düşük yetenek seviyesindeki bireye de 3 puan verme eğiliminde olacaktır. Puanlayıcıların sıklıkla ölçeğin belli noktalarını kullanmaları, puanlamanın ranjını belirgin bir şekilde azaltacağı için hem geçerliğe hem de güvenilirliğe zarar verir (Anastasi, 1988, s. 646).

Merkeze eğilim davranışının birçok sebebi olabilir (Akbari, Eigruber, & Vetschera, 2024; Allred, Crawford, Duffy, & Smith, 2016; Baird, Hayes, Johnson, Johnson, & Lamprianou, 2013; Myford & Wolfe, 2004; Niu, 2021, s. 42; Wolfe, 2004; Xiang, Graeber, Enke, & Gershman, 2021);

- Puanlayıcı ölçekte yer alan en yüksek ve en düşük performans ile ilgili yeterli bilgiye sahip olmadığı için ayırım yapamaz. Bu durumda puanlayıcı ölçeğin orta noktalarını kullanma eğiliminde olur.
- Kriterlerin net bir şekilde ifade edilmediği durumlarda merkeze eğilim davranışına sebep olabilir. Çünkü puanlayıcılar kriterler arasındaki farklılıkları göremez ve uygulayamazlarsa puanlama ölçeğinin orta aralığını kullanma eğiliminde olabilirler.
- Puanlayıcıların yeterince puanlama tecrübelerinin olmaması ve puanlama ölçeğinin kullanımı hakkında herhangi bir eğitim almamış olması merkeze eğilim davranışına sebep olabilir. Tecrübeli puanlayıcılarda merkezi eğilim davranışı daha az gözlenmektedir.
- Bazen puanlayıcılar uçlarda görünmemek (puanlayıcı katılığı/puanlayıcı cömertliği) için puanlama ölçeğinin orta noktalarını kullanma eğiliminde olabilirler.
- Merkeze eğilim davranışının bir diğer sebebi ise bilişsel yükün artmasıdır. Puanlayıcı tarafsız kalmak istediğinde ya da kararsızlık yaşadığı durumlarda ölçeğin orta noktasını kullanma eğiliminde olabilir.

Puanlayıcıların ölçeğin orta noktasını kullanma eğiliminde olması her zaman merkeze eğilim davranışına işaret etmez. Çünkü sınava giren bireyler orta düzey yeteneğe sahip olabilir ve birey yeteneklerinin gerçek dağılımı orta düzeyde olabilir. Araştırmacılar puanlama işleminde merkeze eğilim davranışının belirlenmesi için, ölçülen özelliğin ortalama puanının ölçeğin orta noktasına olan uzaklığını incelerler (DeCoths, 1977; Landy & Farr, 1983). Puanlamalarda merkeze eğilim davranışının belirlenmesinin bir diğer yolu ise merkeze eğilim davranışı sergileme potansiyeli olan puanlayıcının yapmış olduğu puanlamaları

incelemektir. Saal vd. (1980), puanlayıcının bireyleri puanlarken ölçekte yer alan kategorilerde hangi aralığı kullandığına bakılarak merkeze eğilim davranışı hakkında bilgi sahibi olmanın mümkün olduğunu ifade etmiştir. Bu işlem için ilk olarak tüm bireylerin testte yer alan maddelere ait ortalamaları hesaplanarak ortalama puan belirlenmesi gerekmektedir. Sonrasında da her madde için standart sapma belirlenir ve ortalamaya yakın ve standart sapması küçük olan puanlayıcıda merkeze eğilim davranışı vardır denilebilir.

Merkeze eğilim davranışını azaltmak ya da ortadan kaldırma için bazı stratejiler yer almaktadır. Bu stratejiler (Myford & Wolfe, 2003);

- Ölçme aracında yer alan kriterler için gerekli açıklamalar yapılmalıdır. Bu sayede puanlayıcılar her bir kriter hakkında net fikir sahibi olacak ve davranışları daha kolay ayırt edebilecektir. Ölçme aracındaki ardışık kriterlerin arasındaki farklar detaylıca tanımlanmalıdır.
- Ölçeğin uç kategorileri arasında, ölçeğin merkezinde yer alan kategorilere göre daha net farklılıklar olmalıdır.
- Ölçeğin uç kategorileri, puanlayıcıların tercih etmeyeceği kadar sıra dışı olmamalıdır. Çünkü az sayıda puanlayıcı uç kategorileri seçme eğilimindedir.
- Puanlayıcılar, merkezi eğilim davranışı hakkında bilgilendirilmelidir. Bu sayede puanlayıcılar bu davranıştan kaçınabilirler.
- Puanlayıcılardan ölçme aracında yer alan her bir kritere önceden belirlenmiş sayıda puan vermeleri şartı koyularak puanlama yapmaları istenebilir. Ayrıca puanlayıcılardan birey performanslarını sıralaması istenebilir. Bu durum puanlayıcıları puanlama yaparken tüm kriterleri kullanmaya teşvik eder.

Halo Etkisi

Puanlayıcı davranışları arasında en çok araştırma yapılan davranışlardan biri de halo etkisidir (Myford & Wolfe, 2003). Halo etkisi her ne kadar Wells (1907) tarafından tanımlanmış olsa da halo etkisi kavramı ilk olarak Thorndike (1920) tarafından ortaya atılmıştır. Halo etkisi, bireyi performansından bağımsız olarak genel izlenime göre puanlama olarak tanımlanmaktadır (Thorndike, 1920, s. 25). Engelhard (2002); Myford & Wolfe (2004), halo etkisini, puanlayıcıların farklı performanslara benzer puan verme eğilimi olarak ifade etmektedir. DeCoths (1977), halo etkisini puanlayıcının ölçülen özellik konusundaki yetersizliği ya da isteksizliği olarak açıklamaktadır. Puanlayıcılar ölçülen özellik ile ilgili

yeterince bilgiye sahip olmadığı durumlarda puanlamaları genel izlenime dayalı olarak yaparlar (Borman, 1977; DeCoths, 1977).

Halo etkisi davranışını gösteren puanlayıcı ölçekte yer alan kriterler arasında olması gerekenden fazla benzer ya da aynı puan verme eğiliminde olurlar (H. Kim, 2020) . Bir puanlamada halo etkisinin belirlenmesi için dört farklı yöntem önerilmektedir. Bu yöntemler (Saal vd., 1980);

- Ölçülen özelliğe ait kriterlerle ilgili puanlamalar arasındaki korelasyona bakmaktır. Korelasyonun yüksek çıkması ölçülen özellikler arasında ayırım yapılamadığını ve dolayısıyla halo etkisinin varlığını göstermektedir.
- Birey performans değerlendirme sürecinde puanlayıcıların vermiş oldukları puanların faktör analizi sonucunda, varyansın büyük kısmı az sayıda faktör altında toplanırsa bu durum halo etkisinin varlığını göstermektedir.
- Araştırmacılar, her puanlayıcının ölçülen özelliğe ait verdikleri puanlar arasındaki varyansı inceleyebilirler. İnceleme sonucunda puanlardaki küçük standart sapma veya varyans halo etkisine işaret etmektedir.
- *Puanlayıcı x birey* etkileşimi incelenerek varyans analizi gerçekleştirilebilir. Etkileşim istatistiksel açıdan anlamlıysa ve etkileşim puanlamalardaki varyansın büyük bir kısmını açıklıyorsa bu durum halo etkisine işaret etmektedir.

Halo etkisinin varlığının tespiti ile ilgili önerilen yaygın öneri ise çok yüzeyli rasch modelidir (Eckes, 2012, 2015; Engelhard, 1994; Myford & Wolfe, 2004). Halo etkisinin tespiti için çok yüzeyli rasch modelini kullanan araştırmalar oldukça fazladır (Farrokhi & Esfandiari, 2011; Knoch, Read, & von Randow, 2007; Kozaki, 2004; Schaefer, 2008). Puanlayıcıda halo etkisi olduğunda, puanlayıcının kriterlere verdiği puanlar tahmin edilebilir olduğundan model kestirimlerine aşırı uyum gösterirler. Bu sayede halo etkisinin büyüklüğü belirlenebilir (Kim, 2020). Halo etkisinin belirlenmesinde sabit etkili ki-kare istatistiği, ayırma indeksi, ayırma indeksinin güvenilirliği gibi istatistikler kullanılır.

Halo etkisi farklı sebeplerle ortaya çıkabilir. Örneğin, puanlayıcılar birkaç performans inceledikten sonra performans ile ilgili genel bir izlenime varıp sonraki puanlamalarda oluşturduğu izlenimden etkilenebilir. Bazı puanlayıcılar birey performansına odaklanmayı tamamen bırakırken bazı puanlayıcılar da puanlamalarını önceki puanlamalarıyla tutarlı hale getirme eğiliminde olabilirler. Önceden puanlanan bir bireyin standardı puanlamaları etkileyebilir. Ayrıca puanlayıcı ve bireyin birbirini tanıması da halo etkisinin ortaya

çıkmasındaki önemli bir etmendir (Jacobs & Kozlowski, 1985; Kozlowski, Kirsch, & Chao, 1986).

Halo etkisini azaltmak ya da tamamen ortadan kaldırmak için farklı öneriler bulunmaktadır. Bunlar (Cascio, 1982, s. 318);

- Kriterler açık, net, anlaşılır olmalı ve kriterler arasındaki ayrımlar net olmalıdır.
- Puanlayıcıların birden fazla özelliği puanladığı durumlarda zaman zaman ölçekte yer alan kriterlerin sıralaması değiştirilebilir. Örneğin 1 ile 5 arasında puanlanan ölçek, zaman zaman 5 ile 1 şeklinde puanlanacak şekilde kullanıma sunulabilir. Bu sayede puanlayıcının belli bir özelliğe sürekli benzer puan verme eğilimi ortadan kalkabilir.
- Puanlayıcılar halo etkisi ve halo etkisinin puanlamalar üzerindeki etkisi konusunda eğitilebilirler. Bu sayede puanlayıcılar bu davranıştan kaçınabilirler.
- Puanlayıcılar performans değerlendirme sürecinde birden fazla performans yer alsa bile her puanlayıcı öncelikle tüm bireylerin aynı performansını puanlamalıdır. Yani bir puanlayıcı, tüm bireylerin önce birinci performansını değerlendirmeli sonrasında da ikinci performansa geçmelidir.
- Puanlayıcıların puanlaması gereken özellik sayısı fazlaysa kriter sayısı azaltılabilir. Kriter sayısının azalması mümkün değilse her puanlayıcının farklı bir performansı değerlendirmesi sağlanabilir.
- Puanlayıcıların vermiş oldukları puanlar arasındaki ilişkilerin toplam puan üzerindeki etkisini azaltmak için kısmi korelasyon yaklaşımları kullanılmalıdır.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Puanlayıcı Yanlılığı)

Farklılaşan puanlayıcı katılığı ve cömertliği, puanlayıcının bazı öğrencileri puanlarken yüksek, bazı öğrencileri puanlarken de düşük puan verme eğiliminde olmasıdır (Knoch vd., 2007). Kumar (2005) puanlayıcı yanlılığını, puanlayıcının bireyin cinsiyet, yaş gibi özelliklerine göre bazı bireylere daha yüksek veya daha düşük puan verme eğiliminde olması şeklinde ifade etmektedir. Puanlayıcının bu davranışı sergilemesi ölçüm sonuçlarının geçerliğini ve güvenilirliğini olumsuz etkileyecektir.

Puanlayıcı Güvenirliđi

Eđitimde ve sosyal bilim arařtırmalarının merkezindeki kavramlardan biri olan guvenirlik, en basit tanımıyla olęme aracının tesadüfi hatalardan arınık olma düzeyidir. Arařtırma sonucunda elde edilen ölçümlerin guvenirliđi zorunluluktur (Downing & Yudkowsky, 2009, s. 11). Guvenirlik aynı puanlayıcının aynı testi farklı zamanlarda puanladığında aynı sonucu elde etmesidir. Guvenilir bir sınavda bireyler, puanlayıcı kim olursa olsun, ne zaman puanlarsa puanlasın test sonucunun aynı olmasını bekler. Ancak guvenilir olmayan sınavda bireylerin puanları farklı deđiřkenlere bađlı olarak deđiřebilmektedir (Moskal & Leydens, 2000).

Ölęme sonuçlarındaki hatalar, ölęme aracından, puanlayıcıdan, ölçülen yapıdan, ölęme zamanından ve bu kaynakların kendi aralarındaki etkileřiminden kaynaklanabilmektedir (Crocker & Algina, 2006, s. 105-106; Lord & Novick, 2008, s. 39; Tekin, 2013, s. 55; Turgut & Baykul, 2019, s. 120). Puanlayıcının geęerli ve guvenilir deđerlendirme yapabilmesi için amacın dıřındaki tüm varyans kaynaklarından uzak durması gerekmektedir (Craig, 1997, s. 8).

Bireylerin performanslarının deđerlendirilmesi sürecinde birden fazla puanlayıcının kullanılması önerilmektedir (Miller, Linn, & Gronlund, 2009, s. 251). Birden fazla puanlayıcının yer aldıđı deđerlendirme süreçlerinde önemli hata kaynaklarından biri de puanlayıcılardır. Bireyin bir performansını aynı puanlayıcı farklı zamanlarda puanlarsa puanların tutarlılıđı puanlayıcı içi guvenirlik kavramı ile ifade edilirken aynı bireyi birden fazla puanlayıcı puanladığında puanların tutarlılıđı puanlayıcılar arası guvenirlik kavramı ile ifade edilmektedir (Crocker & Algina, 2006, s. 104; Moskal & Leydens, 2000). Ancak birden fazla puanlayıcının yer aldıđı süreçlerde en önemli varyans kaynaklarından birisi puanlayıcı olsa da bařka varyans kaynakları da olabilmektedir. Bu yüzden sadece puanlayıcı içi veya puanlayıcılar arası tutarlılık puanlama guvenirliđi için yeterli deđildir. Olası varyans kaynaklarının ve bu kaynaklar arasındaki etkileřimlerin de göz önünde bulundurulması gerekmektedir. Bu kapsamda ölçmenin temel kuramlarından olan Klasik Test Kuramı (KTK), Genellenebilirlik Kuramı (GK) ve Madde Tepki Kuramı (MTK)'ye dayalı farklı puanlayıcı guvenirliđi belirleme yöntemleri yer almaktadır.

Klasik Test Kuramı Temelli Puanlayıcı Güvenirliği Belirleme Yöntemleri

Klasik Test Kuramı gerçek puan (T) ile hata puanı (E) toplamının gözlenen puana (X) eşit olduğu varsayımına dayanmaktadır ($X = T + E$). KTK'deki hata varyansına sebep olan davranışlardan biri de puanlayıcı davranışlarıdır. KTK temelli puanlayıcı güvenirliği belirleme yöntemleri araştırmanın amacına, yöntemine, araştırmada kullanılacak ölçek türüne ve puanlayıcı sayısına göre farklılık göstermektedir. KTK temelli puanlayıcı güvenirliği belirleme yöntemleri Tablo 2'de verilmiştir.

Tablo 2

KTK Temelli Puanlayıcı Güvenirliği Belirleme Yöntemleri

Yöntem	Ölçek Türü				Puanlayıcı Sayısı	
	Sınıflama	Sıralama	Eşit Oran	Eşit Aralık	2	3 ve fazlası
Pearson momentler çarpımı			+	+	+	
Spearman sıra farkları			+	+	+	
Weighed Kappa İstatistiği					+	
Cohen'in Kappa İstatistiği	+				+	
Fleiss Kappa İstatistiği	+					+
Kendall'in uyum katsayısı		+				+
Kendall τ		+			+	
Krippendorff'un Alpha	+	+	+	+	+	+
Sınıf İçi Korelasyon (ICC)		+	+	+	+	+
Goodman ve Kruskal'in γ		+			+	
Log-linear	+				+	+
Uyum yüzdesi	+				+	+

Kılıç, S. (2009). Ölçümlerin uyumluluğu ve tıptaki uygulamaları. Yüksek Lisans Tezi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ABD, Çukurova Üniversitesi" tarafından yapılan çalışmadan uyarlanmıştır.

Tablo 2'de görüldüğü gibi KTK temelli birçok puanlayıcılar arası güvenirlilik yöntemi bulunmaktadır. KTK temelli puanlayıcı güvenirliliğinin belirlenmesinde birden fazla yöntemin kullanılması önerilmektedir (Graham, Milanowski, & Miller, 2012). Ancak birden fazla yöntem kullanılsa bile KTK'nin puanlayıcı güvenirliliği kestirimlerinin sınırlılıklarını ortadan kaldırmak mümkün olmamaktadır. Çünkü (Barkaoui, 2008, s. 30; Haiyang, 2010; Weir, 2005, s. 31);

- KTK temelli yöntemlerde hata kaynağı tektir. Yani aynı anda birden fazla hata kaynağı göz önünde bulundurulmaz
- KTK'ye göre tüm hatalar tesadüfi hatalardır ve tek boyutludur. Yani sistematik ve tesadüfi hatayı birbirinden ayıramaz.
- KTK temelli yöntemler grup düzeyi hakkında bilgi verirken birey hakkında bilgi vermez.

Bu sınırlılığın ortadan kaldırılması için Cronbach ve arkadaşları genellenebilirlik kuramını (GK) öne sürmüşlerdir (Atılğan, 2005a).

Genellenebilirlik Kuramı Temelli Puanlayıcı Güvenirliği Yöntemleri

KTK'nin birden fazla hata kaynağını bir defada açık şekilde ayırt edemediği için Cronbach, Rajaratman ve Gleser 1963 ve 1965'te yayınladıkları makalelerde genellenebilirlik kuramı (G kuramı)'nın temellerini ortaya koymuşlardır. G kuramı, çoklu hata kaynaklarını ele alabilmektedir ve hem klasik test kuramının hem de varyans analizinin bir uzantısıdır (Güler, Kaya Uyanık, & Taşdelen Teker, 2012, s. 3).

G kuramının KTK'ye üstünlüğü tek bir analiz ile farklı hata varyanslarını aynı anda kestirme fırsatı vermesidir. Ölçme işlemindeki hata kaynaklarının sayısı kadar hata varyansı üretebilmektedir. Örneğin bir çalışmada hata kaynakları sayısı n ise bu çalışmada G kuramı modeli;

$$X = \mu_p + E_1 + E_2 + E_3 + \dots + E_n \quad \text{Eşitlik 1.}$$

şeklinde olacaktır. Eşitlikte “ μ_p ”, KTK'deki gerçek puanla (T) benzer yorumlanan evren puanıdır (Brennan, 2010).

G kuramı, daha güvenilir sonuçlar üretebilmek için ihtiyaç olan test formu, puanlayıcı sayısı gibi hata kaynaklarının da ne kadar olması gerektiği ile ilgili bilgi verir. G kuramı bağlı kararlar için KTK'deki güvenirlilik katsayısına benzer genellenebilirlik katsayısını üretir. Bu katsayı bireyin diğer bireyler arasındaki yerine bağlı olarak hesaplanır. G kuramı mutlak kararlar için ise KTK'ye ek olarak “phi (ϕ)” katsayısı üretmektedir. Phi katsayısı verilen puanların güvenirliliğini farklı durumlarda da kestirebilme imkânı verir. Bu katsayı, bireylerin puan sıralamasındaki ve ham puanlardaki tutarlılık derecesini verir. Belli bir kesme puanı ve üzerindeki puanların (uzman öğretmenlik sınavı, ehliyet sınavı vb.) önemli olduğu durumlarda kullanılması uygundur (Güler, 2010).

KTK ve G kuramının karşılaştırılması tablo 3'te verilmiştir.

Tablo 3

Klasik Test Kuramı ve Genellenebilirlik Kuramlarının Karşılaştırılması

Konu	Klasik test Kuramı	Genellenebilirlik Kuramı
Formlar ve Paralellik	Klasik paralel, eş biçimli, vb.	Rastgele paralel
Gerçek Puan	Formlar üzerinden kestirim	Rastgele paralel formlar üzerinden kestirim
Varsayımları	Zayıf	Zayıf
Avantajları	Basit ve kullanımı yaygın	Kavramsal genişlik, farklı varyans kaynaklarını ayırtabilirlik
Sınırlılıkları	Tek bir hata üretmesi, hataların ayrılması	Kavramsal yoğunluk
Kullanımı	Kolay	Genellikle kolay ancak bazı durumlarda zorlayıcı

Brennan, R. L. (2010). *Generalizability Theory and Classical Test Theory*. Applied Measurement in Education. 24, 1-21

G kuramının KTK'ye üstünlüğü bir örnekle açıklanacak olursa;

Birden fazla puanlayıcının yer aldığı performans değerlendirme sürecinde hata kaynakları birey, madde ve puanlayıcıdır. KTK bir katsayı üretirken G kuramı, birden fazla (birey, madde, puanlayıcı, birey- madde, birey- puanlayıcı, madde- puanlayıcı, birey- madde- puanlayıcı) katsayı üretecektir. G kuramının üreteceği katsayı yüzey sayısına bağlı olarak artabilecektir. Varyans kaynağının ve etkileşimlerin tam sayısına ulaşmak için “n” yüzey sayısı ve yüzey sayısı 2’den fazla olmak şartıyla;

$$2^n - 1$$

Eşitlik 2.

formülü kullanılabilir. Yani 4 yüzeyle bir araştırmada toplam etkileşim sayısı $2^4 - 1 = 16 - 1 = 15$ olacaktır.

G kuramı hata kaynaklarının ve hata kaynakları arasındaki etkileşim hakkında bilgi verse de hangi bireyin ya da hangi maddenin hata ürettiğine dair detaylı bilgi vermez. Daha detaylı bilgi elde edebilmek için madde tepki kuramı temelli yöntemlere başvurmak gerekmektedir.

Madde Tepki Kuramı Temelli Puanlayıcı Güvenirliği Yöntemleri

KTK temelli yöntemlerde madde istatistikleri (ayırta edicilik, madde güçlüğü vs.) testi alan bireylerin özelliklerine bağlıdır. Bu durum KTK'nın en önemli sınırlılıklarındandır (Hambleton, 1995, s. 84; McArthur, 1987, s. 130). Örneğin, iki farklı Türkçe testini alan bir grup bireyi düşünelim. Her iki testte de bireylerin yetenek kestirimleri farklı olacaktır. Ya

da Türkçe testini alan iki farklı grubu düşündüğümüzde de madde istatistikleri her iki grup için de farklı olacaktır. Testten yüksek puan alan bireylerin olduğu grupta madde güçlüğü düşük çıkacakken testten düşük puan alan bireylerin olduğu grupta madde güçlüğü yüksek çıkacaktır. KTK'nin diğer sınırlılıkları ise tüm varyans kaynakları için bir güvenilirlik değeri üretmesi ve birey yeteneklerini tam olarak kestirememesidir (Crocker & Algina, 2006, s. 143; Embretson & Reise, 2000, s. 53; Hambleton, Swaminathan, & Rogers, 1991). KTK'nin bu sınırlılıklarının ortadan kaldırılmasına yönelik olarak Frederic Lord ve George Rasch tarafından 1950'li yıllarda geliştirilen madde tepki modeli (MTK) olarak da ifade edilen örtük özellik teorisi (Domino & Domino, 2006, s. 116), bireyin örtük özellikleri ile maddeye verdiği cevap arasındaki ilişkiyi matematiksel modellere dayalı olarak açıklamaya çalışır (DeMars, 2010, s. 4; Urbina, 2004, s. 101). MTK temelli yöntemler ile kestirilen madde parametreleri, testi alan gruptan bağımsızken kestirilen birey özellikleri de maddelerden bağımsızdır (Bortolotti, Tezza, de Andrade, Bornia, & de Sousa Júnior, 2013; Hambleton, Swaminathan, Cook, Eignor, & Gifford, 1978; Ostini & Nering, 2006, s. 88). Ayrıca MTK testi alan tüm bireyler için ayrı standart hata üretir (Embretson & Reise, 2000, s. 40).

MTK'de boyut sayısı, kategori sayısı ve kategorilerin ölçek türüne göre farklı sınıflandırmalar yapmak mümkündür. Bu açıdan MTK, modeller topluluğu olarak ifade edilebilir (DeVellis, 2003). MTK modellerinden bazıları, madde güçlüğü, madde ayırt ediciliği ve şans parametrelerini içerirken (3 parametrelili model), bazıları madde güçlüğü ve madde ayırt ediciliği parametrelerini (2 parametrelili model) içerir. Diğer MTK modeli ise sadece madde güçlük parametresini içeren (1 parametrelili) modeldir. Bu model Rasch Modeli olarak da ifade edilmektedir. Ayrıca bu modelde madde ayırt ediciliği bir ve şans başarısı sıfır olarak kabul edilmektedir.

Linacre (1989), rasch modeline puanlayıcı etkisi gibi değişkenlik kaynaklarını ekleyip çok yüzeyli rasch modelini geliştirmiştir.

Çok Yüzeyli Rasch Modeli

Rasch modelinde birey puanlarını etkileyen değişkenlik kaynakları madde güçlüğü ve birey yetenekleridir. Rasch modelinde madde güçlükleri ve birey yetenekleri eşit aralıklı bir ölçek üzerinde bulunur ve üretilen sonuçlar örneklemden bağımsızdır (Elhan & Atakurt, 2005; Linacre, 1993b; Sudweeks, Reeve, & Bradshaw, 2004). Çok yüzeyli rasch modeli, Masters (1982) tarafından geliştirilen ve çok kategorili ölçme araçlarının değerlendirilmesini

sağlayan kısmi kredi modelinin bir uzantısı olarak değerlendirilmektedir (Myford, 2002). Bu model, ölçülecek özellikle ilgili tüm değişkenlerin doğal logaritmasını alarak, sıralama ölçeğindeki verileri eşit aralıklı logit cetveline aktaran doğrusal modeldir (Bond & Fox, 2015; Kaya Uyanık, Güler, Taşdelen Teker, & Demir, 2017; Linacre, 1993b, s. 60; Wolfe & Chiu, 1997). Ancak sıralı gözlemler ve ölçümler arasındaki ilişki doğrusal değildir (Linacre, 2023, s. 16). Çok yüzeyli rasch modelinde madde güçlükleri ve birey yetenekleri haricinde ek değişkenler yer almaktadır. Bu modelde, rasch modelinden farklı olarak birey puanlarını etkileyen tüm değişkenlik kaynakları (puanlayıcı, zaman, cinsiyet vb.) göz önünde bulundurulur (Baird vd., 2013, s. 40; Kim, Park, & Kang, 2012; Linacre, 1993a). Bu değişkenler varyans olarak da bilinse de çok yüzeyli rasch modelinde yer alan değişkenlere “yüzey” adı verilmektedir (Koons, 2008, s. 41; Sudweeks vd., 2004). Rasch modelinde olduğu gibi çok yüzeyli rasch modelinde de tüm yüzeyler eşit aralıklı ölçek üzerine yerleştirilebilir (Linacre, 1993b).

Çok yüzeyli rasch modeli, ulusal literatürde “çok yüzeyli rasch ölçme modeli, çok yüzeyli rasch ölçüm modeli, çok yüzeyli rasch modeli” olarak yer alırken (Çetin & İlhan, 2017; Güler, 2014; İlhan, 2016; Karakaya, 2015; Şata & Karakaya, 2020) uluslararası literatürde “many-facet conjoint measurement, FACETS model, multi-facet rasch, many-facet rasch, multi-facet rasch modeling” şeklinde yer almaktadır (Bahrouni, 2016; Eckes, 2015; Engelhard, 1994; Linacre, Engelhard, Tatum, & Myford, 1994; Lunz & Linacre, 1998, s. 47; McNamara, 1996).

Çok yüzeyli rasch modeli araştırmacıya üç tür bilgi sunar. Birincisi, bir testle ilgili değerlendirmelere istatistiksel bir çerçeve sunar (Myford & Wolfe, 2003). Bu sayede maddelerin hangi yetenek seviyesini temsil ettiği, bireylerin genel yetenek düzeyi ya da puanlayıcıların puanlama davranışları ile ilgili genel bir kanıya ulaşılabilir. İkincisi tüm madde tepki kuramı temelli modellerde olduğu gibi çok yüzeyli rasch modelinde de yüzeylere ait bağımsız kestirimler yapılabilir (Kaya Uyanık vd., 2017). Örneğin bir testi alan bireylere ait ölçüm sonuçları, maddelerin güçlük düzeyi, puanlayıcı davranışları (katılık, cömertlik vb.) gibi değişkenlik kaynaklarından bağımsız olarak kestirilebilir. Benzer şekilde bir testte yer alan maddelerin madde parametreleri, testi alan bireylerin ölçülen özelliğe ait yetenek seviyesi, puanlayıcı davranışları (katılık, cömertlik vb.) ve diğer değişkenlik kaynaklarından bağımsız olarak kestirilebilir (Engelhard, 1994). Üçüncüsü çok yüzeyli rasch modeli, modelde yer alan yüzeyleri bağımsız olarak inceleme fırsatı verdiği gibi yüzeyler arasındaki etkileşimleri de inceleme fırsatı verebilmektedir (Abu Kassim, 2007;

Linacre, 2023, s. 16; McNamara, 1996, s. 120). Birey, madde ve puanlayıcı yüzeylerini içeren bir model düşünüldüğünde bu modelde *birey x madde*, *birey x puanlayıcı*, *birey x puanlayıcı x madde* etkileşimleri incelenebilir. Örneğin testi alan bireylerin birey ve madde yüzeylerine ait etkileşimleri puanlayıcı yüzeyinden bağımsız olarak incelenebilir. Bu etkileşim, bireylerin testte yer alan bir madde veya madde grubuna diğer maddelerden farklı yaklaşımda bulunup bulunmadığı belirlemeye ve bu davranışın istatistiksel öneminin tahmin edilmesine imkân tanır. Benzer şekilde birey ve puanlayıcı yüzeylerine ait etkileşimleri de madde yüzeyinden bağımsız olarak incelenebilir. Bu etkileşim, puanlayıcıların testi alan bir bireye veya birey grubuna diğer bireylerden farklı yaklaşımda bulunup bulunmadığını belirlemeye ve bunun istatistiksel öneminin tahmin edilmesine imkân tanır (Linacre, 2023, s. 16). Ayrıca çok yüzeyli rasch modeli ile puanlayıcıların katılık, cömertlik, merkeze eğilim, halo etkisi ve yanlılık gibi davranışları da belirlenebilir (İlhan, 2016; Linacre, 2023, s. 359-360).

Birey ve maddeleri içeren çoktan seçmeli maddeler gibi iki kategorili (1-0) puanlamalara ait standart rasch modeli şu şekilde yazılabilir (Linacre, 2023, s. 17):

$$\log\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i \quad (2.1)$$

Formülde;

P_{ni} : n numaralı bireyin i numaralı maddede başarılı olma olasılığı

B_n : n numaralı bireyin yetenek düzeyi

D_i : i numaralı maddenin güçlüğü

şeklinde ifade edilmektedir.

Yukarıdaki formülde birey ve madde güçlüğü yüzeylerinin olduğu iki yüzeyli (birey ve madde), bir parametrelili (madde güçlüğü) rasch modeli yer almaktadır. Andrich (1978), bir parametrelili modelde yer alan matematiksel özellikleri koruyup geliştirerek tutum ölçekleri gibi çok kategorili puanlamalara uygun hale getirmiştir. İki yüzeyli çok kategorili puanlamalara ait rasch modeli şu şekilde yazılabilir (Linacre, 2023, s. 17):

$$\log\left(\frac{P_{nik}}{P_{ni(k-1)}}\right) = B_n - D_i - F_k \quad (2.2)$$

Formülde;

P_{nik} : n numaralı bireyin i numaralı maddede k puanı gözlenme olasılığı

$P_{ni(k-1)}$: n numaralı bireyin i numaralı maddede k-1 puan gözlenme olasılığı

B_n : n numaralı bireyin yetenek düzeyi

D_i : i numaralı maddenin güçlüğü

F_k : k-1 puanına göre k puanı almanın zorluğu

şeklinde ifade edilmektedir.

Yukarıdaki formülde birey ve madde güçlüğü yüzeylerinin olduğu iki yüzeyli (birey ve madde), çok kategorili puanlamalara ait rasch modeli yer almaktadır. Bu modelin birey ve maddeleri içeren çoktan seçmeli maddeler gibi iki kategorili (1-0) puanlamalara ait standart rasch modelinden farkı, birey puanı belirlenirken maddelerin güçlük düzeylerinin de hesaplama dahil edilmesidir.

Ancak araştırmalarda her zaman iki yüzey yeterli olmayabilir. Puanlayıcı davranışlarının incelendiği çalışmalarda birey ve madde yüzeylerine ek olarak puanlayıcı yüzeyi de eklenmektedir. Araştırma kapsamında ele alınan değişkenlik kaynağına göre (zaman, cinsiyet, vb.) yüzey sayısı da artabilir. Bu yüzden ikiden fazla yüzeyli modellere de ihtiyaç olabilmektedir. Üç yüzeye sahip çok kategorili puanlamalara ait standart rasch modeli şu şekilde yazılabilir (Linacre, 2023, s. 17):

$$\log\left(\frac{P_{nij k}}{P_{nij(k-1)}}\right) = B_n - D_i - C_j - F_k(2.3)$$

Formülde;

$P_{nij k}$: j numaralı puanlayıcının n numaralı bireyin i numaralı maddesine k puan verme olasılığı

$P_{nij(k-1)}$: j numaralı puanlayıcının n numaralı bireyin i numaralı maddesine k-1 puan verme olasılığı

B_n : n numaralı bireyin yetenek düzeyi

D_i : i numaralı maddenin güçlüğü

C_j : j numaralı puanlayıcının katılık ya da cömertlik düzeyi

F_k : k-1 puanına göre k puanı almanın zorluğu

şeklinde ifade edilmektedir.

Yukarıdaki formülde birey ve madde güçlüğü yüzeylerinin olduğu üç yüzeyli (birey, madde ve puanlayıcı), çok kategorili puanlamalara ait rasch modeli yer almaktadır. Bu modelin birey ve maddeleri içeren çoktan seçmeli maddeler gibi iki kategorili (1-0) puanlamalara ait standart rasch modelinden farkı, birey puanı belirlenirken maddelerin güçlük düzeyleri ve puanlayıcı katılık ve cömertliğinin de hesaplamaaya dahil edilmesidir. Yani temel rasch modelinin matematiksel özellikleri korunurken ölçüm durumuna göre bir ya da birden fazla değişken formüle eklenir (Linacre, 2023, s. 17). Mevcut araştırma kapsamında bir grup öğrencinin açık uçlu maddelere vermiş oldukları cevaplar bir grup puanlayıcı tarafından puanlanmıştır. Bu yüzden araştırmadaki yüzey sayısı üçtür (birey, madde ve puanlayıcı). Araştırmaya ait rasch modeli şu şekildedir:

$$\log\left(\frac{P_{bmpk}}{P_{bmp(k-1)}}\right) = B_b - D_m - C_p - F_k \quad (2.4)$$

Formülde;

P_{bmpk} : p numaralı puanlayıcının b numaralı bireyin m numaralı maddesine k puan verme olasılığı

$P_{bmp(k-1)}$: p numaralı puanlayıcının b numaralı bireyin m numaralı maddesine k-1 puan verme olasılığı

B_b : b numaralı bireyin yetenek düzeyi

D_m : m numaralı maddenin güçlüğü

C_p : p numaralı puanlayıcının katılık ya da cömertlik düzeyi

F_k : k-1 puanına göre k puanı almanın zorluğu

Yukarıdaki formülde araştırma kapsamındaki birey ve madde güçlüğü yüzeylerinin olduğu üç yüzeyli (birey, madde ve puanlayıcı), çok kategorili puanlamalara ait rasch modeli yer almaktadır. Araştırmada kapsamında yapılan işlemler oluşturulan model üzerinden yürütülmüştür. Çok yüzeyli rasch modeli, tek bir analizde tüm yüzeyleri bağımsız olarak analiz etmektedir. Yüzeyler, birbirinden bağımsız bir şekilde doğrusal ölçek üzerine yerleştirilir. Bu doğrusal ölçek logit cetveli olarak adlandırılmaktadır. Tüm yüzeylerin tek bir ölçek üzerine yerleştirilmesi ölçek üzerinde tüm yüzeylerin gözlenmesini sağlar. Birey,

madde ve puanlayıcı yüzeylerini içeren bir modele ait logit cetvelinde birey yeteneklerini, madde güçlüklerini ve puanlayıcı davranışlarını gözlemek mümkündür. Ayrıca analiz sonucunda logit değeri, standart hata ve uygunluk istatistikleri gibi bireysel istatistikler üretilmektedir (Linacre, 2023, s. 16; Myford & Wolfe, 2003). Bireysel istatistiklerden ilki olan *logit değeri*, sıralı gözlemlerden elde edilen doğrusal bir ölçü birimidir ve öğeleri kalibre edip kişilerin ölçülmesini sağlar (Linacre, 2023, s. 309, 317). Bir başka deyişle logit değeri, sıralı gözlemlerin eşit aralıklı ölçeğe dönüştürülmüş halidir. Logit değerleri eşit aralıklı ölçekte yer aldıkları için bu değerler arasında dört işlem yapılabilir. Çok yüzeyli rasch analizi sonucunda logit değerleri logit cetvelinde yer alır. Analizde yüzeyler pozitif olarak tanımlanırsa, logit değeri arttıkça birey yeteneği ve madde güçlüğü artar. Ayrıca puanlayıcının pozitif logit değerine sahip olması cömert davrandığına, negatif logit değerine sahip olması da katı davrandığına işaret etmektedir. İkinci bireysel istatistik ise standart hatadır. Sıralı gözlemler eşit aralıklı hale dönüştürülürken dönüşümün anlamlı bir etkiye sahip olabilmesi için gereken minimum değere standart hata adı verilir (Linacre, 2023, s. 320). Çok yüzeyli rasch modelinde her logit değeri için ayrı bir standart hata değeri kestirilmektedir. Logit değeri ve standart hata puanlayıcı davranışlarının belirlenmesinde kullanılır. Puanlayıcıların bireysel düzeyde katı ya da cömertliklerini belirlemek için standart bir logit değeri aralığı olmadığı için logit değerleri t-değerine dönüştürülmektedir (Myford & Wolfe, 2003). Logit ölçüsünün t- değerine dönüştürülmesi için kullanılan denklem:

$$t\text{- değeri} = \frac{\text{puanlayıcıya ait logit değeri} - \text{logit değerlerinin ortalaması}}{\text{logit değerinin standart hatası}} \quad (2.5)$$

şeklindedir.

Hesaplanan t değeri, t dağılım tablosundaki t değerinden büyük yani istatistiksel olarak anlamlı ise puanlayıcıda katılık ya da cömertlik davranışı olduğu söylenebilir. Hesaplanan t değerinin işaretinin pozitif olması puanlayıcı cömertliğini, negatif olması ise puanlayıcı katılığına işaret etmektedir.

Üçüncü bireysel istatistik ise uygunluk istatistikleridir. Uygunluk istatistikleri, gözlenen puan ile beklenen puan arasındaki farktır ve sıra dışı gözlemlerin belirlenmesini sağlar (Linacre, 2023, s. 315). Bir başka deyişle uygunluk istatistikleri, veri setinin tasarlanan modele ne kadar uyum sağladığını belirlemeyi sağlar. Uygunluk istatistikleri standartlaştırılmış artıkların kareleri ile hesaplanır (Myford & Wolfe, 2003). Gözlenen puan ile beklenen puan arasındaki farka artık adı verilmektedir. Bu değer büyük olması, model veri uyumunun iyi olmadığını göstergesidir. Standart artıklar, uygunluk istatistiklerine göre

örneklem büyüklüğüne daha fazla duyarlı oldukları için model veri uyumunun incelenmesinde standart artıklar yerine uygunluk istatistiklerinden yararlanılması önerilir (Brentari & Golia, 2008). Uygunluk istatistikleri uygunluk içi (infit) ve uygunluk dışı (outfit) istatistikleri olmak üzere iki çeşittir. Uygunluk içi (infit) istatistiği, modelde yer alan yüzeyin performansı hakkında bilgi veren uygunluk istatistiğidir (Linacre, 2023, s. 316). Uç değerlere duyarlı olan uyum istatistiğine ise uygunluk dışı istatistiği adı verilmektedir (Linacre, 2023, s. 318). Uygunluk dışı istatistiği uç değerlere daha hassas olduğu için model veri uyumunda uygunluk içi istatistiğinin incelenmesi önerilmektedir (Bond & Fox'dan ileten İlhan, 2015, s. 70). Uygunluk içi ve uygunluk dışı istatistiklerinin hesaplamalarındaki temel fark, uygunluk istatistikleri hesaplanırken doğru yanıt verme olasılığı %50'ye yakın olan bireylere daha yüksek ağırlıklandırma yapılmasıdır (Liu, 2010, s. 52). Uygunluk içi ve uygunluk dışı istatistikleri 0 ile $+\infty$ aralığında değerler alabilirken beklenen değeri birdir (Engelhard, 1994; Myford & Wolfe, 2003). Uygunluk istatistiklerinin bire eşit olması mükemmel uyuma işaret etmektedir. Ancak gerçek pratikte yapılan ölçme durumları için mükemmel uyumun görülmesi neredeyse imkansızdır (Brentari & Golia, 2008). Bu yüzden araştırmacılar uygunluk istatistiklerinin kabul edilebilir aralığı ile ilgili farklı görüşler ortaya atmışlardır. Bazı araştırmacılar uygunluk istatistiklerinin kabul edilebilir aralığının 0,5 ile 1,5 arasında olduğunu ve kabul edilebilir aralığın dışında olan birey, madde ya da puanlayıcının modelden çıkarılmasını önermektedir (Brentari & Golia, 2008; Myford & Wolfe, 2003; Wright & Linacre, 1994). Lunz, Wright, & Linacre (1990), uygunluk istatistikleri için kabul edilebilir aralığın 0,6 ile 1,5 olduğunu belirtirken Wright & Linacre (1994), uygunluk istatistikleri için kabul edilebilir aralığın 0,6 ile 1,4 arası olduğunu, 1,5 ve üzerindeki uygunluk istatistiklerinin ise ölçüm için uygun olmadığını belirtmiştir. Myford & Wolfe (2003), 2'ye kadar olan uygunluk istatistiklerinin kabul edilebilir aralıkta olduğunu ve 1,5 ile 2 arasındaki değerlerin ölçüm için zararlı ya da faydalı olmayacağını belirtmiştir. (Sudweeks vd., 2004), uygunluk istatistiklerinin 2'den büyük olmasının ölçüm için zararlı olduğunu ifade etmiştir. Uygunluk istatistikleri ile ilgili görüşler incelendiğinde istatistiklerin 0,5'ten küçük, 2'den de büyük olmaması gerektiği görülmektedir. Uygunluk içi ve uygunluk dışı istatistiklerinin 0,5'ten küçük olması ya da 2'den büyük olması farklı anlamlar ifade etmektedir. Araştırma kapsamında incelenen herhangi bir yüzeye ait bir bileşenin uygunluk istatistiklerinin 0,5'ten küçük olması, bu bileşenin incelenen yüzeye ait diğer bileşenlere göre farklı bir bilgi sağlamadığını ve yüzeye ait bazı alt bileşenler arasında yüksek korelasyon olduğu anlamına gelmektedir (İlhan, 2015). Örneğin puanlayıcı yüzeyini oluşturan bazı puanlayıcılara ait uygunluk istatistiklerinin 0,5'ten küçük olması, bazı

puanlayıcıların testte yer alan maddeleri bağımsız olarak puanlayamadığı anlamına gelmektedir (Engelhard, 2011). Benzer şekilde madde yüzeyini oluşturan bazı alt bileşenler arasındaki korelasyon yüksek olduğunda uygunluk istatistikleri 0,5'ten düşük çıkmakta ve bu durum yerel bağımsızlık varsayımını ihlal etmektedir (Chan, Chien, Chen-Yi, & Shun-Jin, 2009). Araştırma kapsamında değerlendirilen yüzeye ait bazı alt bileşenlerde uygunluk istatistiklerinin 2'den büyük olması bazı alt bileşenlerin diğer alt bileşenlerle aynı yapıyı ölçmediği anlamına gelmektedir. Örneğin madde yüzeyini oluşturan bazı alt bileşenlere ait uygunluk istatistiklerinin 2'den büyük olması, bu bileşenlerin diğer bileşenlerle aynı özelliği ölçmediği anlamına gelmektedir. Bu durum tek boyutluluk varsayımını ihlal edebilir. Puanlayıcılara ait uygunluk istatistiklerinin kabul edilebilir aralığın dışında olması puanlayıcıların aşırı tutarlı ya da aşırı tutarsız puanlama yaptıklarına işaret etmektedir (Moore, 2009). İlgili literatür incelendiğinde araştırma kapsamında uygunluk istatistikleri için kabul edilebilir aralık 0,5 ile 2 olarak benimsenmiştir.

Çok yüzeyli rasch modeli sadece bireysel düzeyde istatistikler üretmez, aynı zamanda her bir yüzey için ki-kare istatistiği (fixed/all same chi-squared, random/normal chi-squared), ayırma istatistiği (ayırma indeksi/separation index, ayırma oranı/separation ratio ve güvenilirlik indeksi/the reliability of separation index) üretir. Her bir yüzey için yüzeye ilgili istatistikler de hesaplanmaktadır. Örneğin madde yüzeyi için kategori istatistikleri, puanlayıcı yüzeyi için puanlayıcı güvenilirlik istatistiği ve yanlılık için de yanlılık indeksi hesaplanmaktadır.

Grup istatistiklerinden sabit etkili ki-kare istatistiği (fixed/all same chi-squared), ölçme hatası göz önünde bulundurulduğunda, öğenin bileşenlerinin benzer özelliklere sahip olup olmadığını test eder (Linacre, 2023, s. 216). Bir başka deyişle, ki-kare istatistiği bir yüzeye ait öğeler arasındaki farkın istatistiksel olarak değerlendirilmesini sağlar (Engelhard, 1994). Sabit etkili ki-kare testinin hipotezi, yüzeye ait bileşenlerin istatistiksel olarak birbirine benzer olduğudur. Örneğin puanlayıcı yüzeyine ait sabit etkili ki-kare istatistiği, puanlayıcıların katılıklarının benzer olup olmadığını test ederken, madde yüzeyine ait ki-kare istatistiği madde güçlüklerinin benzer olup olmadığını, birey yüzeyine ait ki-kare istatistiği de birey yetenek düzeylerinin benzer olup olmadığını test eder. Sabit etkili ki-kare istatistiğinin istatistiksel açıdan anlamlı olması her yüzey için farklı anlam ifade etmektedir. Örneğin puanlayıcı yüzeyi için anlamlı ki-kare istatistiği, puanlayıcıların istatistiksel olarak birbirlerinden farklı davranışlar sergilediğine işaret ederken, madde yüzeyi için anlamlı ki-kare istatistiği, madde güçlüklerinin istatistiksel açıdan fark olduğu anlamına gelir. Birey

yüzeyi için anlamlı ki-kare istatistiği ise birey yetenek düzeylerinin istatistiksel açıdan farklı olduğuna işaret eder. Çok yüzeyli rasch modeli çalışmalarında bir yüzeye ait ki-kare istatistiğinin yanında diğer istatistiklerin de sunulması gerekir. Çünkü ki-kare istatistiği örneklem büyüklüğüne oldukça hassastır (Myford & Wolfe, 2004; Weigle, 1998). Sabit etkili ki-kare istatistiğinin hesaplama formülü (Linacre, 2023, s. 216);

$$w_i = \frac{1}{SH_i^2}, \quad i = 1, 2, 3, \dots, L \quad \text{olmak üzere;} \quad (2.6)$$

$$\chi^2 = \sum (w_i \cdot D_i^2) - \frac{\sum (w_i \cdot D_i)^2}{\sum (w_i)} \quad (2.7)$$

şeklindedir.

Formülde yer alan değişkenlerin türü hesaplanacak yüzeye göre değişmektedir. Örneğin madde yüzeyi için yapılacak hesaplamada D, madde gücüğü iken L, madde sayısıdır. Puanlayıcı yüzeyi için yapılacak hesaplamada D, puanlayıcının katılık ölçüsü iken L, puanlayıcı sayısıdır.

Grup istatistiklerinden hesaplanan bir diğer ki-kare istatistiği normal (random/normal chi-squared) ki-kare istatistiğidir. Bu istatistik yüzeylerin normal dağılıma sahip dağılımdan rastgele seçilmiş bir örneklem olup olmadığını test eder. Hipotezin anlamlı olmaması beklenir. Yani seçilen örneklemin evreni temsil eden normal dağılıma sahip bir dağılıma sahip olması beklenir. Normal ki-kare istatistiğinin hesaplama formülü (Linacre, 2023, s. 216);

$$\text{var}(D) = \frac{S(D_i - D_{\text{mean}})^2}{L-1} - \frac{(SSH_i^2)}{L} \quad (2.8)$$

$$w_i = \frac{1}{(\text{var}(D) + SH_i^2)} \quad (2.9)$$

$$\chi = \frac{S(w_i \times D_i^2) - (S w_i \times D_i)^2}{S w_i} \quad \text{serbestlik derecesi} = L - 2 \quad (2.10)$$

şeklindedir.

Formülde sabit etkili ki-kare testinde olduğu gibi değişkenler yüzeye göre adlandırılmaktadır ve adlandırma işlemi sabit etkili ki-kare istatistiği ile aynıdır.

Grup düzeyinde hesaplanan diğer istatistikler de ayırma istatistikleridir. Bu istatistiklerden birincisi *ayırma oranıdır*. Ayırma oranı, yüzeye ait hesaplanan ölçümlerin kesinliklerine göre yayılmasının bir ölçüsüdür ve +1 ile $+\infty$ arasında değer alır (Linacre, 2023, s. 216; Sudweeks vd., 2004). Bir yüzey için hesaplanan ayırma oranı, yüzeyin bileşenleri arasındaki

farkın hesaplanan hatadan ne kadar farklı olduğunu gösterir. Ayırma oranı “G” harfi ile gösterilir (Eckes, 2015, s. 62). Ayırma oranının hesaplama formülü (Linacre, 2023, s. 216);

$$\text{Gerçek Standart Sapma} = \text{Gözlenen Standart Sapma} - \text{Standart Hata Ortalaması} \quad (2.11)$$

$$G = \frac{\text{Gerçek Standart Sapma}}{\text{RMSE}} \quad (2.12)$$

Grup düzeyinde hesaplanan ayırma istatistiklerinden bir diğeri de ayırma oranı ile hesaplanan *ayırma indeksi*dir. Ayırma indeksi, tahminlerin yayılmasının bir ölçüsüdür (Linacre, 2023, s. 216). Yani yüzeye ait bileşenlerin kaç düzeyden meydana geldiğini göstermektedir. Örneğin, birey yüzeyine ait hesaplanan ayırma indeksinin 9,89 olması birey yeteneklerinin istatistiksel olarak 10 farklı düzeyde olduğunu gösterirken puanlayıcı yüzeyine ait hesaplanan ayırma indeksinin 3,83 olması puanlayıcıların katılık cömertlik durumları açısından istatistiksel olarak dört farklı düzeyde olduğunu gösterir. Madde yüzeyine ait hesaplanan ayırma indeksinin 2,77 olması ise madde güçlüklerinin istatistiksel olarak 3 farklı düzeyde olduğunu gösterir. Ayırma indeksi “H” harfi ile gösterilir ve hesaplanırken kullanılan formül (Eckes, 2015, s. 63; Linacre, 2023, s. 216);

$$H = \frac{(4 \times \text{Separation}) + 1}{3} \quad (2.13)$$

şeklindedir.

Grup düzeyinde hesaplanan diğeri istatistik de *ayırma indeksinin güvenilirliği*dir. Ayırma indeksinin güvenilirliği, yüzeye ait bileşenlerin ne kadar iyi ayrıldığı hakkında bilgi verir (Engelhard, 1994). Bir başka deyişle ayırma indeksinin güvenilirliği, KR-20 veya Cronbach alfa istatistiğinin rasch modelindeki karşılığıdır (Bond & Fox, 2015, s. 80; Engelhard, 1994; Linacre, 2023, s. 362; Myford & Wolfe, 2003). Ayırma indeksi güvenilirliği 0 ile 1 arasında değer alır. Ayırma indeksi güvenilirliğinin yüksek olması (1’e yakın) birey ve madde yüzeyi için istenen durum olsa da puanlayıcı yüzeyi için düşük olması (0’a yakın) istenir. Çünkü puanlayıcı yüzeyine ait bileşenlerin ayırma indeksi güvenilirliğinin yüksek olması, ilgisiz varyansın da yüksek olmasına işaret etmektedir (Sudweeks vd., 2004). Puanlayıcı eğitimi ile ilgili çalışmalarda puanlayıcı davranışlarının benzer olmasının beklendiği düşünüldüğünde, ayırma indeksi güvenilirliğinin yüksek olması (1’e yakın olması) puanlayıcıların benzer davranışlara sahip olmadığı anlamına gelmekte olup istenen bir durum değildir. Ayırma indeksi güvenilirliği hesaplanırken kullanılan formül;

$$R = \frac{G^2}{1 + G^2} \quad (2.14)$$

şeklindedir.

Puanlayıcı davranışlarının incelendiği çok yüzeyli rasch modelinin kullanıldığı araştırmalarda incelenen bir diğer grup istatistiği de *yanlılık indeksi*dir (Engelhard, 2002). Araştırma kapsamındaki çok yüzeyli rasch modeline (denklem 2.4) *puanlayıcı x birey* etkileşiminin eklenmesiyle yanlılık modeli oluşturulmuş olacaktır. Araştırma kapsamındaki yanlılık için oluşturulan yeni model;

$$\log\left(\frac{P_{bmk}}{P_{bmk-1}}\right) = B_b - D_m - C_p - F_k - I_{pb} \quad (2.15)$$

şeklinde olacaktır. Denklem 2.15'te yer alan modelde I_{pb} , *puanlayıcı x birey* etkileşime ait yanlılık terimidir.

Çok yüzeyli rasch modelinde bazı istatistikler sadece bireysel ya da sadece grup düzeyinde hesaplanırken bazı istatistikler hem bireysel hem de grup düzeyinde hesaplanabilir. Hem bireysel hem de grup düzeyindeki istatistikler uyum yüzdeleri ve kategori istatistikleridir.

Puanlayıcılar arası güvenirliliğin çok yüzeyli rasch modeli ile analizi yapılırken her bir puanlayıcı için gözlenen ve beklenen uyum yüzdeleri hesaplanır. Bu istatistik hem bireysel hem de grup düzeyinde bir istatistiktir. Bir puanlayıcının beklenen ve gözlenen uyum yüzdeleri eşit olması, puanlayıcının puanlayıcı davranışı göstermediği ve mükemmel puanlama yaptığını belirtir (Eckes, 2015, s. 91). Bir puanlayıcının gözlenen ve beklenen puan arasındaki fark sıfır olmasa da birbirine yakın olması istenen bir durumdur. Puanlayıcıya ait gözlenen ve beklenen puanların uyum yüzdesi Rasch -Cohen's Kappa istatistiği ile hesaplanır. Rasch -Cohen's Kappa istatistiği hesaplanırken kullanılan formül (Linacre, 2023, s. 326);

$$\text{Rasch - Cohen's Kappa} = \frac{\text{gözlenen uyum yüzdesi} - \text{beklenen uyum yüzdesi}}{100 - \text{beklenen uyum yüzdesi}} \quad (2.16)$$

şeklindedir. Rasch -Cohen's Kappa istatistiğinin beklenen değeri 0,00'dır. Bu değer in sıfırdan büyük olması, yüksek puanlayıcılar arası uyuma işaret ederken sıfırdan küçük olması, puanlayıcılar arasında düşük uyum olduğunu göstermektedir. Rasch -Cohen's Kappa istatistiği -1 ile +1 arasında değerler alır (Eckes, 2015, s. 92).

Hem bireysel hem de grup düzeyinde hesaplanan diğer istatistik *kategori istatistikleridir*. Performans değerlendirmede kullanılan ölçekteki kriterlerin uygunluğuna karar vermek için kategori istatistikleri kullanılır. Ayrıca puanlayıcılarda merkeze eğilim ve ranj daralması gibi

davranışların olup olmadığının tespit edilmesinde kategori istatistiklerinden faydalanılır (Myford & Wolfe, 2004).

Çok yüzeyli rasch modeli analizleri yapılabilmesi için bazı varsayımları karşılaması gerekmektedir. Çok yüzeyli rasch modeli rasch modelinin uzantısı olduğundan rasch modelinin varsayımları karşılaması gerekir (Eckes, 2015, s. 124; Farrokhi vd., 2011; Farrokhi, Esfandiari, & Schaefer, 2012). Bu varsayımlar, model veri uyumu, tek boyutluluk ve yerel bağımsızlıktır. Mevcut araştırma kapsamında elde edilen verilerin varsayımları sağlayıp sağlamadığının test edilmesi yöntem bölümünde yer aldığı için çok yüzeyli rasch modeline ait varsayımlar yöntem bölümünde ele alınmış ve raporlanmıştır.

Çok Yüzeyli Rasch Modeli Puanlama Desenleri

Birden fazla puanlayıcının yer aldığı performans değerlendirme sürecinde, puanlayıcıların tüm sınavları değerlendirip değerlendirmeyeceği, puanlayıcıların kaç görevi puanlayacağı, her puanlayıcının kaç bireyi puanlayacağı gibi sorular ortaya çıkmaktadır. Bu yüzden birden fazla puanlayıcının yer aldığı değerlendirmelerde puanlama deseni önemlidir (Eckes, 2015). Burada dikkat edilmesi gereken şey tek bir referans çerçevesinde belirsizlik olmadan tüm yüzeyler arasında etkileşimin olmasıdır (Linacre, 1989, s. 13). Literatür incelendiğinde puanlama desenlerinin farklı adlandırıldığı ya da farklı puanlama desenlerinin yer aldığı görülmektedir. Schumacker (1999) puanlama desenlerini çaprazlanmış, yuvalanmış ve karma desen olarak ifade ederken, Linacre (2014b), çaprazlanmış, yuvalanmış, karma ve puanlayıcı çifti deseni olarak ifade etmiştir. Eckes (2015) ise tam çaprazlanmış desen, tamamlanmamış bağlantılı desen, tamamlanmamış bağlantılı olmayan desen ve tamamlanmamış bağlantılı sarmal desen olarak ifade etmiştir.

Çaprazlanmış Desen

Tümüyle çaprazlanmış desenlerde, puanlayıcıların tamamı tüm bireyleri ve tüm maddeleri puanlar. Bu sayede tüm parametrelerin kesin ve doğru ölçülmesine imkân tanır (Linacre, 1989, s. 14). Beş bireyin (B1, B2, B3, B4 ve B5) üç görevini (A, B ve C) dört puanlayıcının (J1, J2, J3, J4) puanlandığı ve tümüyle çaprazlanmış puanlama deseni ile desenlenen bir puanlama sürecinde bireyler ve puanlayıcı örnek dağılımı tablo 4'te verilmiştir.

Tablo 4
Tümüyle Çaprazlanmış Desen

Görev Birey	Puanlayıcılar											
	J1			J2			J3			J4		
	A	B	C	A	B	C	A	B	C	A	B	C
B1	x	x	x	x	x	x	x	x	x	x	x	x
B2	x	x	x	x	x	x	x	x	x	x	x	x
B3	x	x	x	x	x	x	x	x	x	x	x	x
B4	x	x	x	x	x	x	x	x	x	x	x	x
B5	x	x	x	x	x	x	x	x	x	x	x	x

Schumacker, R. E. (1999). Many-facet Rasch analysis with crossed, nested and mixed designs. *Journal of Outcome Measurement*, 3(4), 323-338.

Tümüyle çaprazlanmış puanlama deseni parametrelerin doğru ölçülmesinde en etkili yöntem olsa da görev sayısının çok olduğu, zaman sıkıntısının olduğu ya da geniş ölçekli değerlendirmelerde ekonomik değildir.

Yuvalanmış Desen

Uygulamadaki sorunlardan dolayı her zaman tümüyle çaprazlanmış desenlerin kullanılması mümkün olmamaktadır. Yuvalanmış desenler, sınırlı zaman, geniş katılımlı sınavlarda puanlayıcıların yükünü azaltmaktadır. Bu durum yuvalanmış deseni kullanılabilir hale getirmektedir (Brennan, 1983). Farklı yuvalanmış desenler yer alsın da en sık kullanılan desen, maddelerin puanlayıcılar içinde yuvalandığı bireylerin ise maddeler ve puanlayıcılarla çaprazlandığı yuvalanmış desendir. Bu desende tüm bireyler tüm maddeleri yanıtlar, tüm puanlayıcılar tüm bireyleri puanlar ancak puanlayıcılar testte yer alan farklı maddeleri puanlamaktadırlar (Doğan & Anadolu, 2017). Beş bireyin (B1, B2, B3, B4 ve B5) dört görevden (A, B, C ve D) oluşan sınavının, dört puanlayıcı (J1, J2, J3 ve J4) tarafından puanlandığı ve yuvalanmış desen ile desenlenen bir değerlendirme sürecinde bireyler ve puanlayıcı örnek dağılımı tablo 5'te verilmiştir.

Tablo 5
Yuvalanmış Desen

Görev Birey	Puanlayıcılar															
	J1				J2				J3				J4			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
B1	X	X			X	X					X	X			X	X
B2	X	X			X	X					X	X			X	X
B3	X	X			X	X					X	X			X	X
B4	X	X			X	X					X	X			X	X
B5	X	X			X	X					X	X			X	X

Schumacker, R. E. (1999). Many-facet Rasch analysis with crossed, nested and mixed designs. *Journal of Outcome Measurement*, 3(4), 323-338.

Araştırma puanlayıcı davranışlarının belirlenmesi amaçlandığı için tüm parametrelerin kesin ve doğru ölçülmesi gerekmektedir. Bu yüzden araştırma kapsamında tümüyle çaprazlanmış desen kullanılmıştır.

Puanlayıcı Davranışlarını Benzer Hale Getirmeye Yönelik Uygulamalar

Birden fazla puanlayıcının yer aldığı değerlendirme sürecinde en büyük sınırlılık puanlayıcı güvenilirliğidir. Literatürde bu davranışları ortadan kaldırmak ya da azaltmak için bazı öneriler yer almaktadır. Bu öneriler; birden fazla puanlayıcı kullanma (Gronlund, 1977, s. 85; Kubiszyn & Borich, 2013, s. 170), dereceli puanlama anahtarı kullanma (Dunbar vd., 2006; Ebel & Frisbie, 1991, s. 194; Kutlu vd., 2014, s. 51; Oosterhof, 2003, s. 81) ve puanlayıcı eğitimidir (Bernardin & Buckley, 1981; Haladyna, 1997, s. 143; Hauenstein & McCusker, 2017; İlhan & Çetin, 2014; Lumley & McNamara, 1995). Bu çalışmada da puanlayıcı davranışlarını benzer hale getirmek için birden fazla puanlayıcıdan, dereceli puanlama anahtarından ve puanlayıcı eğitiminden faydalanılmıştır. Bu kapsamda bu yöntemler açıklanmıştır.

Birden Fazla Puanlayıcı Kullanılması

Açık uçlu maddelerin değerlendirilmesi sürecinde birden fazla puanlayıcının yer alması önemlidir. Puanlayıcı davranışlarının benzer hale getirip objektif değerlendirmenin sağlanması için en basit yol birden fazla puanlayıcının kullanılmasıdır (Gronlund, 1977, s. 85-86). Bu yöntemde puanlayıcıların verdiği puanların ortalaması alınır (Ebel, 1951). Birden fazla puanlayıcının yer aldığı değerlendirme sürecinin sınırlılıkları arasında pratik olmaması

(birden fazla puanlayıcı bulmak güçtür) ve bireyin yeteneğini yüksek güvenilirlikte kestirebilmek için gerekli olan puanlayıcı sayısının net olmaması sayılabilir (Sudweeks vd., 2004).

Dereceli Puanlama Anahtarı

Dereceli puanlama anahtarı, birey performansının alt ölçütlerini ve performans seviyelerini gösteren, bireylerin kriterlere göre ilgili performansa ne kadar sahip olduklarını belirleyen şemalardır (Mertler, 2001; Moskal, 2000). Dereceli puanlama anahtarları bireylerin yapmış oldukları çalışmanın hangi kriterine göre değerlendirileceğini ve hangi performansın hangi düzeye karşılık geleceğini gösteren bir kılavuздur (Goodrich, 1997; Khitab, Zaman, & Ghaffar, 2014; Kutlu vd., 2017, s. 51). Dereceli puanlama anahtarında bireyden beklenen tüm davranışlar tanımlanır, bireyin performansına göre düzeyine karar verilir (Bıkmaz Bilgen, 2019, s. 197). İyi hazırlanmış dereceli puanlama anahtarı geçerli ve güvenilir puanların elde edilmesini sağlar (Jonsson & Svingby, 2007).

Dereceli puanlama anahtarının puanlayıcı ve puanlanan açısından avantajları vardır. Puanlayıcılar açısından, puanlama sürecinin objektif hale gelmesine ve puanlama sürecinin sistematik hale gelmesine katkıda bulunur, kullanımı kolaydır. Birey açısından, bireyden beklenen performansın farkına varmasını, bireylerin eksik ya da güçlü yönlerinin farkına varmalarını ve bireylerin öğrenme sorumluluğunu almalarını sağlar (Johnson, Penny, & Gordon, 2009, s. 145; Mertler, 2001).

Geçerli ve güvenilir dereceli puanlama anahtarı geliştirmek için izlenmesi gereken bazı adımlar vardır (Danielson & Hansen, 2016, s. 59; Mertler, 2001; Moskal, 2000; Popham, 2001);

Amacın belirlenmesi: Dereceli puanlama anahtarının hangi amaçla kullanılacağına (izleme, yetiştirme, değer biçme) karar verilmelidir.

Puanlama türünün belirlenmesi: Ölçülecek özelliğin doğasına uygun olarak puanlama türü belirlenmelidir. Ölçülecek özellik bütün olarak değerlendirilmek istendiğinde bütünsel, ölçülecek özelliğin sahip olduğu boyutlar ayrı ayrı değerlendirilmek istendiğinde analitik dereceli puanlama anahtarı kullanılmalıdır.

Dereceli puanlama anahtarı kullanım şekline göre genel ve göreve özgü dereceli puanlama anahtarı olmak üzere ikiye ayrılır. Göreve özgü dereceli puanlama anahtarı her bir görev için

özel olarak oluşturulurken genel dereceli puanlama anahtarı aynı özelliği temsil eden farklı görevler için tercih edilebilir. Örneğin 20 sorudan oluşan fen sınavında her soruya özel dereceli puanlama anahtarı geliştirilirse, bu dereceli puanlama anahtarı göreve özgü dereceli puanlama anahtarı olacaktır. Ancak 20 soruyu kapsayan bir dereceli puanlama anahtarı kullanılırsa bu da genel dereceli puanlama anahtarı olacaktır.

Ölçütlerin (kriterlerin) belirlenmesi: Ölçüt, birey performansına yönelik karar vermede kullanılan kurallardır. Ayrıca bireyin verilen görevi yerine getirirken sergilemesi gereken ya da beklenen durumlardır. Ölçütler performansın bütün yönlerini kapsamalı, performansı temsil etmelidir. Ölçütler, herkesin anlayacağı şekilde, açık, net ve anlaşılır olmalıdır. Ölçüt sayısı belirlenirken ilgili performansın yapısı, bireylerin yaşı ve seviyesi göz önünde bulundurulmalıdır. Ölçüt sayısının gerektiğinden az olması ölçütlerin binişik olmasına, gerektiğinden uzun olması da puanlayıcının çabuk yorulmasına ve puanlayıcıların ölçütleri birbirinden ayıramamasına sebep olabilir. Bu durum dereceli puanlama anahtarının geçerlik ve güvenilirliğine zarar verecektir.

Performans düzeyinin belirlenmesi: Ölçütlerin belirlenmesi sonrasında performans düzeylerine karar vermek gerekmektedir. Tüm performans düzeylerine göre özellikler ayrıntılı yazılmalıdır. Performans düzeyi ile ilgili özellikler yazılırken en yüksek performansın ve en düşük performansın öncelikle yazılması süreci kolaylaştıracaktır. Arada kalan düzeyler tanımlanırken de bireylerin yapabilecekleri hatalar göz önüne alınabilir (Popham, 2011). Literatürde performans düzeyi belirleme sürecinde kaç düzey olması gerektiği ile ilgili net bir sayı olmasa da puanlama sürecinde “merkeze eğilim” gösterilmemesi için çift sayıda düzey belirlenmesi önerilmektedir (Bıkmaz Bilgen, 2019, s. 200). Düzey sayısının çok olması bireylere ayrıntılı dönüt verme imkânı sunar. Ancak her düzey için özellik tanımlamak zordur ve puanlayıcıyı yorabilir. İdeal düzey sayısı 4-6 arası olduğu söylenebilir (Danielson & Hansen, 2016, s. 62).

Deneme uygulaması: İlgili görev küçük bir gruba verilerek bireylerin performans değerlendirme sürecinde geliştirilen dereceli puanlama anahtarı kullanılır. Bu sayede oluşabilecek aksaklıklar belirlenebilir ve asıl uygulamadan önce bu aksaklıklar giderilebilir.

Nihai formun oluşturulması: Deneme uygulamasından sonra öğrenci görüşleri alınabilir. Bireylerin kendilerini ya da arkadaşlarını değerlendirmeleri istenebilir. Bu sayede bireyler kendilerinden neler beklendiği ya da neler yapmaları gerektiği konusunda daha bilinçli olabilirler. Deneme uygulaması sonuçları ve bireylerden gelen dönütlere göre gerekli düzeltmeler yapılarak forma son hali verilir.

Dereceli puanlama anahtarı geliştirme sürecinde bu adımlara ek olarak geçerlik kanıtları toplamak gerekmektedir. Geliştirilen ölçme aracı hedef konuyu ve birey yanıtlarını kapsamalıdır. Dereceli puanlama anahtarının düşük kapsam geçerliğine sahip olması, ilgili performansın değerlendirilmesi sürecinde geçerli sonuçlar elde edilmesine engel olur. Çünkü puanlayıcıların dereceli puanlama anahtarında yer almayan performansın alt boyutunu değerlendirmeye çalışmaları zordur. Bu durum puanlayıcı davranışlarının artmasına sebep olabilir. Kapsam geçerliği yüksek bir dereceli puanlama anahtarı birey yanıtlarını kapsar. Kapsam geçerliği yüksek dereceli puanlama anahtarı geliştirmek için alan uzmanlarından görüş alınmalı ve görüşlere uygun düzeltmeler yapılmalıdır (Moskal & Leydens, 2000).

Dereceli puanlama anahtarı amacına göre bazı kaynaklar bütünsel, bağıl (göreceli) ve analitik dereceli puanlama anahtarı olmak üzere üçe ayırırken (Morrow vd., 2015) bazı kaynaklar da bütünsel (holistik) ve analitik dereceli puanlama anahtarı olmak üzere ikiye ayırmaktadır (Kutlu vd., 2017, s. 54; Mertler, 2001; Moskal, 2000). Araştırmada bütünsel ve analitik dereceli puanlama anahtarı ifadeleri kullanılmıştır.

Bütünsel dereceli puanlama anahtarı

Bütünsel dereceli puanlama anahtarı, ölçülecek performansın parçalara ayrılmadan tek bir puanla puanlandığı ölçme aracıdır (Güler, 2019, s. 82; Kutlu vd., 2017, s. 56). Bütünsel dereceli puanlama anahtarı, birey performansını ya da öğrenci çalışmalarını bütün olarak ele almayı hedefler (Korkmaz, 2004, s. 234). Ürünün ya da son performansın önemli olduğu durumlarda, değerlendirme süresinin kısa olduğu zamanlarda ve performansın bağımsız parçalara ayrılmasının zor olduğu durumlarda bütünsel dereceli puanlama anahtarının kullanılması önerilmektedir (Mertler, 2001; Moskal, 2000).

Analitik dereceli puanlama anahtarı

Analitik dereceli puanlama anahtarı, ölçülecek performansın her aşaması hakkında bilgi veren puanlama aracıdır (Kutlu vd., 2017, s. 60). Birey performansı, performansı oluşturan boyutlara göre değerlendirilir. Bu sayede bireyin güçlü ve zayıf yönleri görülmekte ve birey performansı hakkında daha zengin ve ayrıntılı bilgi elde edilmektedir (Gronlund, 1998; Kutlu vd., 2017, s. 60). Performans değerlendirmede performansın çok boyutlu ve anlamlı parçalardan oluştuğu durumlarda, sürecin önemli olduğu durumlarda, bireylerin

performansa dair güçlü ve zayıf yönlerinin belirlenmesi istendiğinde, yeterli zaman olduğunda analitik dereceli puanlama anahtarının kullanılması önerilmektedir (Kutlu vd., 2017, s. 64; Mertler, 2001). Kutlu vd. (2017, s. 61) analitik dereceli puanlama anahtarının bütünsel dereceli puanlama anahtarına göre daha geçerli ve güvenilir olduğunu ifade etmişlerdir. Ancak Klein vd. (1998) yapmış oldukları çalışmada bütünsel dereceli puanlama anahtarının analitik dereceli puanlama anahtarı kadar güvenilir olduğunu ifade etmişlerdir.

Puanlayıcı Eğitimi

Puanlayıcı davranışlarını benzer hale getirmeye yönelik uygulamalardan bir tanesi de puanlayıcı eğitimidir. Puanlayıcı eğitim programları puanlayıcıların yeterliliklerine odaklanan eğitim yöntemleri olduğu için geçerlik ve güvenilirlik açısından oldukça önemlidir (Robertson, Park, Gillman, & Vergis, 2020; Şata, 2022, s. 88). Puanlayıcı eğitiminde amaç, puanlama sürecini standartlaştırma, puanlayıcıların puanlama davranışlarını benzer hale getirme, puanlayıcılarla ortak anlayış geliştirme ve puanlama işlemi esnasında beklenmedik durumlara karşı hazırlıklı olmaktır (Brijmohan, 2016, s. 11; Cellar vd., 1989; Fahim & Bijani, 2011; Taylor, 2004, s. 2). Puanlayıcı eğitimi her ne kadar puanlayıcı davranışlarını ortadan kaldırmayı hedeflese de bu durum her zaman gerçekleşmez (McIntyre, 1993). Ancak puanlayıcı eğitimi, puanlayıcıların puanlamalarındaki ranjı düşürebilir. Puanlayıcı eğitiminin etkileri tam olarak bilinmese de bazı araştırmacılar puanlayıcı eğitiminin puanlayıcılar arası güvenilirliğe etkisinin olmadığını ya da çok az etkisinin olduğunu ifade etmektedir (Cook vd., 2009; Eckes, 2011; Fahim & Bijani, 2011; Gingerich, Kogan, Yeates, Govaerts, & Holmboe, 2014; Gingerich vd., 2011; Holmboe vd., 2004; Sebok & Syer, 2015). Bazı araştırmacılar ise tam aksine puanlayıcı eğitiminin puanlayıcı içi tutarlığı artırabileceğini, puanlayıcıların ön yargılarını azalttığını, puanlama kalitesini artırdığını ifade etmektedirler (Elder, Barkhuizen, Knoch, & von Randow, 2007; Lim, 2011; McNamara, 1996; Weigle, 1994, 1998).

Puanlayıcı eğitim programlarında performansın bileşenlerine ait eğitim, eğitimin kapsamı (eğitim stratejisi) ve geri bildirim olmak üzere üç temel öge bulunmaktadır (Hauenstein & McCusker, 2017). Performansın bileşenlerine ait eğitim ve geri bildirim öğeleri eğitim içeriğinden bağımsızdır. Performansın bileşenlerine ait eğitim öğesinde, puanlayıcılara performans değerlendirme sürecinde dikkat edilmesi gerekenler adım adım anlatmalı ve olası puanlayıcı davranışları açıklanmalıdır. Performansa dair açıklamalar eksiksiz olmalıdır (Flippo, 1984). Puanlayıcılara her boyut ve her performans düzeyi için farklı örnekler

verilmelidir. Ayrıca performans bileşenlerine ait eğitimde araştırmada kullanılacak ölçek kullanılmalıdır (McNamara, 1996, s. 125-127). Geri bildirim, puanlayıcıların puanlamalarına dair geri bildirimleri içermektedir. Roberts (1998) ve Hodgetts (1993) puanlayıcı eğitiminde iletişim becerilerinin geliştirilmesi ve puanlayıcıların dereceli puanlama anahtarını kullanma pratiğinin gelişmesi için geri bildirim aşamasına bir iki gün ayrılması gerektiğini ifade etmektedir. Tüm eğitim programları ölçülecek performansın boyutları ve boyutlara ait göstergelerin öğretilmesi ile başlamaktadır. Bu durum puanlayıcıların puanlama tutarlılıklarını artırmaktadır (Şata, 2022, s. 89). Puanlayıcı eğitimi esnasında ölçülecek performansın doğru tanımlanmaması halinde puanlayıcı eğitiminin ölçümlerin geçerliği ve güvenilirliğine etkisini belirlemek oldukça zor olacaktır (Hauenstein & McCusker, 2017).

Literatürde puanlayıcı eğitiminin kapsamı, puanlayıcı eğitimi stratejisi olarak da adlandırılmaktadır (Hauenstein & McCusker, 2017). Odak noktasına göre farklılık gösteren puanlayıcı stratejileri olsa da en sık kullanılan puanlayıcı eğitimi stratejileri (Kumar, 2005; Woehr & Huffcutt, 1994; Zedeck & Cascio, 1982) şunlardır;

- Öz Liderlik Eğitimi (SLT),
- Davranış Gözlem Eğitimi (BOT),
- Performans Boyutu Eğitimi (PDT),
- Puanlayıcı Hatası Eğitimi (RET),
- Referans Çerçevesi Eğitimi (FORT)
- Puanlayıcı Değişkenliği Eğitimi (RVT),

Öz Liderlik Eğitimi

Sosyal bilişsel literatürden ve öz denetimlerle ilgili çalışmalardan üretilen öz liderlik, bireyin kendini gerçekleştirme amacıyla ihtiyacı olan öz motivasyonu ve öz düzenlemeyi oluşturma sürecidir (Neck, 1996). Bir başka deyişle öz liderlik, bireyin düşünce ve davranışlarında kendi kendini etkileme süreci olarak da ifade edilebilir (Neck & Houghton, 2006). Öz liderlik eğitiminin temel amacı, bireylerin öz farkındalıklarını artırarak görevlerle başa çıkma becerilerini güçlendirmektir (Neck & Houghton, 2006).

Öz liderlik eğitimi temelde davranış odaklı stratejiler, doğal ödül stratejileri ve yapıcı düşünce kalıpları olmak üzere üç öğeden oluşmaktadır (Houghton & Neck, 2002). Davranış odaklı stratejiler, kişisel hedef, kendini gözlemleme, ödüllendirme, cezalandırma gibi

özellikleri kapsarken doğal ödül stratejileri içsel motivasyonu kapsamaktadır. Yapıcı düşünce kalıpları ise kendi kendine değerlendirme yapmayı ifade etmektedir (Furtner, Tutzer, & Sachse, 2018). Davranış temelli stratejilerde istenen durum ile birey arasında fark varsa öz düzenleme stratejileri ile bu farkı azaltmaya yönelik önlemler alınmalıdır (Furtner, Sachse, & Exenberger, 2012). Puanlayıcının sahip olduğu bilişsel yapının doğruluğu ya da yanlışlığı puanlama davranışını etkileyeceği için alınacak önlemler bireye özgü olmalıdır (Neck vd., 1995).

Öz liderlik eğitiminde lider, merkezi bir role sahip olduğu için rol modellik yapar (Ogunfowora, 2014). Öz liderlik stratejisine dayalı puanlayıcı eğitiminde eğitmen, merkezi bir role sahiptir. Bu durum puanlayıcıların deneme yanılma yapmanın olası zararlı sonuçlarına engel olarak daha kalıcı öğrenmelerin oluşmasına destek olur (Sims & Manz, 1991).

Yapılan çalışmalar öz liderlik eğitiminin bireylerin stres yönetimi ve strese dayanıklılığı ile olumlu ilişkiye sahip olduğunu göstermektedir (Altenhöner, Köhler, Philippi, & Alaze, 2014; Houghton, Wu, Godwin, Neck, & Manz, 2012; Lovelace, Manz, & Alves, 2007; Pietilä, 2019; Schraub, 2011). Bu durum puanlayıcıların puanlayıcı davranışlarını olumlu etkileyebilir.

Davranış Gözlem Eğitimi

Puanlayıcıların gözlem yapma becerilerini geliştirmek, bir görevin farklı boyutlarını fark etmek ve her bir boyuta odaklanmak objektif puanlama açısından önemlidir (Noonan & Sulsky, 2001). Davranış gözlem eğitimi, puanlama sürecinin ilk aşaması olan gözleme odaklanmaktadır (Woehr & Huffcutt, 1994). Davranış gözlem eğitiminin temel amacı, puanlayıcıyı performansı doğru bir şekilde gözlemleyebilecek yetkinliğe ulaştırmaktır (Noonan & Sulsky, 2001).

Davranış gözlem eğitim stratejisi ile tasarlanan eğitim modelleri farklı şekillerde sunulsa da genel olarak eğitim modelleri incelendiğinde odaklanmış gözlem ve not alma ön plana çıkmaktadır. Odaklanmış gözlem ve not alma sayesinde puanlayıcının davranışsal hatırlama yeteneği artar ve bu sayede puanlama doğruluğu artabilir (Roch & O'Sullivan, 2003). Davranışsal gözlem eğitiminde puanlayıcılara örnek performanslar da paylaşılabilmektedir. Bu durum puanlayıcı puanlama doğruluğunu artırmaktadır (Woehr & Feldman, 1993). Ayrıca puanlayıcıların gözlem yeteneklerini artırmak için gözlem kayıtları da

kullanılmaktadır (Antonioni & Woehr, 2001, s. 124). Thornton & Zorich (1980) iki aşamalı davranış gözlem eğitim stratejisi önermişlerdir. Bu eğitim stratejisinde puanlayıcı davranışlarına odaklanması ve performans gözlenirken not alınması önerilmiş ve sistematik gözlem hatalarından kaçınılması gerektiği belirtilmiştir. Araştırma sonucunda puanlayıcıların puanlanma doğruluğunun arttığı ve hatırlama becerilerinin geliştiği gözlenmiştir. Benzer şekilde Hedge & Kavanagh (1988)'ın tasarlamış oldukları eğitim programında puanlayıcılara dikkatli gözlem yapmalarını ve gerektiğinde not alabileceklerini ifade etmişlerdir. Araştırmacılar araştırma sonucunda puanlama doğruluklarının arttığını belirlemişlerdir. Ayrıca bazı çalışmalar davranış gözlem eğitiminin puanlayıcılardaki halo etkisini de azalttığını göstermiştir (Bernardin & Walter, 1977; Latham, Wexley, & Pursell, 1975). Ancak bazı araştırmacılar davranış gözlem eğitiminin puanlama doğruluğunu artırmak için tek başına yeterli olmadığını ve diğer eğitim stratejileri ile birlikte kullanılması gerektiğini ifade etmektedir (Pulakos, 1986).

Performans Boyutu Eğitimi

Performans boyutu eğitimi bilgiyi işlemeye odaklanmasıyla ortaya çıkan stratejilerdendir. Puanlayıcının bilgiyi nasıl işlediği anlaşılması halinde puanlamanın etkinliği de artacaktır (Feldman, 1981; Landy & Farr, 1983). Performans boyutu eğitimi, puanlayıcıların birey performanslarını gözlemlemeden önce performansa ait boyutlara aşina olmasını hedefler. Çünkü puanlayıcılar bir davranış gözlemlendiğinde kanaat oluştururlar (DeNisi vd., 1989; Hastie & Park, 1986; Keller, 1987; Lichtenstein & Srull, 1987; Murphy, Philbin, & Adams, 1989; Woehr & Feldman, 1993). Puanlayıcıların performansa ait aşinalığı iki şekilde sağlanabilmektedir. Birincisi puanlayıcıların puanlama işlemi öncesinde dereceli puanlama anahtarını incelemesiyle ikincisi ise dereceli puanlama anahtarı geliştirilmesi sürecinde puanlayıcıların da sürece dahil olmasıyla gerçekleştirilir. Bu sayede puanlayıcıların puanlama kalitesi artacaktır (Woehr, 1992).

Performans boyutu eğitiminde puanlayıcılara ölçekte yer alan her boyut için uygun performans sunulur. Sunulan performans örnekleri oyun, video ya da yazılı metin, hikaye olarak verilebilir (Feldman vd., 2012). Burada temel amaç hedef performans ile ilgisiz puanlayıcı davranışlarını ortadan kaldırmak ya da bu davranışları azaltarak puanlama doğruluğunu artırmaktır. Bir başka deyişle performans boyutu eğitiminin odak noktası puanlama hataları değil puanlama doğruluğudur (Woehr & Huffcutt, 1994).

Yapılan puanlamalarının doğruluğu, puanlayıcının verdiği puanların uzman puanlayıcılara ait puanlamalar ile karşılaştırılmasıyla hesaplanır. Burada uzman puanlayıcıların yapmış oldukları puanlar gerçek puanlar olarak kabul edilir ve puanlayıcıların yapmış oldukları puanlamalar ne kadar gerçek puanlara yakınsa o kadar doğru puanlama yaptığı kabul edilir (Borman, 1977; Cronbach, 1955; Sulsky & Balzer, 1988).

Puanlayıcı Hatası Eğitimi

Puanlayıcı hatası eğitim stratejisinin temel odağı öznel performans değerlendirme sürecinde ortaya çıkan psikometrik özelliklerdir (Woehr & Huffcutt, 1994). Sık karşılaşılan puanlayıcı katılığı, puanlayıcı cömertliği, merkeze eğilim ve halo etkisi gibi puanlayıcı davranışları üzerine araştırmalar yapıldığında (Cooper, 1981; Landy & Farr, 1980; Saal vd., 1980) performans değerlendirme sürecinde güvenilirliğin düşük olduğu belirtilmiştir (Pearlman, Schmidt, & Hunter, 1980). Puanlayıcı hatası eğitimi, performans değerlendirme sürecinde yaygın olarak ortaya çıkan puanlayıcı davranışlarını (puanlayıcı katılığı, puanlayıcı cömertliği, merkeze eğilim, halo etkisi) tanımayı, bu davranışlardan kaçınmayı ve dolayısıyla puanlama hatalarının azaltılmasını ya da tamamen ortadan kaldırılmasını hedeflemektedir (Woehr & Huffcutt, 1994). Bir başka deyişle puanlayıcı hatası eğitiminin amacı, puanlayıcıların sık karşılaşılan puanlayıcı davranışlarıyla ilgili farkındalıklarını artırarak bu davranışlardan kaçınmalarını sağlamak ve puanlamaların doğruluğunu artırmaktır. Puanlayıcı hatası eğitim stratejisinin etkili olması için puanlayıcılar sürece aktif olarak dahil olmalı ve eğitim esnasında puanlama performanslarına ilişkin sürekli olarak geri bildirim almalıdırlar (Latham & Wexley, 1981). Geri bildirim sayesinde puanlayıcılar hem davranışı hem de ölçülecek performansı içselleştirme imkânı bulacaktır. Ancak yapılan çalışmalar incelendiğinde puanlayıcı hatası eğitim stratejisinin puanlayıcılar arası güvenilirlikte pozitif bir etkiye sahip olmadığı belirtilmiştir (Bernardin & Pence, 1980; Borman, 1975). Ayrıca bu eğitim stratejisinin puanlayıcıların katılık ve cömertlik gibi davranışlarında azalmayı sağlarken puanlama doğruluğunda da azalmaya sebep olduğuna yönelik bulgular da yer almaktadır (Bernardin & Pence, 1980). Puanlayıcı eğitim stratejilerinde en temel amacın puanlayıcı doğruluğu olduğu göz önüne alındığında bazı araştırmacılar puanlayıcı eğitiminde bu stratejinin kullanılmasını önermemektedir (Woehr & Huffcutt, 1994).

Referans Çerçevesi Eğitimi

Puanlayıcı hatası eğitim stratejisinin puanlayıcı davranışlarındaki tartışmalı etkileri üzerine Bernardin & Buckley (1981) referans çerçevesi eğitim stratejisini önermişlerdir. Referans çerçevesi eğitimi, puanlayıcılara performansın tüm boyutları için standartlar sunmayı hedeflemektedir (Woehr & Huffcutt, 1994). Referans çerçevesi eğitim stratejisinin temel amacı, puanlayıcıların performansa ait tüm boyutlar ile ilgili ortak zihinsel model oluşturmalarını ve bu zihinsel modeli puanlama esnasında kullanmalarını sağlamaktır (Athey & McIntyre, 1987; Woehr, 1994). Diğer eğitim stratejileri geri bildirim, standart oluşturma ya da puanlayıcı motivasyonunu artırma gibi yöntemlerle puanlayıcı güvenilirliğini artırmayı hedeflerken referans çerçevesi eğitim stratejisi, puanlayıcı davranışlarını azaltmak ve puanlama doğruluğunu artırmak amacıyla farklı performans düzeyleri için ortak zihinsel model oluşturmaya odaklanır. Bu durum referans çerçevesi eğitim stratejisini diğer eğitim stratejilerinden ayırmakta ve bu eğitim stratejisinin kullanımını artırmaktadır (Aguinis vd., 2009; Athey & McIntyre, 1987; Gorman & Rentsch, 2009; Raczynski, Cohen, Engelhard, & Lu, 2015; Roch vd., 2012; Uggerslev & Sulsky, 2008; Woehr, 1994).

Referans çerçevesi eğitim stratejisi, puanlayıcı güvenilirliğini ve doğruluğunu artırmada en başarılı yöntemlerden biridir (Roch vd., 2012). Bu strateji, puanlayıcılara performansın birden fazla alt boyuta sahip olduğunu ve bu boyutların belli standartlara sahip olduğunu öğretir. Referans çerçevesi eğitim stratejisinde eğitimi verecek uzman ya da uzmanlar ölçülecek performansa tamamen hâkim olan, performansı iyi tanıyan ve yeterli tecrübeye sahip bireyler olarak tanımlanmaktadır. Uzmanlar eğitim öncesinde ölçülecek performansa ait alt boyutlara ilişkin kesin davranışları, standartları belirler ve bu performanslarla ilgili materyaller oluşturur (Sulsky & Balzer, 1988). Referans çerçevesi eğitim stratejisi temelli eğitim programında uzmanlar puanlayıcı davranışlarını benzer hale getirmek için sırasıyla performansa şu adımları izlerler: 1) ölçülecek performansa ait tüm alt boyutların tanımlarını yapmak, 2) puanlama anahtarının referans noktasını belirlemek, 3) her bir boyuta ait davranış göstergelerini belirlemek, 4) puanlama uygulaması yapmak, 5) puanlayıcıların puanlama davranışları ile ilgili puanlayıcılara geri bildirim vermektir (Cascio & Aguinis, 2014, s. 376).

Referans çerçevesi eğitimi, puanlayıcılara performansı oluşturan her boyuta ait standartlaştırılmış örnekler sunduğu için puanlayıcıların öznel davranışlarını azaltan etkili bir yoldur (Athey & McIntyre, 1987; Dierdorff, Surface, & Brown, 2010; Gorman & Rentsch, 2009; Woehr, 1994). Puanlayıcılar vermiş oldukları puanları standartlarla

karşılaştırdığında daha doğru ve güvenilir sonuçlar elde edebilir (Day & Sulsky, 1995; Noonan & Sulsky, 2001). Puanlayıcılara ait iç tutarlılığın geliştirilmesiyle puanlayıcılar arası tutarlılık da artar (Schleicher & Day, 1998).

Literatür incelendiğinde, birçok çalışma referans çerçevesi eğitiminin puanlama doğruluğunu artırmada etkili olduğunu göstermiştir (Athey & McIntyre, 1987; Bernardin & Pence, 1980; Cardy & Keefe, 1994; Day & Sulsky, 1995; Hauenstein & Foti, 1989; S. Kim vd., 2022; Kogan, Conforti, & Holmboe, 2023; R. M. McIntyre, Smith, & Hassett, 1984; Noonan & Sulsky, 2001; Pulakos, 1984, 1986; Schleicher & Day, 1998; Stamoulis & Hauenstein, 1993; Sulsky & Day, 1992, 1994; Woehr, 1994).

Puanlayıcı Değişkenliği Eğitimi

Performans değerlendirme sürecinde puanlama doğruluğunu artırmayı hedefleyen eğitim stratejilerinden birisi de puanlayıcı değişkenliği eğitimidir. Puanlayıcı değişkenliği eğitimi, referans çerçevesi eğitimi ile puanlayıcı hatası eğitiminin harmanlanmış hali olsa da temelinde referans çerçevesi eğitimi yer almaktadır. Bazı araştırmacılar referans çerçevesi eğitiminin puanlama doğruluğunu artırmadığını belirtirken (Murphy & Balzer, 1989) bazı araştırmacılar ise referans çerçevesi eğitiminin puanlama doğruluğunu artırdığını ancak puanlayıcı hafızasını engellediğini ifade etmektedir (Sulsky & Day, 1992). Referans çerçevesi eğitimi ile ilgili bu tartışmalar ve olası sınırlılıklar göz önüne alındığında başka eğitim stratejileri ile harmanlanması fikri gündeme gelmiştir. Stamoulis & Hauenstein (1993) tarafından geliştirilen puanlayıcı değişkenliği eğitimi, referans çerçevesi eğitim stratejisindeki her bir performansa ait örnek sunumunu ve puanlayıcıya geri bildirimini; puanlayıcı hatası eğitiminin de puanlama tekrarlarını kapsamaktadır. Bu eğitim stratejisi gözlenen puanlardaki değişkenliği gerçek puanlardaki değişkenliğe karşılık gelecek şekilde değiştirmeye çalışmayı hedeflediği için puanlayıcı değişkenliği eğitimi adını almıştır (Stamoulis & Hauenstein, 1993).

Puanlayıcı değişkenliği eğitim stratejisinin referans çerçevesi ve puanlayıcı hatası eğitim stratejilerinden temel farkı geri bildirimdir (Mubeena, 2010). Referans çerçevesi eğitiminde puanlayıcı eğitimi veren uzman, puanlayıcının vermiş olduğu puanlar ile hedef arasındaki farkı vurgular ve puanlayıcının puanlayıcı grubuna göre kendi standartlarını oluşturmasını sağlar. Puanlayıcı hatası eğitiminde puanlayıcıda gözlenen davranışın puanlayıcılar tarafından tartışılması ve puanlayıcıların benzer davranışlara sahip olması teşvik edilir.

Puanlayıcı deęiřkenlięi eęitiminde ise puanlayıcı eęitimini veren uzman, puanlayıcıların yapmış oldukları puanlamaları belli bir standartla karşılaştırır ve puanlayıcılar arasındaki farkı ön plana çıkarır. Puanlayıcı deęiřkenlięi eęitiminde temel amaç performans deęerlendirme sürecinde ölçülen özellięe ait standartların puanlayıcılar tarafından iyi bir şekilde tanımlanmasını saęlamaktır. Performansa ait standartlar tanımlanırken her bir boyuta ait örnekler toplanır ve puanlayıcılara sunulur (Hauenstein & McCusker, 2017).

Literatür incelendięinde puanlayıcı deęiřkenlięi eęitimi ile ilgili çalışmaların dięer eęitim stratejilerine göre daha az olduęu görölmektedir. Bu durumun sebebi puanlayıcı deęiřkenlięi eęitiminin temelinde referans çerçevesi eęitim stratejisi olması ve bu eęitim stratejisinin referans çerçevesi eęitimine çok benziyor olmasıdır (Şata, 2022).

Referans çerçevesi eęitiminin, gelişimsel süreçlerde kullanılması daha uygundur ancak deęerlendirmeler idari amaçlı kullanılacaksa puanlayıcı deęiřkenlięi eęitimi kullanılması önerilmektedir (Hauenstein, 1998).

BÖLÜM III

İLGİLİ ARAŞTIRMALAR

Bu bölümde öncelikle açık uçlu soruların puanlanmasında ortaya çıkan puanlayıcı davranışlarının incelenmesi ile ilgili araştırmalara, sonrasında ise puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisinin incelendiği araştırmalara yer verilmiştir.

Puanlayıcı Davranışları ile İlgili Araştırmalar

Görgülü Öztürk (2023) yapmış olduğu çalışmada puanlayıcı eğitiminin 10. sınıf öğrencilerinin matematik açık uçlu maddeleri puanlama davranışlarına etkisini incelemiştir. Araştırmaya 65 öğrenci dahil edilirken araştırmada dört tane açık uçlu madde yer almaktadır. Araştırma sonucunda puanlayıcı eğitiminin puanlayıcıların açık uçlu maddeleri puanlama davranışlarında anlamlı bir etkisi olmadığı görülmüştür.

Aydoğmuş (2021) tarafından yapılan çalışmada puanlama deneyimine sahip olan ve sahip olmayan öğretmenlerin analitik ve bütünsel dereceli puanlama anahtarı kullanılarak yapılan puanlamalardan elde edilen güvenilirlik değerleri karşılaştırılmıştır. MEB Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından yürütülen “Yazılı Anlatım Becerilerinin Ölçülmesi ve Değerlendirilmesi” çalışması için uygulanan “Yazma Becerileri Testi” verileri kullanılarak yapılan çalışmada 8 açık uçlu soru yer almaktadır. Çalışmada farklı illerden 240 birey ve puanlayıcı deneyimine sahip olup olmama durumuna göre iki ayrı grupta 12 sınıf, 12 Türkçe ve 12 Türk Dili ve Edebiyatı öğretmeni yer almaktadır. Puanlama sonuçları Genellebilirlik Kuramı, Kappa Tekniği ve Krippendorff Alpha katsayısı ile kestirim yapılarak puanlamaların tutarlılıkları incelenmiştir. Yapılan analizler sonucunda puanlama deneyiminin tek başına güvenilirliğe belirgin bir etkisi olduğuna dair bir bulguya rastlanmadığı belirtilmiştir.

Arslan Mancar (2019) tarafından yapılan betimsel çalışmada farklı sayıdaki puanlayıcılardan elde edilen puanların tutarlılığı klasik test kuramı ve genellenebilirlik kuramı ile karşılaştırılmıştır. Araştırmada Uluslararası Bakalorya Diploma Programı'ndan mezun olan öğrencilerin biyoloji dersi için hazırladığı 20 performans çalışması ve Uluslararası Bakalorya Örgütü uzmanları tarafından geliştirilen analitik dereceli puanlama anahtarı kullanılmıştır. Bu çalışmalar farklı sayıdaki puanlayıcılar tarafından puanlanmıştır. Araştırmada puanlayıcı güvenilirliğini belirlemek için klasik test kuramına dayalı olarak Kappa Tekniği ve Krippendorff Alpha istatistikleri kullanılırken ek olarak da genellenebilirlik kuramı kullanılmıştır. Araştırmada yapılan analizler ve incelemeler sonucunda Kappa Tekniği ve Krippendorff Alpha istatistikleri sonucunda elde edilen değerlerin hata kaynakları ve oranları hakkında yeterli bilgi vermediği, genellenebilirlik kuramı kapsamında yapılan analizlerde de puanlayıcı sayısının artmasının puanlama işleminin tutarlılığını artıracakı belirtilmiştir. Ayrıca genellenebilirlik kuramına dayalı tekniklerin daha güvenilir sonuçlar elde edilmesine katkı sağladığı belirtilmiştir.

Şata (2019) tarafından yapılan araştırmada performans değerlendirme sürecinde puanlayıcı davranışları ve puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelenmiştir. Araştırmacı literatürde 30'dan fazla puanlayıcı davranışı olduğunu ancak araştırmasında puanlayıcı katılığı, puanlayıcı cömertliği, merkeze eğilim, halo, farklılaşan puanlayıcı katılığı ve cömertliği olmak üzere altı davranışı incelediğini belirtmiştir. Araştırmada deney ve kontrol gruplu ön test ve son test yarı deneysel desen kullanılmıştır. Araştırmada yer alan 45 puanlayıcıdan 22'si deney 23'ü de kontrol grubunda yer almaktadır. Araştırmada veri toplama aracı olarak araştırmacı tarafından geliştirilen analitik dereceli puanlama anahtarı ve Uluslararası İngilizce Dil Test Sistemi tarafından hazırlanan düşünce temelli yazma görevi ve kişisel bilgi formu kullanılmıştır. Veri toplama araçlarından elde edilen veriler çok yüzeyli rasch modeli, genellenebilirlik kuramı ve bağımsız örneklem t testi kullanılarak incelenmiştir. Çok yüzeyli rasch modeli ve genellenebilirlik kuramı analizlerinde tümüyle çaprazlanmış desenler kullanılmıştır. Deneysel işlem süreci 4 hafta sürmüş, işlem sırasında puanlayıcı eğitimi stratejilerinden puanlayıcı hatası ve referans çerçevesi eğitimi stratejileri bir arada kullanılmıştır. Puanlayıcı eğitimi öncesinde bazı puanlayıcı davranışlarının hem bireysel hem de grup düzeyinde ortaya çıktığı gözlemlenmiştir. Puanlayıcı eğitimi sonrasında yapılan inceleme sonucunda puanlayıcı davranışlarının tekrar ortaya çıktığı ancak puanlayıcı davranışlarının ölçme sonuçlarına karışma düzeyinin deney grubunda kontrol grubuna göre istatistiksel olarak daha az görüldüğü belirtilmiştir. Ayrıca araştırmada

genellenebilirlik kuramı sonuçları ile çok yüzeyli rasch modeli sonuçlarının uyumlu olduğu belirtilmiştir.

Bijani (2018) tarafından yapılan çalışmada bireylerin sözlü değerlendirilmeleri esnasında ortaya çıkan puanlayıcı davranışlarını belirlemek ve farklı puanlayıcı davranışlarını ortadan kaldırmak için puanlayıcı eğitimi verilmiştir. Araştırma karma desenle desenlenmiş ve araştırmada 300 öğrenci ve 20 puanlayıcı yer almaktadır. Veri toplama sürecinde anket, görüşme formu ve gözlem tekniği kullanılmıştır. Puanlayıcı eğitiminden önce ve sonra puanlayıcıların tutumları, ön yargıları karşılaştırılmıştır. Puanlayıcı eğitiminin farklı puanlayıcı davranışlarını azalttığı ve puanlayıcı tutarlılıklarının artırdığı ifade edilmiştir. Ayrıca puanlayıcı eğitime karşı olumlu tutuma sahip olan bireylerin daha tutarlı ve daha az hatalı puanlamalar yaptığı belirtilmiştir. Araştırmacı son olarak puanlayıcı tutarlılığını artırmak ve puanlayıcı davranışlarını azaltmak için puanlayıcı eğitimi verilmesi gerektiğini ifade etmiştir.

Hauenstein & McCusker (2017) üç farklı puanlayıcı deseninin puanlayıcı doğruluğuna etkisini incelemiştir. Araştırmada kontrol grubuna performans boyutları eğitimi, deney grubuna ise puanlayıcı hataları eğitimi, referans çerçevesi eğitimi, puanlayıcı değişkenliği eğitimi verilmiştir. Eğitimlerde farklı sayılarda bireyler yer almaktadır. Performans boyutları eğitiminde (kontrol grubu) 21, puanlayıcı hataları eğitiminde 22, referans çerçevesi eğitiminde 24, puanlayıcı değişkenliği eğitiminde 22 olmak üzere toplam 89 birey yer almaktadır. Bireyler gruplara rastgele dağıtılmış ve üç kez ölçüm yapılmıştır. Yapılan analizler sonucunda referans çerçevesi eğitiminin puanlayıcı doğruluğunu artırmadaki en etkili desen olduğu belirtilmiştir.

Davis (2016) puanlayıcı deneyiminin ve puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelemiştir. Araştırmada 20 puanlayıcı ve TOEFL sınavında yer alan 100 konuşma kaydı yer almaktadır. Tümüyle çaprazlanmış desen ile yapılandırılan çalışmada puanlayıcılar eğitimden önce ve sonra kayıtları puanlamışlardır. Araştırma sonucunda puanlayıcı eğitiminin katı puanlama davranışı üzerinde az da olsa etkili olduğu, puanlayıcı tutarlılığına ise anlamlı etki ettiği ifade edilmiştir.

İlhan (2015) tarafından yapılan araştırmada farklı taksonomilere dayanan dereceli puanlama anahtarının açık uçlu matematik soruları puanlanırken kullanılmasında puanlayıcı etkilerini çok yüzeyli rasch modeli ile incelemiştir. Araştırma 8. sınıfta öğrenimine devam eden 104 öğrenci ve yedi puanlayıcı ile yürütülmüştür. Araştırmada veri toplama sürecinde araştırmacılar tarafından geliştirilen 8 tane açık uçlu matematik başarı testi, standart dereceli

puanlama anahtarları, SOLO taksonomisine dayalı dereceli puanlama anahtarları, standart dereceli puanlama anahtarları ile ilgili düşünceler (SRDA) ve SOLO taksonomisine dayalı dereceli puanlama anahtarı ile ilgili düşünceler anketleri kullanılmıştır. Araştırmanın ilk aşamasında puanlayıcılardan standart dereceli puanlama anahtarı kullanarak puanlama yapmaları sağlanmıştır. Sonraki aşamada ise puanlayıcılara aynı testleri SOLO taksonomisine dayalı dereceli puanlama anahtarı kullanarak puanlama yapmaları sağlanmıştır. Elde edilen veriler çok yüzeyli rasch modeli kullanılarak analiz edilmiştir. Analiz sonucunda her iki dereceli puanlama anahtarının da farklı yetenek düzeyindeki bireyleri yüksek güvenirlikte ayırt ettiği, standart dereceli puanlama anahtarından elde edilen sonuçlarda bazı puanlayıcı davranışlarının görüldüğü, SOLO taksonomisine dayalı dereceli puanlama anahtarından elde edilen sonuçlarda ise hiçbir farklı puanlayıcı davranışı olmadığı ve puanlayıcı güvenirliğinin yüksek olduğu belirtilmiştir.

Güler (2014) tarafından yapılan araştırmada, performans değerlendirme sürecinde ortaya çıkan puanlayıcı davranışları çok yüzeyli rasch modeli kullanılarak incelenmiştir. Araştırmada veri toplama aracı olarak 10 açık uçlu maddeden oluşan istatistik testi kullanılmıştır. Araştırmada 55 üniversite öğrencisi ve üç puanlayıcı yer almaktadır. Açık uçlu maddeler değerlendirilirken 11 kategoriye sahip dereceli puanlama anahtarı kullanılmıştır. Çok yüzeyli rasch modeli analizi sonucunda puanlayıcıların benzer davranışlarda bulundukları ve puanlayıcı güvenirliğinin yüksek olduğu belirtilmiştir. Ayrıca dereceli puanlama anahtarındaki 11 kategoriden genelde üçünün tercih edildiği bu yüzden de yapılacak çalışmalarda üç kategori kullanılmasının daha uygun olacağı ifade edilmiştir.

Fahim & Bijani (2011) ikinci dil kompozisyon yazma değerlendirilmesinde puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisi incelenmiştir. Yarı deneysel desen kullanılarak yapılan araştırmada 12 puanlayıcı ve 60 kompozisyon değerlendirilmiştir. Puanlayıcı eğitiminden önce 45, puanlayıcı eğitiminden sonra da 15 kompozisyon puanlayıcılar tarafından değerlendirilmiştir. Araştırmada veri toplama aracı olarak IELTS tarafından geliştirilen dereceli puanlama ölçeği kullanılmıştır. Araştırmanın sonucu olarak puanlayıcı eğitiminin puanlayıcı davranışlarını büyük oranda azalttığı belirtilmiştir. Ayrıca puanlayıcılar arası tutarlılığın da arttığı ifade edilmiştir.

Güler (2008) tarafından yapılan çalışmada matematik başarısını ölçmek için elde edilen puanlar klasik test kuramı, genellenebilirlik kuramı ve çok yüzeyli rasch modeli ile incelenmiştir. Elde edilen puanların güvenirlikleri her üç kurama göre de hesaplanmış ve sonuçlar karşılaştırılmıştır. Çalışmada TIMSS 1999'da yer alan açık uçlu matematik

sorularından 24 tanesi 203 öğrenciye uygulanmıştır. Çalışma iki aşamadan oluşmaktadır. İlk aşamada öğrencilerin cevapları bütünsel dereceli puanlama anahtarı ile 4 puanlayıcı tarafından puanlanmıştır. Araştırmanın ikinci aşamasında puanlayıcılardan elde edilen puanlar farklı kuramlara göre karşılaştırılmıştır. Karşılaştırılma yapılırken klasik test kuramında Cronbach Alpha, Kendall'ın konkordans, puanlayıcılar arası korelasyon katsayısı ve puanlayıcı puanlarının ortalamaları arasında fark olup olmadığı F Testi ile incelenmiştir. Genellenebilirlik kuramında tümüyle çaprazlanmış desen ($b \times g \times p$) kullanılmış, genellenebilirlik ve güvenirlik katsayıları araştırılmıştır. Çok yüzeyli rasch modeli ile de birey madde ve puanlayıcı yüzeylerine ilişkin ayrı ayrı güvenirlik hesaplamaları yapılmıştır. Araştırmada sonucunda ölçme aracının, matematik başarısını belirlemede güvenilir sonuçlar verdiği görülmüştür. Matematik başarısının ölçülmesinde güvenirlik belirlenirken faydalanılacak kuramlardan hangisinin seçileceği, puanların hangi amaç için kullanılacağına bağlı olarak değişebileceği belirtilmiş, en az iki kuramdan faydalanmanın da daha uygun olacağı vurgulanmıştır.

BÖLÜM IV

YÖNTEM

Bu bölümde araştırmanın deseni, çalışma grubu, veri toplama araçları hakkında bilgiler yer almaktadır.

Araştırmanın Deseni

Bu araştırmada, açık uçlu soruların puanlanmasında puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisini incelenmesi amaçlandığı için araştırma denk olmayan kontrol gruplu desen ile tasarlanmıştır (İlhan & Gezer, 2023, s. 120). Araştırmalar çoğu zaman gerçek deneysel desene göre tasarlanamaz. Gerçek deneysel desenlere en uygun alternatif yarı deneysel desenlerdir (Cottrell & Mckenzie, 2011, s. 188). Araştırmada kullanılacak olan denk olmayan kontrol gruplu desen gerçek deneysel araştırmalardaki ön test son test kontrol gruplu desene benzer. Burada katılımcılar gruplara rastgele atanmaz gruplar araştırma öncesinde vardır. Ancak araştırmacı var olan grupların hangisinin deney hangisinin kontrol grubu olacağına karar verir. Gruplara uygulanacak ön test puanları benzerse ve son test puanları arasında fark varsa bu farkın deneysel işleminden kaynaklandığı düşünülür (İlhan & Gezer, 2023, s. 120).

Çalışma Grubu

Araştırmanın çalışma grubu açık uçlu maddeleri cevaplandıran öğrenciler ve öğrenci kağıtlarını değerlendiren puanlayıcılardan oluşmaktadır.

Araştırmanın öğrenci grubu 2022-2023 eğitim- öğretim yılı Bahar dönemi Sivas ili Merkez ilçesinde beş farklı okulun 8. sınıfında öğrenimine devam eden ve araştırmaya gönüllülük

esasına dayalı olarak katılan 70 öğrenciden oluşmaktadır. Ancak öğrenci kağıtlarından uygun olmayan beş tanesi araştırma kapsamı dışında bırakılmıştır. Öğrenci kağıtlarının incelenmesinin ardından kağıtlar iyi, orta ve geliştirilmeli olmak üzere 3 grupta gruplandırılmıştır. Gruplandırma işleminin ardından tüm gruplardan 10 tane olmak üzere toplam 30 kâğıt bütünsel dereceli puanlama anahtarı hazırlanmak üzere analiz dışında tutulmuştur.

Araştırmmanın puanlayıcı grubu ise gönüllülük esasında dayalı olarak araştırmaya dahil edilen 20 Türkçe Öğretmenliği, 24 Matematik Öğretmenliği bölümünde öğrenimine devam eden öğretmen adaylarından oluşmaktadır. Araştırmaya dahil olan öğretmen adaylarının ölçme ve değerlendirme dersinden başarılı olmuş olmaları, daha önce puanlayıcı eğitime katılmamış olmaları ve puanlama deneyimlerinin aynı olmasına dikkat edilmiştir. Araştırmada ön teste katılan ancak son teste katılmayan 4 Türkçe öğretmen adayı, 3 Matematik öğretmen adayı araştırma kapsamından çıkarılmıştır. Ön test puanlamaları çok yüzeysel rasch modeli ile analiz edilmiş ve analiz sonucunda tüm bireylere aynı puan veren iki Türkçe öğretmen adayı, bir Matematik öğretmen adayı analiz dışı bırakılmıştır. Yapılan işlemlerin ardından araştırmaya 14 Türkçe öğretmenliği, 20 Matematik öğretmenliği bölümlerinde öğrenimine devam eden puanlayıcılarla devam edilmiştir.

Veri Toplama Araçları

Araştırmada veri toplama aracı olarak araştırmacı tarafından geliştirilen ortaokul Türkçe ve Matematik dersleri kazanımlarını kapsayan beş Türkçe ve altı Matematik olmak üzere toplam 11 açık uçlu maddeden oluşan başarı testi kullanılmıştır. Başarı testi 70 öğrenciye uygulanmış ve tüm öğrenciler tüm soruları cevaplamıştır. Öğrencilerin verdikleri cevaplar 14 Türkçe ve 20 Matematik öğretmen adayı tarafından puanlanmıştır. Bütünsel dereceli puanlama anahtarı, her bir madde için özel tanımlamaların yer aldığı dört kriterden oluşmaktadır.

Başarı Testlerinin Hazırlanması

Türkçe ve matematik derslerinden oluşan başarı testleri 5 aşamada oluşturulmuştur.

- 1- *Belirtke tablosunun oluşturulması:* Başarı testlerinin geliştirilmesi sürecinde ilk olarak belirtke tabloları oluşturulmuştur. Burada hangi sınıf düzeyinde hangi kazanımlarla ilgili açık uçlu madde hazırlanacağı bilgisi yer almaktadır.
- 2- *Açık uçlu soruların oluşturulması:* Belirtke tablosuna göre açık uçlu maddeler oluşturulmuştur.
- 3- *Uzman görüşü:* Belirtke tablosuna göre hazırlanan açık uçlu maddelerin sınıf düzeyine, kazanımlara uygunluğunun değerlendirilmesi, soruların anlaşılabilirlik durumunun belirlenmesi amacıyla uzman görüşleri alınmıştır. Uzmanlardan testte yer alan her bir açık uçlu maddeyi incelemeleri ve maddeler hakkında uygun, düzeltilmeli, uygun değil şeklinde görüş bildirmeleri istenmiştir. Ayrıca düzeltme önerilen açık uçlu maddelere öneri yazmaları da istenmiştir.
- 4- *Redaksiyon-Uzman görüşü:* Uzman görüşleri sonucunda Türkçe dersi ile ilgili hazırlanan açık uçlu maddelerin sınıf seviyesine ve kazanımlara uygun olduğu belirtilmiştir. Matematik dersinden hazırlanan açık uçlu maddelerden bir tanesinde düzeltme önerilmiş ve düzeltme uygulanarak tekrar uzman görüşü alınmıştır. Düzeltmenin sonucunda hazırlanan Matematik dersi açık uçlu maddelerin de sınıf seviyesine ve kazanımlara uygun olduğu belirtilmiştir.
- 5- *Ön uygulama:* Hazırlanan taslak testin öğrenciler için anlaşılabilirliğini, testin çözülme süresini, yönergeyi ve çözüm için ayrılan alanı değerlendirmek için 30 tane 8. sınıf öğrencisine ön uygulama yapılmıştır. Yapılan uygulama sonucunda testlerin sayısal ve sözel olarak farklı oturumlarda uygulanmasına karar verilmiştir. Ayrıca öğrencilere Türkçe testi için 20 dakika, matematik testi için de 30 dakika süre verilmesine karar verilmiştir.

Dereceli Puanlama Anahtarının Hazırlanması

Açık uçlu maddelerin uygun bir şekilde puanlanabilmesi için bütünsel dereceli puanlama anahtarı (BDPA) geliştirilmiştir. Dereceli puanlama anahtarı geliştirirken sistematik düzen takip edilmelidir. Aksi takdirde dereceli puanlama anahtarı amacından uzaklaşmakta, geçerli ve güvenilir sonuçlar elde edilememektedir (Kutlu vd., 2017, s. 65).

Bıkmaz Bilgen (2019, s. 197), Danielson & Hansen (2016, s. 59), Goodrich (1997), Haladyna (1997), Kutlu vd. (2017, s. 65), Mertler (2001), Moskal (2000)'ın önerdikleri dereceli puanlama anahtarı geliştirme aşamalarının ortak noktaları incelendiğinde;

1. Amacın belirlenmesi,
2. Puanlama anahtarı türüne (analitik, bütünsel) karar verilmesi,
3. Kriterlerin belirlenmesi,
4. Puanlamalara karşılık gelen davranışları belirlenmesi,
5. Ön uygulama
6. Redaksiyon
7. Nihai formun oluşturulması

şeklinde olduğu görülmektedir. Bu çalışmada da ortak öneriler göz önüne alınarak dereceli puanlama anahtarı geliştirilmiştir. Açık uçlu maddelerin cevabı öğrenciler tarafından yapılandırıldığından tek bir cevabı yoktur. Birden çok yanıtın olduğu durumlarda bütünsel dereceli puanlama anahtarı kullanımı önerilmektedir (Moskal, 2000). Hızlı puanlama yapılması gereken ulusal ve uluslararası sınavlarda yer alan açık uçlu soruların puanlanmasında da bütünsel puanlama anahtarı kullanılması önerilmektedir (Arter & McTighe, 2001, s. 49; Karakaya & Şata, 2022, s. 31; Mertler, 2001; Nitko & Brookhart, 2014, s. 220; Reynolds, Livingston, & Willson, 2006, s. 234). Bu çalışmada maddeye özgü bütünsel dereceli puanlama anahtarı oluşturulmuştur. Puanlama anahtarı oluşturulurken öğrenci performanslarının tanımlanmasına ve kategorilere ayrılmasına katkı sağlaması amacıyla öğrenci yanıtları incelenmiştir (Russell & Airasian, 2011, s. 231; Stevens & Levi, 2023, s. 49). Öğrencilerin her bir maddeye vermiş oldukları yanıtlar branş öğretmenleri tarafından belirli kriterler kapsamında genel izlenime dayalı olarak iyi, orta ve geliştirilmeli şeklinde gruplandırılmıştır. Gruplandırma işlemi sonrasında her gruptan 10 öğrenci yanıt seçkisiz olarak seçilmiştir. Toplam 30 öğrencinin yanıt kâğıdı bütünsel dereceli puanlama anahtarı hazırlanması için analize dahil edilmemiştir. Literatür incelendiğinde bütünsel dereceli puanlama anahtarının 4- 6 arasında düzeyden oluşturulabileceği ifade edilmektedir (Danielson & Hansen, 2016, s. 62; Kutlu vd., 2017, s. 67). Bu çalışmada da kriterler dört düzeye (1: en düşük - 4: en yüksek) bölünmüş ve tüm düzeyler doğru ve eksiksiz bir şekilde belirtilerek taslak form oluşturulmuştur. Araştırma kapsamında verilerin analize uygun olması için dereceli puanlama anahtarının en düşük puanı bir olarak belirlenmiştir.

Açık uçlu maddelerin değerlendirilmesi için hazırlanan taslak form 11 Türkçe alan uzmanı, 12 Matematik alan uzmanının görüşüne sunulmuştur. Ayrıca hem Türkçe hem de matematik derslerine ait taslak formlar üç ölçme ve değerlendirme uzmanı tarafından incelenmiştir. Uzman görüş formunda uzmanlardan her bir kriter için (1) uygun, (2) düzeltilmeli, (3) uygun değil şeklinde değerlendirme yapmaları istenmiştir. Ayrıca düzeltme uygun görülen kriterler için öneriler yazmaları da istenmiştir. Uzman görüşleri doğrultusunda kapsam geçerliği de incelenmiştir. Kapsam geçerliği incelenirken Lawshe tekniğinden faydalanılmıştır. Lawshe tekniğinde kapsam geçerliği belirlenirken;

$$KGO = \frac{N_G}{N/2} - 1 \quad (4.1.)$$

formülü kullanılmaktadır. Bu formülde KGO kriterin kapsam geçerlik oranını, N_G ilgili kriterin “uygun” olduğunu ifade eden uzman sayısını, N kriter hakkında görüş bildiren uzman sayısını göstermektedir. KGO, +1 ile -1 arasında değer almaktadır.

Uzman görüşlerine göre her kritere KGO hesaplama formülü uygulanmış ve her bir kriterin KGO değeri belirlenmiştir. Elde edilen değerler Veneziano & Hooper (1997) tarafından meydana getirilen tablo ile karşılaştırılmıştır (Yurdugül, 2005). Tablo 6’da kapsam geçerlik oranı (KGO) için gereken minimum değerler yer almaktadır.

Tablo 6
Kapsam Geçerlik Oranı (KGO) Minimum Değerleri

Uzman Sayısı (N)	Minimum Değer
5	0,99
6	0,99
7	0,99
8	0,78
9	0,75
10	0,62
11	0,59
12	0,56
13	0,54
14	0,51
15	0,49
20	0,42
25	0,37
30	0,33
35	0,31
40	0,29

Veneziano L. ve Hooper J. (1997). “A method for quantifying content validity of health-related questionnaires”. American Journal of Health Behavior, 21(1):67-70.

Tablo 6 incelendiğinde, örneğin 8 uzmanın görüş bildirdiği kriter için KGO değeri 0,78 veya daha büyük ise kriterin kapsam geçerliğine sahip, ölçülen özelliikle ilgili önemli bir kriter olduğu ve ölçme aracında kullanılabileceği söylenebilir. Bu çalışmada 11 Türkçe ve 12 Matematik uzmanı görüş bildirdiği için her bir kriterin KGO eşik değeri Türkçe için 0,59 ve Matematik için 0,56 olarak seçilmiştir. KGO değerleri eşik değerin üzerinde olan kriterlere nihai formda yer verilmiştir.

Ölçme aracındaki kriterlerin kapsam geçerlik oranları hesaplandıktan sonra ölçme aracının tamamı için kapsam geçerlik indeksi (KGİ) incelenir (Lawshe, 1975). Formun son haline alınacak maddelerin KGO değerlerinin ortalamaları ile KGİ hesaplanır (Yurdugül, 2005). KGİ'nin hesaplanırken;

$$KGİ = \frac{\sum_i^m KGO}{m} \quad (4.2)$$

formülü kullanılır. Formülde KGO, kapsam geçerlik oranını, m ise formda yer alan madde sayısını ifade etmektedir.

Açık Uçlu Maddelerin Geçerlik Çalışmaları

Türkçe testinde yer alan açık uçlu maddelere ilişkin kapsam geçerlik oranları Tablo 7'de sunulmuştur.

Tablo 7
Türkçe Dersi Açık Uçlu Maddelere Ait KGO Değerleri

Maddeler	Uygun	Düzeltilmeli	Uygun değil	KGO
1	10	1	0	0,818
2	9	2	0	0,636
3	11	0	0	1,000
4	10	1	0	0,818
5	10	0	1	0,818

Tablo 7 incelendiğinde, taslak formun Türkçe bölümünde yer alan beş açık uçlu maddenin KGO değerinin eşik değerinden (0,591) büyük olduğu görülmektedir.

Türkçe bölümünde yer alan açık uçlu maddelerin KGİ değeri 4.2 numaralı denklem ile hesaplanmış ve 0,818 olarak bulunmuştur. Bu değer 0.591'den büyük olduğu için ölçme aracının kapsam geçerliğinin istatistiksel açıdan anlamlı olduğu söylenebilir.

Matematik testinde yer alan açık uçlu maddelere ilişkin kapsam geçerlik oranları Tablo 8'de sunulmuştur.

Tablo 8

Matematik Dersi Açık Uçlu Maddelere Ait KGO Değerleri

Maddeler	Uygun	Düzeltilmeli	Uygun değil	KGO
1	12	0	0	1,000
2	11	1	0	0,833
3	11	1	0	0,833
4	10	2	0	0,667
5	10	1	1	0,667
6	12	0	0	1,000

Tablo 8 incelendiğinde, taslak formun Matematik bölümünde yer alan altı açık uçlu maddenin KGO değerinin eşik değerinden (0,561) büyük olduğu görülmektedir.

Matematik bölümünde yer alan açık uçlu maddelerin KGİ değeri 4.2 numaralı denklem ile hesaplanmış ve 0,833 olarak bulunmuştur. Bu değer 0.561'den büyük olduğu için ölçme aracının kapsam geçerliğinin istatistiksel açıdan anlamlı olduğu söylenebilir.

Bütünsel Dereceli Puanlama Anahtarının (BDPA) Geçerlik Çalışmaları

Bu bölümde BDPA'nın kapsam geçerliği ve güvenirlik çalışmalarına dair bilgiler yer almaktadır.

Kapsam geçerliği

Türkçe BDPA taslak formunda yer alan kriterlere ilişkin kapsam geçerlik oranları Tablo 9 ve Matematik BDPA taslak formunda yer alan kriterlere ilişkin kapsam geçerlik oranları Tablo 10'da verilmiştir.

Tablo 9

Türkçe BDPA Taslak Formu Kriterlere Ait KGO Değerleri

Madde	Kriterler	Uygun	Düzeltilmeli	Uygun değil	KGO
1	Tam doğru	11	0	0	1,000
	Yakın doğru	9	2	0	0,636
	Uzak doğru	10	1	0	0,818
	Yanlış/Boş	11	0	0	1,000
2	Tam doğru	11	0	0	1,000
	Yakın doğru	10	1	0	0,818
	Uzak doğru	10	1	0	0,818
	Yanlış/Boş	11	0	0	1,000
3	Tam doğru	10	1	0	0,818
	Yakın doğru	10	1	0	0,818
	Uzak doğru	10	1	0	0,818
	Yanlış/Boş	11	0	0	1,000
4	Tam doğru	11	0	0	1,000
	Yakın doğru	10	0	1	0,818
	Uzak doğru	9	2	0	0,636
	Yanlış/Boş	10	1	0	0,818
5	Tam doğru	11	0	0	1,000
	Yakın doğru	9	2	0	0,636
	Uzak doğru	9	2	0	0,636
	Yanlış/Boş	10	1	0	0,818

Tablo 9 incelendiğinde, beş maddeye ait kriterlerin KGO değerlerinin eşik değerinden (0,591) büyük olduğu görülmektedir.

Türkçe testine ait BDPA'daki kriterlerin KGİ değeri 4.2 numaralı denklem ile hesaplanmış ve 0,8453 olarak bulunmuştur. Bu değer 0.591'den büyük olduğu için ölçme aracının kapsam geçerliğinin istatistiksel açıdan anlamlı olduğu söylenebilir.

Tablo 10

Matematik BDPA Taslak Formu Kriterlere Ait KGO Değerleri

Madde	Kriterler	Uygun	Düzeltilmeli	Uygun değil	KGO
1	Tam doğru	11	1	0	0,833
	Yakın doğru	10	2	0	0,667
	Uzak doğru	11	1	0	0,833
	Yanlış/Boş	12	0	0	1,000
2	Tam doğru	12	0	0	1,000
	Yakın doğru	10	1	1	0,667
	Uzak doğru	10	2	0	0,667
	Yanlış/Boş	12	0	0	1,000
3	Tam doğru	12	0	0	1,000
	Yakın doğru	11	1	0	0,833
	Uzak doğru	10	2	0	0,667
	Yanlış/Boş	12	0	0	1,000
4	Tam doğru	12	0	0	1,000
	Yakın doğru	12	0	0	1,000
	Uzak doğru	12	0	0	1,000
	Yanlış/Boş	12	0	0	1,000
5	Tam doğru	12	0	0	1,000
	Yakın doğru	11	1	0	0,833
	Uzak doğru	10	2	0	0,667
	Yanlış/Boş	12	0	0	1,000
6	Tam doğru	12	0	0	1,000
	Yakın doğru	11	1	0	0,833
	Uzak doğru	11	1	0	0,833
	Yanlış/Boş	12	0	0	1,000

Tablo 10 incelendiğinde, altı maddeye ait kriterlerin KGO değerinin eşik değerinden (0,561) büyük olduğu görülmektedir.

Matematik testine ait BDPA'daki kriterlerin KGİ değeri 4.2 numaralı denklem ile hesaplanmış ve 0,842 olarak bulunmuştur. Bu değer 0.561'den büyük olduğu için ölçme aracının kapsam geçerliğinin istatistiksel açıdan anlamlı olduğu söylenebilir.

Puanlayıcı Değerlendirme Formu

Puanlayıcıların, demografik bilgilerinin elde edilmesi ve her bir açık uçlu sorulara verdikleri puanların yazılması amacıyla bu form hazırlanmıştır.

Verilerin Eldesi ve Deneysel İşlem

Bu bölümde veri toplama süreci ve puanlayıcı eğitimi ile ilgili açıklamalara yer verilmiştir.

2022-2023 eğitim öğretim yılı bahar döneminde 8. sınıfta öğrenimine devam eden öğrencilere başarı testi uygulanmıştır. Başarı testleri öğrencilere uygulanmadan önce öğrencilere araştırmanın amacı ve elde edilen verilerin nasıl kullanılacağı hakkında bilgi verilmiştir. Öğrencilere araştırmaya gönüllülük esasına dayalı olarak katılım sağlanacağı ve testin not ile değerlendirilmeyeceği, sadece çalışma için kullanılacağı bildirilmiştir. Ancak öğrencilerden maddeleri cevaplarken ciddi bir sınavı cevaplandırıyormuş gibi cevaplamaları istenmiştir. Ayrıca öğrencilerden cevaplarını yazarken mümkün olduğunca detaylı yazmaları istenmiştir.

Araştırmacı tarafından geliştirilen, tamamı açık uçlu sorulardan oluşan Türkçe ve Matematik başarı testleri 8. sınıf öğrencilerine iki oturumda uygulanmıştır. İlk oturum Türkçe testi ile başlamış ve öğrencilere 20 dakika süre verilmiştir. İlk oturumdan sonra 20 dakika ara verilmiştir. İkinci oturumda Matematik testi uygulanmış ve Matematik testi için öğrencilere 30 dakika süre verilmiştir.

Başarı testlerinin cevaplandırılmasının ardından öğrencilerin cevap kağıtları incelenerek Türkçe ve Matematik testleri kendi içinde iyi, orta, geliştirilmeli olarak ayrılmıştır. Türkçe ve Matematik testlerinden iyi, orta ve geliştirilmeli gruplarının her birinden on tane kâğıt olmak üzere toplam 60 kâğıt (30 Türkçe, 30 Matematik) seçkisiz olarak seçilmiş ve bütünsel dereceli puanlama anahtarı oluşturulmasında kolaylık sağlaması amacıyla ayrılmıştır. Geriye kalan kâğıtlar numaralandırılmış ve bilgisayar ortamına aktarılmıştır. Başarı testi veri toplama süreci tamamlandıktan ve bütünsel dereceli puanlama anahtarı hazırlandıktan sonra puanlayıcı gruplarına bütünsel dereceli puanlama anahtarı hakkında bilgi verilmiş ve başarı testinde yer almayan açık uçlu sorularla kısa örnek uygulama yapılmıştır. Bu sayede grupların benzer deneyimlere sahip olması hedeflenmiş ve yapıyla ilişkisiz varyans azaltılmaya çalışılmıştır.

Öğrenci cevaplarını içeren kâğıtlar bilgisayar ortamında Türkçe ve Matematik öğretmen adaylarından oluşan puanlayıcılar ile paylaşılmış ve puanlayıcılardan yaptıkları değerlendirmeleri puanlayıcı değerlendirme formuna (Bakınız Ek 5 ve Ek 8) yazmaları istenmiştir. Puanlayıcılara deney ya da kontrol grubunda oldukları hakkında bilgi verilmemiştir. Puanlayıcılar öğrenci cevaplarını içeren kâğıtları araştırmacı tarafından geliştirilen bütünsel dereceli puanlama anahtarı ile değerlendirmiştir. İlk değerlendirme

süreci için iki hafta süre verilmiştir. İki haftanın sonunda Türkçe öğretmen adaylarından oluşan 14 kişilik puanlayıcı grubu ve Matematik öğretmen adaylarından oluşan 20 kişilik puanlayıcı grubu puanlayıcı değerlendirme formunu araştırmacıya teslim etmiştir. Gruplardan elde edilen değerlendirmeler bilgisayar ortamına aktarılmış ve veriler analiz edilmiştir. Analizler sonucunda Türkçe ve matematik öğretmen adaylarından oluşan puanlayıcı gruplarının kendi içlerinde istatistiksel olarak benzer puanlayıcı davranışlarına sahip olduğu belirlenmiştir. Puanlayıcı gruplarının kendi içlerinde benzer davranışlara sahip oldukları belirlendikten sonra deney gruplarına puanlayıcı eğitimi verilmiştir.

Puanlayıcı Eğitimi

Açık uçlu soruların değerlendirilmesinde/puanlanmasında puanlayıcı davranışlarının azaltılması ve puanlayıcılar arasında ortak anlayışın geliştirilmesi amacıyla puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi harmanlanarak verilmiştir. Puanlayıcı hatası eğitimi yaygın karşılaşılan puanlayıcı davranışlarına ve bu davranışları azaltmaya odaklanmaktadır (Woehr & Huffcutt, 1994). Ancak yapılan çalışmalar puanlayıcı hatası eğitiminin puanlayıcılar arası güvenilirliğe anlamlı etki yapmadığını ve bazı puanlayıcı davranışlarını azaltırken puanlama doğruluğunu azalttığını göstermektedir (Bernardin & Pence, 1980; Borman, 1975). Puanlayıcı hatası eğitiminin bu sınırlılığından dolayı puanlayıcı hatası eğitimine puanlayıcı doğruluğunu artıran referans çerçevesi eğitimi entegre edilmiştir (Hedge & Kavanagh, 1988; Stamoulis & Hauenstein, 1993). Referans çerçevesi eğitimi puanlayıcıların puanlama için gerekli standartları benimsemesini sağlar ve puanlayıcıların vermiş olduğu puanlar ile gerçek puanlar arasında ilişki kurmalarını sağlar (Woehr & Huffcutt, 1994).

Deney grubunda bulunan puanlayıcılara verilen puanlayıcı eğitimi haftada bir saat olmak üzere dört hafta sürmüştür. Verilen puanlayıcı eğitiminin genel yapısı şekil 1’de verilmiştir.

Öğretmen Adayları İçin Puanlayıcı Eğitimi				
Süre	Konu	Hedef Davranış	Açıklama	Yöntem
20 dakika	Puanlayıcı Eğitimine Giriş	1. Dereceli puanlama anahtarı ve kullanımını bilir. 1. Puanlayıcı eğitimini, puanlayıcı eğitiminin amacını ve önemini bilir/kavrar.	1. Puanlayıcılara hedef grup (7. sınıf) ile ilgili bilgi verilecektir. 2. Puanlayıcılara puanlayıcı eğitimi ile ilgili bilgiler verilecektir.	Anlatım, Soru Cevap
40 dakika	Puanlayıcı Hatası Eğitimi	1. Puanlayıcı hata türlerini bilir. 2. Yaygın karşılaşılan puanlayıcı hata türlerine örnekler verir. 3. Puanlayıcı hatalarının sebeplerini bilir/kavrar. 4. Puanlayıcı hatalarının etkilerini sonuçlarını bilir. 5. Puanlayıcı hatalarını önleme ya da azaltmanın yollarını bilir.	1. Yaygın olarak karşılaşılan katılık, cömertlik ve puanlayıcı yanlılığı davranışları ile ilgili örnekler verilmesi beklenmektedir.	Anlatım, Soru Cevap
60 dakika	Puanlama Kriterleri ve Yöntemleri	1. Puanlamanın nasıl yapılacağını bilir. 2. Puanlamada kullanılacak kriterleri (rubriklerde yer almaktadır) bilir. 3. Örnek bir puanlama yapar.	1. Puanlamada kullanılacak puanlama anahtarının genel yapısı (kaç kriterin yer aldığı, hangi durumlarda hangi puanlamaların yapılacağı, vb.) hakkında bilgi verilecektir. 2. Rubriklerde yer alan durumlar ve açıklamalar puanlayıcı grubu ile tartışılır ve puanlayıcı grubunun tam olarak anlaması sağlanır. 3. Puanlama esnasında kullanılacak kriterlerle ilgili örnekler paylaşılabilecektir. Uygulama sonrasında puanlayıcıların sahip oldukları puanlayıcı davranışları ile ilgili açıklamalar yapılır.	Anlatım, Soru Cevap, Grup İçeri Tartışma
40 dakika	Uyumun Değerlendirilmesi ve Geri Bildirim	1. Puanlayıcı grubu ile ortak anlayış geliştirir.	1. Puanlayıcı grubu örnek birkaç maddeyi birlikte puanlar. Bu sayede ortak anlayışın geliştirilmesi sağlanır. 2. Yapılan değerlendirme sonuçlarından elde edilen veriler ile Rasch analizi uygulaması yapılacaktır.	Grup İçeri Tartışma
60 dakika	Örnek Uygulama ve Geri Bildirim	1. Verilen açık uçlu maddeleri amacına uygun değerlendirir.	1. Puanlayıcılar örnek bir maddeyi bireysel olarak değerlendirirler ve değerlendirme sonucu ile ilgili geri bildirim verilir.	Dönüt Verme

Şekil 1. Puanlayıcı eğitimi genel yapısı

Puanlayıcı eğitimi Şekil 1’de belirtilen puanlayıcı eğitim planı kapsamında yürütülmüştür. Puanlayıcı eğitiminin ilk haftasında puanlayıcılara puanlayıcı eğitiminin tanımı, amacı ve

önemi, puanlayıcı hata türleri, yaygın karşılaşılan puanlayıcı hata türleri hakkında bilgi verilmiştir. Puanlayıcılara hedef puanlayıcı davranışlarıyla ilgili örnekler sunulmuş ve puanlayıcılardan örnekler vermeleri beklenmiştir. Sonraki aşamada puanlayıcı hatalarının sebepleri, sonuçları ve puanlayıcı hatalarını önleme yolları hakkında bilgi verilmiştir. Sonrasında puanlayıcılardan sorular alınmış ve sorular cevaplandıktan sonra eğitimin ilk haftası tamamlanmıştır.

Puanlayıcı eğitiminin ikinci haftasında kısa bir tekrarın ardından puanlayıcılara puanlamada kullanılacak puanlama anahtarı hakkında detaylı bilgi verilmiştir. Bu kapsamda puanlama anahtarında kaç kriterin yer aldığı, hangi durumlarda hangi puanlamanın yapılacağı gibi konularda açıklama yapılmış ve tartışma ortamı sağlanmıştır. Devamında puanlayıcılara hangi durumlarda hangi puanın verilmesi gerektiği ile ilgili bilgi verilmiş ve ortak anlayışın geliştirilmesi amaçlanmıştır. Puanlayıcı grubunun soruları cevaplandırıldıktan sonra testte yer alan maddelerin kriterleri ile ilgili örnekler sunulmuştur.

Puanlayıcı eğitiminin üçüncü haftasında kısa bir tekrarın ardından araştırma kapsamında olmayan açık uçlu soruların puanlayıcılar tarafından puanlanması istenmiş ve puanlayıcılardan kaç puan verdikleri ve neden bu puanı verdikleri sorulmuştur. Verilen puanların grupta paylaşılması istenmiş ve grup içi tartışma ortamı oluşturulmuştur. Bu sayede puanlayıcıların ortak anlayış geliştirmeleri hedeflenmiştir. Örnek bir Rasch analizi yapıp puanlayıcı davranışları hakkında gruba bilgi verilmiştir. Gerekli dönüt düzeltmenin ardından puanlama işlemi tekrarlanmış ve tekrar Rasch analizi yapılmıştır. Analiz sonuçları grupta paylaşılmış ve durum değerlendirilmesi yapılmıştır.

Puanlayıcı eğitiminin dördüncü haftasında kısa bir hatırlatmanın ardından araştırma kapsamındaki açık uçlu sorulardan bazılarının puanlayıcılar tarafından puanlanması istenmiş ve puanlayıcılardan kaç puan verdikleri ve neden bu puanı verdikleri sorulmuştur. Verilen puanların grupta paylaşılması istenmiş ve grup içi tartışma ortamı oluşturulmuştur. Bu sayede puanlayıcıların araştırmada yer alan sorular ile ilgili ortak anlayış geliştirmeleri hedeflenmiştir. Ardından puanlayıcılara açık uçlu maddelerin her bir kriterini temsil eden açık uçlu sorular yöneltilmiş ve bu soruların puanlanması istenmiştir. Bu sayede kriterlerin somutlaştırılması hedeflenmiştir.

Puanlayıcı eğitimi tamamlandıktan sonra, ön testte olduğu gibi Türkçe ve Matematik puanlayıcı deney ve kontrol gruplarına bilgisayar ortamında öğrenci kağıtları, bütünsel dereceli puanlama anahtarı ve puanlayıcı değerlendirme formu verilmiştir. Öğrenci kağıtları

ön testteki sırayla puanlayıcılara verilmiş ve aynı sıra ile puanlanması sağlanmıştır. Puanlayıcı grubuna on gün süre verilmiş ve veri toplama süreci tamamlanmıştır.

Verilerin Analizi

Araştırmanın amacı kapsamında veri analizinde çok yüzeyli rasch modeli, Mann Whitney U testi ve Wilcoxon İşaretli Sıralar testi kullanılmıştır. Çok yüzeyli rasch modeli analizleri için FACET paket programı kullanılırken Mann Whitney U testi ve Wilcoxon İşaretli Sıralar testi için de SPSS paket programı kullanılmıştır. Araştırmada birey, puanlayıcı ve madde olmak üzere üç yüzey bulunmaktadır. Çok yüzeyli rasch modeli modellemelerinde tüm öğrenciler tüm maddeleri cevaplandığı ve tüm puanlayıcılar tüm öğrencileri puanladığı için tümüyle çaprazlanmış desen kullanılmıştır. Çok yüzeyli rasch modeli analizlerine geçmeden önce, analizin tutarlı sonuçlar üretebilmesi için karşılanması gereken varsayımlar kontrol edilmiştir. Çok yüzeyli rasch modeli, rash modellerinden olduğu için rash modeli varsayımlarını karşılamalıdır (Baker & Kim, 2004, s. 115; Eckes, 2015, s. 124; Embretson & Reise, 2000, s. 65; Farrokhi vd., 2011; Farrokhi vd., 2012). Bu varsayımlar tek boyutluluk, yerel bağımsızlık ve model veri uyumudur.

Tek boyutluluk

Madde tepki kuramı temelli modellerde karşılanması gereken varsayımlardan biri ölçülecek yapının tek boyutlu olmasıdır. Tek boyutluluk varsayımı, testin sadece bir özelliği ölçmesidir (Harvey & Hammer, 1999). Veri toplama araçları ile ilgili bölümde geliştirilen bütünsel dereceli puanlama anahtarının tek bir yapıyı ölçtüğü belirtilmiştir. Bu durum çok yüzeyli rasch modelinin ilk varsayımının karşılandığını göstermektedir.

Yerel bağımsızlık

Bir maddeye verilen cevabın diğer maddeye/maddelere verilen cevaptan etkilenmemesi yerel bağımsızlık olarak ifade edilir (Crocker & Algina, 2006, s. 343). Tek boyutluluk varsayımının yerel bağımsızlık varsayımı ile ilişkili olduğu ve tek boyutluluk varsayımı sağlandığında yerel bağımsızlık varsayımının da sağlandığı ifade edilmektedir (Hambleton vd., 1991). Bu kapsamda, araştırmada yerel bağımsızlık varsayımı test edilmemiş ve tek boyutluluk varsayımı karşılandığı için yerel bağımsızlık varsayımının da karşılandığı kabul edilmiştir.

Model veri uyumu

Model veri uyumu, ± 2 aralığı dışında kalan standartlaştırılmış artıkların sayısının toplam veri sayısının %5'i ve ± 3 aralığı dışında kalan standartlaştırılmış artık sayısının da toplam veri sayısının %1'i geçmemesini ifade etmektedir (Linacre, 2014b, s. 181). Çok yüzeyli rasch modeli analizlerinde de standartlaştırılmış artık değerleri (StRes) incelenerek belirlenmektedir.

Araştırmada Türkçe öğretmen adaylarından oluşan deney ve kontrol gruplarında 14 puanlayıcı (yedi deney, yedi kontrol) yer almaktadır. Puanlayıcılar 35 bireyin beş maddeye vermiş oldukları cevapları puanlamışlardır. Meydana gelen toplam gözlem sayısı $35 \times 5 \times 14 = 2450$ 'dir. Türkçe öğretmen adaylarından oluşan puanlayıcı grubunun puanlarına ait artık oranları tablo 11'de verilmiştir.

Tablo 11
Türkçe Öğretmen Adayları Standartlaştırılmış Artık Oranları

Test	Aralık	Artık (n)	Artık (%)
Ön Test	± 2	100	%4,08
	± 3	47	%1,91
Son Test	± 2	100	%4,08
	± 3	55	%2,25

Araştırmada matematik öğretmen adaylarından oluşan deney ve kontrol gruplarında 20 puanlayıcı (10 deney, 10 kontrol) yer almaktadır. Puanlayıcılar 35 bireyin altı maddeye vermiş oldukları cevapları puanlamışlardır. Deney grubu ve kontrol gruplarının her birinde oluşan gözlem sayısı $35 \times 6 \times 20 = 4200$ 'dür. Her bir grubun ön test ve son test verilerine ait standartlaştırılmış artık oranları tablo 12'de verilmiştir.

Tablo 12
Matematik Öğretmen Adayları Standartlaştırılmış Artık Oranları

Test	Aralık	Artık (n)	Artık (%)
Ön Test	± 2	100	%2,38
	± 3	51	%1,21
Son Test	± 2	100	%2,38
	± 3	58	%1,38

Tablo 11 ve tablo 12 incelendiğinde Türkçe dersi öğretmen adaylarından ve matematik dersi öğretmen adaylarından oluşan grubun ön test ve son test artık yüzdelerinin ± 2 düzeyinde uygun aralıkta (en fazla %5) olduğu görülmektedir. Ancak hem Türkçe dersi öğretmen adaylarından hem de matematik dersi öğretmen adaylarından oluşan grubun ± 3 düzeyi

dışında kalan artıklarının yüzdesinin %1 sınırını aştığı görülmektedir. Linacre (2014a) tarafından belirlenen sınırlar kesin olmayıp yaklaşık değer olarak sunulmuştur. Oranların %1'e yakın olduğu düşünüldüğünde model veri uyumunun kabul edilebilir olduğu düşünülmektedir. Ayrıca McNamara (1996), ± 2 veya ± 3 aralığı dışında kalan standart artıkların yüzdesinin önemli bir sapma olmadığı sürece çok yüzeyli rasch analizinin kullanılması gerektiğini ifade etmektedir. Çünkü rasch analizinde veriye en uygun model seçilmekte ve seçilen modele göre işlemler yürütülmekteyken çok yüzeyli rasch analizinde alternatif bir model bulunmamaktadır. Bu yüzden model veri uyumu istenen aralıkta olmasa da performans değerlendirme sürecinde çok yüzeyli rasch modelinin kullanılması önerilmektedir (McNamara, 1996). Sonuç olarak ± 3 aralığı dışında kalan standart artıkların yüzdesi çok yüzeyli rasch modelini kullanmaya engel olmayacak büyüklükte olduğu ifade edilebilir. Ayrıca uyum istatistikleri de model veri uyumunun derecesini belirlemek için kullanılmaktadır (Ariffin, 2008). Bir sonraki bölümde (Bulgular ve Yorum) uyum istatistikleri sonuçları yüzeyler bazında değerlendirilecektir.

Varsayımların karşılandığının belirlenmesinin ardından çok yüzeyli rasch analizine geçilmiştir.

BÖLÜM V

BULGULAR VE YORUM

Bu bölümde, araştırma kapsamında elde edilen verilerin, araştırmanın amaçları doğrultusunda çözümlenip ilgili literatür kapsamında yorumlanması yer almaktadır. İlk olarak çok yüzeyli rasch modelinde yer alan tüm yüzeylere (madde, birey ve puanlayıcı) ait uygunluk değerleri ve model ölçüm raporları incelenmiştir. Ölçüm raporlarının incelenmesinin ardından açık uçlu soruların puanlanmasında ortaya çıkan puanlayıcı davranışları belirlenmiştir. Ortaya çıkan puanlayıcı davranışları belirlendikten sonra puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelenmiştir. Yapılan incelemelerde öncelikle grup düzeyinde istatistikler incelenmiş ve gerekli durumlarda bireysel istatistikler incelenmiştir.

Birinci araştırma sorusuna ait bulgular

Araştırmanın birinci araştırma sorusu “Ön testte ve son testte birey, madde ve puanlayıcı yüzeylerinden elde edilen verilerin uygunluk değerleri ve model ölçümleri nasıldır?” şeklindedir. Bu doğrultuda öğrencilerin Türkçe ve Matematik açık uçlu maddelerine vermiş oldukları yanıtlar puanlayıcılar tarafından puanlanmıştır. Türkçe maddeleri 14, matematik maddeleri de 20 puanlayıcı tarafından puanlanmıştır. Elde edilen veriler çok yüzeyli rasch modeli kapsamında analiz edilmiştir. Sonuçlar raporlanırken Türkçe ve matematik ölçüm sonuçları farklı başlıklarda ele alınmıştır.

Türkçe puanlayıcı grubu uygunluk değerleri ve model ölçümleri

Birinci araştırma sorusu kapsamında 35 öğrencinin beş açık uçlu Türkçe maddelerine verdikleri cevaplar 14 puanlayıcı (yedi deney, yedi kontrol) tarafından puanlanmıştır. Elde edilen verilerin çok yüzeyli rasch modeline uygunluk değerleri ve model ölçüm raporları incelenmiştir. Veriler incelenirken önce ön teste ait sonuçlar raporlanmış sonrasında da son test ait sonuçlar verilmiştir. Türkçe puanlayıcı grubu ön test madde yüzeyine ait ölçüm raporu tablo 13'te verilmiştir.

Tablo 13

Türkçe Puanlayıcı Grubu Madde Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Maddeler	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	3,58	3,70	0,77	0,07	1,22	1,59
2	3,33	3,51	0,33	0,05	1,11	1,47
3	3,37	3,54	0,39	0,06	0,98	1,39
4	2,55	2,64	-0,58	0,05	0,74	0,72
5	2,25	2,21	-0,92	0,05	0,77	0,80
Ortalama	3,01	3,12	0,00	0,05	0,96	1,19
S (Evren)	0,52	0,59	0,64	0,01	0,19	0,36
S (Örneklem)	0,58	0,66	0,71	0,01	0,21	0,40
Model, Evren	RMSE= 0,05 Düzeltilmiş S.H.= 0,63 Ayırma Oranı= 11,62 Ayırma İndeksi= 15,83 Güvenirlilik= 0,99					
Model, Örneklem	RMSE= 0,05 Düzeltilmiş S.H.= 0,71 Ayırma Oranı= 13,00 Ayırma İndeksi= 17,67 Güvenirlilik= 0,99					
Model, Ki-kare (Sabit etkili) : 704,6 sd= 4 p= 0,00						
Model, Ki-kare (Normal) : 4,0 sd= 3 p= 0,26						
S: Standart sapma, RMSE: Hata kareleri ortalamasının karekökü, S.H.: Standart hata, sd: Serbestlik derecesi						

Çok yüzeyli rasch modeli madde yüzeyi incelemesi sonucunda elde edilen logit değerinin sıfırdan büyük olması maddenin güçlük düzeyinin yüksek, küçük olması da maddenin güçlük düzeyinin düşük olduğunu ifade etmektedir. Bir başka deyişle logit değeri arttıkça madde kolaylaşırken logit değeri azaldıkça madde zorlaşmaktadır. Tablo 13 incelendiğinde maddelerin logit değerlerinin -0,92 ile 0,77 arasında değiştiği ve maddelerin güçlük düzeyleri arasında 1,69 [0,77-(-0,92)] logit değişim olduğu görülmektedir. Maddelerin logit değerlerinin ortalaması 0,00 ve standart sapması 0,71'dir. Bu değerler ışığında en kolay maddenin birinci madde, en zor maddenin ise beşinci madde olduğu söylenebilir. Madde yüzeyinin uygunluk içi değerleri 0,74 ile 1,22 arasında değişirken ortalaması 0,96'dır. Madde yüzeyinin uygunluk dışı değerleri 0,72 ile 1,59 arasında değişirken ortalaması 1,19'dur. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm maddelerin istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi

olduğunu göstermektedir. Madde yüzeyinin ayırma indeksi değeri 17,67'dir. Bu değerin başlangıç değeri olan 2,00'nin çok üzerinde olması, maddelerin işlevsel olduğu ve farklı yetenek seviyesindeki bireylere de uygulanabileceğini göstermektedir (Nakamura, 2002). Madde yüzeyinin ayırma oranı 13,00 ve güvenilirlik indeksinin 0,99 olduğu görülmektedir. Bu değerlerin yüksek olması testte yer alan maddelerin güçlük düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için ki-kare istatistiğinden faydalanılmaktadır. Grup düzeyindeki çok yüzeyli rasch modeli analizlerinde evren ve örneklem olmak üzere iki farklı model sonucu üretilmektedir. Evren modeli, yüzeyin muhtemel tüm öğeleri modele dahil edilirken, örneklem modeli sadece seçilen öğeleri modele dahil etmektedir (Linacre, 2023, s. 214). Bir araştırmada bir yüzeye ait tüm öğelere ulaşmak çok zor olduğu için yüzeylerden seçilen örneklemlerden faydalanılmaktadır. Bu yüzden bir yüzeye ait grup istatistikleri raporlanırken örneklem modeline ait değerler dikkate alınmıştır. Bu kapsamda sabit etkili ki-kare değeri incelenmiş ve bu değerin istatistiksel olarak anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 704,6 (4)$, $p = 0,00 < 0,01$). Bu durum maddelerin güçlük düzeylerinin arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Madde yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından birey yüzeyine ait ölçüm raporu incelenmiştir. Birey yüzeyine ölçüm raporu tablo 14'te verilmiştir.

Tablo 14

Türkçe Puanlayıcı Grubu Birey Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Birey	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	2,39	2,33	-0,16	0,12	1,48	1,53
2	3,06	3,21	0,60	0,13	0,81	0,92
4	2,91	3,04	0,43	0,13	0,56	0,55
5	2,84	2,95	0,34	0,13	0,63	0,57
6	2,37	2,31	-0,17	0,12	1,11	1,07
7	2,40	2,35	-0,14	0,12	0,67	0,61
8	2,60	2,62	0,07	0,12	0,91	0,89
9	3,13	3,29	0,69	0,14	0,99	0,81
10	2,70	2,76	0,18	0,13	0,68	0,63
11	3,06	3,21	0,60	0,13	0,82	1,48
12	3,80	3,85	2,06	0,25	1,83	2,96
13	3,49	3,62	1,22	0,16	1,16	1,42
14	3,10	3,26	0,65	0,13	1,34	1,77
16	2,04	1,89	-0,53	0,13	1,21	1,58
18	2,77	2,86	0,26	0,13	0,68	0,62
19	2,21	2,10	-0,34	0,12	1,10	0,98
20	3,01	3,16	0,54	0,13	0,68	0,75
22	3,43	3,58	1,12	0,15	1,93	3,06
23	3,76	3,82	1,89	0,23	0,97	1,74
25	3,49	3,62	1,22	0,16	0,96	1,57
26	3,89	3,91	2,56	0,34	0,95	1,78
27	2,87	2,99	0,38	0,13	1,04	0,99
28	3,79	3,84	2,00	0,24	1,27	1,71
29	3,61	3,72	1,48	0,18	0,70	0,53
30	3,44	3,59	1,15	0,16	0,69	0,80
31	2,19	2,07	-0,37	0,12	0,82	0,73
32	2,77	2,86	0,26	0,13	0,64	0,67
33	2,71	2,78	0,20	0,13	0,63	0,59
34	2,53	2,52	-0,01	0,12	0,74	0,69
35	2,84	2,95	0,34	0,13	1,02	1,11
36	2,61	2,64	0,09	0,12	0,62	0,58
37	3,69	3,77	1,66	0,20	1,69	2,61
38	3,37	3,53	1,03	0,15	0,89	1,53
39	3,31	3,48	0,94	0,15	0,79	0,93
40	3,31	3,48	0,94	0,15	0,90	0,94
Ortalama	3,01	3,08	0,66	0,15	0,97	1,19
S (Evren)	0,50	0,57	0,75	0,05	0,34	0,66
S (Örnekleme)	0,51	0,57	0,76	0,05	0,35	0,67
Model, Evren	RMSE= 0,16 Düzeltilmiş S.H.= 0,73 Ayırma Oranı= 4,61 Ayırma İndeksi= 6,48 Güvenirlilik= 0,96					
Model, Örnekleme	RMSE= 0,16 Düzeltilmiş S.H.= 0,74 Ayırma Oranı= 4,68 Ayırma İndeksi= 6,58 Güvenirlilik= 0,96					
Model, Ki-kare (Sabit etkili) :	625,5 sd= 34 p= 0,00					
Model, Ki-kare (Normal) :	31,5 sd= 33 p= 0,54					

Çok yüzeyli rasch modeli birey yüzeyi incelemesi sonucunda elde edilen logit değerinin sıfırdan büyük olması birey yeteneğinin (performansının) yüksek, sıfırdan küçük olması ise birey yeteneğinin düşük olmasını ifade etmektedir. Bir başka deyişle logit değeri arttıkça

birey yeteneđi artarken logit değeri azaldıkça birey yeteneđi azalmaktadır. Tablo 14 incelendiđinde bireylerin yetenek düzeylerinin -0,53 ile 2,56 logit arasında değıştiđi ve bireylerin yetenek düzeyleri arasında 3,09 [2,56-(-0,53)] logit değışim olduđu görölmektedir. Logit değışim aralıđının büyük olması bireylerin performanslarının geniş bir aralıđa sahip olduđunu göstermektedir. Bireylerin logit değeri ortalaması 0,66 ve standart sapması 0,74'tür. Bu değeri ışığında en yüksek yeteneđe sahip birey 26 numaralı birey iken en düşük yetenek seviyesinde olan birey 16 numaralı bireydir. Birey yüzeyi için uygunluk içi değeri 0,56 ile 1,93 arasında değışirken ortalaması 0,97'dir. Birey yüzeyi için uygunluk dışı değeri 0,53 ile 3,06 arasında değışirken ortalaması 1,19'dur. Birey yüzeyinin uygunluk içi ve uygunluk dışı değeri ortalamalarının 1'e yakın olması tüm bireylerin istenen uygunluk değeri sahip olduđu ve model veri uyumunun iyi olduđunu göstermektedir. Birey yüzeyinin ayırma oranı 4,68 ve güvenilirlik indeksinin 0,96 olduđu görölmektedir. Bu değeri yüksek olması testi alan bireylerin yetenek düzeylerinin farklılık gösterdiđine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadıđını test etmek için ki-kare istatistiđinden faydalanılmaktadır. Grup düzeyindeki çok yüzeyli rasch modeli analizlerinde evren ve örneklem olmak üzere iki farklı model sonucu üretilmektedir. Evren modeli, yüzeyin muhtemel tüm öğeleri modele dahil edilirken, örneklem modeli sadece seçilen öğeleri modele dahil etmektedir (Linacre, 2023, s. 214). Bir araştırmada bir yüzeye ait tüm öğelere ulaşmak çok zor olduđu için yüzeylerden seçilen örneklemlerden faydalanılmaktadır. Bu yüzden bir yüzeye ait grup istatistikleri raporlanırken örneklem modeline ait değeri dikkate alınmıştır. Bu kapsamda sabit etkili ki-kare değeri incelenmiş ve bu değeri istatistiksel olarak anlamlı olduđu belirlenmiştir ($\chi^2(sd) = 625,5 (34), p = 0,00 < 0,01$). Bu durum bireylerin yetenek düzeyleri arasında istatistiksel olarak anlamlı bir fark olduđunu göstermektedir.

Birey yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından puanlayıcı yüzeyine ait ölçüm raporu incelenmiştir. Puanlayıcı yüzeyine ait ölçüm raporu tablo 15'te verilmiştir.

Tablo 15

Türkçe Puanlayıcı Grubu Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Puanlayıcı	Logit	Standart	Uygunluk	Uygunluk	Gözlenen	Beklenen	Rasch	t-
	Değeri	Hata	İçi	Dışı	Uyum	Uyum	Kappa	değeri
1	-0,07	0,09	0,99	1,31	61,40	46,00	0,37	1,67
2	0,15	0,09	0,72	0,80	66,70	47,10	0,26	4,70
3	0,47	0,10	0,66	0,67	61,40	47,80	0,39	2,00
4	-0,33	0,08	0,84	1,24	51,70	44,10	0,13	-3,75
5	0,18	0,09	0,81	0,88	68,00	47,20	0,21	-1,22
6	0,40	0,10	1,45	0,96	46,20	47,70	0,31	-2,22
7	-0,30	0,08	1,01	1,81	51,40	44,40	0,18	-5,25
8	0,06	0,09	0,94	0,83	63,90	46,70	0,29	-0,78
9	0,13	0,09	0,88	0,97	66,00	47,00	0,14	-4,13
10	0,14	0,09	0,79	0,76	62,10	47,00	-0,03	4,00
11	-0,11	0,09	0,95	1,64	57,40	45,80	0,32	0,67
12	-0,11	0,09	0,85	1,35	66,90	45,80	0,36	1,44
13	-0,20	0,09	0,83	1,30	62,20	45,20	0,28	1,56
14	-0,42	0,08	1,19	2,00	53,30	43,20	0,39	-1,22
Ortalama	0,00	0,09	0,92	1,19				
S (Evren)	0,26	0,00	0,19	0,51				
S (Örnekleme)	0,27	0,00	0,20	0,53				
Model, Evren	RMSE= 0,09 Düzeltilmiş S.H.= 0,24 Ayırma Oranı= 2,68 Ayırma İndeksi= 3,91 Güvenirlilik= 0,88							
Model, Örneklem	RMSE= 0,09 Düzeltilmiş S.H.= 0,25 Ayırma Oranı= 2,80 Ayırma İndeksi= 4,06 Güvenirlilik= 0,89							
Model, Ki-kare (Sabit Etkili) :	111,8	sd= 13	p= 0,00					
Model, Ki-kare (Normal) :	11,6	sd= 12	p= 0,47					
S: Standart sapma, RMSE: Hata kareleri ortalamasının karekökü, S.H.: Standart hata, sd: Serbestlik derecesi								

Tablo 15 incelendiğinde her puanlayıcıya ait logit değeri, logit değerine ait standart hata, uygunluk değerleri, uyum yüzdeleri, Rasch-kappa ve t-değerine ait bilgilerin yer aldığı görülmektedir. Rasch kappa istatistiği, çok yüzeyli rasch analizinde çıktı olarak verilmemektedir. Ancak analiz sonucu elde edilen uyum yüzdeleri kullanılarak elde edilir. Rasch kappa istatistiğinin hesaplanmasında;

$$\frac{\text{gözlenen uyum} - \text{beklenen uyum}}{100 - \text{beklenen uyum}}$$

formülü kullanılır. Ayrıca t- değeri hesaplanmasında da;

$$\frac{\text{puanlayıcıya ait logit değeri} - \text{puanlayıcıların logit ortalaması}}{\text{logit değerine ait standart hata (ölçmenin standart hatası)}}$$

formülü kullanılır.

Puanlayıcıya ait logit değerinin artı olması puanlayıcının cömertlik, eksi olması ise puanlayıcının katılık davranışını sergilediğini göstermektedir. Tablo 15'te puanlayıcıların logit değerlerinin -0,42 ile 0,47 arasında değiştiği ve puanlayıcıların katılık/cömertliklerine ait aralığın 0,89 [0,47-(-0,42)] olduğu görülmektedir. Puanlayıcı yüzeyi için uygunluk içi değerleri 0,66 ile 1,45 arasında değişirken ortalaması 0,92'dir. Puanlayıcı yüzeyi için uygunluk dışı değerleri 0,63 ile 2,61 arasında değişirken ortalaması 1,19'dur. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm puanlayıcıların istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Puanlayıcı yüzeyine ait ayırma oranının 2,80, ayırma indeksinin 4,06 ve güvenilirlik indeksinin 0,89 olduğu görülmektedir. Bu değerler puanlayıcıların açık uçlu soruları değerlendirme sürecinde farklı davranışlar sergileyebileceklerini göstermektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare değeri incelenmiştir. Yapılan incelemede sabit etkili ki-kare değerinin istatistiksel açıdan anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 111,8 (13), p = 0,00 < 0,01$). Bu durum Türkçe puanlayıcı grubundaki puanlayıcıların açık uçlu soruların değerlendirilmesinde farklı puanlayıcı davranışları sergilediklerini göstermektedir.

Araştırma kapsamında Türkçe puanlayıcı grubuna ait tüm yüzeylerin ön test sonuçlarının ölçüm raporları incelendikten sonra son teste ait ölçüm raporları verilmiştir. Bu kapsamda ön test raporuna benzer olarak yüzeyler madde, birey ve puanlayıcı sıralamasıyla sunulmuştur. Madde yüzeyinin son test ölçümlerine ait ölçüm raporu tablo 16'da sunulmuştur.

Tablo 16

Türkçe Puanlayıcı Grubu Madde Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Maddeler	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	3,79	3,85	1,21	0,10	1,35	1,27
2	3,39	3,54	0,12	0,06	1,13	1,39
3	3,33	3,49	0,02	0,06	0,62	0,58
4	3,33	3,48	0,01	0,06	0,95	1,34
5	2,23	2,19	-1,36	0,05	0,93	1,19
Ortalama	3,21	3,31	0,00	0,06	1,00	1,15
S (Evren)	0,52	0,58	0,82	0,02	0,24	0,30
S (Örneklem)	0,58	0,64	0,91	0,02	0,27	0,33
Model, Evren	RMSE= 0,07 Düzeltilmiş S.H.= 0,81 Ayırma Oranı= 12,12 Ayırma İndeksi= 16,50 Güvenirlilik= 0,99					
Model, Örneklem	RMSE= 0,07 Düzeltilmiş S.H.= 0,91 Ayırma Oranı= 13,56 Ayırma İndeksi= 18,42 Güvenirlilik= 0,99					
Model, Ki-kare (Sabit etkili) : 805,1 sd= 4 p= 0,00						
Model, Ki-kare (Normal) : 4,0 sd= 3 p= 0,26						
S: Standart sapma, RMSE: Hata kareleri ortalamasının karekökü, S.H.: Standart hata, sd: Serbestlik derecesi						

Tablo 16 incelendiğinde maddelerin logit değerlerinin -1,36 ile 1,21 arasında değiştiği ve maddelerin güçlük düzeyleri arasında 2,57 [1,21-(-1,36)] logit değişim olduğu görülmektedir. Maddelerin logit değerlerinin ortalaması 0,00 ve standart sapması 0,91'dir. Madde yüzeyinin uygunluk içi değerleri 0,62 ile 1,35 arasında değişirken ortalaması 1,00'dır. Madde yüzeyinin uygunluk dışı değerleri 0,58 ile 1,39 arasında değişirken ortalaması 1,15'tir. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm maddelerin istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Madde yüzeyinin ayırma indeksi değeri 18,42'dir. Bu değer başlangıç değeri olan 2,00'nin çok üzerinde olması, maddelerin işlevsel olduğu ve farklı yetenek seviyesindeki bireylere de uygulanabileceğini göstermektedir (Nakamura, 2002). Madde yüzeyinin ayırma oranı 13,56 ve güvenirlilik indeksinin 0,99 olduğu görülmektedir. Bu değerlerin yüksek olması testte yer alan maddelerin güçlük düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabiti etkili ki-kare değeri incelenmiş ve bu değer istatistiksel olarak anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 805,1$ (4), $p = 0,00 < 0,01$). Bu durum maddelerin güçlük düzeylerinin arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Madde yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından birey yüzeyine ait ölçüm raporu incelenmiştir. Birey yüzeyine ait ölçüm raporu tablo 17'de verilmiştir.

Tablo 17

Türkçe Puanlayıcı Grubu Birey Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Birey	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	2,43	2,39	-0,14	0,13	2,07	3,13
2	3,19	3,36	0,84	0,15	0,84	0,75
4	3,21	3,39	0,89	0,15	0,57	0,52
5	3,20	3,37	0,86	0,15	0,63	0,52
6	2,54	2,55	-0,01	0,13	0,75	0,72
7	2,89	3,00	0,41	0,14	0,64	0,60
8	3,20	3,37	0,86	0,15	0,89	1,17
9	3,17	3,34	0,82	0,15	0,62	0,54
10	3,09	3,25	0,69	0,15	0,48	0,43
11	3,34	3,51	1,11	0,16	0,93	2,12
12	3,79	3,85	2,28	0,26	1,82	2,54
13	3,69	3,78	1,90	0,21	1,22	1,40
14	3,16	3,33	0,80	0,15	1,26	1,43
16	1,94	1,78	-0,72	0,14	1,34	2,70
18	3,13	3,29	0,75	0,15	0,58	0,53
19	2,70	2,76	0,18	0,13	1,14	0,97
20	3,01	3,16	0,58	0,14	0,40	0,40
22	3,47	3,62	1,36	0,17	1,57	1,51
23	3,64	3,75	1,78	0,20	1,23	1,81
25	3,71	3,80	2,00	0,22	1,25	1,37
26	3,91	3,94	3,18	0,41	0,96	1,09
27	2,89	3,00	0,41	0,14	1,42	1,40
28	3,84	3,89	2,59	0,30	1,08	1,59
29	3,63	3,74	1,74	0,20	0,84	0,66
30	3,46	3,61	1,33	0,17	0,92	0,98
31	2,70	2,76	0,18	0,13	0,78	0,68
32	3,16	3,33	0,80	0,15	0,57	0,48
33	3,20	3,37	0,86	0,15	0,70	0,57
34	2,80	2,89	0,30	0,13	0,61	0,56
35	3,07	3,23	0,67	0,14	0,94	0,87
36	3,19	3,36	0,84	0,15	0,85	0,71
37	3,79	3,85	2,28	0,26	2,17	2,47
38	3,49	3,64	1,39	0,18	1,07	1,82
39	3,40	3,56	1,22	0,17	0,76	0,76
40	3,50	3,65	1,42	0,18	0,59	0,55
Ortalama	3,21	3,33	1,04	0,17	0,99	1,15
S (Evren)	0,43	0,46	0,80	0,06	0,43	0,73
S (Örneklem)	0,43	0,47	0,82	0,06	0,43	0,73
Model, Evren	RMSE= 0,18 Düzeltilmiş S.H.= 0,78 Ayırma Oranı= 4,29 Ayırma İndeksi= 6,05 Güvenirlilik= 0,95					
Model, Örneklem	RMSE= 0,18 Düzeltilmiş S.H.= 0,80 Ayırma Oranı= 4,35 Ayırma İndeksi= 6,14 Güvenirlilik= 0,95					
Model, Ki-kare (Sabit etkili) :	588,9 sd= 34 p= 0,00					
Model, Ki-kare (Normal) :	31,1 sd= 33 p= 0,56					

Tablo 17 incelendiğinde birey yüzeyi için uygunluk içi değerleri 0,40 ile 2,17 arasında değişirken ortalaması 0,99'dur. Birey yüzeyi için uygunluk dışı değerleri 0,40 ile 3,13 arasında değişirken ortalaması 1,15'tir. Uygunluk içi ve uygunluk dışı değerlerinden bazılarının istenen aralığa (0,50 ile 2,00) sahip olmadığı görülse de çalışmanın amacı ön test ve son test sonuçlarını karşılaştırmak olduğu için bu bireyler analizden için çıkarılmamıştır.

Ayrıca uygunluk indeksleri ortalamalarının 1'e yakın olması model veri uyumunun iyi olduğunu göstermektedir. Birey yüzeyinin ayırma oranı 4,35, ayırma indeksi 6,14 ve güvenilirlik indeksinin 0,95 olduğu görülmektedir. Bu değerlerin yüksek olması testi alan bireylerin yetenek düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare istatistiği incelenmiş ve bu değer istatistiksel olarak anlamlı olduğu bulunmuştur ($\chi^2(sd) = 588,9 (34)$, $p = 0,00 < 0,01$). Bu durum bireylerin yetenek düzeyleri arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Madde ve birey yüzeylerine ilişkin ölçüm raporunun incelenmesinin ardından puanlayıcı yüzeyine ait ölçüm raporu incelenmiştir. Puanlayıcı yüzeyine ait ölçüm raporu tablo 18'de verilmiştir.

Tablo 18

Türkçe Puanlayıcı Grubu Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Puanlayıcı	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı	Gözlenen Uyum	Beklenen Uyum	Rasch Kappa	t-değeri
1	0,10	0,10	0,87	0,76	73,20	51,60	0,45	1,00
2	0,07	0,10	0,82	0,84	69,60	51,40	0,37	0,70
3	0,14	0,10	0,85	0,81	74,40	51,80	0,47	1,40
4	-0,30	0,09	0,89	1,44	66,80	48,70	0,35	-3,33
5	-0,10	0,10	0,88	1,60	61,10	50,30	0,22	-1,00
6	0,14	0,10	0,85	0,81	74,40	51,80	0,47	1,40
7	-0,17	0,10	0,94	1,59	54,00	49,80	0,08	-1,70
8	0,08	0,10	1,07	0,97	67,50	51,50	0,33	0,80
9	0,15	0,10	1,01	1,24	67,20	51,80	0,32	1,50
10	0,12	0,10	0,83	0,80	67,10	51,70	0,32	1,20
11	-0,21	0,10	1,14	1,72	60,50	49,50	0,22	-2,10
12	0,30	0,11	0,88	0,88	71,70	52,40	0,41	2,73
13	-0,16	0,10	0,69	0,88	68,80	49,90	0,38	-1,60
14	-0,19	0,10	1,38	1,81	66,10	49,70	0,33	-1,90
Ortalama	0,00	0,10	0,94	1,15				
S (Evren)	0,17	0,00	0,16	0,38				
S (Örneklem)	0,18	0,00	0,17	0,39				
Model, Evren	RMSE= 0,10 Düzeltilmiş S.H.= 0,14 Ayırma Oranı= 1,40 Ayırma İndeksi= 2,20 Güvenirlilik= 0,66							
Model, Örneklem	RMSE= 0,10 Düzeltilmiş S.H.= 0,15 Ayırma Oranı= 1,48 Ayırma İndeksi= 2,31 Güvenirlilik= 0,69							
Model, Ki-kare (Sabit Etkili) :	41,7	sd= 13	p= 0,00					
Model, Ki-kare (Normal) :	9,9	sd= 12	p= 0,62					

Tablo 18 incelendiğinde her puanlayıcıya ait logit değeri, logit değerine ait standart hata, uygunluk değerleri, uyum yüzdeleri, Rasch-kappa ve t-değerine ait bilgilerin yer aldığı görülmektedir. Puanlayıcıların logit değerlerinin -0,30 ile 0,30 arasında değiştiği ve

puanlayıcıların katılık/cömertliklerine ait aralığın 0,60 [0,30-(-0,30)] olduğu görülmektedir. Puanlayıcı yüzeyi için uygunluk içi değerleri 0,69 ile 1,38 arasında değişirken ortalaması 0,94'tür. Puanlayıcı yüzeyi için uygunluk dışı değerleri 0,76 ile 1,81 arasında değişirken ortalaması 1,15'tir. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm puanlayıcıların istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Puanlayıcı yüzeyine ait ayırma oranının 1,48, ayırma indeksinin 2,31 ve güvenilirlik indeksinin 0,69 olduğu görülmektedir. Bu istatistiklerin küçük olması puanlayıcıların katılık ve cömertlik açısından benzer davranışlara sahip olabileceklerini göstermektedir. Ancak kesin yargıya varmak için sabit etkili ki kare istatistiğinin incelenmesi gerekmektedir. Sabit etkili ki kare istatistiği incelendiğinde, ki kare istatistiğinin istatistiksel olarak anlamlı olmadığı görülmektedir ($\chi^2(sd) = 41,7$ (13), $p = 0,00 < 0,01$). Bu durum Türkçe puanlayıcı grubundaki puanlayıcıların açık uçlu soruların değerlendirilmesinde farklı puanlayıcı davranışları sergilediklerini göstermektedir.

Matematik puanlayıcı grubu uygunluk değerleri ve model ölçümleri

Birinci araştırma sorusu kapsamında 35 öğrencinin altı açık uçlu Matematik maddelerine verdikleri cevaplar 20 puanlayıcı (on deney, on kontrol) tarafından puanlanmış ve elde edilen verilerin çok yüzeyli rasch modeline göre uygunluk değerleri ve model ölçüm raporları incelenmiştir. Veriler incelenirken önce ön teste ait sonuçlar raporlanmış sonrasında da son test ait sonuçlar verilmiştir. Madde yüzeyine ait ölçüm raporu tablo 19'da verilmiştir.

Tablo 19

Matematik Puanlayıcı Grubu Madde Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Maddeler	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	3,99	3,99	4,12	0,38	1,01	2,20
2	3,76	3,85	1,00	0,08	0,89	1,13
3	3,11	3,24	-0,48	0,05	0,90	0,92
4	2,19	2,05	-1,63	0,04	0,99	1,10
5	2,62	2,61	-1,12	0,04	0,64	0,62
6	1,98	1,82	-1,88	0,04	1,42	1,37
Ortalama	2,94	2,93	0,00	0,10	0,98	1,22
S (Evren)	0,75	0,83	2,07	0,12	0,23	0,49
S (Örneklem)	0,82	0,91	2,27	0,13	0,25	0,54
Model, Evren	RMSE= 0,16 Düzeltilmiş S.H.= 2,06 Ayırma Oranı= 12,84 Ayırma İndeksi= 17,46 Güvenirlilik= 0,99					
Model, Örneklem	RMSE= 0,16 Düzeltilmiş S.H.= 2,26 Ayırma Oranı= 14,08 Ayırma İndeksi= 19,10 Güvenirlilik= 0,99					
Model, Ki-kare (Sabit etkili) : 1610,3 sd= 5 p= 0,00						
Model, Ki-kare (Normal) : 4,9 sd= 4 p= 0,30						
S: Standart sapma, RMSE: Hata kareleri ortalamasının karekökü, S.H.: Standart hata, sd: Serbestlik derecesi						

Çok yüzeyli rasch modeli madde yüzeyi incelemesi sonucunda elde edilen logit değerinin sıfırdan büyük olması maddenin güçlük düzeyinin yüksek, küçük olması da maddenin güçlük düzeyinin düşük olduğunu ifade etmektedir. Bir başka deyişle logit değeri arttıkça madde kolaylaşırken logit değeri azaldıkça madde zorlaşmaktadır. Tablo 19 incelendiğinde maddelerin logit değerlerinin -1,88 ile 4,12 arasında değiştiği ve maddelerin güçlük düzeyleri arasında 6,00 [4,12-(-1,88)] logit değişim olduğu görülmektedir. Maddelerin logit değerlerinin ortalaması 0,00 ve standart sapması 2,26'dır. Bu değerler ışığında en kolay maddenin birinci madde, en zor maddenin ise altıncı madde olduğu söylenebilir. Madde yüzeyinin uygunluk içi değerleri 0,64 ile 1,42 arasında değişirken ortalaması 0,98'dir. Madde yüzeyinin uygunluk dışı değerleri 0,62 ile 2,20 arasında değişirken ortalaması 1,22'dir. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması maddelerin istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Madde yüzeyinin ayırma indeksi değeri 19,10'dur. Bu değer başlangıç değeri olan 2,00'nin çok üzerinde olması, maddelerin işlevsel olduğu ve farklı yetenek seviyesindeki bireylere de uygulanabileceğini göstermektedir (Nakamura, 2002). Madde yüzeyinin ayırma oranı 14,08 ve güvenirlilik indeksinin 0,99 olduğu görülmektedir. Bu değerlerin yüksek olması testte yer alan maddelerin güçlük düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare istatistiği incelenmiştir ve bu değer istatistiksel olarak anlamlı

Madde yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından birey yüzeyine ait ölçüm raporu incelenmiştir. Birey yüzeyine ölçüm raporu tablo 20’de verilmiştir.

Tablo 20

Matematik Puanlayıcı Grubu Birey Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Birey	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	2,16	2,32	-0,18	0,14	0,71	0,64
2	1,85	1,56	-1,01	0,16	0,66	0,89
4	2,95	3,57	1,25	0,11	1,25	1,04
6	3,21	3,72	1,65	0,12	0,99	3,02
7	3,11	3,67	1,49	0,11	1,21	0,90
8	3,58	3,87	2,36	0,14	1,27	0,99
9	2,68	3,32	0,83	0,12	1,00	0,67
10	2,50	3,06	0,52	0,12	0,96	0,67
11	3,38	3,79	1,93	0,12	0,68	4,60
12	2,89	3,52	1,16	0,11	0,90	1,05
13	3,49	3,84	2,15	0,13	0,73	0,48
14	3,40	3,80	1,98	0,12	0,97	0,96
15	2,46	2,99	0,44	0,12	0,99	0,99
16	2,51	3,08	0,53	0,12	1,33	1,07
17	2,79	3,43	1,00	0,11	1,62	1,20
19	3,36	3,79	1,90	0,12	1,12	3,77
21	3,35	3,78	1,89	0,12	1,10	0,77
22	2,34	2,76	0,22	0,13	1,56	1,53
23	3,03	3,62	1,36	0,11	0,63	0,44
24	3,12	3,68	1,52	0,11	0,88	0,63
25	3,55	3,86	2,28	0,14	0,88	0,50
26	2,92	3,54	1,19	0,11	1,18	2,01
27	2,78	3,42	0,99	0,11	0,81	1,46
28	3,07	3,64	1,43	0,11	0,86	0,77
29	3,08	3,65	1,44	0,11	0,76	0,71
30	3,19	3,71	1,62	0,12	0,61	0,41
31	2,76	3,40	0,95	0,11	1,57	1,37
32	1,93	1,74	-0,76	0,16	0,91	1,07
33	2,97	3,58	1,27	0,11	0,92	2,26
35	3,42	3,81	2,01	0,12	0,82	0,76
36	3,53	3,85	2,22	0,13	0,73	0,58
37	3,52	3,84	2,21	0,13	0,84	0,74
38	2,77	3,41	0,96	0,11	1,10	1,95
39	2,88	3,51	1,14	0,11	0,81	0,63
40	2,47	3,01	0,46	0,12	0,63	1,25
Ortalama	2,94	3,40	1,21	0,12	0,97	1,22
S (Evren)	0,45	0,55	0,81	0,01	0,27	0,92
S (Örneklem)	0,46	0,56	0,82	0,01	0,27	0,93
Model, Evren	RMSE= 0,12 Düzeltilmiş S.H.= 0,80 Ayırma Oranı= 6,47 Ayırma İndeksi= 8,96 Güvenirlilik= 0,98					
Model, Örneklem	RMSE= 0,12 Düzeltilmiş S.H.= 0,81 Ayırma Oranı= 6,57 Ayırma İndeksi= 9,09 Güvenirlilik= 0,98					
Model, Ki-kare (Sabit etkili) :	1190,9 sd= 34 p= 0,00					
Model, Ki-kare (Normal) :	33,0 sd= 33 p= 0,47					

Çok yüzeyli rasch modeli birey yüzeyi incelemesi sonucunda elde edilen logit değerinin sıfırdan büyük olması birey yeteneğinin (performansının) yüksek, sıfırdan küçük olması ise

birey yeteneğinin düşük olmasını ifade etmektedir. Bir başka deyişle logit değeri arttıkça birey yeteneği artarken logit değeri azaldıkça birey yeteneği azalmaktadır. Tablo 20 incelendiğinde bireylerin yetenek düzeylerinin -1,01 ile 2,36 logit arasında değiştiği ve bireylerin yetenek düzeyleri arasında 3,37 [2,36-(-1,01)] logit değişim olduğu görülmektedir. Logit değişim aralığının büyük olması bireylerin performanslarının geniş bir aralığa sahip olduğunu göstermektedir. Bireylerin logit değerlerinin ortalaması 1,21 ve standart sapması 0,81'dir. Bu değerler ışığında en yüksek yeteneğe sahip birey 8 numaralı birey iken en düşük yetenek seviyesinde olan birey 2 numaralı bireydir. Birey yüzeyi için uygunluk içi değerleri 0,61 ile 1,62 arasında değişirken ortalaması 0,97'dir. Birey yüzeyi için uygunluk dışı değerleri 0,41 ile 4,60 arasında değişirken ortalaması 1,22'dir. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm bireylerin istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Birey yüzeyinin ayırma oranı 6,57 ve güvenilirlik indeksinin 0,98 olduğu görülmektedir. Bu değerlerin yüksek olması testi alan bireylerin yetenek düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare istatistiği incelenmiş ve bu değer istatistiksel olarak anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 1190,9$ (34), $p = 0,00 < 0,01$). Bu durum bireylerin yetenek düzeyleri arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Birey yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından puanlayıcı yüzeyine ait ölçüm raporu incelenmiştir. Puanlayıcı yüzeyine ait ölçüm raporu tablo 21'de verilmiştir.

Tablo 21

Matematik Puanlayıcı Grubu Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Puanlayıcı	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı	Gözlenen Uyum	Beklenen Uyum	Rasch Kappa	t-değeri
1	0,00	0,09	0,87	0,66	73,80	54,40	0,43	0,00
2	0,06	0,09	1,14	0,95	73,50	54,50	0,42	0,67
3	0,28	0,09	1,05	0,78	71,30	54,50	0,37	3,11
4	-0,23	0,09	0,74	0,77	71,60	53,90	0,38	-2,56
5	-0,07	0,09	0,85	0,76	71,40	54,30	0,37	-0,78
6	-0,17	0,09	0,81	0,73	71,90	54,10	0,39	-1,89
7	-0,25	0,09	0,89	1,33	65,80	53,80	0,26	-2,78
8	-0,23	0,09	0,79	0,67	66,90	53,90	0,28	-2,56
9	0,01	0,09	1,14	0,93	73,40	54,40	0,42	0,11
10	0,32	0,09	1,36	1,04	72,50	54,50	0,40	3,56
11	0,18	0,09	1,62	8,57	54,00	54,60	-0,01	2,00
12	-0,07	0,09	0,94	0,82	74,80	54,30	0,45	-0,78
13	-0,05	0,09	0,83	0,70	72,80	54,30	0,40	-0,56
14	-0,05	0,09	0,92	0,80	74,80	54,30	0,45	-0,56
15	-0,27	0,09	0,72	0,64	73,20	53,70	0,42	-3,00
16	0,35	0,09	1,24	1,19	73,20	54,50	0,41	3,89
17	-0,04	0,09	0,87	0,74	74,70	54,30	0,45	-0,44
18	0,09	0,09	1,09	0,83	72,30	54,50	0,39	1,00
19	0,41	0,09	1,08	0,89	68,50	54,40	0,31	4,56
20	-0,27	0,09	0,72	0,64	73,20	53,70	0,42	-3,00
Ortalama	0,00	0,09	0,98	1,22				
S (Evren)	0,21	0,00	0,23	1,70				
S (Örneklem)	0,21	0,00	0,23	1,74				
Model, Evren	RMSE= 0,19 Düzeltilmiş S.H.= 0,19 Ayırma Oranı= 2,03 Ayırma İndeksi= 3,05 Güvenirlilik= 0,81							
Model, Örneklem	RMSE= 0,09 Düzeltilmiş S.H.= 0,19 Ayırma Oranı= 2,10 Ayırma İndeksi= 3,13 Güvenirlilik= 0,82							
Model, Ki-kare (Sabit Etkili) :	100,8	sd= 19	p= 0,00					
Model, Ki-kare (Normal) :	16,0	sd= 18	p= 0,59					

Tablo 21 incelendiğinde her puanlayıcıya ait logit değeri, logit değerine ait standart hata, uygunluk değerleri, uyum yüzdeleri, Rasch-kappa ve t-değerine ait bilgilerin yer aldığı görülmektedir. Puanlayıcıların logit değerlerinin -0,27 ile 0,41 arasında değiştiği ve puanlayıcıların katılık/cömertliklerine ait aralığın 0,68 [0,41-(-0,27)] olduğu görülmektedir. Puanlayıcı yüzeyi için uygunluk içi değerleri 0,72 ile 1,62 arasında değişirken ortalaması 0,98'dir. Puanlayıcı yüzeyi için uygunluk dışı değerleri 0,64 ile 8,57 arasında değişirken ortalaması 1,22'dir. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm puanlayıcıların istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Puanlayıcı yüzeyine ait ayırma oranının 2,10,

ayırma indeksinin 3,13 ve güvenilirlik indeksinin 0,82 olduğu görülmektedir. Bu değerler puanlayıcıların açık uçlu soruları değerlendirme sürecinde farklı davranışlar sergileyebileceklerini göstermektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki- kare değeri incelenmiştir. Yapılan incelemede sabit etkili ki-kare değerinin istatistiksel açıdan anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 100,8 (19), p = 0,00 < 0,01$). Bu durum matematik puanlayıcı grubundaki puanlayıcıların açık uçlu soruların değerlendirilmesinde farklı puanlayıcı davranışları sergilediklerini göstermektedir.

Araştırma kapsamında Matematik puanlayıcı grubuna ait tüm yüzeylerin ön test sonuçlarının ölçüm raporları incelendikten sonra son teste ait ölçüm raporları verilmiştir. Bu kapsamda ön test raporuna benzer olarak yüzeyler madde, birey ve puanlayıcı sıralamasıyla sunulmuştur. Matematik puanlayıcı grubu madde yüzeyinin son test ölçümlerine ait ölçüm raporu tablo 22’de sunulmuştur.

Tablo 22

Matematik Puanlayıcı Grubu Madde Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Maddeler	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	4,00	4,00	5,23	0,71	1,00	0,99
2	3,77	3,86	0,95	0,08	0,90	1,11
3	3,24	3,38	-0,46	0,05	1,00	0,93
4	1,82	1,65	-2,32	0,05	0,83	1,04
5	2,57	2,53	-1,39	0,04	0,72	0,74
6	2,06	1,90	-2,00	0,04	1,40	1,40
Ortalama	2,91	2,89	0,00	0,16	0,98	1,03
S (Evren)	0,82	0,92	2,58	0,24	0,21	0,20
S (Örneklem)	0,90	1,01	2,82	0,27	0,24	0,22
Model, Evren	RMSE= 0,29 Düzeltilmiş S.H.= 2,56 Ayırma Oranı= 8,75 Ayırma İndeksi= 12,00 Güvenirlik= 0,99					
Model, Örneklem	RMSE= 0,29 Düzeltilmiş S.H.= 2,81 Ayırma Oranı= 9,60 Ayırma İndeksi= 13,13 Güvenirlik= 0,99					
Model, Ki-kare (Sabit etkili) : 1897,1 sd= 5 p= 0,00						
Model, Ki-kare (Normal) : 4,8 sd= 4 p= 0,31						
S: Standart sapma, RMSE: Hata kareleri ortalamasının karekökü, S.H.: Standart hata, sd: Serbestlik derecesi						

Tablo 22 incelendiğinde maddelerin logit değerlerinin -2,32 ile 5,23 arasında değiştiği ve maddelerin güçlük düzeyleri arasında 7,55 [5,23-(-2,32)] logit değişim olduğu görülmektedir. Maddelerin logit değerlerinin ortalaması 0,00 ve standart sapması 2,81’dir. Bu değerler ışığında en kolay maddenin birinci madde, en zor maddenin ise dördüncü madde olduğu söylenebilir. Madde yüzeyinin uygunluk içi değerleri 0,72 ile 1,40 arasında değişirken ortalaması 0,98’dir. Madde yüzeyinin uygunluk dışı değerleri 0,74 ile 1,40

arasında deęiřirken ortalaması 1,03'tür. Uygunluk içi ve uygunluk dışı deęerlerinin ortalamalarının 1'e yakın olması tüm maddelerin istenen uygunluk deęerlerine sahip olduęu ve model veri uyumunun iyi olduęunu göstermektedir. Madde yüzeyinin ayırma indeksi deęeri 13,13'tür. Bu deęerin başlangıç deęeri olan 2,00'nin çok üzerinde olması, maddelerin işlevsel olduęu ve farklı yetenek seviyesindeki bireylere de uygulanabileceęini göstermektedir (Nakamura, 2002). Madde yüzeyinin ayırma oranı 9,60 ve güvenilirlik indeksinin 0,99 olduęu görölmektedir. Bu deęerlerin yüksek olması testte yer alan maddelerin güçlük düzeylerinin farklılık gösterdięine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadıęını test etmek için sabit etkili ki-kare istatistięi incelenmiřtir ve bu deęerin istatistiksel olarak anlamlı olduęu belirlenmiřtir ($\chi^2(sd) = 1897,1 (5), p = 0,00 < 0,01$). Bu durum maddelerin güçlük düzeylerinin arasında istatistiksel olarak anlamlı bir fark olduęunu göstermektedir.

Son test madde yüzeyine iliřkin ölçüm raporunun incelenmesinin ardından birey yüzeyine ait ölçüm raporu incelenmiřtir. Birey yüzeyine ölçüm raporu tablo 23'te verilmiřtir.

Tablo 23

Matematik Puanlayıcı Grubu Birey Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Birey	Gözlenen ortalama	Düzeltilmiş Ortalama	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı
1	2,08	2,18	-0,29	0,15	0,68	0,41
2	1,83	1,55	-1,06	0,17	0,58	1,31
4	2,96	3,67	1,54	0,12	1,20	1,30
6	3,20	3,79	1,96	0,12	0,98	0,66
7	3,17	3,78	1,90	0,12	1,10	0,84
8	3,49	3,88	2,51	0,13	1,37	0,80
9	2,61	3,35	0,91	0,13	1,00	0,71
10	2,52	3,22	0,73	0,13	1,26	0,80
11	3,20	3,79	1,96	0,12	0,81	1,98
12	2,88	3,62	1,39	0,12	0,83	0,59
13	3,41	3,86	2,34	0,13	0,86	0,50
14	3,37	3,85	2,26	0,13	0,90	0,62
15	2,46	3,12	0,61	0,13	1,47	1,71
16	2,70	3,46	1,08	0,12	1,56	1,34
17	2,73	3,49	1,14	0,12	1,65	6,89
19	3,38	3,85	2,28	0,13	1,19	1,71
21	3,37	3,85	2,26	0,13	0,77	0,46
22	2,47	3,13	0,63	0,13	1,84	1,44
23	2,96	3,67	1,54	0,12	0,76	0,53
24	3,07	3,73	1,72	0,12	0,65	0,46
25	3,58	3,91	2,73	0,14	0,72	0,45
26	2,91	3,64	1,45	0,12	1,01	1,69
27	2,90	3,63	1,44	0,12	0,61	0,48
28	3,11	3,75	1,80	0,12	0,67	0,46
29	3,07	3,73	1,72	0,12	0,60	0,42
30	3,08	3,74	1,74	0,12	0,64	0,48
31	2,81	3,56	1,28	0,12	0,99	0,69
32	2,03	2,04	-0,43	0,15	1,75	1,65
33	2,95	3,67	1,52	0,12	0,77	0,52
35	3,26	3,81	2,06	0,12	0,95	0,60
36	3,29	3,82	2,12	0,12	0,89	0,75
37	3,26	3,81	2,06	0,12	0,85	0,91
38	2,53	3,23	0,74	0,13	0,81	0,53
39	2,78	3,54	1,23	0,12	1,19	0,84
40	2,53	3,23	0,74	0,13	0,69	0,67
Ortalama	2,91	3,49	1,42	0,13	0,99	1,03
S (Evren)	0,42	0,53	0,83	0,01	0,34	1,10
S (Örneklem)	0,42	0,54	0,84	0,01	0,34	1,12
Model, Evren	RMSE= 0,13 Düzeltilmiş S.H.= 0,82 Ayırma Oranı= 6,42 Ayırma İndeksi= 8,90 Güvenirlilik= 0,98					
Model, Örneklem	RMSE= 0,13 Düzeltilmiş S.H.= 0,83 Ayırma Oranı= 6,52 Ayırma İndeksi= 9,02 Güvenirlilik= 0,98					
Model, Ki-kare (Sabit etkili) :	1177,0 sd= 34 p= 0,00					
Model, Ki-kare (Normal) :	33,0 sd= 33 p= 0,47					

Tablo 23 incelendiğinde bireylerin yetenek düzeylerinin -1,06 ile 2,73 logit arasında değiştiği ve bireylerin yetenek düzeyleri arasında 3,79 [2,73-(-1,06)] logit değişim olduğu görülmektedir. Logit değişim aralığının büyük olması bireylerin performanslarının geniş bir

aralığa sahip olduğunu göstermektedir. Bireylerin logit değerlerinin ortalaması 1,42 ve standart sapması 0,83'tür. Bu değerler ışığında en yüksek yeteneğe sahip bireyler 25 numaralı bireyler iken en düşük yetenek seviyesinde olan birey 2 numaralı bireydir. Birey yüzeyi için uygunluk içi değerleri 0,58 ile 1,84 arasında değişirken ortalaması 0,99'dur. Birey yüzeyi için uygunluk dışı değerleri 0,41 ile 6,89 arasında değişirken ortalaması 1,03'tür. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm bireylerin istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Birey yüzeyinin ayırma oranı 6,52 ve güvenilirlik indeksinin 0,98 olduğu görülmektedir. Bu değerlerin yüksek olması testi alan bireylerin yetenek düzeylerinin farklılık gösterdiğine işaret etmektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare istatistiği incelenmiş ve bu değer istatistiksel olarak anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 1177,0 (34)$, $p = 0,00 < 0,01$). Bu durum bireylerin yetenek düzeyleri arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Son test birey yüzeyine ilişkin ölçüm raporunun incelenmesinin ardından puanlayıcı yüzeyine ait ölçüm raporu incelenmiştir. Puanlayıcı yüzeyine ait ölçüm raporu tablo 24'te verilmiştir.

Tablo 24

Matematik Puanlayıcı Grubu Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Puanlayıcı	Logit Değeri	Standart Hata	Uygunluk İçi	Uygunluk Dışı	Gözlenen Uyum	Beklenen Uyum	Rasch Kappa	t-değeri
P1	0,12	0,10	1,15	0,89	72,70	56,20	0,38	1,20
P2	-0,08	0,10	0,92	0,84	74,00	56,10	0,41	-0,80
P3	0,35	0,10	1,21	1,02	71,10	56,00	0,34	3,50
P4	-0,05	0,10	0,85	0,64	74,50	56,10	0,42	-0,50
P5	-0,19	0,10	0,68	0,66	67,40	55,90	0,26	-1,90
P6	-0,12	0,10	0,77	0,65	78,20	56,00	0,50	-1,20
P7	-0,06	0,10	1,01	0,92	70,90	56,10	0,34	-0,60
P8	-0,18	0,10	0,98	0,84	69,90	55,90	0,32	-1,80
P9	0,02	0,10	1,10	0,82	75,30	56,20	0,44	0,20
P10	0,19	0,10	1,15	0,81	72,90	56,20	0,38	1,90
P11	0,03	0,10	1,62	4,35	47,90	56,20	-0,19	0,30
P12	-0,23	0,10	0,86	0,82	68,60	55,80	0,29	-2,30
P13	-0,52	0,10	0,54	1,78	62,00	54,70	0,16	-5,20
P14	0,21	0,10	1,14	0,86	74,10	56,20	0,41	2,10
P15	-0,10	0,10	0,82	0,81	76,70	56,00	0,47	-1,00
P16	0,20	0,10	0,97	0,79	76,30	56,20	0,46	2,00
P17	-0,06	0,10	0,79	0,59	76,20	56,10	0,46	-0,60
P18	0,05	0,10	1,02	0,77	75,50	56,20	0,44	0,50
P19	0,52	0,10	1,24	0,99	68,10	55,70	0,28	5,20
P20	-0,10	0,10	0,82	0,81	76,70	56,00	0,47	-1,00
Ortalama	0,00	0,10	0,98	1,03				
S (Evren)	0,22	0,00	0,23	0,80				
S (Örneklem)	0,23	0,00	0,24	0,82				
Model, Evren RMSE= 0,10 Düzeltilmiş S.H.= 0,20 Ayırma Oranı= 2,10 Ayırma İndeksi= 3,13 Güvenirlilik= 0,81								
Model, Örneklem RMSE= 0,10 Düzeltilmiş S.H.= 0,21 Ayırma Oranı= 2,16 Ayırma İndeksi= 3,22 Güvenirlilik= 0,82								
Model, Ki-kare (Sabit Etkili) : 105,1 sd= 19 p= 0,00								
Model, Ki-kare (Normal) : 16,1 sd= 18 p= 0,58								

Tablo 24 incelendiğinde her puanlayıcıya ait logit değeri, logit değerine ait standart hata, uygunluk değerleri, uyum yüzdeleri, Rasch-kappa ve t-değerine ait bilgilerin yer aldığı görülmektedir. Puanlayıcıların logit değerlerinin -0,52 ile 0,52 arasında değiştiği ve puanlayıcıların katılık/cömertliklerine ait aralığın 1,04 [0,52-(-0,52)] olduğu görülmektedir. Puanlayıcı yüzeyi için uygunluk içi değerleri 0,54 ile 1,62 arasında değişirken ortalaması 0,98'dir. Puanlayıcı yüzeyi için uygunluk dışı değerleri 0,59 ile 4,35 arasında değişirken ortalaması 1,03'tür. Uygunluk içi ve uygunluk dışı değerlerinin ortalamalarının 1'e yakın olması tüm puanlayıcıların istenen uygunluk değerlerine sahip olduğu ve model veri uyumunun iyi olduğunu göstermektedir. Puanlayıcı yüzeyine ait ayırma oranının 2,16,

ayırma indeksinin 3,22 ve güvenilirlik indeksinin 0,82 olduğu görülmektedir. Bu değerler puanlayıcıların açık uçlu soruları değerlendirme sürecinde farklı davranışlar sergileyebileceklerini göstermektedir. Farkın istatistiksel olarak anlamlı olup olmadığını test etmek için sabit etkili ki-kare değeri incelenmiştir. Yapılan incelemede sabit etkili ki-kare değerinin istatistiksel açıdan anlamlı olduğu belirlenmiştir ($\chi^2(sd) = 105,1 (19)$, $p = 0,00 < 0,01$). Bu durum matematik puanlayıcı son test grubundaki puanlayıcıların açık uçlu soruların değerlendirilmesinde farklı davranışlar sergilediklerini göstermektedir.

Araştırma kapsamında Türkçe ve matematik puanlayıcı gruplarının ön ve son test ölçümlerine ait madde, birey ve puanlayıcı yüzeylerinin çok yüzeyli rasch modeli sonuçları incelendiğinde, her iki puanlayıcı grubunun hem ön hem de son test sonuçlarına ait hesaplanan uygunluk değerlerinin ve ölçüm raporlarının kabul edilebilir aralıkta olduğu gözlenmiştir. Birey temelli yapılan incelemelerde de uygunluk istatistikleri için Türkçe puanlayıcı grubunda Myford & Wolfe (2003)'ün önerdiği 0,5 ve 2,0 aralığının dışında kalan ögenin olmadığı ancak matematik grubuna ait uygunluk istatistiklerinde üst sınır olan 2,00 değerini aşan ögelerin yer aldığı gözlenmiştir. Bond & Fox'dan aktaran (İlhan, 2015, s. 70) uygunluk dışı istatistiklerin uç değerlere karşı hassas olduğunu belirtmekte ve bu yüzden model veri uyumu incelemelerinde uygunluk içi istatistiklerinin göz önüne alınmasını önermektedir. Bu kapsamda her iki puanlayıcı grubunda da model veri uyumunu bozan ögenin olmadığı söylenebilir. Tüm bu sonuçlar ele alındığında, her iki puanlayıcı grup için yapılan analizlerin geçerli ve güvenilir olduğu söylenebilir.

İkinci araştırma sorusuna ait bulgular

Araştırmanın ikinci araştırma sorusu “Puanlayıcı eğitimi öncesinde açık uçlu soruların puanlanması sürecinde deney ve kontrol grubu puanlayıcı davranışları nasıldır?” şeklindedir. Bu doğrultuda Türkçe ve matematik puanlayıcı gruplarının puanlayıcı davranışları çok yüzeyli rasch modeli kapsamında incelenmiştir. İncelemelerde öncelikle Türkçe puanlayıcı grubu puanlayıcı davranışları incelenirken sonrasında Matematik puanlayıcı grubu puanlayıcı davranışları incelenmiştir. Ayrıca puanlayıcı davranışları incelenirken önce grup istatistikleri incelenirken sonrasında bireysel istatistikler değerlendirilmiştir. Sonuçlar raporlanırken Türkçe ve matematik ölçüm sonuçları farklı başlıklarda ele alınmıştır.

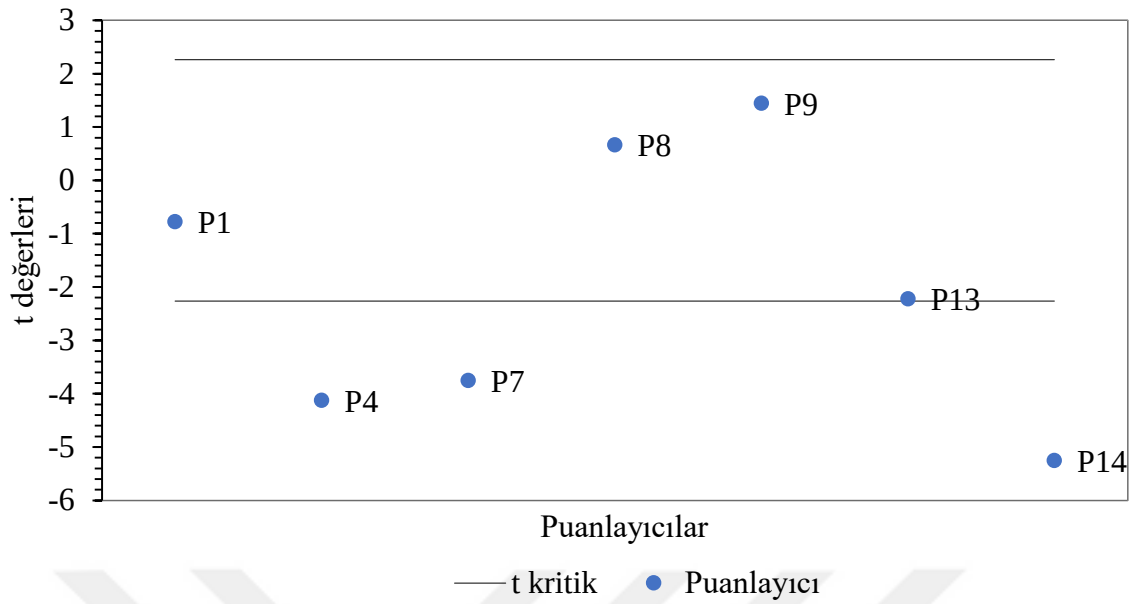
Puanlayıcı Eğitimi Öncesinde Türkçe Puanlayıcı Grubu Puanlayıcı Davranışları

Katılık ve Cömertlik

Puanlayıcı eğitimi öncesinde Türkçe puanlayıcı grubu deney ve kontrol gruplarının katılık ve cömertlik davranışlarının incelenmesi için puanlayıcı yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi gibi grup düzeyindeki istatistikler incelenmiştir. Yapılan incelemede bu değerlerin yüksek olduğu gözlenmiş ve bu durumun puanlayıcıların grup düzeyinde farklı davranışlar sergileyebileceklerini göstermiştir (Tablo 15). Bu durumu istatistiksel açıdan anlamlılığını test etmek için sabit etkili ki kare değeri incelenmiş ve bu değer anlamlı olduğu görülmüştür ($\chi^2(sd) = 111,8 (13)$, $p = 0,00 < 0,01$). Grup düzeyinde farklı davranışların olduğu tespit edildikten sonra farklılığın hangi puanlayıcı ya da puanlayıcılardan kaynaklandığını belirlemek için birey düzeyindeki istatistikler incelenmiştir.

Birey düzeyindeki istatistikler incelenirken t değeri istatistiği kullanılmaktadır. Puanlayıcıya ait t değeri, puanlayıcıya ait logit ölçüsünden tüm puanlayıcıların logit ortalaması farkının standart hataya bölünmesiyle elde edilir. Hesaplanan t değeri, t dağılım tablosundaki kritik t değeri ile karşılaştırılır ve istatistiksel açıdan anlamlılığı test edilir. Puanlayıcıya ait hesaplanan t değeri, kritik t değeri ile karşılaştırılır ve puanlayıcı davranışı belirlenir.

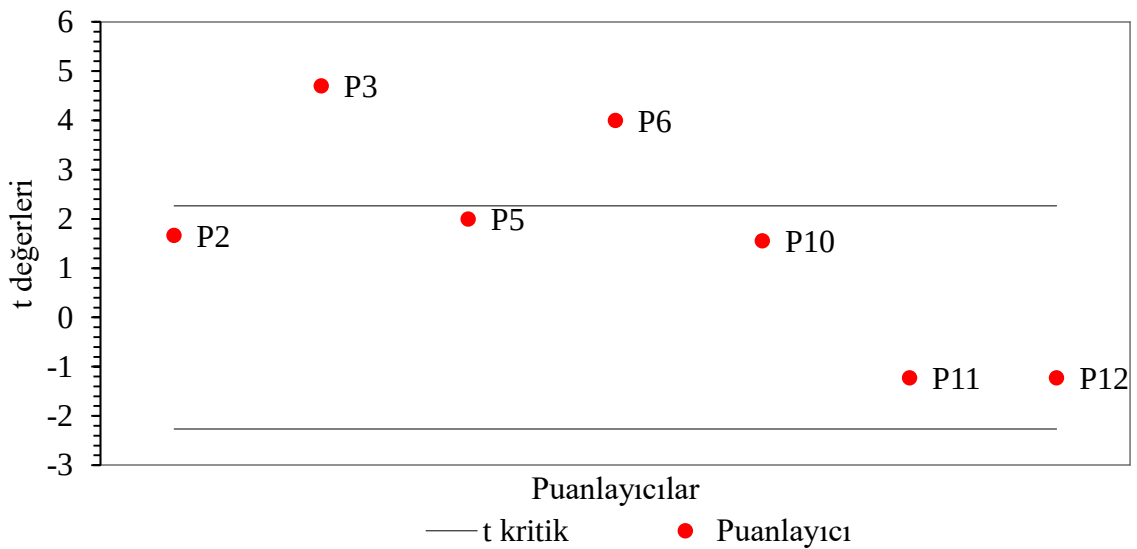
Araştırmada kontrol grubunda yedi, deney grubunda ise yedi puanlayıcı olduğundan serbestlik derecesi altıdır. Serbestlik derecesi altı ve anlamlılık düzeyi 0,05 için t kritik değeri (t_{kritik}) 2,447'dir. Tablo ve şekiller 2,447 değeri göz önünde bulundurularak yapılandırılmıştır. Türkçe deney grubunda yer alan puanlayıcıların ön test puanlarının katılık ve cömertlik durumları şekil 1'de verilmiştir.



Şekil 2. Türkçe puanlayıcı deney grubu ön test verilerinden elde edilen t değerleri.

Şekil 2 incelendiğinde, deney grubunda yer alan puanlayıcılardan P4, P7 ve P14 numaralı puanlayıcıların t değerlerinin kabul edilebilir aralığın dışında olduğu ve puanlayıcı katılımı davranışı sergilediği görülmektedir. Deney grubunda üç puanlayıcıda (%47,86) puanlayıcı katılımı davranışı yer alırken dört puanlayıcıda (%57,14) katılım ya da cömertlik davranışı bulunmamaktadır.

Türkçe puanlayıcı kontrol grubunda yer alan puanlayıcıların ön test puanlarının katılım ve cömertlik durumları şekil 3'te verilmiştir.



Şekil 3. Türkçe puanlayıcı kontrol grubu ön test verilerinden elde edilen t değerleri

Şekil 3 incelendiğinde, kontrol grubunda yer alan puanlayıcılardan P3 ve P6 numaralı puanlayıcıların t değerlerinin kabul edilebilir aralığın dışında olduğu ve puanlayıcı cömertliği davranışı sergilediği görülmektedir. Kontrol grubunda iki puanlayıcıda (%28,57) puanlayıcı cömertliği davranışı yer alırken beş puanlayıcıda (%71,43) katılık ya da cömertlik davranışı bulunmamaktadır.

Türkçe deney ve kontrol gruplarının ön test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Mann-Whitney U testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında kullanılacak en güçlü test Mann-Whitney U testidir (Tekindal, 2022, s. 135). T değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcıların t değerleri karşılaştırılmıştır. Türkçe puanlayıcı grubunun t değerlerinin farklarına ait Mann-Whitney U testi sonuçları tablo 25’te verilmiştir.

Tablo 25

Deney ve Kontrol Gruplarının Ön Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları

Grup	N	Sıra Ortalaması	Sıra Toplamı	U	p
Deney	7	7,57	53,00		
Kontrol	7	7,43	52,00	24,00	0,949
Toplam	14				

Tablo 25 incelendiğinde, Türkçe puanlayıcı deney ve kontrol grupları ön test sonuçları arasında katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olmadığı görülmektedir ($U=24,00$; $p=0,949>0,05$). Bu durum deney ve kontrol gruplarının puanlayıcı eğitimi öncesinde katılık ve cömertlik açısından benzer davranışlara sahip olduklarını göstermektedir.

Merkeze Eğilim

Puanlayıcı eğitimi öncesinde Türkçe puanlayıcı grubu deney ve kontrol gruplarının merkeze eğilim davranışının incelenmesi için kategori istatistikleri, birey yüzeyine ait ayırma oranı, ayırma indeksi, güvenilirlik indeksi, ki kare istatistiği ve madde yüzeyinin uygunluk istatistikleri gibi grup düzeyindeki istatistikler incelenmiştir. Deney ve kontrol gruplarına ait kategori istatistikleri tablo 26’da verilmiştir.

Tablo 26

Deney ve Kontrol Gruplarının Ön Test Sonuçlarına Ait Kategori İstatistikleri

Kategoriler	Frekans (f)	Yüzde (%)	Yığılmalı yüzde (%)	Ortalama logit	Beklenen logit	Uyum Dışı
1	437	18	18	-0,48	-0,42	1,2
2	294	12	30	0,15	0,04	1,6
3	516	21	51	0,63	0,59	1,3
4	1203	49	100	1,22	1,24	0,9

Tablo 26 incelendiğinde 4 numaralı kategorinin çok kullanıldığı, 1, 2 ve 3 numaralı kategorilerin ise benzer sayılarda kullanıldığı görülmektedir. Bu durumda puanlayıcıların merkeze eğilim davranışının olmadığı söylenebilir. Ancak istatistiksel olarak bu bilgi yeterli değildir. Merkeze eğilim davranışının belirlenmesi için birey yüzeyine ait ayırma oranı, ayırma indeksi, güvenilirlik indeksi ve ki kare istatistiği gibi grup düzeyindeki istatistiklerin incelenmesi gerekmektedir (Myford & Wolfe, 2004). Tablo 14'teki deney ve kontrol grubunda yer alan puanlayıcıların birey yüzeyine ait ölçüm raporu incelendiğinde, ayırma oranının, ayırma indeksinin ve güvenilirlik indeksinin yüksek olduğu ve ki kare değerinin de anlamlı çıktığı görülmektedir. Yani birey performanslarının yetenek seviyelerine göre doğru ve istatistiksel olarak anlamlı bir şekilde ayrıldığı söylenebilir. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Merkeze eğilim davranışının bir diğer göstergesi, madde yüzeyine ait uygunluk istatistikleridir. Madde yüzeyine ait ölçüm raporunda maddelerin uygunluk istatistiklerinin kabul edilebilir aralıkta (0,50 ile 2,00) olmaması merkeze eğilim davranışına işaret ederken uygunluk istatistiklerinin kabul edilebilir aralıkta olması merkeze eğilim davranışının olmadığını göstermektedir. Tablo 13'te yer alan deney ve kontrol gruplarının ön test madde yüzeyine ait ölçüm raporları incelendiğinde tüm maddelerin kabul edilebilir uygunluk istatistiklerine sahip olduğu görülmektedir. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Grup düzeyinde istatistikler incelendiğinde hem deney hem de kontrol grubunda bulunan puanlayıcıların açık uçlu soruların değerlendirilmesinde grup düzeyinde merkeze eğilim davranışını göstermedikleri belirlenmiştir.

Grup düzeyindeki merkeze eğilim davranışı incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel düzeyde merkeze eğilim davranışının belirlenmesinde her bir puanlayıcıya ait uyum istatistikleri ve her bir puanlayıcıya ait kategori istatistiklerinin incelenmesi gerekmektedir. Her bir puanlayıcıya ait uyum istatistiklerinin yer aldığı tablo 15 incelendiğinde, tüm puanlayıcıların uygunluk içi ve uygunluk dışı değerlerinin kabul

edilebilir aralıkta (0,50 ile 2,00) olduğu görülmektedir. Bu durum bireysel düzeyde merkeze eğilim davranışının olmadığı bir göstergesidir. Ancak bu sonuç tek başına yeterli değildir. Bireysel düzeyde merkeze eğilim davranışının incelenmesinde kullanılan bir diğer istatistik her bir puanlayıcıya ait kategori istatistiklerinin incelenmesidir. Kategori istatistikleri incelenirken uygunluk dışı istatistiği incelenmelidir. Deney ve kontrol gruplarının ön testlerinin her bir puanlayıcı için hesaplanan kategori istatistikleri tablo 27’de verilmiştir.

Tablo 27

Deney ve Kontrol Grubu Ön Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri

Grup	Puan- layıcı	Kategori İstatistiği (%)				Uygunluk Dışı				Eşik değer				Merkeze Eğilim
		1	2	3	4	1	2	3	4	1	2	3	4	
Deney	P1	25	9	23	43	0,5	2,4	1,4	0,8	--	0,61	-0,75	0,14	VAR
Kontrol	P2	18	6	25	52	0,9	4,9	1,0	0,8	--	0,19	-1,27	1,08	VAR
Kontrol	P3	10	6	31	53	0,4	0,9	2,7	1,1	--	-0,54	-1,54	2,08	VAR
Deney	P4	19	19	31	31	1,1	1,2	1,2	0,9	--	-0,88	-0,36	1,24	YOK
Kontrol	P5	19	7	23	51	0,6	3,8	1,1	0,8	--	0,31	-0,98	0,68	VAR
Kontrol	P6	7	10	36	48	1,1	1,2	0,9	0,9	--	-0,19	-0,71	0,90	YOK
Deney	P7	17	23	24	36	1,1	0,8	1,3	1,1	--	-0,62	0,21	0,42	YOK
Deney	P8	32	10	7	52	0,5	1,7	4,5	0,5	--	0,29	0,56	-0,85	VAR
Deney	P9	24	14	11	51	0,6	3,4	1,8	0,7	--	0,04	0,49	-0,54	VAR
Kontrol	P10	18	16	17	49	1,4	1,4	1,4	0,9	--	-0,52	0,28	0,24	YOK
Kontrol	P11	16	21	18	45	1,3	0,7	1,7	0,9	--	-0,52	0,40	0,12	YOK
Kontrol	P12	22	13	19	45	1,1	2,4	1,0	0,8	--	0,14	-0,19	0,05	VAR
Deney	P13	23	15	24	38	0,9	1,5	0,9	0,9	--	-0,16	-0,33	0,49	YOK
Deney	P14	25	14	28	33	1,1	1,2	1,1	0,9	--	0,27	-0,59	0,31	YOK

Tablo 27 incelendiğinde, 14 puanlayıcıdan yedi puanlayıcıda dereceli puanlama anahtarının bazı kategorilerine eğilim davranışı olduğu görülmektedir. Tablo 22’ye göre P2, P3, P5, P8 ve P9 numaralı puanlayıcıların bazı kategorileri diğer kategorilere göre daha yoğun kullandıkları belirlenmiştir. Dereceli puanlama anahtarının kullanım örüntüsü incelendiğinde puanlayıcıların ölçeğin merkez bölümünü kullanma eğilimlerinin daha az olduğu görülmektedir. Ancak istatistiksel olarak incelendiğinde, merkeze eğilim davranışı, deney grubunda P1, P8 ve P9 numaralı puanlayıcılarda olmak üzere üç, kontrol grubunda da P2, P3, P5 ve P12 numaralı puanlayıcılarda olmak üzere dört puanlayıcıda görülmektedir.

Puanlayıcıların merkeze eğilim davranışına yatkınlıklarının karşılaştırılması için ortalama eşik analizi yapılmıştır (Rahman, Ahmad, Yasin, & Hanafi, 2017). Bir numaralı puanlayıcının ortalama eşik değerinin 0,68 $[[0,61-(-0,75)] / 2]$, iki numaralı puanlayıcının ortalama eşik değerinin 1,17 $[[1,08-(-1,27)] / 2]$, üç numaralı puanlayıcının ortalama eşik değerinin 1,81 $[[2,08-(-1,54)] / 2]$, beş numaralı puanlayıcının ortalama eşik değerinin 0,83 $[[0,68-(-0,98)] / 2]$, sekiz numaralı puanlayıcının ortalama eşik değerinin 0,70 $[[0,56-(-0,85)] / 2]$, dokuz numaralı puanlayıcının ortalama eşik değerinin 0,51 $[[0,49-(-0,54)] / 2]$, on iki numaralı puanlayıcının ortalama eşik değerinin 0,16 $[[0,14-(-0,19)] / 2]$ olduğu görülmektedir. Bu durum üç numaralı puanlayıcının merkeze eğilim davranışına diğer puanlayıcılara göre daha yatkın olduğunu göstermektedir.

Halo Etkisi

Puanlayıcı eğitimi öncesinde Türkçe puanlayıcı grubu deney ve kontrol gruplarının halo etkisinin incelenmesi için madde yüzeyine ait ayırma oranı, ayırma indeksi ve sabit etkili ki kare istatistiği gibi grup düzeyindeki istatistikler incelenmiştir. Madde yüzeyine ait ölçüm raporunun yer aldığı tablo 13 incelendiğinde; ayırma oranının 13,64, ayırma indeksinin 18,52, güvenilirlik indeksinin 0,99 olduğu ve sabit etkili ki kare istatistiğinin anlamlı olduğu ($\chi^2(sd) = 766,2 (4), p = 0,00 < 0,01$) görülmektedir. Madde yüzeyine ait istatistiklerin yüksek ve sabit etkili ki kare değerinin anlamlı olması puanlayıcıların yapmış oldukları puanlamalarda grup düzeyinde halo etkisi davranışının olmadığını göstermektedir. Grup düzeyinde halo etkisi davranışı incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel istatistikler incelenirken, madde güçlükleri arasındaki fark büyük olduğunda puanlayıcıya ait uygunluk istatistikleri 1'den çok büyük olduğunda, madde güçlükleri arasındaki fark küçük olduğunda da puanlayıcıya ait uygunluk istatistikleri 1'den çok küçük olduğunda halo etkisi davranışının sergilendiği ifade edilmiştir (Myford & Wolfe, 2004). Türkçe puanlayıcı grubu ön test madde yüzeyine ait ölçüm raporunun yer aldığı tablo 13 incelendiğinde maddelerin güçlük düzeyleri arasındaki farkın 1,69 $(0,77-[-0,92])$ olduğu görülmektedir. Bu fark büyük olduğundan puanlayıcılara ait uygunluk istatistiklerinin 1'den çok büyük olanların halo etkisi davranışı gösterdikleri kabul edilmiştir. Türkçe puanlayıcı grubu ön test puanlayıcı yüzeyine ait uygunluk istatistiklerinin yer aldığı Tablo 15 incelendiğinde, dört puanlayıcıda halo etkisi davranışı tespit edilmiştir. Halo etkisi deney grubunda P7 ve P14 numaralı puanlayıcılarda belirlenirken kontrol grubunda P11 numaralı puanlayıcıda belirlenmiştir.

Halo etkisinin bulunup bulunmadığına karar vermek için önerilen bir diğer yöntem ise madde güçlüklerinin birbirine eşitleyerek çok yüzeyli rasch analizini tekrar edip puanlayıcılara ait uyum istatistiklerini incelemektir (Linacre, 2023, s. 360). Madde güçlükleri eşitlenerek oluşturulan yeni modele mükemmel uyum gösteren (uygunluk içi=1,00 veya uygunluk dışı=1,00) puanlayıcıların halo davranışını sergiledikleri kabul edilmektedir (Linacre, 2023, s. 360). Bu yüzden Türkçe testindeki beş açık uçlu sorunun kriter güçlükleri sıfıra eşitlenerek çok yüzeyli rasch analizi tekrar edilmiştir. Ardından uygunluk istatistikleri incelenmiştir. Kriter güçlükleri eşitlendikten sonra Türkçe puanlayıcı grubunun puanlayıcı yüzeyine ait ölçüm raporu Tablo 28’de verilmiştir.

Tablo 28

Türkçe Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Grup	Puanlayıcı	Uygunluk İçi	Uygunluk Dışı
Deney	P1	1,00	0,87
Kontrol	P2	1,02	0,89
Kontrol	P3	0,95	0,84
Deney	P4	0,82	0,78
Kontrol	P5	1,04	0,88
Kontrol	P6	0,94	1,33
Deney	P7	0,88	0,94
Deney	P8	1,22	1,09
Deney	P9	1,13	0,97
Kontrol	P10	1,06	0,95
Kontrol	P11	1,01	1,04
Kontrol	P12	1,01	1,01
Deney	P13	0,96	1,00
Deney	P14	0,99	1,03

Tablo 28’de Türkçe puanlayıcı grubuna ait uygunluk içi ve uygunluk dışı gibi bireysel istatistikleri yer almaktadır. Puanlayıcılara ait bireysel istatistikler incelendiğinde, deney grubundaki P1 ve P13 numaralı puanlayıcıların mükemmel uyum gösterdiği ve dolayısıyla halo davranışını sergiledikleri belirlenmiştir.

Tüm bireysel istatistiklerin incelenmesi sonucunda, deney grubunda P1, P7, P13 ve P14 numaralı puanlayıcılarda olmak üzere dört, kontrol grubunda da P11 numaralı puanlayıcıda olmak üzere bir puanlayıcıda halo etkisi belirlenmiştir.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)

Farklılaşan puanlayıcı katılığı ve cömertliği davranışı incelenirken modelde yer alan yüzeyler arasındaki etkileşimler incelenmektedir. Mevcut çalışma kapsamında birey (b), puanlayıcı (p) ve madde (m) yüzeyi olmak üzere üç yüzey bulunmaktadır. Bu yüzeyler arasındaki etkileşimler de *puanlayıcı x birey (pxb)*, *puanlayıcı x madde (pxm)*, *birey x madde (bxm)* ve *birey x puanlayıcı x madde (bxpxm)* şeklindedir. Mevcut çalışmada puanlayıcıların maddelerle ilgili davranışlarını incelemek amaçlanmadığı için *puanlayıcı x madde (pxm)* etkileşimi çalışmaya dahil edilmemiştir. Benzer şekilde çalışmada puanlayıcı davranışları incelendiği ve bireylerin maddeleri cevaplandırma seviyeleri de incelenmediği için *birey x madde (bxm)* etkileşimi de çalışmaya dahil edilmemiştir. Ayrıca puanlayıcıların hem madde hem de bireye göre davranışları araştırma kapsamında olmadığı için *birey x puanlayıcı x madde (bxpxm)* etkileşimi de çalışmaya dahil edilmemiştir. Çalışma kapsamında Türkçe açık uçlu soruların puanlanması sürecinde meydana gelen yanlılık davranışlarının tespiti için *puanlayıcı x birey (pxb)* etkileşimi incelenmiştir.

Puanlayıcı x birey (pxb) etkileşiminde yer alan istatistiksel olarak anlamlı olan etkileşimlerin tespiti için t değerlerinden faydalanılmaktadır. Linacre (2023, s. 190), ± 2 aralığı dışında kalan t değerlerinin istatistiksel olarak anlamlı olduğunu ifade etmektedir. Mevcut çalışmada 35 öğrencinin Türkçe açık uçlu sorulara vermiş oldukları tüm cevaplar yedi deney, yedi kontrol olmak üzere toplam 14 puanlayıcı tarafından puanlanmıştır. Dolayısıyla iki yüzey arasında 490 (14 x 35) olası etkileşim varken 245 (7 x 35) etkileşim deney grubunda, 245 (7 x 35) etkileşim de kontrol grubunda meydana gelmiştir. Analiz sonuçları incelenirken öncelikle grup düzeyindeki istatistikler incelenmiş ardından bireysel istatistiklere geçilmiştir. Grup düzeyindeki istatistikler incelendiğinde ki kare istatistiğinin istatistiksel açıdan anlamlı olmadığı bulunmuştur (χ^2 (sd)=244,5 (490), $p > 0,01$). Bu durum puanlayıcıların grup düzeyinde yanlılık davranışı sergilemedikleri şeklinde yorumlanabilir. Grup düzeyindeki istatistikler incelendikten sonra bireysel istatistikler incelenmiştir. Yapılan incelemede, her grup için meydana gelen olası 245 etkileşimden kontrol grubunda istatistiksel olarak anlamlı dört tane etkileşim bulunurken deney grubunda istatistiksel olarak anlamlı iki tane etkileşim yer almaktadır. Yani Türkçe puanlayıcı grubunda toplam altı puanlayıcı farklılaşan puanlayıcı katılığı ve cömertliği davranışı sergilemiştir. Bir başka deyişle altı puanlayıcıda yanlılık davranışı vardır. Grupların etkileşimlerine ait istatistiklerin tamamı Ek 9'da yer alırken etkileşimlerin tamamını sunmak mümkün olmadığı için sadece istatistiksel olarak anlamlı olan etkileşimler tablo 29'da verilmiştir.

Tablo 29

Türkçe Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri

Grup	Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Bias (logit)	Standart Hata	t
Deney	P4	35	18	12,76	1,40	0,66	2,12
Kontrol	P6	28	16	19,35	-1,63	0,51	-3,17
Kontrol	P6	37	15	19,05	-1,54	0,49	-3,16
Kontrol	P6	22	13	18,22	-1,45	0,46	-3,14
Kontrol	P6	16	18	12,07	1,55	0,66	2,34
Deney	P8	35	10	14,50	-0,96	0,47	-2,07
χ^2 (sd)=244,5 (490), $p > 0,01$							

Tablo 29 incelendiğinde, kontrol grubunda yer alan puanlayıcılardan P6 numaralı puanlayıcıda farklılaşan puanlayıcı katılımı ve cömertliği davranışı olduğu görülmektedir. Örneğin P6 numaralı puanlayıcı 22, 28 ve 37 numaralı bireyleri katı puanlarken 16 numaralı bireyi cömert puanlamıştır. Ayrıca deney grubunda yer alan puanlayıcılardan P4 ve P8 numaralı puanlayıcılarda farklılaşan puanlayıcı katılımı ve cömertliği davranışı olduğu görülmektedir. Örneğin P4 numaralı puanlayıcı 35 numaralı bireyi puanlarken cömert davranırken P8 numaralı puanlayıcı 35 numaralı bireyi puanlarken katı davranmıştır. Özetle farklılaşan puanlayıcı katılımı ve cömertliği davranışı deney grubunda iki puanlayıcıda (P4 ve P8) gözlenirken kontrol grubunda bir puanlayıcıda (P6) gözlenmiştir.

İkinci araştırma sorusu kapsamında puanlayıcı eğitimi öncesinde Türkçe puanlayıcı grubunun puanlayıcı davranışları incelenmiştir. Türkçe puanlayıcı grubunun puanlayıcı davranışları incelendikten sonra puanlayıcı eğitimi öncesinde Matematik puanlayıcı grubunun puanlayıcı etkileri incelenmiştir.

Puanlayıcı Eğitimi Öncesinde Matematik Puanlayıcı Grubu Puanlayıcı Davranışları

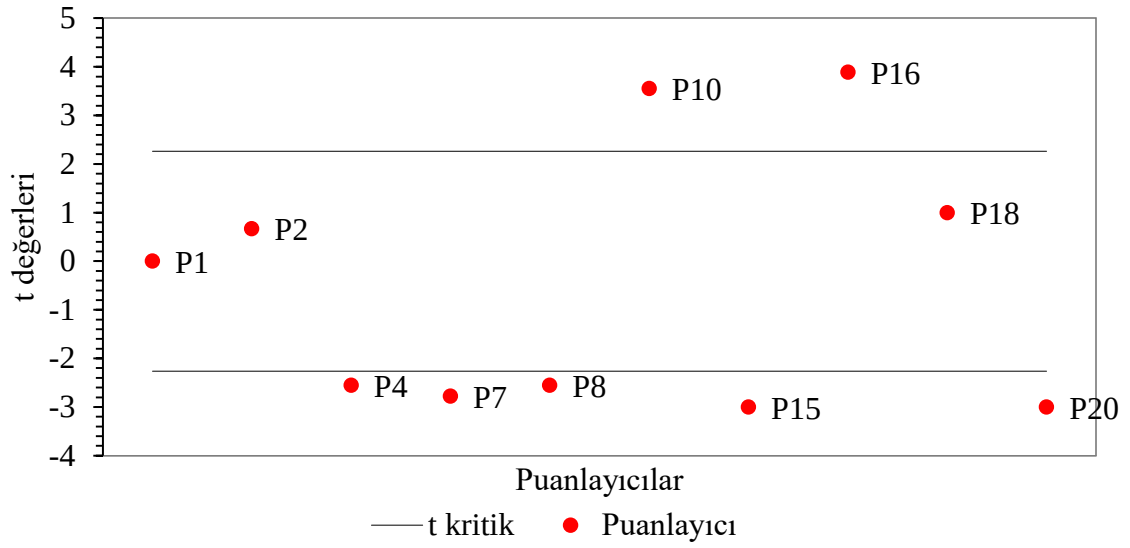
Matematik puanlayıcı grubu puanlayıcı etkileri incelenirken öncelikle grup sonrasında da bireysel düzeyde etkiler raporlanmıştır.

Katılık ve Cömertlik

Puanlayıcı eğitimi öncesinde Matematik puanlayıcı grubu deney ve kontrol gruplarının katılık ve cömertlik davranışlarının incelenmesi için puanlayıcı yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi gibi grup düzeyindeki istatistikler incelenmiştir. Yapılan incelemede puanlayıcı yüzeyine ait istatistiklerin yüksek olduğu gözlenmiş ve bu durumun puanlayıcıların grup düzeyinde farklı davranışlar sergileyebileceklerini göstermiştir (Tablo 21). Bu durumu istatistiksel açıdan anlamlılığını test etmek için sabit etkili ki kare değeri incelenmiş ve bu değerin anlamlı olduğu görülmüştür ($\chi^2(sd) = 100,8$ (19), $p = 0,00 < 0,01$). Grup düzeyinde farklı davranışların olduğu tespit edildikten sonra farklılığın hangi puanlayıcı ya da puanlayıcılardan kaynaklandığını belirlemek için birey düzeyindeki istatistikler incelenmiştir.

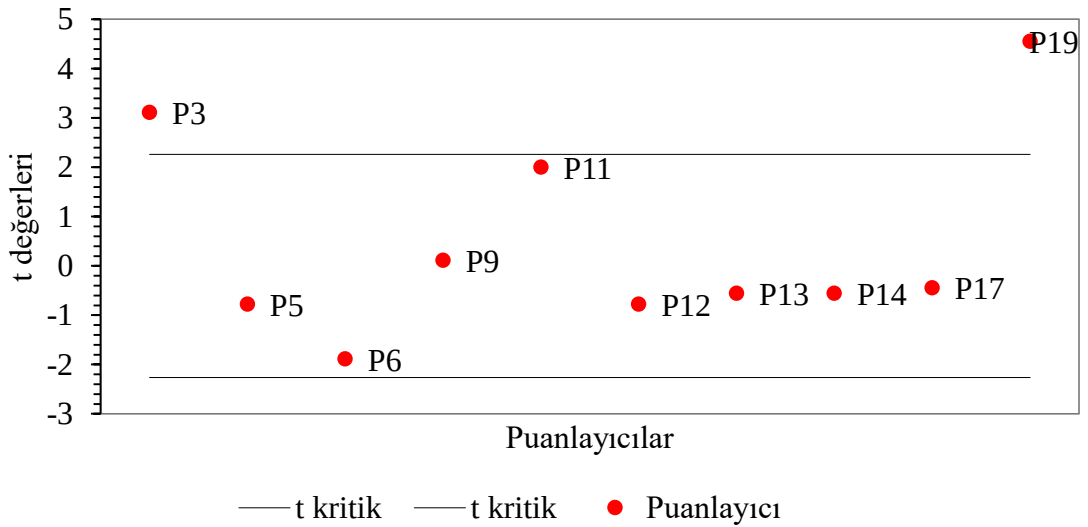
Birey düzeyindeki istatistikler incelenirken t değeri istatistiği kullanılmaktadır. Puanlayıcıya ait t değeri, puanlayıcıya ait logit ölçüsünden tüm puanlayıcıların logit ortalaması farkının standart hataya bölünmesiyle elde edilir. Hesaplanan t değeri, t dağılım tablosundaki kritik t değeri ile karşılaştırılır ve istatistiksel açıdan anlamlılığı test edilir. Puanlayıcıya ait hesaplanan t değeri, kritik t değeri ile karşılaştırılır ve puanlayıcı davranışı belirlenir.

Araştırmada kontrol grubunda on, deney grubunda ise on puanlayıcı olduğundan serbestlik derecesi dokuzdur. Serbestlik derecesi dokuz ve anlamlılık düzeyi 0,05 için t kritik değeri (t_{kritik}) 2,262'dir. Tablo ve şekiller 2,262 değeri göz önünde bulundurularak yapılandırılmıştır. Matematik deney grubunda yer alan puanlayıcıların ön test puanlarının katılık ve cömertlik durumlarını gösteren t değerleri şekil 4'te verilmiştir. Matematik puanlayıcı kontrol grubunda yer alan puanlayıcıların ön test puanlarının katılık ve cömertlik durumlarını gösteren t değerleri şekil 5'te verilmiştir.



Şekil 4. Matematik puanlayıcı deney grubu ön test verilerinden elde edilen t değerleri.

Şekil 4 incelendiğinde, deney grubunda yer alan puanlayıcılardan P4, P7, P8, P10, P15, P16 ve P20 numaralı puanlayıcıların t değerlerinin kabul edilebilir aralığın dışında olduğu görülmektedir. Matematik puanlayıcı deney grubunda beş puanlayıcıda (%50,00) puanlayıcı katılımı davranışı görülürken iki puanlayıcıda (%20,00) da puanlayıcı cömertliği davranışı görülmektedir. Matematik puanlayıcı deney grubunda yer alan puanlayıcılardan P4, P7, P8, P15 ve P20 numaralı puanlayıcıların puanlayıcı katılımı davranışı sergilediği, P10 ve P16 numaralı puanlayıcıların ise puanlayıcı cömertliği davranışı sergilediği görülmektedir.



Şekil 5. Matematik puanlayıcı kontrol grubu ön test verilerinden elde edilen t değerleri.

Şekil 5 incelendiğinde, kontrol grubunda yer alan puanlayıcılardan P3 ve P19 numaralı puanlayıcıların t değerlerinin kabul edilebilir aralığın dışında olduğu görülmektedir. Matematik puanlayıcı kontrol grubunda iki puanlayıcıda (%20,00) puanlayıcı cömertliği davranışı görülürken puanlayıcı katılığı davranışı yer almamaktadır. Matematik puanlayıcı kontrol grubunda yer alan puanlayıcılardan P3 ve P19 numaralı puanlayıcıların puanlayıcı cömertliği davranışı sergilediği görülmektedir.

Matematik puanlayıcı deney ve kontrol gruplarının ön test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Mann Whitney U testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında kullanılacak en güçlü test Mann-Whitney U testidir (Tekindal, 2022, s. 135). T değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcıların t değerleri karşılaştırılmıştır. Matematik puanlayıcı grubunun t değerlerinin farklarına ait Mann-Whitney U testi sonuçları tablo 30’da verilmiştir.

Tablo 30

Deney ve Kontrol Gruplarının Ön Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları

Grup	N	Sıra Ortalaması	Sıra Toplamı	U	p
Deney	10	12,30	123,00	32,00	0,173
Kontrol	10	8,70	87,00		
Toplam	20				

Tablo 30 incelendiğinde, matematik puanlayıcı deney ve kontrol grupları ön test sonuçları arasında katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olmadığı görülmektedir ($U=32,00$; $p=0,173>0,05$). Bu durum matematik puanlayıcı deney ve kontrol gruplarının puanlayıcı eğitimi öncesinde katılık ve cömertlik açısından benzer davranışlara sahip olduklarını göstermektedir.

Merkeze Eğilim

Puanlayıcı eğitimi öncesinde Matematik puanlayıcı grubu deney ve kontrol gruplarının merkeze eğilim davranışının incelenmesi için kategori istatistikleri, birey yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi ve ki kare istatistiği ve madde yüzeyinin uygunluk istatistikleri gibi grup düzeyindeki istatistikler incelenmiştir. Matematik puanlayıcı grubuna ait kategori istatistikleri tablo 31’de verilmiştir.

Tablo 31

Matematik Puanlayıcı Grubuna Ait Ön Test Sonuçlarına Ait Kategori İstatistikleri

Kategoriler	Frekans (f)	Yüzde (%)	Yığılmalı yüzde (%)	Ortalama logit	Beklenen logit	Uyum Dışı
1	938	20	20	-0,91	-0,90	0,90
2	635	14	34	-0,25	-0,22	1,0
3	688	17	52	0,56	0,48	1,6
4	2059	48	100	2,75	2,76	1,2

Tablo 31 incelendiğinde 4 numaralı kategorinin çok kullanıldığı, 1, 2 ve 3 numaralı kategorilerin ise benzer sayılarda kullanıldığı görülmektedir. Bu durum, birey performanslarından kaynaklı olabileceği gibi puanlayıcıların merkeze eğilim davranışlarının da bir sonucu olabilir. Bu durumun belirlenmesi için birey yüzeyine ait ayırma oranı, ayırma indeksi, güvenilirlik indeksi ve ki kare istatistiği gibi grup düzeyindeki istatistiklerin incelenmesi gerekmektedir (Myford & Wolfe, 2004). Tablo 20’deki deney ve kontrol grubunda yer alan puanlayıcıların birey yüzeyine ait ölçüm raporu incelendiğinde, ayırma oranının, ayırma indeksinin ve güvenilirlik indeksinin yüksek olduğu ve sabit etkili ki kare değerinin de anlamlı çıktığı görülmektedir. Yani birey performanslarının yetenek seviyelerine göre doğru ve istatistiksel olarak anlamlı bir şekilde ayrılmaktadır. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Merkeze eğilim davranışının bir diğer göstergesi, madde yüzeyinin uygunluk istatistikleridir. Madde yüzeyine ait ölçüm raporunda maddelerin uygunluk istatistiklerinin kabul edilebilir aralıkta (0,50 ile 2,00) olmaması merkeze eğilim davranışına işaret ederken uygunluk istatistiklerinin kabul edilebilir aralıkta olması merkeze eğilim davranışının olmadığını göstermektedir. Tablo 19’da yer alan deney ve kontrol gruplarının ön test madde yüzeyine ait ölçüm raporları incelendiğinde bir maddenin kabul edilebilir uygunluk istatistikleri dışında bir değere sahip olduğu görülmektedir. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olabileceğini göstermektedir. Tüm grup düzeyindeki istatistikler bir arada ele alındığında açık uçlu soruların değerlendirilmesinde grup düzeyinde merkeze eğilim davranışının olmadığı söylenebilir.

Grup düzeyindeki merkeze eğilim davranışı incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel düzeyde merkeze eğilim davranışının belirlenmesinde her bir puanlayıcıya ait uyum istatistikleri ve her bir puanlayıcıya ait kategori istatistiklerinin incelenmesi gerekmektedir. Her bir puanlayıcıya ait uyum istatistiklerinin yer aldığı tablo 21 incelendiğinde, 11 numaralı puanlayıcı haricinde tüm puanlayıcıların uygunluk içi ve

uygunluk dışı değerlerinin kabul edilebilir aralıkta (0,50 ile 2,00) olduğu görülmektedir. Bu durum 11 numaralı puanlayıcının merkeze eğilim davranışı sergilediğinin, diğer puanlayıcıların bireysel düzeyde merkeze eğilim davranışının sergilemediğinin bir göstergesidir. Ancak bu sonuç tek başına yeterli değildir. Bireysel düzeyde merkeze eğilim davranışının incelenmesinde kullanılan bir diğer istatistik her bir puanlayıcıya ait kategori istatistiklerinin incelenmesidir. Kategori istatistikleri incelenirken uygunluk dışı istatistiği incelenmelidir. Deney ve kontrol gruplarının ön testlerinin her bir puanlayıcı için hesaplanan kategori istatistikleri tablo 32’de verilmiştir.

Tablo 32

Deney ve Kontrol Grubu Ön Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri

Grup	Puan- layıcı	Kategori İstatistiği (%)				Uygunluk Dışı				Eşik değer				Merkeze Eğilim
		1	2	3	4	1	2	3	4	1	2	3	4	
Deney	P1	28	15	18	39	0,7	0,7	2,4	1,1	--	-0,38	-0,10	0,48	VAR
Deney	P2	24	15	23	38	0,8	1,4	0,9	1,1	--	-0,34	-0,35	0,70	YOK
Kontrol	P3	23	16	16	45	0,7	1,6	0,6	1,0	--	-0,23	0,10	0,13	YOK
Deney	P4	27	22	21	31	0,8	1,5	0,8	1,2	--	-0,97	0,06	0,91	YOK
Kontrol	P5	25	17	23	34	1,1	0,8	0,9	1,2	--	0,82	0,21	1,04	YOK
Kontrol	P6	27	19	22	31	1,0	1,0	1,1	1,0	--	-0,80	-0,21	1,01	YOK
Deney	P7	23	16	20	41	1,2	0,9	0,9	1,1	--	-0,66	-0,19	0,85	YOK
Deney	P8	25	22	23	31	1,0	0,8	1,0	1,2	--	-0,99	-0,03	1,01	YOK
Kontrol	P9	26	17	18	39	0,8	1,4	1,4	0,9	--	-0,54	0,02	0,52	YOK
Deney	P10	23	11	16	49	0,7	1,2	1,8	0,9	--	0,27	-0,14	-0,13	YOK
Kontrol	P11	15	16	21	48	0,7	0,4	1,7	1,2	--	-0,20	0,02	0,17	VAR
Kontrol	P12	27	15	22	36	0,9	0,8	1,0	1,3	--	-0,40	-0,28	0,68	YOK
Kontrol	P13	25	17	22	36	1,1	0,9	0,7	1,2	--	-0,87	-0,18	1,05	YOK
Kontrol	P14	26	17	19	37	0,9	0,7	1,0	1,3	--	-0,61	0,00	0,61	YOK
Deney	P15	30	18	20	32	1,0	1,0	1,3	0,8	--	-1,07	-0,14	1,20	YOK
Deney	P16	18	16	21	46	0,7	1,7	0,9	1,0	--	-0,18	-0,02	0,20	YOK
Kontrol	P17	26	16	21	37	1,1	0,9	1,2	1,2	--	-0,67	-0,15	0,83	YOK
Deney	P18	22	17	22	39	0,7	1,7	0,8	0,9	--	-0,78	-0,15	0,94	YOK
Kontrol	P19	18	18	21	43	0,8	0,6	1,7	0,9	--	-0,51	0,03	0,49	YOK
Deney	P20	30	18	20	32	1,0	1,0	1,3	0,8	--	-1,07	-0,14	1,20	YOK

Tablo 32 incelendiğinde, puanlayıcıların genel olarak puanlama anahtarının tüm kategorilerini kullandıkları görülmektedir. Tablo 32’ye göre P1, P3, P9, P10, P11, P14, P18 ve P19 numaralı puanlayıcıların bazı kategorileri diğer kategorilere göre daha yoğun kullandıkları belirlenmiştir. Ancak dereceli puanlama anahtarının kullanım örüntüsü incelendiğinde, deney grubunda yer alan P1 numaralı puanlayıcının ve kontrol grubunda yer alan P11 numaralı puanlayıcının merkeze eğilim davranışını sergilediği görülmektedir. Merkeze eğilim davranışı, deney grubunda bir ve kontrol grubunda bir olmak üzere toplam

iki puanlayıcıda belirlenmiştir. Puanlayıcıların merkeze eğilim davranışına yatkınlıklarının karşılaştırılması için ortalama eşik analizi yapılmıştır (Rahman vd., 2017). Bir numaralı puanlayıcının ortalama eşik değerinin 0,43 $[[0,48-(-0,38)]/2]$, on bir numaralı puanlayıcının ortalama eşik değerinin 0,19 $[[0,17-(-0,20)]/2]$ olduğu görülmektedir. Bu durum bir numaralı puanlayıcının merkeze eğilim davranışına on bir numaralı puanlayıcıya göre daha yatkın olduğunu göstermektedir.

Halo Etkisi

Puanlayıcı eğitimi öncesinde Matematik puanlayıcı grubu deney ve kontrol gruplarının halo etkisi davranışlarının incelenmesi için madde yüzeyine ait ayırma oranı, ayırma indeksi ve sabit etkili ki kare değeri gibi grup düzeyindeki istatistikler incelenmiştir. Madde yüzeyine ait ölçüm raporunun yer aldığı tablo 19 incelendiğinde; ayırma oranının 14,08, ayırma indeksinin 19,10, güvenirlik indeksinin 0,99 olduğu ve sabit etkili ki kare istatistiğinin anlamlı olduğu ($\chi^2(sd) = 1610,3 (5), p = 0,00 < 0,01$) görülmektedir. Madde yüzeyine ait istatistiklerin yüksek ve sabit etkili ki kare değerinin anlamlı olması puanlayıcıların yapmış oldukları puanlamalarda grup düzeyinde halo etkisinin olmadığını göstermektedir. Grup düzeyinde halo etkisi incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel istatistikler incelenirken, madde güçlükleri arasındaki fark büyük olduğunda puanlayıcıya ait uygunluk istatistikleri 1'den çok büyük olduğunda, madde güçlükleri arasındaki fark küçük olduğunda da puanlayıcıya ait uygunluk istatistikleri 1'den çok küçük olduğunda halo etkisinin olduğunu ifade edilmiştir (Myford & Wolfe, 2004). Matematik puanlayıcı grubu ön test madde yüzeyine ait ölçüm raporunun yer aldığı tablo 19 incelendiğinde maddelerin güçlük düzeyleri arasındaki farkın 6,00 $[4,12-(-1,88)]$ olduğu görülmektedir. Bu fark büyük olduğundan puanlayıcılara ait uygunluk istatistiklerinin 1'den çok büyük olanların halo etkisine sahip olduğu kabul edilmiştir. Matematik puanlayıcı grubu ön test puanlayıcı yüzeyine ait uygunluk istatistiklerinin yer aldığı Tablo 21 incelendiğinde, bir puanlayıcıda halo etkisi tespit edilmiştir. Halo etkisi deney grubunda gözlenmezken kontrol grubunda 11 numaralı puanlayıcıda belirlenmiştir.

Halo etkisinin bulunup bulunmadığına karar vermek için önerilen bir diğer yöntem ise madde güçlüklerinin birbirine eşitleyerek çok yüzeyli rasch analizini tekrar edip puanlayıcılara ait uyum istatistiklerini incelemektir (Linacre, 2023, s. 360). Madde güçlükleri eşitlenerek oluşturulan yeni modele mükemmel uyum gösteren (uygunluk

İçerik=1,00 veya uygunluk dışı=1,00) puanlayıcıların halo etkisinin olduğu kabul edilmektedir (Linacre, 2023, s. 360). Bu yüzden Matematik testindeki altı açık uçlu sorunun kriter güçlükleri sıfıra eşitlenerek çok yüzeyli Rasch analizi tekrar edilmiştir. Ardından uygunluk istatistikleri incelenmiştir. Kriter güçlükleri eşitlendikten sonra Matematik puanlayıcı grubunun puanlayıcı yüzeyine ait ölçüm raporu Tablo 33’te verilmiştir.

Tablo 33

Matematik Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Ön Test Ölçümlerine Ait Ölçüm Raporu

Grup	Puanlayıcı	Uygunluk İçerik	Uygunluk Dışı
Deney	P1	1,00	0,97
Deney	P2	1,01	1,00
Kontrol	P3	1,01	0,97
Deney	P4	1,03	1,04
Kontrol	P5	1,01	1,04
Kontrol	P6	1,05	1,07
Deney	P7	0,94	0,94
Deney	P8	0,93	0,93
Kontrol	P9	1,02	0,99
Deney	P10	1,03	0,97
Kontrol	P11	0,93	0,88
Kontrol	P12	1,04	1,05
Kontrol	P13	1,01	1,02
Kontrol	P14	1,03	1,05
Deney	P15	1,06	1,06
Deney	P16	0,93	0,90
Kontrol	P17	1,03	1,04
Deney	P18	1,00	0,97
Kontrol	P19	0,92	0,90
Deney	P20	1,06	1,06

Tablo 33’te Matematik puanlayıcı grubuna ait uygunluk içerik ve uygunluk dışı gibi bireysel istatistikleri yer almaktadır. Puanlayıcılara ait bireysel istatistikler incelendiğinde, deney grubundaki P1, P2 ve P18 numaralı puanlayıcıların mükemmel uyum gösterdiği ve dolayısıyla halo etkisinin olduğu belirlenmiştir. Kontrol grubunda mükemmel uyum sağlayan puanlayıcı yer almamaktadır.

Tüm bireysel istatistiklerin incelenmesi sonucunda, deney grubunda P1, P2 ve P18 numaralı puanlayıcılarda olmak üzere üç, kontrol grubunda da P11 numaralı puanlayıcıda olmak üzere bir puanlayıcıda halo etkisi belirlenmiştir. Sonuç olarak halo etkisinin merkeze eğilim davranışı gibi grup düzeyinde ortaya çıkmasa da bireysel düzeyde ortaya çıktığı belirlenmiştir.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)

Farklılaşan puanlayıcı katılığı ve cömertliği davranışı incelenirken modelde yer alan yüzeyler arasındaki etkileşimler incelenmektedir. Mevcut çalışma kapsamında birey (b), puanlayıcı (p) ve madde (m) yüzeyi olmak üzere üç yüzey bulunmaktadır. Bu yüzeyler arasındaki etkileşimler de *puanlayıcı x birey (pxb)*, *puanlayıcı x madde (pxm)*, *birey x madde (bxm)* ve *birey x puanlayıcı x madde (bxpxm)* şeklindedir. Mevcut çalışmada puanlayıcıların maddelerle ilgili davranışlarını incelemek amaçlanmadığı için *puanlayıcı x madde (pxm)* etkileşimi çalışmaya dahil edilmemiştir. Benzer şekilde çalışmada puanlayıcı davranışları incelendiği ve bireylerin maddeleri cevaplandırma seviyeleri de incelenmediği için *birey x madde (bxm)* etkileşimi de çalışmaya dahil edilmemiştir. Ayrıca puanlayıcıların hem madde hem de bireye göre davranışları araştırma kapsamında olmadığı için *birey x puanlayıcı x madde (bxpxm)* etkileşimi de çalışmaya dahil edilmemiştir. Çalışma kapsamında Matematik açık uçlu soruların puanlanması sürecinde meydana gelen yanlılık davranışlarının tespiti için *puanlayıcı x birey (pxb)* etkileşimi incelenmiştir.

Puanlayıcı x birey (pxb) etkileşiminde yer alan istatistiksel olarak anlamlı olan etkileşimlerin tespiti için t değerinden faydalanılmaktadır. Linacre (2023, s. 190), ± 2 aralığı dışında kalan t değerlerinin istatistiksel olarak anlamlı olduğunu ifade etmektedir. Mevcut çalışmada 35 öğrencinin Matematik açık uçlu sorulara vermiş oldukları tüm cevaplar 10 deney, 10 kontrol olmak üzere toplam 20 puanlayıcı tarafından puanlanmıştır. Dolayısıyla iki yüzey arasında 700 (20 x 35) olası etkileşim varken 350 (10 x 35) etkileşim deney grubunda, 350 (10 x 35) etkileşim de kontrol grubunda meydana gelmiştir. Analiz sonuçları incelenirken öncelikle grup düzeyindeki istatistikler incelenmiş ardından bireysel istatistiklere geçilmiştir. Grup düzeyindeki istatistikler incelendiğinde ki kare istatistiğinin istatistiksel açıdan anlamlı olmadığı bulunmuştur (χ^2 (sd)=296,2 (700), $p>0,01$). Bu durum puanlayıcıların grup düzeyinde yanlılık davranışı sergilemedikleri şeklinde yorumlanabilir. Grup düzeyindeki istatistikler incelendikten sonra bireysel istatistikler incelenmiştir. Yapılan incelemede, her grup için meydana gelen olası 350 etkileşimden kontrol grubunda istatistiksel olarak anlamlı bir tane etkileşim bulunurken deney grubunda istatistiksel olarak anlamlı iki tane etkileşim yer almaktadır. Yani Matematik puanlayıcı grubunda toplam üç puanlayıcı farklılaşan puanlayıcı katılığı ve cömertliği davranışı sergilemiştir. Bir başka deyişle üç puanlayıcıda yanlılık davranışı vardır. Grupların etkileşimlerine ait istatistiklerin tamamı Ek 11’de yer alırken etkileşimlerin tamamını sunmak mümkün olmadığı için sadece istatistiksel olarak anlamlı olan etkileşimler tablo 34’te verilmiştir.

Tablo 34

Matematik Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri

Grup	Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Bias (logit)	Standart Hata	t
Deney	P18	22	18	14,30	1,01	0,50	2,01
Deney	P2	8	18	21,69	-1,09	0,50	-2,18
Kontrol	P11	22	19	14,57	1,18	0,51	2,31

 χ^2 (sd)=296,2 (700), $p > 0,01$

Matematik puanlayıcı grubu ön test sonuçlarına ait puanlayıcı x birey etkileşimlerinin yer aldığı tablo 34 incelendiğinde, üç puanlayıcıda farklılaşan puanlayıcı katılığı ve cömertliği davranışı olduğu görülmektedir.

Deney grubunda yer alan puanlayıcılardan P2 ve P18 numaralı puanlayıcılarda ve kontrol grubunda yer alan P11 numaralı puanlayıcıda farklılaşan puanlayıcı katılığı davranışı olduğu görülmektedir. Kontrol grubunda yer alan P11 numaralı puanlayıcının 22 numaralı bireye olması gerekenden daha fazla puan verdiği dolayısıyla farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Deney grubunda yer alan P18 numaralı puanlayıcının 22 numaralı bireye olması gerekenden daha fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Benzer şekilde deney grubunda yer alan P2 numaralı puanlayıcının 8 numaralı bireye olması gerekenden az puan verdiği yani farklılaşan puanlayıcı katılığı davranışı sergilediği görülmektedir. Bir başka deyişle deney grubunda bir puanlayıcıda (P2) farklılaşan puanlayıcı katılığı davranışı yer alırken bir puanlayıcıda (P18) da farklılaşan puanlayıcı cömertliği davranışı yer almaktadır. Kontrol grubunda bir puanlayıcıda (P11) farklılaşan puanlayıcı cömertliği davranışı yer alırken farklılaşan puanlayıcı katılığı davranışı sergileyen puanlayıcı yer almamaktadır.

Araştırma sonuçlarını destekleyen çalışmalar

Bu bölümde ikinci araştırma sorusu kapsamında elde edilen bulgulara ait sonuçları destekleyen araştırmalar her bir davranış için ayrı ayrı raporlanmıştır. Çalışmalar raporlanırken katılık ve cömertlik, merkeze eğilim, halo etkisi ve yanlılık şeklinde ayrı ayrı başlıklar altında sunulmuştur.

Katılık ve Cömertlik

Araştırmanın ikinci araştırma sorusu kapsamında Türkçe ve matematik puanlayıcı grubunda yer alan öğretmen adaylarının hem grup düzeyinde hem de bireysel düzeyde katılık ve cömertlik davranışı sergiledikleri bulunmuştur. Literatür incelendiğinde, performans değerlendirme sürecinde genellikle puanlayıcı etkisinin karıştığı görülmekte olup en sık araştırılan ve karşılaşılan puanlayıcı etkilerinden birinin puanlayıcı katılığı ve cömertliği olduğu görülmektedir. Bu kapsamda mevcut araştırma sonuçlarını destekleyen çalışmalardan bazıları raporlanmıştır.

Örneğin; Baharudin, Maskor, & Matore (2022) yapmış oldukları çalışmada yazma performansının değerlendirilmesinde puanlayıcı etkisini incelemiştir. Araştırma sonucunda puanlayıcıların katılık ve cömertlik davranışı sergilediklerini tespit etmişlerdir. Li (2022), yapmış olduğu çalışmada yabancı dil olarak İngilizce yazma becerilerinin değerlendirilmesinde derecelendirme ölçeğinin öğretmenler ve öğrenciler (akran) tarafından nasıl kullanıldığını araştırmıştır. Çok yüzeyli Rasch modelinin kullanıldığı araştırmada öğretmenlerin öğrencilere (akran) göre daha katı puanlama yaptıkları tespit edilmiştir. Ayrıca puanlayıcı içi güvenilirlik öğretmenlerde, puanlayıcılar arası güvenilirlik ise öğrencilerde daha yüksek çıkmıştır. Esfandiari (2021) yapmış olduğu çalışmada 150 yabancı dil öğretmeni öğrencilerin yazmış olduğu ve beş paragraftan oluşan 10 makaleyi iki derecelendirme ölçeği kullanarak değerlendirmişlerdir. Araştırma sonucunda öğretmenlerin makaleleri değerlendirirken farklı ciddiyette (katılık ve cömertlik) olduklarını tespit etmişlerdir. Erguvan & Aksu Dünya (2020), üniversite birinci sınıf öğrencilerinin kompozisyonlarını değerlendiren puanlayıcıların davranışlarını incelemiştir. Araştırma sonucunda puanlayıcıların katılık ve cömertlik açısından birbirlerinden farklı oldukları tespit edilmiştir. Yılmaz (2017), öğretmen adaylarının puanlayıcı davranışlarının incelenmesi amacıyla yapmış olduğu çalışmada puanlayıcı grubunun grup düzeyinde katılık cömertlik davranışı sergilemediği ancak bireysel düzeyde katılık cömertlik davranışının olduğunu belirlemiştir. Abu Kassim (2007), 12 öğrenci tarafından yazılan metnin 35 İngilizce öğretmeni tarafından değerlendirilme sonuçlarını çok yüzeyli Rasch modeli ile incelemiştir. Araştırma sonucunda puanlayıcıların katılık ve cömertlik davranışı sergiledikleri tespit edilmiştir.

Merkeze Eğilim

Araştırmanın ikinci araştırma sorusu kapsamında Türkçe ve matematik puanlayıcı grubunda yer alan öğretmen adaylarının grup düzeyinde merkeze eğilim davranışı sergilemedikleri ancak bireysel düzeyde merkeze eğilim davranışı sergiledikleri bulunmuştur. Elde edilen araştırma sonucu literatürde yer alan çalışmalarla uyum göstermektedir. Bu kapsamda mevcut araştırma sonuçlarını destekleyen çalışmalardan bazıları raporlanmıştır. Örneğin; A'laudden, Hanafî, & Rahman (2023) öğretmen adaylarının performanslarının değerlendirilmesi sürecinde puanlayıcıların merkeze eğilim davranışlarını incelemiştir. Çalışmaya 170 öğretmen adayı ve 194 puanlayıcı dahil edilmiştir. Yapılan çalışma sonucunda puanlayıcıların grup düzeyinde merkeze eğilim davranışına sahip olmadıkları ancak bireysel düzeyde bu davranışın gözlemlendiği belirlenmiştir. Li (2022), yapmış olduğu çalışmada yabancı dil olarak İngilizce yazma becerilerinin değerlendirilmesinde derecelendirme ölçeğinin öğretmenler ve öğrenciler (akran) tarafından nasıl kullanıldığını araştırmıştır. Çok yüzeyli Rasch modelinin kullanıldığı araştırmada öğretmenlerin öğrencilere (akran) göre daha katı puanlama yaptıkları ancak her iki grupta da merkeze eğilim etkisi olduğu tespit edilmiştir. Ayrıca puanlayıcı içi güvenilirlik öğretmenlerde, puanlayıcılar arası güvenilirlik ise öğrencilerde daha yüksek çıkmıştır. Erman Aslanoğlu, Şata ve Karakaya (2020)'nin üniversite öğrencilerinin öz ve akran değerlendirme sürecindeki puanlayıcı davranışlarının belirlenmesi amacıyla yapmış oldukları çalışmada puanlayıcıların grup düzeyinde merkeze eğilim davranışı göstermezken bireysel düzeyde 18 puanlayıcıda merkeze eğilim davranışı gösterdiği belirlenmiştir. Linlin (2020) yapmış olduğu araştırmada puanlayıcıların grup düzeyindeki davranışlarını incelemiş ve puanlayıcıların genelde üst kategorileri tercih ettiğini ve puanlayıcıların grup düzeyinde merkeze eğilim davranışı göstermediklerini belirlemiştir. Jones & Bergin (2019) yapmış oldukları çalışmada ABD'deki öğretmenlik uygulamalarının değerlendirilmesi sürecinde okul müdürlerinin davranışlarını incelemiştir. Çalışmaya 421 müdürün yapmış olduğu değerlendirmeler dahil edilmiştir. Araştırma sonucunda müdürlerin grup düzeyinde merkeze eğilim davranışlarının olmadığı ancak bireysel düzeyde merkeze eğilim davranışının olduğunu belirlemiştir. Ayrıca müdürlerin dereceli puanlama anahtarının daha çok üst kısmını kullanma (%75) eğiliminde olduklarını da belirtmişlerdir. Esfandiari (2015) yazma becerilerinin değerlendirilmesinde ortaya çıkan puanlayıcı davranışlarını incelemek için çok yüzeyli Rasch modelini kullanmıştır. Araştırma kapsamında merkeze eğilim ve halo etkisi incelenmiştir. Araştırma sonucunda puanlayıcıların grup düzeyinde merkeze eğilim

göstermediğini ancak bireysel düzeyde puanlayıcılarda merkeze eğilim davranışının olduğu tespit edilmiştir. Literatürde yer alan çalışmalar, mevcut çalışma sonuçlarını desteklemektedir.

Halo Etkisi

Araştırmanın ikinci araştırma sorusu kapsamında Türkçe ve matematik puanlayıcı grubunda yer alan öğretmen adaylarının grup düzeyinde halo etkisi olmadığı ancak bireysel düzeyde halo etkisi olduğu bulunmuştur. Elde edilen araştırma sonucu literatürde yer alan çalışmalarla uyum göstermektedir. Bu kapsamda mevcut araştırma sonuçlarını destekleyen çalışmalardan bazıları raporlanmıştır. Örneğin; Ayoobiyan & Ahmadi (2023) yapmış oldukları çalışmada puanlayıcılardan bir tanesinde halo etkisi olduğunu belirlemişlerdir. Eckes & Jin (2022), yanıltıcı hale etkisini mixed Rasch modeli ile araştırmışlardır. Araştırma sonucunda puanlayıcılarda halo etkisinin olduğu tespit edilmiştir. Jansen, Vögelin, Machts, Keller, & Möller (2021) yapmış oldukları çalışmada ikinci dil olarak İngilizce hazırlanan makalelerin değerlendirilmesi sürecinde deneysel olarak yazım yanlışı olan makalelerde daha fazla halo etkisinin yer aldığını belirlemişlerdir. Jones & Bergin (2019) yapmış oldukları çalışmada ABD'deki öğretmenlik uygulamalarının değerlendirilmesi sürecinde okul müdürlerinin davranışlarını incelemişlerdir. Çalışmaya 421 müdürün yapmış olduğu değerlendirmeler dahil edilmiştir. Araştırma sonucunda müdürlerin grup düzeyinde halo etkisinin olmadığını ancak bireysel düzeyde müdürlerin halo etkisi belirlemişlerdir. Sonnert, Hazari, & Sadler (2018) yapmış oldukları çalışmada matematik öğretmenlerinin değerlendirilmesinde lisans öğrencileri ve uzman değerlendirme performansları karşılaştırılmıştır. Çalışma kapsamında matematik öğretmenleri öğretim zenginliği ve matematiksel doğruluk olmak üzere iki boyutta değerlendirilmiştir. Çalışmaya 85 lisans öğrencisi katılmış ve çalışma sonucunda öğrencilerin puanlamalarında halo etkisi tespit edilmiştir. Yılmaz (2017), öğretmen adaylarının puanlayıcı davranışlarının incelenmesi amacıyla yapmış olduğu çalışmada puanlayıcı grubunun grup düzeyinde olmasa da bireysel düzeyde halo etkisinin olduğunu belirlemiştir. Esfandiari (2015) yazma becerilerinin değerlendirilmesinde ortaya çıkan puanlayıcı davranışlarını incelemek için çok yüzeyli Rasch modelini kullanmıştır. Araştırma kapsamında merkeze eğilim ve halo etkisi incelenmiştir. Araştırma sonucunda puanlayıcılarda grup düzeyinde halo etkisi olmadığını ancak bireysel düzeyde puanlayıcılarda halo etkisinin olduğu tespit edilmiştir. Literatürde yer alan çalışmalar mevcut çalışma sonuçlarını desteklemektedir.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)

Araştırmanın ikinci araştırma sorusu kapsamında Türkçe ve matematik puanlayıcı grubunda yer alan öğretmen adaylarının grup düzeyinde farklılaşan puanlayıcı katılığı ve cömertliği davranışı olmadığı ancak bireysel düzeyde farklılaşan puanlayıcı katılığı ve cömertliği davranışı olduğu bulunmuştur. Elde edilen araştırma sonucu literatürde yer alan çalışmalarla uyum göstermektedir. Bu kapsamda mevcut araştırma sonuçlarını destekleyen çalışmalardan bazıları raporlanmıştır. Örneğin; Esfandiari (2021) yapmış olduğu çalışmada 150 yabancı dil öğretmeni öğrencilerin yazmış olduğu ve beş paragraftan oluşan 10 makaleyi iki derecelendirme ölçeği kullanarak değerlendirmişlerdir. Araştırma sonucunda öğretmenlerin makaleleri değerlendirirken bazı öğretmenlerde farklılaşan puanlayıcı katılığı ve cömertliği davranışının olduğunu tespit etmiştir. Jones & Bergin (2019) yapmış oldukları çalışmada ABD’deki öğretmenlik uygulamalarının değerlendirilmesi sürecinde okul müdürlerinin davranışlarını incelemişlerdir. Çalışmaya 421 müdürün yapmış olduğu değerlendirmeler dahil edilmiştir. Araştırma sonucunda müdürlerinde farklılaşan puanlayıcı katılığı ve cömertliği davranışının olduğu belirlenmiştir. Yılmaz (2017), öğretmen adaylarının puanlayıcı davranışlarının incelenmesi amacıyla yapmış olduğu çalışmada, bazı puanlayıcılarda farklılaşan puanlayıcı katılığı ve cömertliği davranışı olduğunu belirtmiştir. Winke, Gass, & Myford (2013), 107 puanlayıcının 432 TOEFL konuşma örneklerini değerlendirme sonuçlarını incelemiştir. Araştırma sonucunda puanlayıcıların bazılarında farklılaşan puanlayıcı katılığı ve cömertliği davranışının yer aldığı tespit edilmiştir. Mastergeorge & Martínez (2010) yaptıkları çalışmada engeli olan ve engeli olmayan öğrencilerin değerlendirilmesi sürecinde puanlayıcı davranışlarını incelemişlerdir. Araştırmaya üçüncü, beşinci ve dokuzuncu sınıftaki engelli ve engelli olmayan öğrencilerin dil ve matematik performanslarına verilen puanlar incelenmiştir. Puanlayıcıların engeli olan ve olmayan öğrencileri bilmesi ya da bilmemesi durumunda yapılan puanlamalarda farklılıklar yer almaktadır. Araştırma sonuçları, puanlayıcıların engelli öğrencilere puan verirken farklılaşan puanlayıcı katılığı ve cömertliği davranışını sergileyebileceklerini göstermektedir.

Üçüncü araştırma sorusuna ait bulgular

Araştırmanın üçüncü araştırma sorusu “Puanlayıcı eğitimi sonrasında açık uçlu soruların puanlanması sürecinde deney ve kontrol grubu puanlayıcı davranışları nasıldır?” şeklindedir. Bu doğrultuda Türkçe ve matematik puanlayıcı gruplarının puanlayıcı davranışları çok

yüzeyli rasch modeli kapsamında incelenmiştir. İncelemelerde öncelikle Türkçe puanlayıcı grubu puanlayıcı davranışları incelenirken sonrasında Matematik puanlayıcı grubu puanlayıcı davranışları incelenmiştir. Ayrıca puanlayıcı davranışları incelenirken önce grup istatistikleri incelenirken sonrasında bireysel istatistikler değerlendirilmiştir. Sonuçlar raporlanırken Türkçe ve matematik ölçüm sonuçları farklı başlıklarda ele alınmıştır.

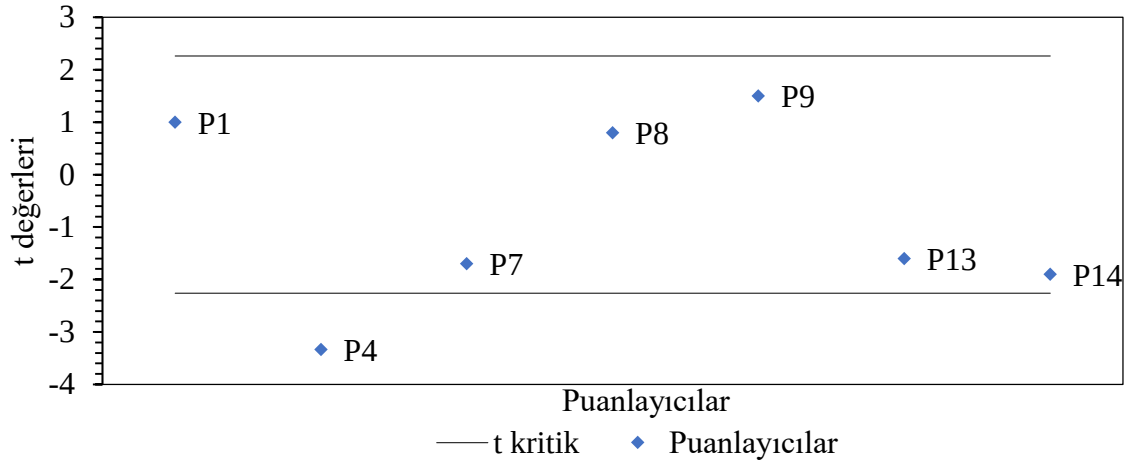
Puanlayıcı Eğitimi Sonrasında Türkçe Puanlayıcı Grubu Puanlayıcı Davranışları

Katılık ve Cömertlik

Puanlayıcı eğitimi sonrasında Türkçe puanlayıcı grubu deney ve kontrol gruplarının katılık ve cömertlik davranışlarının incelenmesi için ayırma oranı, ayırma indeksi ve güvenilirlik indeksi gibi grup düzeyindeki istatistikler incelenmiştir. Yapılan incelemede ayırma oranı 1,48, ayırma indeksi 2,31 ve güvenilirlik indeksi 0,69 olarak belirlenmiştir (Tablo 18). Bu istatistiklerin küçük olması puanlayıcıların katılık ve cömertlik açısından benzer davranışlara sahip olabileceklerini göstermektedir. Ancak kesin yargıya varmak için sabit etkili ki kare istatistiğinin incelenmesi gerekmektedir. Sabit etkili ki kare istatistiği incelendiğinde, ki kare istatistiğinin istatistiksel olarak anlamlı olmadığı görülmektedir ($\chi^2(sd) = 41,7 (13), p = 0,00 < 0,01$). Bu durum puanlayıcıların grup düzeyinde katılık ve cömertlik davranışına sahip olduklarını göstermektedir. Bu farklılığın hangi puanlayıcıdan ya da puanlayıcılardan kaynaklandığını belirlemek için bireysel istatistikler incelenmiştir.

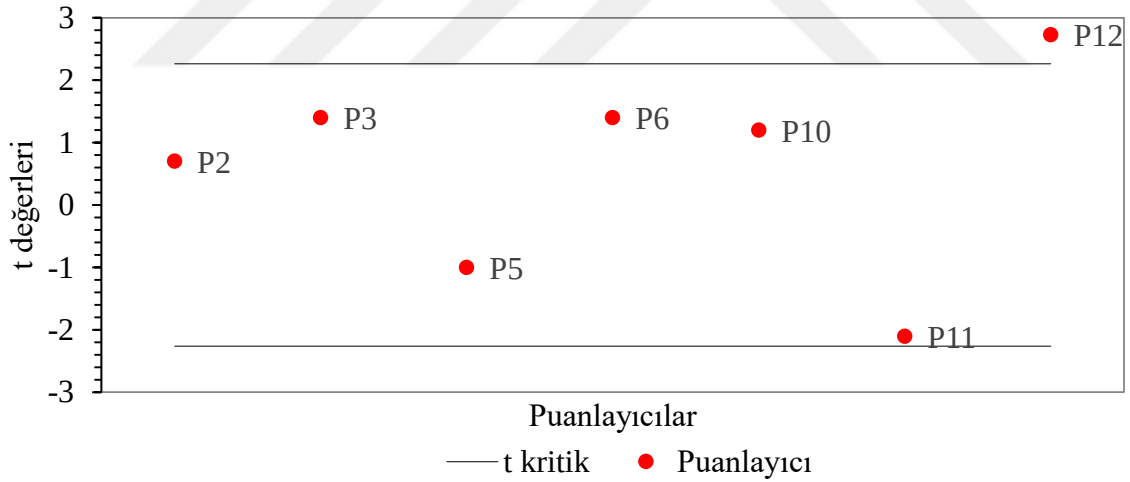
Birey düzeyindeki istatistikler incelenirken puanlayıcılara ait t değeri hesaplanmaktadır. Puanlayıcının t değeri, puanlayıcıya ait logit ölçüsünden tüm puanlayıcıların logit ortalaması farkının standart hataya bölünmesiyle elde edilir. Hesaplanan t değeri, t dağılım tablosundaki kritik t değeri ile karşılaştırılır ve istatistiksel açıdan anlamlılığı test edilir. Puanlayıcıya ait hesaplanan t değeri, kritik t değeri ile karşılaştırılır ve puanlayıcı davranışı belirlenir.

Araştırmada deney ve kontrol grubunda yedi puanlayıcı olduğundan serbestlik derecesi altıdır. Serbestlik derecesi altı ve anlamlılık düzeyi 0,05 için t kritik değeri (t_{kritik}) 2,447'dir. Tablo ve şekiller 2,447 değeri göz önünde bulundurularak yapılandırılmıştır. Türkçe deney ve kontrol grubunda yer alan puanlayıcıların ön test puanlarının katılık ve cömertlik durumları şekil 6 ve şekil 7'de verilmiştir.



Şekil 6. Türkçe puanlayıcı deney grubu son test verilerinden elde edilen t değerleri.

Şekil 6 incelendiğinde, deney grubunda yer alan puanlayıcılardan P4 numaralı puanlayıcının t değerinin kabul edilebilir aralığın dışında olduğu ve puanlayıcı katılığı davranışı sergilediği görülmektedir. Deney grubunda bir puanlayıcıda (%14,29) puanlayıcı katılığı davranışı yer alırken altı puanlayıcıda (%85,71) katılık ya da cömertlik davranışı bulunmamaktadır.



Şekil 7. Türkçe puanlayıcı kontrol grubu son test verilerinden elde edilen t değerleri.

Şekil 7 incelendiğinde, kontrol grubunda yer alan puanlayıcılardan P12 numaralı puanlayıcının t değerinin kabul edilebilir aralığın dışında olduğu ve puanlayıcı cömertliği davranışı sergilediği görülmektedir. Kontrol grubunda bir puanlayıcıda (%14,29) puanlayıcı cömertliği davranışı yer alırken altı puanlayıcıda (%85,71) katılık ya da cömertlik davranışı bulunmamaktadır.

Türkçe puanlayıcı deney ve kontrol gruplarının son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Mann-Whitney U testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında kullanılacak en güçlü test Mann-Whitney U testidir (Tekindal, 2022, s. 135). Ayrıca Türkçe puanlayıcı gruplarının grup içi ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için de Wilcoxon İşaretli Sıralar testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında ilişkili iki gruba ait ortalamalar arasındaki fark test edilirken kullanılan alternatif yöntem Wilcoxon İşaretli Sıralar testidir (Büyüköztürk, Çokluk, Köklü, 2019, s. 197). T değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcıların t değerleri karşılaştırılmıştır. Ancak t değerleri karşılaştırılırken deneysel işlemin etkililiğini daha iyi gözlemlemek için puanlayıcıların son test t değerlerinden ön test t değerlerinin farkı alınmıştır. Puanlayıcıların t değerlerinin farklarına ait Mann-Whitney U testi sonuçları tablo 35'te verilmiştir.

Tablo 35

Deney ve Kontrol Gruplarının Son Test t İstatistikleri Farklarına Ait Mann-Whitney U Testi Sonuçları

Grup	N	Sıra Ortalaması	Sıra Toplamı	U	p	η^2
Deney	7	4,00	28,00	0,00	0,002	0,602
Kontrol	7	11,00	77,00			
Toplam	14					

Tablo 35 incelendiğinde, deney ve kontrol gruplarının son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olduğu görülmektedir ($U=0,00$; $p= 0,002<0,05$). Bu durum deney ve kontrol gruplarının puanlayıcı eğitimi sonrasında katılık ve cömertlik açısından farklı davranışlara sahip olduklarını göstermektedir. Ayrıca puanlayıcıların katılık ve cömertlik davranışlarındaki değişikliğin %60,2'sinin ($U=0,00$; $p= 0,002<0,05$; $\eta^2= 0,602$) puanlayıcı eğitimi ile açıklanabildiği belirlenmiştir. Puanlayıcı eğitimi sonrasında deney grubunda katılık cömertlik durumu %42,86'dan %14,29'a gerilemiştir.

Deney grubunun puanlayıcı eğitimi öncesinde ve puanlayıcı eğitimi sonrasında katılık ve cömertlik açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Wilcoxon İşaretli Sıralar testi yapılmıştır. t değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcılara ait t değerleri karşılaştırılmıştır. Ancak t değerleri

karşılaştırılırken deneysel işlemin etkililiğini daha iyi gözlemlemek için puanlayıcıların son test t değerleri ile ön test t değerlerinin farkı alınmıştır. Deney grubunda yer alan puanlayıcıların t değerlerine ait Wilcoxon İşaretli Sıralar testi tablo 36’da verilmiştir.

Tablo 36

Deney Grubu Ön Test ve Son Test t Değerlerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları

Son test- ön test	n	Sıra Ortalaması	Sıra Toplamı	z	p
Negatif Sıralar	7	4,00	28,00		
Pozitif Sıralar	0	0,00	0,00	-2,366	0,018
Eşit Sıralar	0				

Tablo 36 incelendiğinde, deney grubu ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olduğu görülmektedir ($z=-2,366$; $p=0,018<0,05$). Bu durum, puanlayıcı eğitimi öncesinde ve sonrasında Türkçe puanlayıcı grubunun katılık ve cömertlik davranışları arasında anlamlı fark olduğunu göstermektedir. Yani deney grubunda puanlayıcı eğitimi sonrasında katılık cömertlik durumundaki değişim istatistiksel olarak anlamlıdır.

Kontrol grubunda ikinci uygulamada katılık cömertlik durumu %28,57’den %14,29’a gerilemiştir. Kontrol grubunun birinci uygulama ve ikinci uygulama arasında katılık ve cömertlik açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Wilcoxon İşaretli Sıralar Testi yapılmıştır. t değeri katılık ve cömertlik davranışının bir göstergesi olduğu için ön ve son testteki puanlayıcılara ait t değerleri karşılaştırılmıştır. Kontrol grubunda yer alan puanlayıcıların t değerlerine ait Wilcoxon İşaretli Sıralar testi sonuçları tablo 37’de verilmiştir.

Tablo 37

Kontrol Grubu Birinci ve İkinci Uygulama t Değerlerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları

Son test- on test	n	Sıra Ortalaması	Sıra Toplamı	z	p
Negatif Sıralar	5	4,20	21,00		
Pozitif Sıralar	2	3,50	7,00	-1,183	0,237
Eşit Sıralar	0				

Tablo 37 incelendiğinde, kontrol grubu ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olmadığı görülmektedir ($z=-1,183$; $p=0,237>0,05$). Bu durum, birinci uygulama ve ikinci uygulama arasında Türkçe puanlayıcı kontrol grubunun katılık ve cömertlik davranışları arasında anlamlı fark olmadığını

göstermektedir. Yani Türkçe puanlayıcı kontrol grubunun katılık ve cömertlik açısından birinci ve ikinci uygulama arasındaki değişim istatistiksel olarak anlamlı değildir.

Merkeze Eğilim

Puanlayıcı eğitimi sonrasında Türkçe puanlayıcı grubu deney ve kontrol gruplarının merkeze eğilim davranışının incelenmesi için kategori istatistikleri, birey yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi ve ki kare istatistiği ve madde yüzeyinin uygunluk istatistikleri gibi grup düzeyindeki istatistikler incelenmiştir. Türkçe puanlayıcı grubuna ait kategori istatistikleri tablo 38’de verilmiştir.

Tablo 38

Türkçe Puanlayıcı Grubuna Ait Son Test Sonuçlarına Ait Kategori İstatistikleri

Kategoriler	Frekans (f)	Yüzde (%)	Yığılmalı yüzde (%)	Ortalama logit	Beklenen logit	Uyum Dışı
1	293	12	12	-0,47	-0,47	1,2
2	230	9	21	0,14	0,15	1,5
3	585	24	45	0,84	0,82	1,0
4	1342	55	100	1,62	1,62	1,1

Tablo 38 incelendiğinde 4 numaralı kategorinin çok kullanıldığı, 1, 2 ve 3 numaralı kategorilerin ise benzer sayılarda kullanıldığı görülmektedir. Bu durum, birey performanslarından kaynaklı olabileceği gibi puanlayıcıların merkeze eğilim davranışlarının da bir sonucu olabilir. Bu durumun belirlenmesi için birey yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi ve ki kare istatistiği gibi grup düzeyindeki istatistiklerin incelenmesi gerekmektedir (Myford & Wolfe, 2004). Tablo 17’deki deney ve kontrol grubunda yer alan puanlayıcıların birey yüzeyine ait ölçüm raporu incelendiğinde, ayırma oranının, ayırma indeksinin ve güvenilirlik indeksinin yüksek olduğu ve sabit etkili ki kare değerinin de anlamlı çıktığı görülmektedir. Yani birey performanslarının yetenek seviyelerine göre doğru ve istatistiksel olarak anlamlı bir şekilde ayrılmaktadır. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Merkeze eğilim davranışının bir diğer göstergesi, madde yüzeyinin uygunluk istatistikleridir. Madde yüzeyine ait ölçüm raporunda maddelerin uygunluk istatistiklerinin kabul edilebilir aralıkta (0,50 ile 2,00) olmaması merkeze eğilim davranışına işaret ederken uygunluk istatistiklerinin kabul edilebilir aralıkta olması merkeze eğilim davranışının olmadığını göstermektedir. Tablo 16’da yer alan deney ve kontrol gruplarının son test madde yüzeyine ait ölçüm raporları incelendiğinde tüm maddelerin kabul edilebilir uygunluk

istatistiklerine sahip olduđu gör÷lmektedir. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Grup düzeyinde istatistikler incelendiğinde hem deney hem de kontrol grubunda bulunan puanlayıcıların açık uçlu soruların değerlendirilmesinde grup düzeyinde merkeze eğilim davranışını göstermedikleri belirlenmiştir.

Grup düzeyindeki merkeze eğilim davranışı incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel düzeyde merkeze eğilim davranışının belirlenmesinde her bir puanlayıcıya ait uyum istatistikleri ve her bir puanlayıcıya ait kategori istatistiklerinin incelenmesi gerekmektedir. Her bir puanlayıcıya ait uyum istatistiklerinin yer aldığı tablo 18 incelendiğinde tüm puanlayıcıların uygunluk içi ve uygunluk dışı değerlerinin kabul edilebilir aralıkta (0,50 ile 2,00) olduğu gör÷lmektedir. Bu durum puanlayıcıların bireysel düzeyde merkeze eğilim davranışının sergilemediğinin bir göstergesidir. Ancak bu sonuç tek başına yeterli değildir. Bireysel düzeyde merkeze eğilim davranışının incelenmesinde kullanılan bir diğer istatistik her bir puanlayıcıya ait kategori istatistiklerinin incelenmesidir. Kategori istatistikleri incelenirken uygunluk dışı istatistiğı incelenmelidir. Deney ve kontrol gruplarının ön testlerinin her bir puanlayıcı için hesaplanan kategori istatistikleri tablo 39’da verilmiştir.

Tablo 39

Deney ve Kontrol Grubu Son Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri

Grup	Puan- layıcı	Kategori İstatistiği (%)				Uygunluk Dışı				Eşik değer				Merkeze Eğilim
		1	2	3	4	1	2	3	4	1	2	3	4	
Deney	P1	13	6	25	56	0,8	2,5	0,4	2,3	--	0,18	-1,09	0,91	VAR
Kontrol	P2	12	7	31	51	0,5	2,4	0,9	1,3	--	0,09	-1,21	1,12	VAR
Kontrol	P3	12	5	27	55	0,6	2,9	0,8	1,5	--	0,25	-1,30	1,05	VAR
Deney	P4	16	12	32	41	1,1	1,5	0,8	1,0	--	-0,30	-0,70	1,00	YOK
Kontrol	P5	9	10	38	42	0,4	1,6	0,9	1,0	--	-0,50	-0,89	1,39	VAR
Kontrol	P6	12	5	27	55	0,6	2,9	0,8	1,5	--	0,25	-1,30	1,05	VAR
Deney	P7	5	23	27	45	1,5	0,5	1,4	1,0	--	-1,41	0,54	0,87	YOK
Deney	P8	21	9	11	59	0,6	1,4	1,8	0,8	--	0,10	0,12	-0,22	YOK
Deney	P9	16	11	14	59	0,6	1,7	1,1	0,9	--	-0,03	0,23	-0,20	YOK
Kontrol	P10	11	13	24	52	1,5	1,1	0,9	0,9	--	-0,76	-0,09	0,85	YOK
Kontrol	P11	12	16	23	48	1,0	0,7	1,4	0,9	--	-0,53	0,09	0,44	YOK
Kontrol	P12	10	6	30	55	0,4	1,4	0,7	1,3	--	-0,08	-1,26	1,35	VAR
Deney	P13	14	12	34	41	0,8	2,1	0,8	1,0	--	-0,78	-0,58	1,36	VAR
Deney	P14	20	7	18	55	1,0	1,7	0,8	0,8	--	0,84	-0,59	-0,25	YOK

Tablo 39 incelendiğinde, puanlayıcıların genel olarak puanlama anahtarının tüm kategorilerini kullandıkları görülmektedir. Tablo 39'a göre P1, P2, P3, P6, P8, P9, P10, P12 ve P14 numaralı puanlayıcıların bazı kategorileri diğer kategorilere göre daha yoğun kullandıkları belirlenmiştir. Dereceli puanlama anahtarının kullanım örüntüsü incelendiğinde, kontrol grubunda yer alan P2, P3, P5, P6 ve P12 numaralı puanlayıcıların merkeze eğilim davranışı sergilediği görülmektedir. Deney grubunda ise P1 ve P13 numaralı puanlayıcıların merkeze eğilim davranışı sergilediği görülmektedir. Özetle kontrol grubunda beş, deney grubunda iki puanlayıcıda merkeze eğilim davranışı belirlenmiştir. Puanlayıcıların merkeze eğilim davranışına yatkınlıklarının karşılaştırılması için ortalama eşik analizi yapılmıştır (Rahman vd., 2017). Bir numaralı puanlayıcının ortalama eşik değerinin 1,00 $[[0,91-(-1,09)] / 2]$, iki numaralı puanlayıcının ortalama eşik değerinin 1,17 $[[1,12-(-1,21)] / 2]$, üç numaralı puanlayıcının ortalama eşik değerinin 1,18 $[[1,05-(-1,30)] / 2]$, beş numaralı puanlayıcının ortalama eşik değerinin 1,14 $[[1,39-(-0,89)] / 2]$, altı numaralı puanlayıcının ortalama eşik değerinin 1,18 $[[1,05-(-1,30)] / 2]$, on iki numaralı puanlayıcının

ortalama eşik değerinin 1,30 $[[1,35-(-1,26)] / 2]$, on üç numaralı puanlayıcının ortalama eşik değerinin 1,07 $[[1,36-(0,78)] / 2]$ olduğu görülmektedir. Bu durum on iki numaralı puanlayıcının merkeze eğilim davranışına diğer puanlayıcılara göre daha yatkın olduğunu göstermektedir.

Halo Etkisi

Puanlayıcı eğitimi sonrasında Türkçe puanlayıcı grubu deney ve kontrol gruplarının halo etkisi davranışlarının incelenmesi için madde yüzeyine ait ayırma oranı, ayırma indeksi ve sabit etkili ki kare değeri gibi grup düzeyindeki istatistikler incelenmiştir. Madde yüzeyine ait ölçüm raporunun yer aldığı tablo 16 incelendiğinde; ayırma oranının 13,56, ayırma indeksinin 18,42, güvenilirlik indeksinin 0,99 olduğu ve sabit etkili ki kare istatistiğinin anlamlı olduğu ($\chi^2(sd) = 805,1 (4), p = 0,00 < 0,01$) görülmektedir. Madde yüzeyine ait istatistiklerin yüksek ve sabit etkili ki kare değerinin anlamlı olması puanlayıcıların yapmış oldukları puanlamalarda grup düzeyinde halo etkisinin olmadığını göstermektedir. Grup düzeyinde halo davranışı incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel istatistikler incelenirken, madde güçlükleri arasındaki fark büyük olduğunda puanlayıcıya ait uygunluk istatistikleri 1'den çok büyük olduğunda, madde güçlükleri arasındaki fark küçük olduğunda da puanlayıcıya ait uygunluk istatistikleri 1'den çok küçük olduğunda halo davranışının sergilendiği ifade edilmiştir (Myford & Wolfe, 2004). Türkçe puanlayıcı grubu son test madde yüzeyine ait ölçüm raporunun yer aldığı tablo 16 incelendiğinde maddelerin güçlük düzeyleri arasındaki farkın 2,57 $(1,21-[-1,36])$ olduğu görülmektedir. Bu fark büyük olduğundan puanlayıcılara ait uygunluk istatistiklerinin 1'den çok büyük olanların halo davranışı gösterdikleri kabul edilmiştir. Türkçe puanlayıcı grubu son test puanlayıcı yüzeyine ait uygunluk istatistiklerinin yer aldığı Tablo 18 incelendiğinde, dört puanlayıcıda halo davranışı tespit edilmiştir. Halo etkisi deney grubunda P7 ve P14 numaralı puanlayıcılarda olmak üzere iki puanlayıcıda tespit edilirken kontrol grubunda P5 ve P11 numaralı puanlayıcılarda olmak üzere iki puanlayıcıda tespit edilmiştir.

Halo etkisinin bulunup bulunmadığına karar vermek için önerilen bir diğer yöntem ise madde güçlüklerinin birbirine eşitleyerek çok yüzeyli rasch analizini tekrar edip puanlayıcılara ait uyum istatistiklerini incelemektir (Linacre, 2023, s. 360). Madde güçlükleri eşitlenerek oluşturulan yeni modele mükemmel uyum gösteren (uygunluk içi=1,00 veya uygunluk dışı=1,00) puanlayıcıların halo davranışını sergiledikleri kabul

edilmektedir (Linacre, 2023, s. 360). Bu yüzden Türkçe testindeki beş açık uçlu sorunun kriter güçlükleri sıfıra eşitlenerek çok yüzeyli rasch analizi tekrar edilmiştir. Ardından uygunluk istatistikleri incelenmiştir. Kriter güçlükleri eşitlendikten sonra Türkçe son test puanlayıcı grubunun puanlayıcı yüzeyine ait ölçüm raporu Tablo 40'ta verilmiştir.

Tablo 40

Türkçe Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Grup	Puanlayıcı	Uygunluk İçi	Uygunluk Dışı
Deney	P1	1,03	0,90
Kontrol	P2	0,92	0,82
Kontrol	P3	1,00	0,88
Deney	P4	0,90	0,95
Kontrol	P5	0,89	1,03
Kontrol	P6	1,00	0,88
Deney	P7	0,83	1,00
Deney	P8	1,24	1,12
Deney	P9	1,15	1,00
Kontrol	P10	0,99	0,90
Kontrol	P11	1,05	1,22
Kontrol	P12	1,02	0,90
Deney	P13	0,85	0,80
Deney	P14	1,26	1,37

Tablo 40'ta kriter güçlüğü eşitlendikten sonra Türkçe puanlayıcı grubuna ait uygunluk içi ve uygunluk dışı gibi bireysel istatistikleri yer almaktadır. Puanlayıcılara ait bireysel istatistikler incelendiğinde, deney grubunda P7 ve P9, kontrol grubunda ise P3 ve P6 numaralı puanlayıcıların mükemmel uyum gösterdiği, dolayısıyla halo davranışını sergilediği belirlenmiştir.

Tüm bireysel istatistiklerin incelenmesi sonucunda, deney grubunda P7, P9 ve P14 numaralı puanlayıcılarda olmak üzere üç, kontrol grubunda da P3, P5, P6 ve P11 numaralı puanlayıcılarda olmak üzere dört puanlayıcıda halo etkisi belirlenmiştir.

Sonuç olarak halo etkisinin merkeze eğilim davranışı gibi grup düzeyinde ortaya çıkmasa da bireysel düzeyde ortaya çıktığı belirlenmiştir.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)

Farklılaşan puanlayıcı katılığı ve cömertliği davranışı incelenirken modelde yer alan yüzeyler arasındaki etkileşimler incelenmektedir. Mevcut çalışma kapsamında birey (b), puanlayıcı (p) ve madde (m) yüzeyi olmak üzere üç yüzey bulunmaktadır. Bu yüzeyler

arasındaki etkileşimler de *puanlayıcı x birey (pxb)*, *puanlayıcı x madde (pxm)*, *birey x madde (bxm)* ve *birey x puanlayıcı x madde (bxpxm)* şeklindedir. Mevcut çalışmada puanlayıcıların maddelerle ilgili davranışlarını incelemek amaçlanmadığı için *puanlayıcı x madde (pxm)* etkileşimi çalışmaya dahil edilmemiştir. Benzer şekilde çalışmada puanlayıcı davranışları incelendiği ve bireylerin maddeleri cevaplandırma seviyeleri de incelenmediği için *birey x madde (bxm)* etkileşimi de çalışmaya dahil edilmemiştir. Ayrıca puanlayıcıların hem madde hem de bireye göre davranışları araştırma kapsamında olmadığı için *birey x puanlayıcı x madde (bxpxm)* etkileşimi de çalışmaya dahil edilmemiştir. Çalışma kapsamında Türkçe açık uçlu soruların puanlanması sürecinde meydana gelen yanlışlık davranışlarının tespiti için *puanlayıcı x birey (pxb)* etkileşimi incelenmiştir.

Puanlayıcı x birey (pxb) etkileşiminde yer alan istatistiksel olarak anlamlı olan etkileşimlerin tespiti için t değerinden faydalanılmaktadır. (Linacre, 2023, s. 190), ± 2 aralığı dışında kalan t değerlerinin istatistiksel olarak anlamlı olduğunu ifade etmektedir. Mevcut çalışmada 35 öğrencinin Türkçe açık uçlu sorulara vermiş oldukları tüm cevaplar yedi deney, yedi kontrol olmak üzere toplam 14 puanlayıcı tarafından puanlanmıştır. Dolayısıyla iki yüzey arasında 490 (14 x 35) olası etkileşim varken 245 (7 x 35) etkileşim deney grubunda, 245 (7 x 35) etkileşim de kontrol grubunda meydana gelmiştir. Analiz sonuçları incelenirken öncelikle grup düzeyindeki istatistikler incelenmiş ardından bireysel istatistiklere geçilmiştir. Grup düzeyindeki istatistikler incelendiğinde ki kare istatistiğinin istatistiksel açıdan anlamlı olmadığı bulunmuştur (χ^2 (sd)=242,4 (490), $p > 0,01$). Bu durum puanlayıcıların grup düzeyinde yanlışlık davranışı sergilemedikleri şeklinde yorumlanabilir. Grup düzeyindeki istatistikler incelendikten sonra bireysel istatistikler incelenmiştir. Yapılan incelemede, her grup için meydana gelen olası 490 etkileşimden kontrol grubunda istatistiksel olarak anlamlı altı tane etkileşim bulunurken deney grubunda istatistiksel olarak anlamlı bir tane etkileşim yer almaktadır. Yani Türkçe puanlayıcı grubunda toplam yedi puanlayıcı farklılaşan puanlayıcı katılımı ve cömertliği davranışı sergilemiştir. Bir başka deyişle yedi puanlayıcıda yanlışlık davranışı vardır. Grupların etkileşimlerine ait istatistiklerin tamamı Ek 10'da yer alırken etkileşimlerin tamamını sunmak mümkün olmadığı için sadece istatistiksel olarak anlamlı olan etkileşimler tablo 41'de verilmiştir.

Tablo 41

Türkçe Puanlayıcı grubuna ait anlamlı puanlayıcı x birey etkileşimleri

Grup	Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Bias (logit)	Standart Hata	t
Kontrol	P5	12	16,00	18,83	-1,32	0,57	-2,34
Kontrol	P5	23	15,00	18,05	-1,12	0,53	-2,11
Kontrol	P5	16	15,00	9,31	1,38	0,53	2,61
Kontrol	P11	25	15,00	18,27	-1,24	0,53	-2,34
Kontrol	P11	27	19,00	13,64	2,13	0,99	2,16
Kontrol	P12	16	16,00	10,93	1,28	0,57	2,25
Deney	P14	27	19,00	13,71	2,11	0,99	2,14

 $\chi^2(sd) = 242,4 (490), p > 0,01$

Türkçe puanlayıcı grubu son test sonuçlarına ait puanlayıcı x birey etkileşimlerinin yer aldığı tablo 41 incelendiğinde, yedi puanlayıcıda farklılaşan puanlayıcı katılımı ve cömertliği davranışı olduğu görülmektedir.

Deney grubunda yer alan puanlayıcılardan P14 numaralı puanlayıcıda ve kontrol grubunda yer alan P5, P11 ve P12 numaralı puanlayıcılarda farklılaşan puanlayıcı katılımı ve cömertliği davranışı olduğu görülmektedir. Deney grubunda yer alan P14 numaralı puanlayıcının 27 numaralı bireye olması gerekenden fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Kontrol grubunda yer alan P5 numaralı puanlayıcının 12 ve 23 numaralı bireylere olması gerekenden az puan verdiği yani farklılaşan puanlayıcı katılımı davranışı sergilediği görülmektedir. Ayrıca P5 numaralı puanlayıcının 16 numaralı bireye de olması gerekenden fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Kontrol grubunda yer alan P11 numaralı puanlayıcının 25 numaralı bireye olması gerekenden az puan verdiği yani farklılaşan puanlayıcı katılımı davranışı sergilediği görülürken 27 numaralı bireye de olması gerekenden fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Benzer şekilde kontrol grubunda yer alan P12 numaralı puanlayıcının 16 numaralı bireye olması gerekenden fazla puan verdiği ve dolayısıyla farklılaşan puanlayıcı cömertliği davranışı sergilediği görülmektedir. Özetle, kontrol grubunda iki puanlayıcıda (P5 ve P11) farklılaşan puanlayıcı katılımı ve cömertliği davranışı, bir puanlayıcıda (P12) farklılaşan puanlayıcı cömertliği davranışı tespit edilirken deney grubunda bir puanlayıcıda (P14) farklılaşan puanlayıcı cömertliği davranışı tespit edilmiştir.

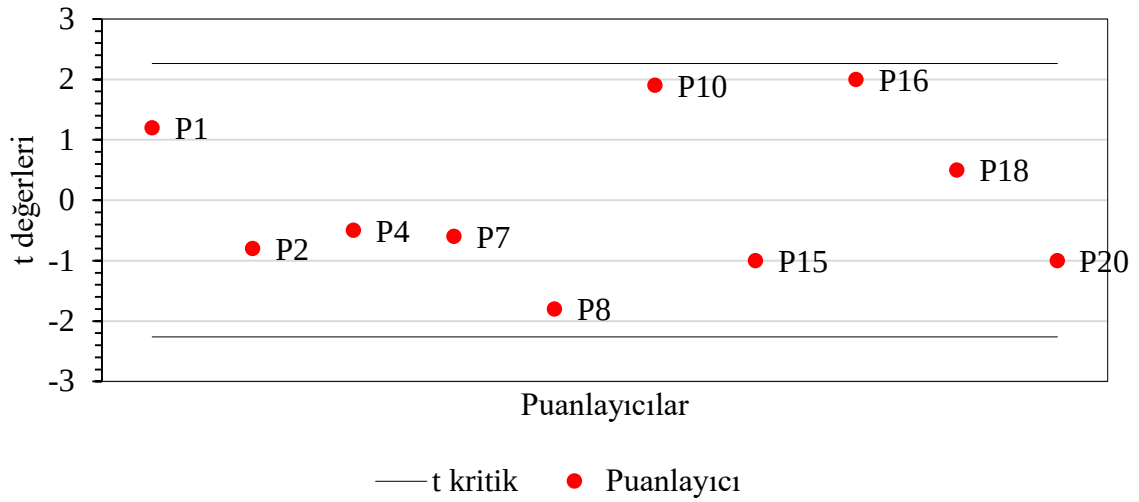
Puanlayıcı Eğitimi Sonrasında Matematik Puanlayıcı Grubu Puanlayıcı Davranışları

Katılık ve Cömertlik

Puanlayıcı eğitimi sonrasında Matematik puanlayıcı grubu deney ve kontrol gruplarının katılık ve cömertlik davranışlarının incelenmesi için puanlayıcı yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi gibi grup düzeyindeki istatistikler incelenmiştir. Yapılan incelemede ayırma oranı 2,16, ayırma indeksi 3,22 ve güvenilirlik indeksi 0,82 olarak belirlenmiştir (Tablo 24). Bu istatistiklerin küçük olması puanlayıcıların katılık ve cömertlik açısından benzer davranışlara sahip olabileceklerini göstermektedir. Ancak kesin yargıya varmak için sabit etkili ki kare istatistiğinin incelenmesi gerekmektedir. Sabit etkili ki kare istatistiği incelendiğinde, ki kare istatistiğinin istatistiksel olarak anlamlı olduğu görülmektedir ($\chi^2(sd) = 105,1$ (19), $p = 0,00 < 0,01$). Bu durum puanlayıcıların grup düzeyinde katılık ve cömertlik davranışının olduğunu göstermektedir. Grup düzeyindeki istatistikler incelendikten sonra bireysel istatistiklere geçilmiştir.

Birey düzeyindeki istatistikler incelenirken t değeri istatistiği kullanılmaktadır. Puanlayıcının t değeri, puanlayıcıya ait logit ölçüsünden tüm puanlayıcıların logit ortalaması farkının standart hataya bölünmesiyle elde edilir. Hesaplanan t değeri, t dağılım tablosundaki kritik t değeri ile karşılaştırılır ve istatistiksel açıdan anlamlılığı test edilir. Puanlayıcıya ait hesaplanan t değeri, kritik t değeri ile karşılaştırılır ve puanlayıcı davranışı belirlenir.

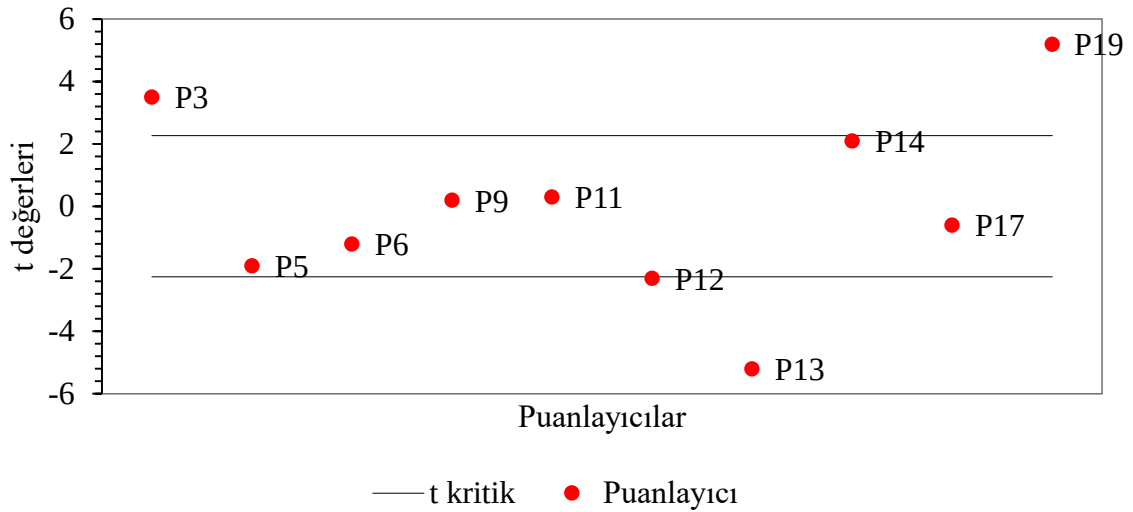
Araştırmada kontrol grubunda on, deney grubunda ise on puanlayıcı olduğundan serbestlik derecesi dokuzdur. Serbestlik derecesi dokuz ve anlamlılık düzeyi 0,05 için t kritik değeri (t_{kritik}) 2,262'dir. Tablo ve şekiller 2,262 değeri göz önünde bulundurularak yapılandırılmıştır. Matematik deney grubunda yer alan puanlayıcıların son test puanlarının katılık ve cömertlik durumlarını gösteren t değerleri şekil 8 ve şekil 9'da verilmiştir. Matematik puanlayıcı deney grubunda yer alan puanlayıcıların son test puanlarının katılık ve cömertlik durumlarını gösteren t değerleri şekil 8'de verilmiştir.



Şekil 8. Matematik puanlayıcı deney grubu son test verilerinden elde edilen t değerleri.

Şekil 8 incelendiğinde, deney grubunda yer alan tüm puanlayıcıların t değerlerinin kabul edilebilir aralıkta olduğu görülmektedir. Yani deney grubunda yer alan puanlayıcılarda katılık ya da cömertlik davranışı bulunmamaktadır.

Matematik puanlayıcı kontrol grubunda yer alan puanlayıcıların son test puanlarının katılık ve cömertlik durumlarını gösteren t değerleri şekil 9’da verilmiştir.



Şekil 9. Matematik puanlayıcı kontrol grubu son test verilerinden elde edilen t değerleri.

Şekil 9 incelendiğinde, kontrol grubunda yer alan puanlayıcılardan P3, P12, P13 ve P19 numaralı puanlayıcıların t değerlerinin kabul edilebilir aralığın dışında olduğu görülmektedir. Matematik puanlayıcı kontrol grubunda iki puanlayıcıda (%20,00) puanlayıcı katılığı davranışı görülürken iki puanlayıcıda (%20,00) da puanlayıcı cömertliği

davranışı görülmektedir. Matematik puanlayıcı kontrol grubunda yer alan puanlayıcılardan P12 ve P13 numaralı puanlayıcıların puanlayıcı katılığı davranışı sergilediği, P3 ve P19 numaralı puanlayıcıların ise puanlayıcı cömertliği davranışı sergilediği görülmektedir.

Matematik puanlayıcı deney ve kontrol gruplarının son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Mann Whitney U testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında kullanılacak en güçlü test Mann-Whitney U testidir (Tekindal, 2022, s. 135). Ayrıca matematik puanlayıcı gruplarının grup içi ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için de Wilcoxon İşaretli Sıralar testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında ilişkili iki gruba ait ortalamalar arasındaki fark test edilirken kullanılan alternatif yöntem Wilcoxon İşaretli Sıralar testidir (Büyüköztürk, Çokluk, Köklü, 2019, s. 197). T değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcıların t değerleri karşılaştırılmıştır. Ancak t değerleri karşılaştırılırken deneysel işlemin etkililiğini daha iyi gözlemlemek için puanlayıcıların son test t değerlerinden ön test t değerlerinin farkı alınmıştır. Puanlayıcıların t değerlerinin farklarına ait Mann-Whitney U testi sonuçları tablo 42’de verilmiştir.

Tablo 42

Deney ve Kontrol Gruplarının Son Test t Değerleri Farklarına Ait Mann-Whitney U Testi Sonuçları

Grup	N	Sıra Ortalaması	Sıra Toplamı	U	p	η^2
Deney	10	5,85	58,50	3,50	0,00	0,578
Kontrol	10	15,15	151,50			
Toplam	20					

Tablo 42 incelendiğinde, matematik puanlayıcı deney ve kontrol gruplarının son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olduğu görülmektedir ($U=3,50$; $p= 0,00<0,05$). Bu durum deney ve kontrol gruplarının puanlayıcı eğitimi sonrasında katılık ve cömertlik açısından farklı davranışlara sahip olduklarını göstermektedir. Ayrıca puanlayıcıların katılık ve cömertlik davranışlarındaki değişikliğin %57,8’inin ($U=3,50$; $p= 0,00<0,05$; $\eta^2= 0,578$) puanlayıcı eğitimi ile açıklanabildiği belirlenmiştir. Puanlayıcı eğitimi öncesinde deney grubunda katılık cömertlik oranı %70,00 iken puanlayıcı eğitimi sonrasında tamamen ortadan kalkmıştır. Matematik puanlayıcı deney grubunun puanlayıcı eğitimi öncesinde ve puanlayıcı eğitimi

sonrasında katılık ve cömertlik açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Wilcoxon İşaretli Sıralar Testi yapılmıştır. Parametrik testlerin varsayımları karşılanmadığında ilişkili iki gruba ait ortalamalar arasındaki fark test edilirken kullanılan alternatif yöntem Wilcoxon İşaretli Sıralar testidir (Büyüköztürk, Çokluk, Köklü, 2019, s. 197). t değeri katılık ve cömertlik davranışının bir göstergesi olduğu için puanlayıcılara ait t değerleri karşılaştırılmıştır. Ancak t değerleri karşılaştırılırken deneysel işlemin etkililiğini daha iyi gözlemlemek için puanlayıcıların son test t değerleri ile ön test t değerlerinin farkı alınmıştır. Matematik puanlayıcı deney grubunda yer alan puanlayıcıların t değerlerine ait Wilcoxon İşaretli Sıralar testi sonuçları tablo 43'te verilmiştir.

Tablo 43

Deney Grubu Ön Test ve Son Test t İstatistiklerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları

Son test- on test	n	Sıra Ortalaması	Sıra Toplamı	z	p
Negatif Sıralar	10	5,50	55,00	-2,807	0,005
Pozitif Sıralar	0	0,00	0,00		
Eşit Sıralar	0				

Tablo 43 incelendiğinde, deney grubu ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olduğu görülmektedir ($z=-2,807$; $p=0,005<0,05$). Bu durum, puanlayıcı eğitimi öncesinde ve sonrasında matematik puanlayıcı grubunun katılık ve cömertlik davranışları arasında anlamlı fark olduğunu göstermektedir. Yani deney grubunda puanlayıcı eğitimi sonrasında katılık cömertlik durumundaki değişim istatistiksel olarak anlamlıdır.

Matematik puanlayıcı kontrol grubunda ikinci uygulamada katılık cömertlik durumu %20,00'den %40,00'a çıkmıştır. Kontrol grubunun birinci uygulama ve ikinci uygulama arasında katılık ve cömertlik açısından istatistiksel olarak birbirlerinden farklı olup olmadıklarının test edilmesi için Wilcoxon İşaretli Sıralar testi yapılmıştır. t değeri katılık ve cömertlik davranışının bir göstergesi olduğu için ön ve son testteki puanlayıcılara ait t değerleri karşılaştırılmıştır. Kontrol grubunda yer alan puanlayıcıların t değerlerine ait Wilcoxon İşaretli Sıralar testi sonuçları tablo 44'te verilmiştir.

Tablo 44

Kontrol Grubu Ön Test ve Son Test t İstatistiklerine Ait Wilcoxon İşaretli Sıralar Testi Sonuçları

Son test- on test	n	Sıra Ortalaması	Sıra Toplamı	z	p
Negatif Sıralar	2	7,00	14,00	-1,376	0,169
Pozitif Sıralar	8	5,13	41,00		
Eşit Sıralar	0				

Tablo 44 incelendiğinde, matematik puanlayıcı kontrol grubu ön test ve son test sonuçlarının katılık ve cömertlik davranışları açısından istatistiksel olarak anlamlı fark olmadığı görülmektedir ($z=-1,376$; $p=0,169>0,05$). Bu durum, birinci uygulama ve ikinci uygulama arasında matematik puanlayıcı kontrol grubunun katılık ve cömertlik davranışları arasında anlamlı fark olmadığını göstermektedir. Yani matematik puanlayıcı kontrol grubunun katılık ve cömertlik açısından birinci ve ikinci uygulama arasındaki değişim istatistiksel olarak anlamlı değildir.

Merkeze Eğilim

Puanlayıcı eğitimi sonrasında Matematik puanlayıcı grubu deney ve kontrol gruplarının merkeze eğilim davranışının incelenmesi için kategori istatistikleri, birey yüzeyine ait ayırma oranı, ayırma indeksi ve güvenilirlik indeksi ve ki kare istatistiği ve madde yüzeyinin uygunluk istatistikleri gibi grup düzeyindeki istatistikler incelenmiştir. Matematik puanlayıcı grubuna ait kategori istatistikleri tablo 45'te verilmiştir.

Tablo 45

Matematik Puanlayıcı Grubuna Ait Son Test Sonuçlarına Ait Kategori İstatistikleri

Kategoriler	Frekans (f)	Yüzde (%)	Yığılmalı yüzde (%)	Ortalama logit	Beklenen logit	Uyum Dışı
1	883	21	21	-1,06	-1,05	0,9
2	628	15	36	-0,35	-0,30	1,0
3	669	16	52	0,63	0,48	1,0
4	2020	48	100	3,31	3,34	1,2

Tablo 45 incelendiğinde 4 numaralı kategorinin çok kullanıldığı, 1, 2 ve 3 numaralı kategorilerin ise benzer sayılarda kullanıldığı görülmektedir. Bu durum, birey performanslarından kaynaklı olabileceği gibi puanlayıcı etkisinden de kaynaklanabilir. Ancak puanlayıcıların kategorileri kullanma durumları incelendiğinde merkeze eğilim davranışının yer almadığı söylenebilir. Ancak bu etkinin belirlenmesi için birey yüzeyine ait

ayırma oranı, ayırma indeksi ve güvenilirlik indeksi ve ki kare istatistiği gibi grup düzeyindeki istatistiklerin incelenmesi gerekmektedir (Myford & Wolfe, 2004). Tablo 23'teki deney ve kontrol grubunda yer alan puanlayıcıların birey yüzeyine ait ölçüm raporu incelendiğinde, ayırma oranının, ayırma indeksinin ve güvenilirlik indeksinin yüksek olduğu ve sabit etkili ki kare değerinin de anlamlı çıktığı görülmektedir ($\chi^2(sd) = 1177,0 (34), p = 0,00 < 0,01$). Yani birey performansları yetenek seviyelerine göre doğru ve istatistiksel olarak anlamlı bir şekilde ayrılmaktadır. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışı olmadığını göstermektedir. Merkeze eğilim davranışının bir diğer göstergesi, madde yüzeyinin uygunluk istatistikleridir. Madde yüzeyine ait ölçüm raporunda maddelerin uygunluk istatistiklerinin kabul edilebilir aralıkta (0,50 ile 2,00) olmaması merkeze eğilim davranışına işaret ederken uygunluk istatistiklerinin kabul edilebilir aralıkta olması merkeze eğilim davranışının olmadığını göstermektedir. Tablo 22'de yer alan deney ve kontrol gruplarının son test madde yüzeyine ait ölçüm raporları incelendiğinde tüm maddelerin kabul edilebilir uygunluk istatistiklerine sahip olduğu görülmektedir. Bu durum puanlayıcıların grup düzeyinde merkeze eğilim davranışına sahip olmadıklarını göstermektedir.

Grup düzeyinde istatistikler incelendikten sonra puanlayıcı eğitimi sonrasında hem deney hem de kontrol grubunda bulunan puanlayıcıların açık uçlu soruların değerlendirilmesinde bireysel düzeyde merkeze eğilim davranışları incelenmiştir. Bireysel düzeyde merkeze eğilim davranışının belirlenmesinde her bir puanlayıcıya ait uyum istatistikleri ve her bir puanlayıcıya ait kategori istatistiklerinin incelenmesi gerekmektedir. Her bir puanlayıcıya ait uyum istatistiklerinin yer aldığı tablo 24 incelendiğinde, 11 numaralı puanlayıcı haricinde tüm puanlayıcıların uygunluk içi ve uygunluk dışı değerlerinin kabul edilebilir aralıkta (0,50 ile 2,00) olduğu görülmektedir. Bu durum 11 numaralı puanlayıcının merkeze eğilim davranışı sergilediğinin, diğer puanlayıcıların bireysel düzeyde merkeze eğilim davranışının sergilemediğinin bir göstergesidir. Ancak bu sonuç tek başına yeterli değildir. Bireysel düzeyde merkeze eğilim davranışının incelenmesinde kullanılan bir diğer istatistik her bir puanlayıcıya ait kategori istatistiklerinin incelenmesidir. Kategori istatistikleri incelenirken uygunluk dışı istatistiği incelenmelidir. Deney ve kontrol gruplarının ön testlerinin her bir puanlayıcı için hesaplanan kategori istatistikleri tablo 46'da verilmiştir.

Tablo 46

Deney ve Kontrol Grubu Son Test Her Bir Puanlayıcıya Ait Kategori İstatistikleri

Grup	Puanlayıcı	Kategori İstatistiği (%)				Uygunluk Dışı				Eşik değer				Merkeze Eğilim
		1	2	3	4	1	2	3	4	1	2	3	4	
Deney	P1	26	15	18	41	0,7	0,6	1,7	1,1	--	-0,14	-0,11	0,25	YOK
Deney	P2	29	17	15	39	0,7	1,3	1,5	1,1	--	-0,47	0,08	0,39	YOK
Kontrol	P3	22	15	19	44	0,7	1,1	1,2	0,9	--	-0,16	-0,07	0,23	YOK
Deney	P4	29	15	17	39	1,0	1,4	1,2	0,8	--	-0,58	-0,13	0,71	YOK
Kontrol	P5	23	26	21	30	1,0	1,7	0,9	1,0	--	-1,60	0,00	1,60	YOK
Kontrol	P6	28	17	21	34	0,9	0,6	1,1	1,2	--	-0,65	-0,27	0,93	YOK
Deney	P7	29	16	17	39	0,7	1,6	1,1	1,0	--	-0,30	-0,02	0,33	YOK
Deney	P8	26	18	23	33	0,9	0,6	1,0	1,4	--	-0,64	-0,25	0,89	YOK
Kontrol	P9	25	16	16	42	0,9	0,9	1,5	1,0	--	-0,48	0,13	0,35	YOK
Deney	P10	27	12	13	48	0,7	1,1	1,7	1,0	--	0,09	0,11	-0,20	YOK
Kontrol	P11	13	24	20	43	1,0	0,6	1,4	1,2	--	-0,88	0,41	0,47	YOK
Kontrol	P12	27	25	14	34	0,9	1,7	0,9	1,3	--	-1,03	0,53	0,50	YOK
Kontrol	P13	22	23	25	30	0,9	0,9	1,0	1,2	--	-2,21	-0,36	2,57	YOK
Kontrol	P14	26	14	16	44	0,6	1,0	1,3	0,9	--	-0,27	-0,01	0,28	YOK
Deney	P15	27	17	22	34	0,8	1,0	1,1	1,2	--	-0,74	-0,29	1,03	YOK
Deney	P16	24	14	24	39	0,7	0,8	1,3	1,0	--	-0,30	-0,39	0,79	YOK
Kontrol	P17	26	19	19	36	0,9	0,8	1,0	1,1	--	-0,88	-0,01	0,89	YOK
Deney	P18	25	18	15	41	0,9	0,7	1,9	1,0	--	-0,81	0,37	0,44	YOK
Kontrol	P19	16	19	21	44	0,8	0,6	1,7	0,9	--	-0,67	0,14	0,53	YOK
Deney	P20	27	17	22	34	0,8	1,0	1,1	1,2	--	-0,74	-0,29	1,03	YOK

Tablo 46 incelendiğinde, puanlayıcıların genel olarak puanlama anahtarının tüm kategorilerini kullandıkları görülmektedir. Tablo 46'ya göre P1, P2, P3, P10 ve P19 numaralı puanlayıcıların bazı kategorileri diğer kategorilere göre daha yoğun kullandıkları belirlenmiştir. Ancak hem deney hem de kontrol grubunda yer alan puanlayıcılarda merkeze eğilim davranışının olmadığı görülmektedir. Özetle puanlayıcı eğitimi sonrasında deney grubunda yer alan puanlayıcılarda merkeze eğilim davranışının yer almadığı, kontrol grubunda ise P11 numaralı puanlayıcının merkeze eğilim davranışı sergilediği görülmektedir. Puanlayıcıların merkeze eğilim davranışına yakınlıklarının karşılaştırılması için ortalama eşik analizi yapılmıştır (Rahman vd., 2017). On bir numaralı puanlayıcının ortalama eşik değerinin $0,68 \left[\frac{0,47 - (-0,88)}{2} \right]$ olduğu görülmektedir.

Halo Etkisi

Puanlayıcı eğitimi sonrasında Matematik puanlayıcı grubu deney ve kontrol gruplarının halo etkisi davranışlarının incelenmesi için madde yüzeyine ait ayırma oranı, ayırma indeksi ve

sabit etkili ki kare değeri gibi grup düzeyindeki istatistikler incelenmiştir. Madde yüzeyine ait ölçüm raporunun yer aldığı tablo 22 incelendiğinde; ayırma oranının 9,60, ayırma indeksinin 13,13, güvenilirlik indeksinin 0,99 olduğu ve sabit etkili ki kare istatistiğinin anlamlı olduğu ($\chi^2(sd) = 1897,1 (5)$, $p = 0,00 < 0,01$) görülmektedir. Madde yüzeyine ait istatistiklerin yüksek ve sabit etkili ki kare değerinin anlamlı olması puanlayıcıların yapmış oldukları puanlamalarda grup düzeyinde halo etkisinin olmadığını göstermektedir. Grup düzeyinde halo etkisi incelendikten sonra bireysel istatistikler incelenmiştir. Bireysel istatistikler incelenirken, madde güçlükleri arasındaki fark büyük olduğunda puanlayıcıya ait uygunluk istatistikleri 1'den çok büyük olduğunda, madde güçlükleri arasındaki fark küçük olduğunda da puanlayıcıya ait uygunluk istatistikleri 1'den çok küçük olduğunda halo etkisinin olduğu ifade edilmiştir (Myford & Wolfe, 2004). Matematik puanlayıcı grubu son test madde yüzeyine ait ölçüm raporunun yer aldığı tablo 22 incelendiğinde maddelerin güçlük düzeyleri arasındaki farkın 7,55 (5,23-[-2,32]) olduğu görülmektedir. Bu fark büyük olduğundan puanlayıcılara ait uygunluk istatistiklerinin 1'den çok büyük olanların halo etkisine sahip oldukları kabul edilmiştir. Matematik puanlayıcı grubu son test puanlayıcı yüzeyine ait uygunluk istatistiklerinin yer aldığı Tablo 24 incelendiğinde, P11 ve P13 numaralı puanlayıcılarda halo etkisi tespit edilmiştir. Halo etkisi, deney grubunda gözlenmezken kontrol grubunda iki puanlayıcıda (P11 ve P13) tespit edilmiştir.

Halo etkisinin bulunup bulunmadığına karar vermek için önerilen bir diğer yöntem ise madde güçlüklerini birbirine eşitleyerek çok yüzeyli rasch analizini tekrar edip puanlayıcılara ait uyum istatistiklerini incelemektir (Linacre, 2023, s. 360). Madde güçlükleri eşitlenerek oluşturulan yeni modele mükemmel uyum gösteren (uygunluk içi=1,00 veya uygunluk dışı=1,00) puanlayıcıların halo etkisine sahip oldukları kabul edilmektedir (Linacre, 2023, s. 360). Bu yüzden Matematik testindeki altı açık uçlu sorunun kriter güçlükleri sıfıra eşitlenerek çok yüzeyli rasch analizi tekrar edilmiştir. Ardından uygunluk istatistikleri incelenmiştir. Kriter güçlükleri eşitlendikten sonra Matematik puanlayıcı grubunun puanlayıcı yüzeyine ait ölçüm raporu Tablo 47'de verilmiştir.

Tablo 47

Matematik Puanlayıcı Grubu Kriter Güçlüğü Eşitlendikten Sonra Puanlayıcı Yüzeyinin Son Test Ölçümlerine Ait Ölçüm Raporu

Grup	Puanlayıcı	Uygunluk İçi	Uygunluk Dışı
Deney	P1	0,99	0,95
Deney	P2	1,10	1,09
Kontrol	P3	0,95	0,91
Deney	P4	1,07	1,06
Kontrol	P5	0,97	0,99
Kontrol	P6	1,02	1,01
Deney	P7	1,10	1,10
Deney	P8	0,94	0,94
Kontrol	P9	1,02	0,99
Deney	P10	1,06	1,02
Kontrol	P11	0,96	0,95
Kontrol	P12	1,07	1,09
Kontrol	P13	0,84	0,85
Kontrol	P14	1,07	1,04
Deney	P15	1,03	1,04
Deney	P16	0,98	0,95
Kontrol	P17	0,96	0,95
Deney	P18	0,99	0,96
Kontrol	P19	0,86	0,81
Deney	P20	1,03	1,04

Tablo 47’de Matematik son test puanlayıcı grubuna ait uygunluk içi ve uygunluk dışı gibi bireysel istatistikleri yer almaktadır. Puanlayıcılara ait bireysel istatistikler incelendiğinde, mükemmel uyum gösteren puanlayıcının yer almadığı ve dolayısıyla halo etkisinin olmadığı belirlenmiştir.

Tüm bireysel istatistiklerin incelenmesi sonucunda, deney grubunda halo etkisi gözlenmezken, kontrol grubunda P11 ve P13 numaralı puanlayıcılarda halo etkisi tespit edilmiştir.

Farklılaşan Puanlayıcı Katılığı ve Cömertliği (Yanlılık)

Farklılaşan puanlayıcı katılığı ve cömertliği davranışı incelenirken modelde yer alan yüzeyler arasındaki etkileşimler incelenmektedir. Mevcut çalışma kapsamında birey (b), puanlayıcı (p) ve madde (m) yüzeyi olmak üzere üç yüzey bulunmaktadır. Bu yüzeyler arasındaki etkileşimler de *puanlayıcı x birey (pxb)*, *puanlayıcı x madde (pxm)*, *birey x madde (bxm)* ve *birey x puanlayıcı x madde (bpxpm)* şeklindedir. Mevcut çalışmada puanlayıcıların maddelerle ilgili davranışlarını incelemek amaçlanmadığı için *puanlayıcı x madde (pxm)*

etkileşimi çalışmaya dahil edilmemiştir. Benzer şekilde çalışmada puanlayıcı davranışları incelendiği ve bireylerin maddeleri cevaplandırma seviyeleri de incelenmediği için *birey x madde (bxm)* etkileşimi de çalışmaya dahil edilmemiştir. Ayrıca puanlayıcıların hem madde hem de bireye göre davranışları araştırma kapsamında olmadığı için *birey x puanlayıcı x madde (bxpxm)* etkileşimi de çalışmaya dahil edilmemiştir. Çalışma kapsamında Matematik açık uçlu soruların puanlanması sürecinde meydana gelen yanlılık davranışlarının tespiti için *puanlayıcı x birey (pxb)* etkileşimi incelenmiştir.

Puanlayıcı x birey (pxb) etkileşiminde yer alan istatistiksel olarak anlamlı olan etkileşimlerin tespiti için t değerinden faydalanılmaktadır. Linacre (2023, s. 190), ± 2 aralığı dışında kalan t değerlerinin istatistiksel olarak anlamlı olduğunu ifade etmektedir. Mevcut çalışmada 35 öğrencinin Matematik açık uçlu sorulara vermiş oldukları tüm cevaplar 10 deney, 10 kontrol olmak üzere toplam 20 puanlayıcı tarafından puanlanmıştır. Dolayısıyla iki yüzey arasında 700 (20 x 35) olası etkileşim varken 350 (10 x 35) etkileşim deney grubunda, 350 (10 x 35) etkileşim de kontrol grubunda meydana gelmiştir. Analiz sonuçları incelenirken öncelikle grup düzeyindeki istatistikler incelenmiş ardından bireysel istatistiklere geçilmiştir. Grup düzeyindeki istatistikler incelendiğinde ki kare istatistiğinin istatistiksel açıdan anlamlı olmadığı bulunmuştur (χ^2 (sd)=366,6 (700), $p > 0,01$). Bu durum puanlayıcıların grup düzeyinde yanlılık davranışı sergilemedikleri şeklinde yorumlanabilir. Grup düzeyindeki istatistikler incelendikten sonra bireysel istatistikler incelenmiştir. Yapılan incelemede, her grup için meydana gelen olası 350 etkileşimden kontrol grubunda istatistiksel olarak anlamlı yedi tane etkileşim bulunurken deney grubunda istatistiksel olarak anlamlı bir tane etkileşim yer almaktadır. Yani Matematik puanlayıcı grubunda toplam sekiz puanlayıcı farklılaşan puanlayıcı katılığı ve cömertliği davranışı sergilemiştir. Bir başka deyişle sekiz puanlayıcıda yanlılık davranışı vardır. Grupların etkileşimlerine ait istatistiklerin tamamı Ek 12’de yer alırken etkileşimlerin tamamını sunmak mümkün olmadığı için sadece istatistiksel olarak anlamlı olan etkileşimler tablo 48’de verilmiştir.

Tablo 48

Matematik Puanlayıcı Grubu Son Test Sonuçlarına Ait Anlamlı Puanlayıcı x Birey Etkileşimleri

Grup	Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Bias (logit)	Standart Hata	t
Deney	P1	25	12	16,35	-1,63	0,69	-2,35
Kontrol	P3	8	16	20,39	-1,29	0,56	-2,32
Kontrol	P3	21	12	16,26	-1,60	0,69	-2,31
Kontrol	P11	2	12	15,59	-1,39	0,69	-2,01
Kontrol	P11	32	19	15,23	1,12	0,54	2,08
Kontrol	P12	10	14	10,98	1,37	0,61	2,25
Kontrol	P14	8	19	14,41	1,40	0,54	2,60
Kontrol	P19	29	19	12,20	2,29	0,54	4,28

χ^2 (sd)=366,6 (700), $p > 0,01$

Matematik puanlayıcı grubu son test sonuçlarına ait puanlayıcı ve birey etkileşiminin yer aldığı tablo 48 incelendiğinde, sekiz puanlayıcıda farklılaşan puanlayıcı katılığı ve cömertliği davranışı olduğu görülmektedir.

Deney grubunda yer alan puanlayıcılardan P1 numaralı puanlayıcıda ve kontrol grubunda yer alan P3 ve P11 numaralı puanlayıcılarda farklılaşan puanlayıcı katılığı davranışı olduğu görülmektedir. Kontrol grubunda yer alan P11, P12, P14 ve P19 numaralı puanlayıcılarda farklılaşan puanlayıcı cömertliği davranışı olduğu görülmektedir. Deney grubunda yer alan P1 numaralı puanlayıcının 25 numaralı bireye olması gerekenden daha az puan verdiği yani farklılaşan puanlayıcı katılığı davranışı sergilediği görülmektedir. Kontrol grubunda yer alan P3 numaralı puanlayıcının 8 ve 21 numaralı bireylere ve P11 numaralı puanlayıcının 2 numaralı bireye olması gerekenden daha az puan verdiği yani farklılaşan puanlayıcı katılığı davranışı sergiledikleri görülmektedir. Kontrol grubunda yer alan P12 numaralı puanlayıcının 10 numaralı bireye ve P14 numaralı puanlayıcının da 8 numaralı bireye olması gerekenden daha fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışını sergiledikleri görülmektedir. Benzer şekilde kontrol grubunda yer alan P11 numaralı puanlayıcının 32 numaralı bireye ve P19 numaralı puanlayıcının 29 numaralı bireye olması gerekenden daha fazla puan verdiği yani farklılaşan puanlayıcı cömertliği davranışını sergiledikleri görülmektedir. Bir başka deyişle deney grubunda bir puanlayıcıda (P1) farklılaşan puanlayıcı katılığı davranışı yer alırken farklılaşan puanlayıcı cömertliği davranışı sergileyen puanlayıcı bulunmamaktadır. Kontrol grubunda iki puanlayıcıda (P3, P11) farklılaşan puanlayıcı katılığı davranışı yer alırken dört puanlayıcıda (P11, P12, P14 ve P19) farklılaşan puanlayıcı cömertliği davranışı yer almaktadır. Ayrıca kontrol grubunda yer

alan P11 numaralı puanlayıcıda hem farklılaşan puanlayıcı katılığı hem de farklılaşan puanlayıcı cömertliği davranışlarının yer aldığı görülmektedir.

Araştırma sonuçlarını destekleyen çalışmalar

Bu bölümde üçüncü araştırma sorusu kapsamında elde edilen bulgulara ait sonuçları destekleyen araştırmalar raporlanmıştır. Ancak ikinci araştırma sorusunda olduğu gibi her bir davranış ayrı başlıklar altında sunulmamıştır. Çünkü, puanlayıcı davranışları (katılık, cömertlik vs.) puanlayıcıdan puanlayıcıya değişiklik gösteren, değişmez bir özellik olduğundan puanlayıcı eğitiminin puanlayıcı etkisini etkileyip etkilemeyeceği halen belirsizdir (Barrett, 2001). Literatürde puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelenmiş ancak az sayıda çalışma puanlayıcı eğitiminde ön test son test yöntemini kullanmıştır. Ayrıca puanlayıcı eğitiminde ön test son test yöntemini kullanan araştırmacılar daha çok katılık, cömertlik ve yanlılık davranışlarına ve tutarlılığa odaklanırken halo etkisi, merkeze eğilim davranışlarına odaklanan araştırma sayısı oldukça azdır. Bu yüzden araştırma sonucunu destekleyen araştırmalar sırasıyla bir başlık altında sunulmuştur.

Borman (1975) halo etkisini azaltmak için tasarlamış olduğu eğitimin puanlayıcıların halo etkisi davranışlarına etkisini incelemiştir. Araştırma sonucunda puanlayıcı eğitiminin halo etkisini azalttığı ancak puanlamaların güvenilirliğinin daha düşük olduğunu tespit etmiştir.

Pulakos (1984), eğitim stratejilerinden biri olan puanlayıcı hatası eğitiminin puanlayıcıların halo etkisini azaltmada etkili olduğunu tespit etmiştir.

Sulsky & Day (1992) referans çerçevesi eğitiminin performans değerlendirme sürecine etkisini araştırmıştır. Araştırmada 93 üniversite öğrencisi üç bireyi üç boyutta değerlendirilmiştir. Araştırma, referans çerçevesi eğitimi alan grubun daha doğru değerlendirme ve sınıflama yapabildiğini göstermiştir.

Woehr & Huffcutt (1994), Athey & McIntyre (1987), Woehr (1994), Gorman & Rentsch (2009) yapmış oldukları çalışmalarda referans çerçevesi eğitiminin performans değerlendirme sürecinde verilen puanların doğruluğunu artırmada etkili olduğu ifade edilmiştir.

Weigle (1998), puanlayıcı deneyiminin puanlayıcı eğitimi öncesi ve sonrası puanlayıcı davranışlarına etkisini araştırmıştır. Araştırmada sekiz deneyimli sekiz deneyimsiz olmak üzere 16 puanlayıcı 30 makaleyi değerlendirmiştir. Araştırma sonucunda deneyimsiz

puanlayıcıların deneyimli puanlayıcılara göre daha az tutarlı oldukları ve daha katı davrandıkları belirlenmiştir. Puanlayıcı eğitiminden sonra ise deneyimli ve deneyimsiz puanlayıcılar arasındaki farklar azalmış ve tutarlılıkları artmıştır. Ancak farklı deneyimlere sahip puanlayıcılar arasında katılık ve cömertlik arasındaki farklılıklar tamamen ortadan kalkmamıştır. Weigle (1998), bu durumu puanlayıcı eğitiminin puanlayıcıların benzer puanları vermelerinden ziyade puanlayıcıların daha öngörülebilir puanlar vermelerini sağladığı belirtmiştir. Bir başka deyişle puanlayıcı eğitimi puanlayıcılar arası güvenilirlikten ziyade puanlayıcı içi güvenilirliğe daha fazla katkı sağlamaktadır.

May (2008) yapmış olduğu çalışmada akran değerlendirmelerinde sosyal tarz yanlılığının azaltılmasında puanlayıcı eğitiminin etkisini araştırmıştır. Araştırma sonucunda puanlayıcı eğitiminin öğrencilerin yanlılık davranışında azalmayı sağladığı belirtilmiştir.

Tajeddin & Alemi (2014) yapmış oldukları çalışmada puanlayıcı eğitiminin puanlama doğruluğunu artırdığını ancak bazı puanlayıcılarda yanlılık davranışına sebep olduğunu belirtmiştir.

İlhan (2015), açık uçlu matematik sorularının değerlendirilmesi sürecinde standart ve SOLO taksonomisine dayalı hazırlanan rubriklerin puanlayıcı davranışlarına etkisini incelemiştir. Araştırmada yedi puanlayıcı 104 öğrencinin matematik açık uçlu sorularına vermiş oldukları cevapları iki farklı taksonomiye göre hazırlanan rubriğe göre değerlendirmiştir. Araştırma sonucunda standart rubriklerle yapılan puanlamalarda puanlayıcı davranışları oldukça fazlayken (halo etkisi, tutarsızlık, yanlılık) SOLO taksonomisine dayalı hazırlanmış rubriklerde puanlayıcı davranışlarının azaldığı belirlenmiştir.

Attali (2016), deneyimli puanlayıcılar ile puanlayıcı eğitimi almış puanlayıcıların davranışlarını incelemiştir. Araştırma sonucunda puanlayıcı eğitimi alan grubun deneyimli puanlayıcılara benzer doğrulukta puanlama yapabildiklerini göstermiştir. Bu durum puanlayıcı eğitiminin puanlama kalitesini artırdığını göstermektedir.

Idris & Zakaria (2016) puanlayıcı eğitiminin akran değerlendirme becerilerine etkisini araştırmıştır. Karma desen olarak tasarlanan araştırmada nicel boyutunda rasch modeli kullanılırken nitel boyutunda nicel sonuçları desteklemek amacıyla yarı yapılandırılmış görüşme kullanılmıştır. Araştırmada puanlayıcı eğitiminin değerlendirmeyi daha tutarlı ve doğru hale getirdiği tespit edilmiştir.

Bijani & Khabiri (2017), sözlü performans değerlendirme sürecinde puanlayıcı eğitiminin etkisini incelemiştir. Araştırmaya 200 bireyin sözlü performansını değerlendiren 20

puanlayıcı dahil edilmiştir. Araştırmada puanlayıcı eğitimi öncesinde ve sonrasında puanlayıcılardan elde edilen puanlar çok yüzeyli rasch modeli ile incelenmiştir. Araştırma sonuçları, puanlayıcı eğitiminin puanlayıcıların yanlılık davranışlarında azalma, tutarlılığında ise artış sağladığını göstermiştir. Araştırmacılar puanlayıcı eğitiminin puanlayıcı değişkenliğini tamamen ortadan kaldıramayacağını ve deneyimsiz puanlayıcıların deneyimli puanlayıcılara göre puanlayıcı eğitiminden daha fazla faydalandıklarını göstermiştir.

Bijani (2018), sözlü değerlendirme sürecinde puanlayıcı etkisini belirlemeyi ve bu etkileri azaltmaya yönelik puanlayıcı eğitiminin etkisini gözlemlemeyi amaçlamıştır. Araştırmaya 300 öğrenci ve 21 puanlayıcı dahil edilmiş ve araştırma sonucunda puanlayıcı eğitiminin puanlayıcıların katılık ve yanlılık gibi davranışlarının azalmasına katkı sağladığı belirtilmiştir. Ayrıca puanlayıcı eğitiminin puanlayıcıların tutarlılıklarının artmasına sebep olduğunu da belirlemiştir.

Knoch, Fairbairn, Myford, & Huisman (2018) puanlayıcı eğitiminin çevrim içi ve yüz yüze olması durumunun puanlayıcı davranışlarına etkisini araştırdıkları çalışma, 24 puanlayıcı, 40 öğrenci performansı ve dört görev üzerinden yürütülmüştür. Araştırmacılar, deneyimsiz puanlayıcıları araştırmaya dahil etmiş ve puanlayıcılardan eğitim öncesinde hazırlanan performans görevlerini değerlendirmeleri istemiştir. Puanlayıcı eğitimi öncesinde grupların puanlayıcı davranışları arasında fark olmadığı tespit edildikten sonra puanlayıcılar iki gruba ayrılmıştır. Birinci gruba 11 puanlayıcı seçilmiş ve bu gruba çevrim içi puanlayıcı eğitimi verilmiştir. İkinci gruba ise 13 puanlayıcı seçilmiş ve bu gruba da yüz yüze puanlayıcı eğitimi verilmiştir. Araştırma sonucunda her iki puanlayıcı grubunun da istatistiksel olarak benzer olduklarını, puanlayıcıların katılık ve cömertlik durumlarının benzer olduğunu tespit etmişlerdir.

Ahmadi (2019), tartışma temelli puanlayıcı eğitiminin, puanlayıcıların daha hızlı ve güvenli bir şekilde değerlendirme yapmasını sağladığını ve puanlayıcıların katılık, cömertlik, tutarlılık ve merkeze eğilim davranışlarını da daha iyi hale getirdiğini tespit etmiştir.

Kang, Rubin, & Kermad (2019), deneyimsiz puanlayıcıların ikinci dil olarak konuşulan İngilizce performanslarının değerlendirilmesine puanlayıcı geçmişi, tutum ve puanlayıcı eğitiminin puanlayıcı davranışlarına etkisini araştırmışlardır. Araştırmada 82 deneyimsiz puanlayıcı 112 bireyin TOEFL sınavındaki konuşma becerilerini değerlendirmiştir. Araştırma sonucunda puanlardaki sapmanın yaklaşık %20'sinin puanlayıcı geçmişi ve

tutumundan kaynaklandığı belirlenmiştir. Ancak puanlayıcı eğitiminin puanlayıcı geçmişi, tutum etkisini ve yanlılığı önemli ölçüde azalttığı belirlenmiştir.

Şata (2019), 45 puanlayıcının dahil olduğu araştırmada puanlayıcı eğitiminin performans değerlendirme sürecinde ortaya çıkan puanlayıcı davranışlarına etkisini araştırmıştır. Araştırma sonucunda puanlayıcı eğitimi öncesinde katılık, cömertlik ve yanlılık davranışlarının hem grup düzeyinde hem de bireysel düzeyde ortaya çıktığı, halo etkisi ve merkeze eğilim davranışlarının ise grup düzeyinde olmasa da bireysel düzeyde ortaya çıktığı belirlenmiştir. Puanlayıcı eğitimi sonrasında ise puanlayıcı davranışlarının tamamen ortadan kalkmasa da istatistiksel olarak azaldığı belirlenmiştir.

Doosti & Ahmadi Safa (2021), sözlü dil değerlendirme sürecinde puanlayıcılar arası güvenilirliğin artırılmasında puanlayıcı eğitiminin etkisini ve puanlayıcı eğitimi ile öğrencilerin beklentilerinin göz önünde bulundurulmasının öğrencilerin adil değerlendirilmeye ilişkin algıları üzerindeki etkisini araştırmışlardır. Araştırmada dört puanlayıcı ve 31 öğrencinin sözlü performansları değerlendirilmiştir. Araştırma, puanlayıcı eğitiminin puanlayıcılar arası güvenilirliği artırdığını göstermiştir. Ayrıca araştırmada öğrencilerin beklentilerinin göz önüne alınmasının öğrencilerin sınavın adil yapılmasına olan algılarını da iyileştirdiği sonucuna ulaşılmıştır.

Rethinasamy (2021), yaygın olarak kullanılan puanlayıcı eğitim modüllerinin puanlama doğruluğuna etkisini araştırdığı çalışmada puanlayıcı eğitiminin puanlama doğruluğunu artırdığı belirlenmiştir.

Bijani, Hashempour, & Said Bani Orabah (2022) puanlayıcı eğitiminin puanlayıcı etkisini tamamen ortadan kaldırmaya da puanlayıcıların katılık, cömertlik ve yanlılık davranışlarını azaltmaya katkı sağladığını belirtmiştir.

Bijani & Said Bani Orabah (2022) performans değerlendirme sürecinde puanlayıcı eğitiminin ve geri bildirimin etkisini inceledikleri çalışmaya 300 öğrenci ve 20 puanlayıcı dahil edilmiştir. Araştırma sonucunda puanlayıcı eğitiminin puanlayıcıların katılık, cömertlik ve yanlılık davranışlarında azalma sağladığını belirtmiştir. Ancak puanlayıcı eğitimi bireysel farklılığı ortadan kaldırmadığı için puanlayıcı etkisini tamamen ortadan kaldırmamıştır. Ayrıca puanlayıcı eğitiminin deneyimsiz puanlayıcılarda daha etkili olduğu ve puanlayıcı eğitiminden sonra uzun zaman geçtiğinde eğitimin etkisinin azalabileceğini belirtilmiştir.

Mohd Noh & Mohd Matore (2022), sözlü sınavları değerlendirme sürecinde puanlayıcı deneyimi, puanlayıcı eğitimi ve mesleki deneyim gibi değişkenlerin puanlayıcı davranışlarına etkisini çok yüzeyli rash modeli ile araştırmıştır. Araştırmada 164 İngilizce öğretmeni altı öğrencinin performansını kelime bilgisi, dil bilgisi ve iletişimsel yeterlik olmak üzere üç boyutta incelemiştir. Araştırmacılar, öğretmenlerin puanlama deneyimi ve mesleki deneyimin puanlamaları daha kaliteli hale getirdiğini ifade ederken puanlayıcı eğitiminin puanlama kalitesinde anlamlı katkı sağlamadığını belirtmişlerdir.



BÖLÜM VI

SONUÇLAR

Bu çalışma, açık uçlu soruların değerlendirilmesi sürecinde puanlayıcı etkisinin belirlenmesi, azaltılması ya da ortadan kaldırılması için kullanılan yöntemlerden biri olan puanlayıcı eğitiminin etkisini belirlemek için yapılmıştır. Araştırma sonucunda elde edilen sonuçlar araştırma sorularının sırasına göre verilmiştir. Ayrıca elde edilen sonuçlara göre ilerde yapılacak çalışmalar için öneriler de bu bölümde sunulmuştur.

Araştırma sonuçları sunulurken önce Türkçe puanlayıcı grubu sonuçları raporlanırken sonrasında matematik puanlayıcı grubuna ait sonuçlar raporlanmıştır

Türkçe puanlayıcı grubuna ait araştırma sonuçları sırasıyla şu şekildedir;

Araştırma kapsamında kullanılmak üzere beş tane yedinci sınıf Türkçe dersi kazanımlarını kapsayan açık uçlu soru hazırlanmıştır. Hazırlanan soruların geçerlik ve güvenirlik çalışmaları yapılmış ve öğrencilere uygulanabileceği belirlenmiştir. Sonrasında açık uçlu soruların değerlendirilmesinde kullanılmak üzere bütünsel dereceli puanlama anahtarları geliştirilmiştir. Geliştirilen bütünsel dereceli puanlama anahtarlarının geçerli ve güvenilir olduğu ve açık uçlu soruların değerlendirilmesinde kullanılabileceği belirlenmiştir. Sonuç olarak öğrencilerin Türkçe dersi performanslarının değerlendirilmesinde kullanılacak beş açık uçlu soru ve geçerli ve güvenilir bütünsel dereceli puanlama ölçekleri geliştirilmiştir.

Araştırma için hazırlanan açık uçlu sorular öğrencilere dağıtılmış ve öğrencilerin cevapları hazırlanan bütünsel dereceli puanlama anahtarları ile araştırma kapsamında değerlendirilmek üzere çoğaltılmıştır.

Araştırmada birey, puanlayıcı ve madde olmak üzere üç yüzey bulunmaktadır ve bu üç yüzeye ait model ölçümleri ve uyum istatistikleri çok yüzeyli rasch modeli kapsamında incelenmiştir. Yapılan inceleme sonucunda model veri uyumunu bozan beş birey, iki

puanlayıcı araştırmadan çıkarılmıştır. Tüm yüzeylerin model veri uyumunu sağladığı tespit edildikten sonra puanlayıcı eğitimi öncesinde Türkçe dersi açık uçlu soruların değerlendirilmesinde meydana gelen puanlayıcı davranışları (katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği) incelenmiştir. Puanlayıcı davranışları incelenirken önce grup düzeyindeki istatistikler incelenirken sonrasında bireysel istatistikler incelenmiştir.

Puanlayıcı eğitimi öncesinde Türkçe açık uçlu soruların değerlendirilmesi sürecinde ön test ölçümlerine ait katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği davranışları incelenmiştir. İncelemede grupların istatistiksel benzer oldukları belirlenmiştir. Katılık ve cömertlik davranışı Türkçe puanlayıcı ön test grubunda hem grup düzeyinde hem de bireysel düzeyde tespit edilmiştir. Merkeze eğilim, halo etkisi ve yanlılık davranışlarının ise grup düzeyinde ortaya çıkmadığı ancak bireysel düzeyde ortaya çıktığı belirlenmiştir. Katılık ve cömertlik davranışı deney grubunda üç puanlayıcıda (%42,86) tespit edilirken kontrol grubunda iki puanlayıcıda (%28,57) belirlenmiştir. Merkeze eğilim davranışı deney grubunda üç puanlayıcıda (%42,86) tespit edilirken kontrol grubunda dört puanlayıcıda (%57,14) tespit edilmiştir. Halo etkisi deney grubunda dört puanlayıcıda (%57,14) tespit edilirken kontrol grubunda bir puanlayıcıda (%14,29) tespit edilmiştir. Yanlılık davranışı ise deney grubunda iki puanlayıcıda (%28,57) tespit edilirken kontrol grubunda bir puanlayıcıda (%14,29) tespit edilmiştir.

Ortaya çıkan puanlayıcı davranışlarını ortadan kaldırmak ya da azaltmak için Türkçe puanlayıcı grubuna puanlayıcı eğitimi verilmiştir. Puanlayıcı eğitimi içeriği, puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi modellerinin harmanlanmasıyla oluşturulmuştur. Oluşturulan puanlayıcı eğitimi puanlayıcılara haftada bir saat olmak üzere toplam dört saat verilmiştir. Türkçe açık uçlu soruların değerlendirilmesi sürecinde son test ölçümlerine ait katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği davranışları incelenmiştir. Puanlayıcı eğitimi sonrasında puanlayıcı davranışları tamamen ortadan kalmasa da puanlayıcı davranışlarında azalma olduğu görülmüştür. Bu durum puanlayıcı kaymasından meydana gelebilmektedir (Moore, 2009, s. 11). Katılık ve cömertlik davranışı Türkçe puanlayıcı son test grubunda hem grup düzeyinde hem de bireysel düzeyde tespit edilmiştir. Merkeze eğilim, halo etkisi ve yanlılık davranışlarının ise grup düzeyinde ortaya çıkmadığı ancak bireysel düzeyde ortaya çıktığı belirlenmiştir. Katılık ve cömertlik davranışı deney grubunda üçten bire (%14,29), merkeze eğilim davranışı deney grubunda üçten ikiye (%28,57), halo etkisi deney grubunda dörtten üçe (%42,86) ve yanlılık davranışı

deney grubunda ikiden bire (%14,29) düşmüştür. Ayrıca Türkçe puanlayıcı deney grubu ön test ve son test puanları arasında istatistiksel olarak da anlamlı fark bulunmuştur. Kontrol grubunda ise katılık ve cömertlik davranışı ikiden bire düşerken merkeze eğilim davranışı, halo etkisi ve yanlılık davranışlarında artış olmuştur. Ancak Türkçe puanlayıcı kontrol grubu ön test ve son test puanları arasında istatistiksel olarak anlamlı fark bulunmamıştır.

Puanlayıcı eğitimi öncesinde Türkçe puanlayıcı deney grubu ön test sonuçları ile Türkçe puanlayıcı kontrol grubu ön test sonuçları arasında istatistiksel olarak anlamlı fark bulunmamaktadır. Ancak puanlayıcı eğitimi sonrasında Türkçe puanlayıcı deney grubu son test sonuçları ile Türkçe puanlayıcı kontrol grubu son test sonuçları arasında ve Türkçe puanlayıcı deney grubu son test sonuçları ile Türkçe puanlayıcı deney grubu ön test sonuçları arasında istatistiksel olarak anlamlı fark bulunmuştur.

Sonuç olarak puanlayıcı eğitiminin Türkçe açık uçlu soruların puanlanmasında puanlayıcıların katılık, cömertlik, merkeze eğilim, halo etkisi ve yanlılık davranışlarında istatistiksel olarak anlamlı bir etkiye sahip olduğu bulunmuştur.

Son olarak puanlayıcıların katılık ve cömertlik davranışlarındaki değişikliğin %60,2'sinin puanlayıcı eğitimi ile açıklanabileceği bulunmuştur.

Matematik puanlayıcı grubuna ait araştırma sonuçları sırasıyla şu şekildedir;

Araştırma kapsamında kullanılmak üzere altı tane yedinci sınıf matematik dersi kazanımlarını kapsayan açık uçlu soru hazırlanmıştır. Hazırlanan soruların geçerlik ve güvenirlik çalışmaları yapılmış ve öğrencilere uygulanabileceği belirlenmiştir. Sonrasında açık uçlu soruların değerlendirilmesinde kullanılmak üzere bütünsel dereceli puanlama anahtarları geliştirilmiştir. Geliştirilen bütünsel dereceli puanlama anahtarlarının geçerli ve güvenilir olduğu ve açık uçlu soruların değerlendirilmesinde kullanılabileceği belirlenmiştir. Sonuç olarak öğrencilerin matematik dersi performanslarının değerlendirilmesinde kullanılacak altı açık uçlu soru ve geçerli ve güvenilir bütünsel dereceli puanlama ölçekleri geliştirilmiştir.

Araştırma için hazırlanan açık uçlu sorular öğrencilere dağıtılmış ve öğrencilerin cevapları hazırlanan bütünsel dereceli puanlama anahtarları ile araştırma kapsamında değerlendirilmek üzere çoğaltılmıştır.

Araştırmada birey, puanlayıcı ve madde olmak üzere üç yüzey bulunmaktadır ve bu üç yüzeye ait model ölçümleri ve uyum istatistikleri çok yüzeyli Rasch modeli kapsamında incelenmiştir. Yapılan inceleme sonucunda model veri uyumunu bozan beş birey, bir

puanlayıcı araştırmadan çıkarılmıştır. Tüm yüzeylerin model veri uyumunu sağladığı tespit edildikten sonra puanlayıcı eğitimi öncesinde matematik dersi açık uçlu soruların değerlendirilmesinde meydana gelen puanlayıcı davranışları (katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği) incelenmiştir. Puanlayıcı davranışları incelenirken önce grup düzeyindeki istatistikler incelenirken sonrasında bireysel istatistikler incelenmiştir.

Puanlayıcı eğitimi öncesinde matematik açık uçlu soruların değerlendirilmesi sürecinde ön test ölçümlerine ait katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği davranışları incelenmiştir. İncelemede grupların istatistiksel benzer oldukları belirlenmiştir. Katılık ve cömertlik davranışı matematik puanlayıcı ön test grubunda hem grup düzeyinde hem de bireysel düzeyde tespit edilmiştir. Merkeze eğilim, halo etkisi ve yanlışlık davranışlarının ise grup düzeyinde ortaya çıkmadığı ancak bireysel düzeyde ortaya çıktığı belirlenmiştir. Katılık ve cömertlik davranışı deney grubunda yedi puanlayıcıda (%70) tespit edilirken kontrol grubunda iki puanlayıcıda (%20) belirlenmiştir. Merkeze eğilim davranışı deney grubunda bir puanlayıcıda (%10) tespit edilirken kontrol grubunda bir puanlayıcıda (%10) tespit edilmiştir. Halo etkisi deney grubunda üç puanlayıcıda (%30) tespit edilirken kontrol grubunda bir puanlayıcıda (%10) tespit edilmiştir. Yanlışlık davranışı ise deney grubunda iki puanlayıcıda (%20) tespit edilirken kontrol grubunda bir puanlayıcıda (%10) tespit edilmiştir.

Ortaya çıkan puanlayıcı davranışlarını ortadan kaldırmak ya da azaltmak için matematik puanlayıcı grubuna puanlayıcı eğitimi verilmiştir. Puanlayıcı eğitimi içeriği, puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi modellerinin harmanlanmasıyla oluşturulmuştur. Oluşturulan puanlayıcı eğitimi puanlayıcılara haftada bir saat olmak üzere toplam dört saat verilmiştir. Matematik açık uçlu soruların değerlendirilmesi sürecinde son test ölçümlerine ait katılık, cömertlik, merkeze eğilim, halo etkisi, farklılaşan puanlayıcı katılığı ve cömertliği davranışları incelenmiştir. Puanlayıcı eğitimi sonrasında katılık, cömertlik, merkeze eğilim, halo etkisi davranışları tamamen ortadan kalkarken yanlışlık davranışında da azalma olduğu görülmüştür. Puanlayıcı eğitimi sonrasında puanlayıcı davranışlarının tamamen ortadan kalkmaması puanlayıcı kaymasından kaynaklanabilmektedir (Moore, 2009, s. 11). Katılık ve cömertlik davranışı matematik puanlayıcı son test grubunda sadece grup düzeyinde tespit edilmiştir. Merkeze eğilim ve halo etkisi davranışlarının ise hem grup düzeyinde hem de bireysel düzeyde olmadığı tespit edilmiştir. Yanlışlık davranışının ise grup düzeyinde yer almadığı ancak bireysel düzeyde bir puanlayıcıda (%10) olduğu tespit edilmiştir. Katılık ve

cömertlik davranışı deney grubunda yediden sıfıra, merkeze eğilim davranışı deney grubunda birden sıfıra, halo etkisi deney grubunda üçten sıfıra ve yanlılık davranışı deney grubunda ikiden bire (%10) düşmüştür. Ayrıca matematik puanlayıcı deney grubu ön test ve son test puanları arasında istatistiksel olarak da anlamlı fark bulunmuştur. Kontrol grubunda ise katılık ve cömertlik davranışı ikiden bire (%10) düşerken merkeze eğilim davranışı, halo etkisi ve yanlılık davranışlarında artış olmuştur. Ancak matematik puanlayıcı kontrol grubu ön test ve son test puanları arasında istatistiksel olarak anlamlı fark bulunmamıştır.

Puanlayıcı eğitimi öncesinde matematik puanlayıcı deney grubu ön test sonuçları ile matematik puanlayıcı kontrol grubu ön test sonuçları arasında istatistiksel olarak anlamlı fark bulunmamaktadır. Ancak puanlayıcı eğitimi sonrasında matematik puanlayıcı deney grubu son test sonuçları ile matematik puanlayıcı kontrol grubu son test sonuçları arasında ve matematik puanlayıcı deney grubu son test sonuçları ile matematik puanlayıcı deney grubu ön test sonuçları arasında istatistiksel olarak anlamlı fark bulunmuştur.

Sonuç olarak puanlayıcı eğitiminin matematik açık uçlu soruların puanlanmasında puanlayıcıların katılık, cömertlik, merkeze eğilim, halo etkisi ve yanlılık davranışlarında istatistiksel olarak anlamlı bir etkiye sahip olduğu bulunmuştur.

Son olarak puanlayıcıların katılık ve cömertlik davranışlarındaki değişikliğin %57,8'inin puanlayıcı eğitimi ile açıklanabileceği bulunmuştur.

Hem Türkçe hem de matematik kontrol puanlayıcı gruplarında puanlayıcı etkisinde artış meydana gelmiştir. Kontrol grubundaki puanlayıcı etkilerindeki artışın, puanlayıcı eğitiminin önemini ortaya koyduğu söylenebilir. Çünkü puanlayıcı eğitimi almayan puanlayıcılar puanlama sürecinde çevresel faktörlerden etkilenecek puanlama yapmış olabilirler. Bu durum sonucunda puanlayıcı etkisi artmış olabilir. Ayrıca kontrol grubunda meydana gelen değişiklikler John Henry etkisi ile de açıklanabilir. John Henry etkisi; kontrol grubundaki bireylerin deney grubuna karşı rekabete girmesi ve yüksek performans gösterme eğiliminde olmasıdır (Heinich, 1973; Saretsky, 1972). Deney grubu ile rekabet eğiliminde olan puanlayıcılarda puanlayıcı etkileri artmış olabilir. Bir başka deyişle deneye tabi olduklarının farkına varan puanlayıcılar daha ciddi bir performans göstermeye çalışırken daha fazla hata yapmış olabilirler.

Puanlayıcı eğitiminin matematik puanlayıcı grubunda Türkçe puanlayıcı grubuna göre istatistiksel olarak daha etkili olduğu bulunmuştur. Türkçe dersindeki performansların soyut olması, puanlayıcı eğitiminin etkisini azaltırken matematik dersindeki performansların

somut olması, puanlayıcı eğitiminin daha etkili olmasına sebep olmuş olabilir. Ayrıca her ne kadar puanlayıcı eğitimi verilse de Türkçe puanlayıcı grubundaki puanlayıcıların öğrencilerin yazı güzelliği, imla kuralları gibi farklı becerilere odaklandığı da görülmüştür. Bu durum Türkçe puanlayıcı grubunda puanlayıcı eğitiminin etkisini azaltmış olabilir. Son olarak her iki puanlayıcı grubu için de puanlayıcı eğitiminin veriliş zamanı, eğitimi ortamı, puanlayıcı eğitiminin tasarımı ve araştırmacı kaynaklı eksiklikler gibi değişkenler de puanlayıcı eğitiminin etkisini azaltmış olabilir.

Öneriler

Bu araştırma, Türkçe ve matematik dersleri kazanımlarını kapsayan açık uçlu soruların değerlendirilmesi sürecinde ortaya çıkan puanlayıcı davranışlarının incelenmesi ve puanlayıcı eğitiminin bu davranışlara etkisini belirlemek amacıyla yapılmıştır. Araştırmada elde edilen sonuçlar doğrultusunda hem araştırmacılara hem de uygulamacılara öneriler sunulmuştur. Öneriler iki farklı başlıkta sırasıyla sunulmuştur.

Araştırmacılara Öneriler

Araştırma kapsamında puanlayıcı eğitiminin matematik puanlayıcı grubunda Türkçe puanlayıcı grubuna göre daha etkili olduğu görülmüştür. Bu durumdan hareketle matematik puanlayıcı grubu için verilen puanlayıcı eğitimi süresinin yeterli olduğu ancak Türkçe puanlayıcı eğitim süresinin yetersiz olduğu söylenebilir. Türkçe puanlayıcı grubu için eğitim süresinin arttırılması durumunda puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelenebilir.

Araştırmada referans çerçevesi ve puanlayıcı hatası eğitim stratejilerinin Türkçe ve matematik puanlayıcı grubunun puanlayıcı davranışlarına etkisi incelenmiştir. Yapılan araştırma sonucunda oluşturulan eğitim deseninin matematik puanlayıcı grubunda daha etkili olduğu belirlenmiştir. Türkçe puanlayıcı grubu için farklı eğitim desenleri ile benzer araştırmalar yapılabilir.

Araştırmada farklı derslere (Türkçe ve matematik) aynı puanlayıcı eğitim stratejileri (puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi) kullanılarak puanlayıcı eğitimi verilmiştir. Yapılacak çalışmalarda aynı derse farklı puanlayıcı eğitim stratejileri kullanılarak puanlayıcı eğitiminin puanlayıcı davranışlarına etkisi incelenebilir.

Araştırmada puanlayıcı eğitimi sonrasında hem Türkçe hem de matematik puanlayıcı grubunda bazı davranışların azalsa da ortadan kalkmadığı görülmüştür. Bu durumun sebebinin incelenmesi ve puanlayıcı eğitiminin etkisinin artırılması için farklı davranışlar sergileyen puanlayıcılarla ek olarak görüşmeler yapılabilir. Araştırma sonuçları puanlayıcı görüşleriyle desteklenebilir.

Araştırmada puanlayıcı eğitiminin hem Türkçe hem de matematik puanlayıcı grubunda puanlayıcı davranışlarının azaltılmasında etkili olduğu bulunmuştur. Puanlayıcı eğitiminin etkisinin ne kadar süreceği ile ilgili araştırmalar yapılabilir.

Yapılan araştırmada puanlayıcı eğitimi hem Türkçe hem de matematik puanlayıcı grubunda toplam dört saat verilmiştir. Puanlayıcı eğitiminin daha uzun olması halinde puanlayıcı davranışlarına etkisi incelenebilir.

Araştırmada puanlayıcı eğitiminin Türkçe ve matematik puanlayıcı grubunun puanlayıcı davranışlarına etkisi incelenmiştir. Yapılacak çalışmalarda puanlayıcı eğitiminin farklı derslerdeki (resim, müzik, beden eğitimi, fen bilimleri, vs.) puanlayıcı davranışlarına etkisi incelenebilir.

Puanlayıcı eğitiminin Türkçe ve matematik puanlayıcı grubunun puanlayıcı katılımı, puanlayıcı cömertliği, merkeze eğilim, halo etkisi ve yanlılık davranışlarına etkisi incelenmiştir. Yapılacak çalışmalarda farklı puanlayıcı davranışları da incelenebilir.

Mevcut çalışmada puanlayıcı eğitimi küçük gruplara (Türkçe, n=7; matematik, n=10) uygulanmıştır. Yapılacak çalışmalarda daha büyük gruplarla çalışılıp puanlayıcı eğitiminin etkisi incelenebilir.

Çalışmada Türkçe ve matematik puanlayıcı grubuna puanlayıcı eğitim stratejilerinden puanlayıcı hatası eğitimi ve referans çerçevesi eğitimi verilmiştir. Farklı eğitim stratejileri kullanılarak puanlayıcı davranışlarına etkisi incelenebilir.

Kısa vadede sınıflarda ölçme ve değerlendirme süreçlerinde aktif kullanılması mümkün görünmese de otomatik puanlama süreçlerine yapay zekâ uygulamalarının entegrasyonu söz konusu olabilir. Puanlayıcıların doğru bir şekilde yapmış oldukları puanlamaların yapay zekâ öğrenme süreçlerine katkı sağlayacağı düşünülmektedir. Puanlayıcı eğitimi sırasında beklenen davranışları sergileyen puanlamalar yapay zekâ sistemlerine eklenebilir.

Uygulamacılara Öneriler

MEB'in tüm sınavlarda açık uçlu soruların kullanılmasına yönelik kararı göz önüne alındığında puanlayıcı davranışlarının önemi daha da artmıştır. Puanlayıcı eğitiminin puanlayıcı davranışlarını azalttığı düşünüldüğünde açık uçlu soruların değerlendirilmesinde puanlayıcılara puanlayıcı eğitimi verilmesi gerektiği düşünülmektedir.

Farklı öğrenim düzeylerinde ders geçme ve kalma, işe alım gibi önemli kararların verildiği uygulamalarda puanlayıcılara puanlayıcı eğitimi verilerek sonuçların daha geçerli ve güvenilir olması sağlanabilir.



KAYNAKLAR

- A'laudden, S. W. A. W., Hanafi, N. M., & Rahman, A. A. (2023). Central tendency effects using multi-facet Rasch measurement (MFRM) among raters in assessing teaching training. *AIP Conference Proceedings*, 2582(1).
- Abedi, J. (2006). *Language issues in item development*. New Jersey: Lavrence Erlbaum.
- Abu Kassim, N. L. (2007). *Exploring rater judging behaviour using the many-facet Rasch model*. Paper presented at the Paper Presented in the Second Biennial International Conference on Teaching and Learning of English in Asia:Exploring New Frontiers (TELiA2), Universiti Utara, Malaysia. <https://repo.uum.edu.my/id/eprint/3212/>
- Abu Kassim, N. L. (2011). Judging behaviour and rater errors: an application of the many-facet Rasch model. *GEMA Online Journal of Language Studies*, 3, 179-197.
- Aguinis, H. (2012). Implementing a performance management system. In *Performance management* (3 ed., s. 168-194). New Jersey: Pearson-Prentice Hall.
- Aguinis, H., Mazurkiewicz, M. D., & Heggstad, E. D. (2009). Using web-based frame-of-reference training to decrease biases in personality-based job analysis: An experimental field study. *Personnel Psychology*, 62(2), 405-438.
- Ahmadi, A. (2019). A study of raters' behavior in scoring l2 speaking performance: Using rater discussion as a training tool. *Issues in Language Teaching*, 8(1), 195-224.
- Akbari, K., Eigruher, M., & Vetschera, R. (2024). Risk attitudes: The central tendency bias. *EURO Journal on Decision Processes*, 12, 100042.
- Akın, Ö., & Baştürk, R. (2012). Keman eğitiminde temel becerilerin rasch ölçme modeli ile değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 31(31), 175-187.

- Aktaş, M., & Alıcı, D. (2017). Kontrol listesi, analitik rubrik ve dereceleme ölçeklerinde puanlayıcı güvenilirliğinin genellenebilirlik kuramına göre incelenmesi. *International Journal of Eurasia Social Sciences/Uluslararası Avrasya Sosyal Bilimler Dergisi*, 8(29), 4-25.
- Alderson, J. C., Clapham, C., & Wall, D. (2005). *Language test construction and evaluation* (9 ed.): Cambridge: Cambridge University.
- Alkan, M. (2013). *PISA 2009 okuma becerileri açık uçlu sorularının puanlanmasında genellenebilirlik kuramındaki farklı desenlerin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, 23(6), 1825-1831.
- Altenhöner, T., Köhler, M., Philippi, M., & Alaze, F. (2014). Maßnahmen des betrieblichen Gesundheitsmanagements. *Präv Gesundheitsf*, 9(1), 3-9.
- Anadolu Üniversitesi. (2017). Anadolu Üniversitesi açıköğretim, iktisat ve işletme fakülteleri eğitim-öğretim ve sınav yönetmeliği. <https://www.anadolu.edu.tr/acikogretim/yonetmelikler-ve-esaslar-yonergeler/yonetmelikler/anadolu-universitesi-acikogretim-iktisat-ve-isletme-fakulteleri-egitim-ogretim-ve-sinav-yonetmeligi>, sayfasından erişilmiştir.
- Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives : complete edition*. New York: Addison Wesley Longman.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Antonioni, D., & Woehr, D. J. (2001). *Improving the quality of multisource rater performance*. San Francisco: Jossey-Bass.
- Ariffin, S. R. (2008). *Inovasi dalam pengukuran dan penilaian pendidikan*. Bangi: Universiti Kebangsaan Malaysia.

- Arslan Mancar, S. (2019). *Performansa dayalı durum belirlemede puanlayıcılar arası güvenilirlik tekniklerinin karşılaştırılması*. Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Oaks, California: Corwin.
- Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology*, 72(4), 567-572.
- Atılğan, H. (2005a). Genellenebilirlik kurami ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. *Journal of Educational Sciences & Practices*, 4(7), 14-25.
- Atılğan, H. (2005b). Müzik öğretmenliği özel yetenek seçme sınavının çok yüzeyli Rasch modeli ile analizi (İnönü Üniversitesi örneği). *Eurasian Journal of Educational Measurement* 20, 62-73.
- Attali, Y. (2016). A comparison of newly-trained and experienced raters on a standardized writing assessment. *Language Testing*, 33(1), 99-115.
- Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a Generic Approach in Automated Essay Scoring. *The Journal of Technology, Learning and Assessment*, 10(3), 4-32.
- Aydoğmuş, M. A. (2021). *Puanlayıcı deneyimlerine göre puanlayıcılar arası güvenilirliklerin farklı yöntemlerle incelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Ayoobiyan, H., & Ahmadi, A. (2023). Detecting halo effects across rubric criteria in L2 writing assessment: A many-facet rasch analysis. *Applied Research on English Language*, 12(1), 159-176.
- Babcock, B. G. E. (2009). *Estimating a noncompensatory IRT model using a modified Metropolis algorithm*. (Ph.D.). University of Minnesota, United States Minnesota. <https://www.proquest.com/dissertations-theses/estimating-noncompensatory-irt-model-using/docview/304939866/se-2?accountid=11054> sayfasından erişilmiştir.
- Bachman, L. F. (2004). *Statistical analyses for language assesment*. Cambridge: Cambridge University..

- Bağcan Büyükturan, E., & Çıkrıkçı Demirtaşlı, N. (2013). Comparing the psychometric characteristics of multiple choice tests and structural communication grids. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 46(1), 395-416.
- Baharudin, H., Maskor, Z. M., & Matore, M. E. E. M. (2022). The raters' differences in Arabic writing rubrics through the Many-Facet Rasch measurement model. *Frontiers in Psychology*, 13, 4-35.
- Bahrouni, F. (2016). Using multi-facet Rasch model (MFRM) in rater-mediated assessment. *Journal of Teaching English for Specific and Academic Purposes*, 4(1), 195-212.
- Baird, J., Hayes, M., Johnson, R., Johnson, S., & Lamprianou, I. (2013). Marker effects and examination reliability: A comparative exploration from the perspectives of generalizability theory, Rasch modelling and multilevel modelling. <http://dera.ioe.ac.uk/17683/1/2013-01-21-marker-effects-and-examination-reliability.pdf> sayfasından erişilmiştir.
- Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.
- Barkaoui, K. (2008). *Effects of scoring method and rater experience on esl essay rating processes and outcomes*. Toronto: University of Toronto.
- Barrett, S. (2001). The impact of training on rater variability. *International Education Journal*, 2(1), 49-58.
- Başol, G. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyli rasch ölçme modeli ile değerlendirilmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 1(1), 51-57.
- Bennett, R. E., Ward, W. C., Rock, D. A., & LaHart, C. (1990). Toward a framework for constructed-response items. *ETS Research Report Series*, 1990(1), i-35.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in Rater Training. *Academy of Management Review*, 6(2), 205-212.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, 65(1), 60-66.

- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology*, 62(1), 64-69.
- Bijani, H. (2018). Investigating the validity of oral assessment rater training program: A mixed-methods study of raters' perceptions and attitudes before and after training. *Cogent Education*, 5(1), 1460901.
- Bijani, H., Hashempour, B., & Said Bani Orabah, S. (2022). Development and validation of a training-embedded speaking assessment rating scale: A multifaceted rasch analysis in speaking assessment. *International Journal of Research in English Education*, 7(3), 32-45.
- Bijani, H., & Khabiri, M. (2017). Investigating the effect of training on raters' bias toward test takers in oral proficiency assessment: A facets analysis. *Journal of Asia TEFL*, 14(4), 687.
- Bijani, H., & Said Bani Orabah, S. (2022). Facet variability in the light of rater training in measuring oral performance: a multifaceted rasch analysis. *Issues in Language Teaching*, 11(2), 255-290.
- Bıkmaz Bilgen, Ö. (2019). Tamamlayıcı ölçme ve değerlendirme teknikleri I: Performans değerlendirme. In N. Doğan (Ed.), *Eğitimde ölçme ve değerlendirme* (1 ed., s. 182-218). Ankara: Pegem Akademi.
- Birenbaum, M., & Tatsuoka, K. K. (1987). Open-ended versus multiple-choice response formats—it does make a difference for diagnostic purposes. *Applied Psychological Measurement*, 11(4), 385-395.
- Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences*. New York: Routledge.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, 60(5), 556-560.
- Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance*, 20(2), 238-252.
- Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Quality & Quantity*, 47(4), 2341-2360.

- Brennan, R. L. (1983). *Elements of generalizability theory*: USA: The American College Testing Program.
- Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education*, 24(1), 1-21.
- Brennan, R. L., Gao, X., & Colton, D. A. (1995). Generalizability analyses of work keys listening and writing tests. *Educational and Psychological Measurement*, 55(2), 157-176.
- Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of applied measurement*, 9(1), 45.
- Brijmohan, A. (2016). *A many-facet Rasch measurement analysis to explore rater effects and rater training in medical school admissions*. Toronto: University of Toronto.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Virginia: Alexandria, ASCD.
- Brookhart, S. M. (2014). *How to design questions and tasks to assess student thinking*: Virginia: Alexandria, ASCD.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., & Harris, M. D. (1998). *Automated scoring using a hybrid feature identification technique*. Paper presented at the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1.
- Cardy, R. L., & Keefe, T. J. (1994). Observational purpose and evaluative articulation in frame-of-reference training: The effects of alternative processing modes on rating accuracy. *Organizational Behavior and Human Decision Processes*, 57(3), 338-357.
- Cascio, W. F. (1982). *Applied psychology in personnel management*. Reston: Reston.
- Cascio, W. F., & Aguinis, H. (2014). *Applied psychology in human resource management*. Saint Paul: Pearson.

- Cellar, D. F., Curtis, J. R., Kohlepp, K., Poczapski, P., & Mohiuddin, S. (1989). The effects of rater training, job analysis format and congruence of training on job evaluation ratings. *Journal of Business and Psychology*, 3(4), 387-401.
- Chan, A. L., Chien, T.-W., Chen-Yi, S., & Shun-Jin, L. (2009). Validation of a shortened Taiwanese version of an asthma quality-of-life questionnaire by Rasch model. *Allergy and asthma proceedings*, 30(2), 171.
- Choorapanthiyil, M. J. (2007). *How international teaching assistants conceptualize teaching higher order thinking: A grounded theory approach* (Order No. 3259526). Available from ProQuest Dissertations & Theses Global. (304851726). <https://www.proquest.com/dissertations-theses/how-international-teaching-assistants/docview/304851726/se-2> sayfasından erişilmiştir.
- Coffman, W. E. (1971). On the reliability of ratings of essay examinations in English. *Research in the Teaching of English*, 5(1), 24-36.
- Congdon, P. J., & MeQueen, J. (2000). The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement*, 37(2), 163-178.
- Cook, D. A., Dupras, D. M., Beckman, T. J., Thomas, K. G., & Pankratz, V. S. (2009). Effect of rater training on reliability and accuracy of mini-cex scores: A randomized, controlled trial. *Journal of General Internal Medicine*, 24(1), 74-79.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90(2), 218-244.
- Cottrell, R. R., & McKenzie, J. F. (2011). Quantitative research methods: Experimental. In *Health promotion & education research methods: Using the five-chapter thesis/Dissertation model* (2. ed., s. 171-192). Sudbury, Massachusetts: Jones and Bartlett.
- Craig, D. A. (1997). *Rater's perception of handwriting legibility and word-processing in L2 writing assessment*. Urbana, Illinois: College of the University Urbana.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, 52(3), 177-193.

- Cunningham, G. K. (1998). *Assessment in the classroom: Constructing and interpreting texts*. London: The Falmer.
- Çetin, B. (2019). Yazılı yoklamalar ve açık uçlu maddeler. In B. Çetin (Ed.), *Eğitimde ölçme ve değerlendirme* (1 ed., s. 128). Ankara: Anı.
- Çetin, B., & İlhan, M. (2017). Standart ve solo taksonomisine dayali rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı katılığı ve cömertliğinin incelenmesi. *Eğitim ve Bilim*, 42(189), 10-1.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2021). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (6 ed.). Ankara: Pegem Akademi.
- Danielson, C., & Hansen, P. (2016). *Performance tasks and rubrics for early elementary mathematics* (2 ed.). New York: Routledge.
- Davis, L. (2016). The influence of training and experience on rater performance in scoring spoken language. *Language Testing*, 33(1), 117-135.
- Day, D. V., & Sulsky, L. M. (1995). Effects of frame-of-reference training and information configuration on memory organization and rating accuracy. *Journal of Applied Psychology*, 80(1), 158-167.
- DeCoths, T. A. (1977). An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Performance*, 19(2), 247-266.
- DeMars, C. (2010). *Item response theory*. Madison Avenue, New York: Oxford University.
- DeNisi, A., Robbins, T., & Williams, K. (1989). *Retrieval vs. computational models: What we remember vs. what we use*. Paper presented at the paper presented as part of the relationship between memory and judgment in performance appraisal, a symposium presented at the Conference of the Society for Industrial and Organizational Psychology (D. Woehr, Chair), Boston, MA.
- DeVellis, R. F. (2003). *Scale development theory and applications* (2 ed.). London: Sage.
- Dierdorff, E. C., Surface, E. A., & Brown, K. G. (2010). Frame-of-reference training effectiveness: Effects of goal orientation and self-efficacy on affective, cognitive, skill-based, and transfer outcomes. *Journal of Applied Psychology*, 95(6), 1181-1191.

- Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Learning and instruction, 10*(1), 11-20.
- Doğan, C. D., & Anadol, H. Ö. (2017). Comparing fully crossed and nested designs where items nested in raters in generalizability theory. *Kastamonu Eğitim Dergisi, 25*(1), 361-372.
- Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction*: Cambridge: Cambridge University.
- Doosti, M., & Ahmadi Safa, M. (2021). Fairness in oral language assessment: Training raters and considering examinees' expectations. *International Journal of Language Testing, 11*(2), 64-90.
- Downing, S. M. (2005). Threats to the validity of clinical teaching assessments: What about rater error? *Medical Education, 39*(4), 353-355.
- Downing, S. M., & Yudkowsky, R. (2009). *Assessment in health professions education*. New York: Taylor & Francis.
- Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education, 31*(2), 115-128.
- Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika, 16*(4), 407-424.
- Ebel, R. L., & Frisbie, D. A. (1991). Multiple-choice test items. In *Essentials of educational measurement* (5th ed., s. 154-178). New Delhi: Prentice.
- Eckes, T. (2011). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*. Frankfurt: Peter Lang.
- Eckes, T. (2012). Operational rater types in writing assessment: linking rater cognition to rater behavior. *Language Assessment Quarterly, 9*(3), 270-292.
- Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*. Frankfurt: Peter Lang.
- Eckes, T., & Jin, K.-Y. (2022). Detecting illusory halo effects in Rater-Mediated assessment: A mixture rasch facets modeling approach. *Psychological Test and Assessment Modeling, 64*(1), 87-111.

- Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing*, 24(1), 37-64.
- Elhan, A. H., & Atakurt, Y. (2005). Ölçeklerin değerlendirilmesinde niçin Rasch analizi kullanılmalıdır? *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 58(1), 47-50.
- Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. In *Multivariate Applications Books Series*. Manwah, New Jersey: Lawrence Erlbaum.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Engelhard, G. (2012). Monitoring raters in performance assessments. In *Large-scale assessment programs for all students* (pp. 224-249). New York: Routledge.
- Engelhard, G. (2011). Evaluating the bookmark judgments of standard-setting panelists. *Educational and Psychological Measurement*, 71(6), 909-924.
- Engelhard, G. (2013). Assessments with raters: rater-invariant measurement. In *Invariant measurement using rasch models in the social, behavioral and health sciences*. New York: Psychology.
- Erguvan, I. D., & Aksu Dünya, B. (2020). Analyzing rater severity in a freshman composition course using many facet Rasch measurement. *Language Testing in Asia*, 10(1), 1.
- Erman Aslanoğlu, A. (2022). Üst düzey zihinsel beceriler ve ölçülmesi. In İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* (s. 2-27). Ankara: Pegem Akademi.
- Esfandiari, R. (2015). Rater errors among peer-assessors: Applying the many-facet rasch measurement model. *Iranian Journal of Applied Linguistics*, 18(2), 4-25.
- Esfandiari, R. (2021). Rater-mediated assessment of iranian undergraduate students' college essays: Many-facet rasch modelling. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, 9(1), 93-119.
- Fahim, M., & Bijani, H. (2011). The effects of rater training on raters' severity and bias in second language writing assessment. *International Journal of Language Testing*, 1(1), 1-16.

- Farrokhi, F., & Esfandiari, R. (2011). A many-facet rasch model to detect halo effect in three types of raters. *Theory & Practice in Language Studies*, 1(11), 21-25.
- Farrokhi, F., Esfandiari, R., & Dalili, M. V. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15(11), 76-83.
- Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-101.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66(2), 127-148.
- Feldman, J. M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions*, 32(4), 279-286.
- Flippo, E. B. (1984). *Personnel management*. New York: McGraw Hill.
- Furtner, M. R., Sachse, P., & Exenberger, S. (2012). Learn to influence yourself: Full range self-leadership training. *Journal of the Indian Academy of Applied Psychology*, 38(2), 294-304.
- Furtner, M. R., Tutzer, L., & Sachse, P. (2018). The mindful self-leader: Investigating the relationships between self-leadership and mindfulness. *Social Behavior and Personality: an International Journal*, 46(3), 353-360.
- Galati, D. E. (2007). *A g-theory based adjustment process for calibrating marker scores*. Toronto: University of Toronto.
- Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the 'black box' differently: assessor cognition from three research perspectives. *Medical Education*, 48(11), 1055-1068.
- Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. S. (2014). Seeing the 'black box' differently: assessor cognition from three research perspectives. *Medical Education*, 48(11), 1055-1068.
- Gingerich, A., Regehr, G., & Eva, K. W. (2011). Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Academic Medicine*, 86(10), 1-7.

- Goodrich, H. (1997). Understanding rubrics: The dictionary may define " rubric," but these models provide more clarity. *Educational Leadership*, 54(4), 14-17.
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology*, 94(5), 1336-1344.
- Görgülü Öztürk, B. (2023). *Açık uçlu maddelerin puanlanmasında öğrencilere verilen puanlayıcı eğitiminin puanlayıcı davranışlarına etkisinin çok yüzeyli rasch ölçme modeli ile incelenmesi*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. *Online Submission*. <https://files.eric.ed.gov/fulltext/ED532068.pdf>, sayfasından erişilmiştir.
- Gronlund, N. E. (1977). *Constructing achievement test*. New Jersey: Prentice-Hall..
- Gronlund, N. E. (1998). *Assessment of student achievement*. North Carolina: Eric.
- Guilford, J. P. (1954). *Psychometric methods*, 2nd ed. New York: McGraw-Hill.
- Güler, N. (2008). *Klasik test kurami genellenebilirlik kurami ve rasch modeli üzerine bir araştırma*. Doktora Tezi Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Güler, N. (2010). Genellenebilirlik kurami ve SPSS ile GENOVA programlariyla hesaplanan G ve K çalışmalarına ilişkin sonuçların karşılaştırılması. *Eğitim ve Bilim*, 34(154), 14-25.
- Güler, N. (2014). Analysis of open-ended statistics questions with many facet rasch model. *Eurasian Journal of Educational Research*, 55, 73-90.
- Güler, N. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Güler, N., Kaya Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik kurami*. Ankara: Pegem Akademi.
- Haiyang, S. (2010). An application of classical test theory and many facet Rasch measurement in analyzing the reliability of an English test for non-English major graduates. *Chinese Journal of Applied Linguistics*, 33(2), 87-102.

- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Bacon: Allyn and Bacon.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Taylor & Francis.
- Hambleton, R. K. (1995). Meeting the measurement challenges of the 1990s and beyond new assessment models and methods. In T. Oakland & R. K. Hambleton (Eds.), *International Perspectives on Academic Assessment* (s. 83-104). Dordrecht: Springer Netherlands.
- Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., & Gifford, J. A. (1978). Developments in latent trait theory: models, technical issues, and applications. *Review of Educational Research*, 48(4), 467-510.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. New York: Sage.
- Hancock, C. L. (1995). Implementing the assessment standards for school mathematics: enhancing mathematics learning with open-ended questions. *The Mathematics Teacher*, 88(6), 496-499.
- Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *The Counseling Psychologist*, 27(3), 353-383.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93(3), 258-268.
- Hauenstein, N. M. A. (1998). Training raters to increase the accuracy of appraisals and the usefulness of feedback. *Performance Appraisal: State of the Art in Practice*, 4, 404-442.
- Hauenstein, N. M. A., & Foti, R. J. (1989). From laboratory to practice: Neglected issues in implementing frame-of-reference rater training. *Personnel Psychology*, 42(2), 359-378.
- Hauenstein, N. M. A., & McCusker, M. E. (2017). Rater training: Understanding effects of training content, practice ratings, and feedback. *International Journal of Selection and Assessment*, 25(3), 253-266.

- Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, 73(1), 68-73.
- Heinich, R. (1973). Technology and the management of instruction. *AV Communication Review*, 21(2), 224-224.
- Hodgetts, R. M. (1993). *Blueprints for continuous improvement : lessons from the Baldrige winners*: New York: AMA Membership.
- Holmboe, E. S., Hawkins, R. E., & Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence. *Annals of Internal Medicine*, 140(11), 874-881.
- Houghton, J. D., & Neck, C. P. (2002). The revised self-leadership questionnaire. *Journal of Managerial Psychology*, 17(8), 672-691.
- Houghton, J. D., Wu, J., Godwin, J. L., Neck, C. P., & Manz, C. C. (2012). Effective stress management: a model of emotional intelligence, self-leadership, and student stress coping. *Journal of Management Education*, 36(2), 220-238.
- Idris, M., & Zakaria, M. H. (2016). Gauging ESL learners' CEFR ratings on oral proficiency in rater training. *Man in India*, 96(6), 1675-1682.
- İlhan, M. (2015). *Standart ve solo taksonomisine dayali rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeyli rasch modeli ile incelenmesi*. Doktora Tezi, Gaziantep Üniversitesi Eğitim Bilimleri Enstitüsü, Gaziantep.
- İlhan, M. (2016). Açık uçlu sorularla yapılan ölçmelerde klasik test kuramı ve çok yüzeyli Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(2), 346-368.
- İlhan, M., & Çetin, B. (2014). Performans değerlendirmeye karışan puanlayıcı etkilerini azaltmanın yollarından biri olarak puanlayıcı eğitimleri: kuramsal bir analiz. *Journal of European Education*, 2, 29-38.
- İlhan, M., & Gezer, M. (2023). Araştırmaların sınıflandırılması. In B. Çetin, M. İlhan, & M. G. Şahin (Eds.), *Eğitimde araştırma yöntemleri: Temel kavramları, ilkeler ve süreçler* (2. ed., s. 101-133). Ankara: Pegem Akademi.

- Jackson, S. E., Schuler, R. S., & Werner, S. (2018). Conducting performance manageme. In *Managing human resources*. (12 ed., s. 317). New York: Oxford University.
- Jacobs, R., & Kozlowski, S. W. J. (1985). A closer look at halo error in performance ratings. *Academy of Management Journal*, 28(1), 201-212.
- Jansen, T., Vögelin, C., Machts, N., Keller, S., & Möller, J. (2021). Don't just judge the spelling! The influence of spelling on assessing second-language student essays. *Frontline Learning Research*, 9(1), 44-65.
- Johnson, R. L., Penny, J. A., & Gordon, B. (2009). *Assessing performance: designing, scoring, and validating performance tasks*. New York: Guilford.
- Jones, E., & Bergin, C. (2019). Evaluating teacher effectiveness using classroom observations: A Rasch analysis of the rater effects of principals. *Educational Assessment*, 24(2), 91-118.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Kang, O., Rubin, D., & Kermad, A. (2019). The effect of training and rater differences on oral proficiency assessment. *Language Testing*, 36(4), 481-504.
- Karakaya, İ. (2015). Comparison of self peer and instructor assessments in the portfolio assessment by using many facet rasch model. *Journal of Education and Human Development*, 4(2), 14-37.
- Karakaya, İ. (2022). Açık uçlu maddelerin hazırlanması ve incelenmesi. In İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* (1 ed., s. 40-75). Ankara: Pegem Akademi.
- Karakaya, İ., & Şata, M. (2022). Açık uçlu maddeler. In İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* (1th ed.). Ankara: Pegem Akademi.
- Kaya Uyanık, G., Güler, N., Taşdelen Teker, G., & Demir, S. (2017). Türkiye’de eğitim alanında yayımlanan ölçek geliştirme çalışmalarının uygunluğunun çok yüzeyli rasch modeli ile incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 8(2), 183-199.

- Keller, K. L. (1987). Memory factors in advertising: The effect of advertising retrieval cues on brand evaluations. *Journal of Consumer Research*, 14(3), 316-333.
- Khatti, N. (1995). How performance assessments affect teaching and learning. *Educational Leadership*, 53(3), 80-83.
- Khitab, U., Zaman, A., & Ghaffar, A. (2014). Secondary school teachers' ability to understand scoring rubric and practices. *J. Appl. Environ. Biol. Sci*, 4(8S), 153-159.
- Kim, H. (2020). Effects of rating criteria order on the halo effect in L2 writing assessment: a many-facet Rasch measurement analysis. *Language Testing in Asia*, 10(1), 16.
- Kim, S., Copeland, L., Cohen, E., Galt, J., Terregino, C. A., & Pradhan, A. (2022). Frame-of-reference training for students: promoting a shared mental model for clerkship performance with an online, interactive training module. *Journal of General Internal Medicine*, 37(6), 1575-1577.
- Kim, Y. (2009). *Combining constructed response items and multiple choice items using a hierarchical rater model*. Doktora Tezi, Columbia University, New York. <https://www.proquest.com/dissertations-theses/combining-constructed-response-items-multiple/docview/304862491/se-2?accountid=11054> sayfasından erişilmiştir.
- Kim, Y., Park, I., & Kang, M. (2012). Examining rater effects of the TGMD-2 on children with intellectual disability. *Adapted Physical Activity Quarterly*, 29(4), 346-365.
- Kingsbury, F. A. (1922). Analyzing ratings and training raters. *Journal of Personnel Research*, 1, 377-383.
- Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., Othman, A. R. (1998). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education*, 11(2), 121-137.
- Knoch, U., Fairbairn, J., Myford, C., & Huisman, A. (2018). Evaluating the relative effectiveness of online and face-to-face training for new writing raters. *Papers in Language Testing and Assessment*, 7(1), 61-86.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26-43.

- Kogan, J. R., Conforti, L. N., & Holmboe, E. S. (2023). Faculty perceptions of frame of reference training to improve workplace-based assessment. *J Grad Med Educ*, 15(1), 81-91.
- Koons, H. H. (2008). *The reading -writing connection: An investigation of the relationship between reading ability and writing quality across multiple grades and three writing discourse modes*. (Ph.D.). The University of North Carolina at Chapel Hill, United States North Carolina. <https://www.proquest.com/dissertations-theses/reading-writing-connection-investigation/docview/193673409/se-2?accountid=11054> sayfasından erişilmiştir.
- Korkmaz, H. (2004). *Fen ve teknoloji eğitiminde alternatif değerlendirme yaklaşımları*. Ankara: Yeryüzü.
- Kozaki, Y. (2004). Using Genova and Facets to set multiple standards on performance assessment for certification in medical translation from Japanese into English. *Language Testing*, 21(1), 1-27.
- Kozlowski, S. W., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, rater familiarity, conceptual similarity and halo error: An exploration. *Journal of Applied Psychology*, 71(1), 45-49.
- Köse, İ. A., Usta, H. G., & Yandı, A. (2016). Sunum yapma becerilerinin çok yüzeyli rasch analizi ile değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(4), 1853-1864.
- Kubiszyn, T., & Borich, G. D. (2013). *Educational testing and measurement*. New Jersey: John Wiley & Sons.
- Kumar, D. (2005). Performance appraisal: The importance of rater training. *Journal of the Kuala Lumpur Royal Malaysia Police College*, 4(1), 1-17.
- Kutlu, Ö. (2006). Üst düzey zihinsel süreçleri belirleme yolları: yeni durum belirleme yaklaşımları. *Çağdaş Eğitim*, 335, 14-26.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2010). *Öğrenci başarısının belirlenmesi performans ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2014). *Öğrenci başarısının belirlenmesi: Performans ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi.

- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2017). *Ölçme ve değerlendirme performansa ve portfolyoya dayalı durum belirleme* (5. ed.). Ankara: Pegem Akademi.
- Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the intelligent essay assessor. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (p. 87-112). Mahwah, New Jersey: Lawrence Erlbaum.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72-107.
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance : methods, theory, and applications*. Cambridge, Massachusetts: Academic Press.
- Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*: Saddle River: Addison-Wesley.
- Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology*, 60(5), 550-555.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Levy, F., & Murnane, R. J. (2004). *The New Division of Labor*. Princeton: Princeton University..
- Li, W. (2022). Scoring rubric reliability and internal validity in rater-mediated EFL writing assessment: Insights from many-facet Rasch measurement. *Reading and Writing*, 35(10), 2409-2431.
- Lichtenstein, M., & Srull, T. K. (1987). Processing objectives as a determinant of the relationship between recall and judgment. *Journal of Experimental Social Psychology*, 23(2), 93-118.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543-560.
- Linacre, J. M. (1989). *Many-faceted Rasch measurement*. (Ph.D.). The University of Chicago, United States -- Illinois. <https://www.proquest.com/dissertations-theses/many-faceted-rasch-measurement/docview/252294604/se-2?accountid=11054> sayfasından erişilmiştir.

- Linacre, J. M. (1993a). *Generalizability theory and many-facet rasch measurement*. Atlanta: Paper presented at the Theory Into Practice.
- Linacre, J. M. (1993b). Rasch-based generalizability theory. *Rasch Measurement Transactions*, 7(1), 283-284.
- Linacre, J. M. (1994). *Many-facet rasch measurement*. Chicago: Mesa Press.
- Linacre, J. M. (2014a). A user's guide to Facets: Rasch-model computer programs. Retrieved December, 18, 2018.
- Linacre, J. M. (2014b). A user's guide to FACETS: Rasch-model computer programs.
- Linacre, J. M. (2023). *Facets computer program for many-facet Rasch measurement* [version 3.85.1].
- Linacre, J. M., Engelhard, G., Tatum, D. S., & Myford, C. M. (1994). Measurement with judges: Many-faceted conjoint measurement. *International Journal of Educational Research*, 21(6), 569-577.
- Linlin, C. (2020). Comparison of automatic and expert teachers' rating of computerized English listening-speaking test. *English Language Teaching*, 13(1), 18-30.
- Linn, R. L., & Gronlund, N. E. (1995). *Measurement and Assessment in Teaching*: USA: Merrill.
- Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*: Iap.
- Livingston, S. A. (2009). Constructed-response test questions: Why we use them; How we score them. R&D Connections. Number 11. *Educational Testing Service*.
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*: Charlotte, North Carolina: Age.
- Lovelace, K. J., Manz, C. C., & Alves, J. C. (2007). Work stress and leadership development: The role of self-leadership, shared leadership, physical fitness and flow in managing demands and increasing job control. *Human Resource Management Review*, 17(4), 374-387.
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: implications for training. *Language Testing*, 12(1), 54-71.

- Lunz, M. E., & Linacre, J. M. (1998). Measurement designs using multifacet Rasch modeling. *Modern Methods for Business Research*, 4, 47-77.
- Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345.
- Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and Many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15(2), 158-180.
- Mastergeorge, A. M., & Martínez, J. F. (2010). Rating performance assessments of students with disabilities: a study of reliability and bias. *Journal of Psychoeducational Assessment*, 28(6), 536-550.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- May, G. L. (2008). The Effect of Rater Training on Reducing Social Style Bias in Peer Evaluation. *Business Communication Quarterly*, 71(3), 297-313.
- McArthur, D. L. (1987). The three-parameter logistic model. In D. L. McArthur (Ed.), *Alternative approaches to the assessment of achievement* (P. 129-158). Dordrecht: Springer.
- McIntyre, P. (1993). *The importance and effectiveness of moderation training on the reliability of teacher assessments of ESL writing samples*. Yüksek Lisans Tezi, University of Melbourne, Melbourne.
- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, 69(1), 147-156.
- McNamara, T. F. (1996). *Measuring second language performance*. Harlow: Longman.
- MEB. (2015). Akademik becerilerin izlenmesi ve değerlendirilmesi (ABİDE) projesi. 2015. <https://abide.meb.gov.tr/proje-hakkinda.asp> sayfasından erişilmiştir.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, 7(1), 25.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10th ed.). Upper Saddle River, New Jersey: Pearson Education.

- Mohd Noh, M. F., & Mohd Matore, M. E. E. (2022). Rater severity differences in English language as a second language speaking assessment based on rating experience, training experience, and teaching experience through many-faceted Rasch measurement analysis. *Frontiers in Psychology*, 13, 33-39.
- Moore, B. B. (2009). *A consideration of rater effects and rater design via signal detection theory*. (Ph.D.). Columbia University, United States New York. <https://www.proquest.com/dissertations-theses/consideration-rater-effects-design-via-signal/docview/304862541/se-2?accountid=11054> sayfasından erişilmiştir.
- Morrow, J. R., Mood, D., Disch, J., & Kang, M. (2015). *Measurement and evaluation in human performance* (5 ed.). New York: Human Kinetics.
- Moser, K., Kemter, V., Wachsmann, K., Köver, N. Z., & Soucek, R. (2018). Evaluating rater training with double-pretest one-posttest designs: an analysis of testing effects and the moderating role of rater self-efficacy. *The International Journal of Human Resource Management*, 29(18), 2609-2631.
- Moskal, B. M. (2000). Scoring Rubrics: What, When and How? *Practical Assessment, Research, and Evaluation*, 7(1), 4-12.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment Research and Evaluation*, 7(1), 21-36.
- Mubeena, M. (2010). Development of rater training programs. <https://www.expertbase.org/a386-development-of-rater-training-programs> sayfasından erişilmiştir.
- Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology*, 74(4), 619-624.
- Murphy, K. R., Philbin, T. A., & Adams, S. R. (1989). Effect of purpose of observation on accuracy of immediate and delayed performance ratings. *Organizational Behavior and Human Decision Processes*, 43(3), 336-354.
- Myford, C. M. (2002). Investigating design features of descriptive graphic rating scales. *Applied Measurement in Education*, 15(2), 187-215.
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.

- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- Nakamura, Y. (2002). Teacher assessment and peer assessment in practice. *Educational Studies*, 44, 203-215.
- Neck, C. P. (1996). Thought self-leadership: a self-regulatory approach towards overcoming resistance to organizational change. *The International Journal of Organizational Analysis*, 4(2), 202-216.
- Neck, C. P., & Houghton, J. D. (2006). Two decades of self-leadership theory and research. *Journal of Managerial Psychology*, 21(4), 270-295.
- Neck, C. P., Stewart, G. L., & Manz, C. C. (1995). Thought self-leadership as a framework for enhancing the performance of performance appraisers. *The Journal of Applied Behavioral Science*, 31(3), 278-302.
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (6. ed.). Edinburgh Gate: Pearson.
- Niu, C. (2021). *Rate to measure mathematics teaching: using the many-facet rasch modeling to reevaluate the mathematics classroom observation protocol for practices (MCOP2)*. (Doktora Tezi). University of Kentucky, Lexington, Kentucky.
- Noonan, L. E., & Sulsky, L. M. (2001). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance*, 14(1), 3-26.
- ODS (2022) Öğrenci/Öğretmen Destek Sistemi. https://ods.eba.gov.tr_sayfasından erişilmiştir.
- Ogunfowora, B. (2014). It's all a matter of consensus: Leader role modeling strength as a moderator of the links between ethical leadership and employee outcomes. *Human Relations*, 67(12), 1467-1490.
- Ok, A. B., Bilgiç, H. C., & Bilgiç, R. (2013). Öğretim elemanı değerlendirmesinde kullanılan formatların hale etkisi, cömertlik etkisi ve kullanıcı tepkileri açısından karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 44(44), 237-248.
- Oosterhof, A. (2003). *Developing and using classroom assessments* (3th ed.). Upper Saddle River, New Jersey: Pearson..

- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. London, New Delhi: Sage.
- Othman, N., & Mohamad, K. A. (2014). Thinking skill education and transformational progress in Malaysia. *International Education Studies*, 7(4), 27-32.
- ÖSYM. (2015). Yazılı Sınav (Açık Uçlu Sorularla Sınav). <https://www.osym.gov.tr/TR,721/yazili-sinav-acik-uclu-sorularla-sinav-04022015.html> sayfasından erişilmiştir.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65(4), 373-406.
- Pietilä, A. (2019). *Work Recovery in self-managing organizations*. (Master's Programme in Information Networks). Aalto University. <https://urn.fi/URN:NBN:fi:aalto-201910275919> sayfasından erişilmiştir.
- Popham, W. J. (2001). *Classroom assesment: What teachers need to know*. Boston: Allyn and Bacon.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69(4), 581-588.
- Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior and Human Decision Processes*, 38(1), 76-91.
- Raczynski, K. R., Cohen, A. S., Engelhard, G., & Lu, Z. (2015). Comparing the effectiveness of self-paced and collaborative frame-of-reference training on rater accuracy in a large-scale writing assessment. *Journal of Educational Measurement*, 52(3), 301-318.
- Rahman, A. A., Ahmad, J., Yasin, R. M., & Hanafi, N. M. (2017). Investigating central tendency in competency assessment of design electronic circuit: analysis using many facet Rasch measurement (MFRM). *International Journal of Information and Education Technology*, 7(7), 525-528.

- Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions*. Stillwater, Oklahoma, USA: New Forums.
- Rethinasamy, S. (2021). The effects of different rater training procedures on esl essay raters' rating accuracy. *Pertanika Journal of Social Sciences & Humanities*, 29, 14-27.
- Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education* (2. ed.). New York: Pearson.
- Roberts, G. E. (1998). Perspectives on enduring and emerging issues in performance appraisal. *Public Personnel Management*, 27(3), 301-320.
- Robertson, R. L., Park, J., Gillman, L., & Vergis, A. (2020). The impact of rater training on the psychometric properties of standardized surgical skill assessment tools. *The American Journal of Surgery*, 220(3), 610-615.
- Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time and behavior observation training. *International Journal of Training and Development*, 7(2), 93-107.
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczyńska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, 85(2), 370-395.
- Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Principles and Standards for School Mathematics (NCTM)*, 94(1), 22.
- Royal, K. D., & Hecker, K. G. (2016). Rater errors in clinical performance assessments. *Journal of Veterinary Medical Education*, 43(1), 5-8.
- Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment concepts and applications* (7. ed.). New York: McGraw-Hill.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2), 413-428.
- Saretsky, G. (1972). The OEO P.C. experiment and the John Henry effect. *The Phi Delta Kappan*, 53(9), 579-581.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.

- Schleicher, D. J., & Day, D. V. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision Processes*, 73(1), 76-101.
- Schraub, E. M. (2011). *Affect at Work - The impact of emotion regulation on employees' well-being, proactive and adaptive performance*. Heidelberg: Ruprecht Karl University Heidelberg.
- Schumacker, R. E. (1999). Many-facet Rasch analysis with crossed, nested, and mixed designs. *Journal of Outcome Measurement*, 3, 323-338.
- Sebok, S. S., & Syer, M. D. (2015). Seeing things differently or seeing different things? Exploring raters' associations of noncognitive attributes. *Academic Medicine*, 90(11), 50-55.
- Sims, H., & Manz, C. (1991). Superleadership: Beyond the myth of heroic leadership. *Organ. Dyn*, 19, 18-35.
- Sonnert, G., Hazari, Z., & Sadler, P. M. (2018). Evaluating the quality of middle school mathematics teachers, using videos rated by college students. *Studies in Educational Evaluation*, 58, 60-69.
- Stalnaker, J. M. (1951). The essay type of examination. In E. F. Lindquist (Ed.), *Educational measurement* (Vol. 1, s. 495-530). Wisconsin: George Banta.
- Stamoulis, D. T., & Hauenstein, N. M. A. (1993). Rater training and rating accuracy: Training for dimensional accuracy versus training for ratee differentiation. *Journal of Applied Psychology*, 78(6), 994-1003.
- Stevens, D. D., & Levi, J. A. (2023). How to construct a rubric. In *Introduction to rubrics* (2. ed.). New York: Routledge.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239-261.
- Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology*, 73(3), 497-506.

- Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, 77(4), 501-510.
- Sulsky, L. M., & Day, D. V. (1994). Effects of frame-of-reference training on rater accuracy under alternative time delays. *Journal of Applied Psychology*, 79(4), 535-543.
- Şahan, Ö. (2018). *The impact of rater experience and essay quality on rater behavior and scoring*. Doktora Tezi, Çanakkale Onsekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü, Çanakkale.
- Şata, M. (2019). *Performans değerlendirme sürecinde puanlayıcı eğitiminin puanlayıcı davranışları üzerindeki etkisinin incelenmesi*. Doktora Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Şata, M. (2022). Açık uçlu maddelerin puanlanması. In İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* (Vol. 4, s. 75-101). Ankara: Pegem Akademi.
- Şata, M., & Karakaya, I. (2020). Investigation of the use of electronic portfolios in the determination of student achievement in higher education using the many-facet rasch measurement model. *Educational Policy Analysis and Strategic Research*, 15(1), 7-21.
- Tajeddin, Z., & Alemlı, M. (2014). Pragmatic rater training: does it affect non-native L2 teachers' rating accuracy and bias? *International Journal of Language Testing*, 4(1), 66-83.
- Taylor, L. (2004). *Second language writing assessment: Cambridge ESOL's ongoing research agenda*. <https://www.cambridgeenglish.org/images/23132-research-notes-16.pdf> sayfasından erişilmiştir.
- Tekin, H. (2013). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı.
- Tekindal, M. A. (2022). Mann-whitney u testi: Paket program uygulamaları ve makale örnekleri. In R. Güven & M. A. Tekindal (Eds.), *Temel istatistik yöntemler: SPSS ve JAMOVİ paket program uygulamaları* (s. 135-146). Ankara: Akademisyen.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25-29.

- Thornton, G. C., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, 65(3), 351-354.
- Tobaş, C. (2020). *Performansın değerlendirilmesinde farklılaşan puanlayıcı davranışlarının çok yüzeyli rasch ölçme modeli ile incelenmesi*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Turgut, M. F., & Baykul, Y. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Uggerslev, K. L., & Sulsky, L. M. (2008). Using frame-of-reference training to understand the implications of rater idiosyncrasy for rating accuracy. *Journal of Applied Psychology*, 93(3), 711-719.
- Urbina, S. (2004). *Essentials of psychological testing* (A. S. Kaufman & N. L. Kaufman Eds.). Hoboken, New Jersey: John Wiley & Sons.
- Veneziano, L., & Hooper, J. (1997). A method for quantifying content validity of health-related questionnaires. *American Journal of Health Behavior*, 21(1), 67-70.
- Wang, B. (2010). On Rater Agreement and rater training. *English language teaching*, 3(1), 108-112.
- Ward, W. C., Frederiksen, N., & Carlson, S. B. (1980). Construct validity of free-response and machine-scorable forms of a test. *Journal of Educational Measurement*, 17(1), 11-29.
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11(2), 197-223.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287.
- Weir, C. J. (2005). *Language testing and validation an evidence-based approach*. New York: Palgrave Macmillan.
- Wells, F. L. (1907). *A statistical study of literary merit: With remarks on some new phases of the method*. Cambridge: Science Press.
- Wigglesworth, G. (1993). Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. *Language Testing*, 10(3), 305-319.

- Wind, S. A. (2019). Examining the impacts of rater effects in performance assessments. *Applied Psychological Measurement*, 43(2), 159-171.
- Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30(2), 231-252.
- Woehr, D. J. (1992). Performance dimension accessibility: Implications for rating accuracy. *Journal of Organizational Behavior*, 13(4), 357-367.
- Woehr, D. J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology*, 79(4), 525-534.
- Woehr, D. J., & Feldman, J. M. (1993). Processing objective and question order effects on the causal relation between memory and judgment in performance appraisal: The tip of the iceberg. *Journal of Applied Psychology*, 78(2), 232-241.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67(3), 189-205.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46, 35-51.
- Wolfe, E. W., & Chiu, C. W. T. (1997). *Detecting rater effects with a multi-faceted rating scale model*. Chicago, IL.
- Wolfe, E. W., Jiao, H., & Song, T. (2015). A family of rater accuracy models. *Journal of Applied Measurement*, 16(2), 153-160.
- Wright, B., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 2, 370.
- Wu, M. (2017). Some IRT-based analyses for interpreting rater effects. *Psychological Test and Assessment Modeling*, 59(4), 453-470.
- Xiang, Y., Graeber, T., Enke, B., & Gershman, S. J. (2021). Confidence and central tendency in perceptual judgment. *Attention, Perception, & Psychophysics*, 83(7), 3024-3034.
- Yılmaz, F. N. (2017). Analysis of the rater effects in the rating of diagnostic trees prepared by teacher candidates by the many-facet rasch model. *JEP*, 8(18), 14-21

Yurdugöl, H. (2005). *Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması*. Paper presented at the XIV. Ulusal Eğitim Bilimleri Kongresi, Pamukkale Üniversitesi Eğitim Fakültesi.

Zedeck, S., & Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. *Journal of Applied Psychology*, 67(6), 752-758.



EKLER



EK 1. Türkçe Dersi Açık Uçlu Sorular

1-2 Garanti Belgesi ve Kullanım Kılavuzu

Bir telefonun garanti şartları ve kullanım kılavuzu aşağıdaki gibidir.

GARANTİ ŞARTLARI

1. Garanti süresi ürünün teslim tarihinden itibaren başlar ve 2 (iki) yıldır.
2. Garanti hizmetinden yararlanmak için ürünün yanında faturanın da yer alması gerekmektedir.
3. Ürünün, garanti süresi içinde malzeme, işçilik ve montaj hatalarından dolayı arızalanması durumunda hiçbir ücret talep edilmeden cihazın tamiri yapılacaktır.
4. Ürünün, kullanım kılavuzunda yer alan haller dışında kullanılmasından dolayı ortaya çıkan arızalar garanti kapsamında değildir.

KULLANIM KILAVUZU

1. Telefonunuzu hasarlı güç kablolarına, priz veya fişlerle kullanmayınız.
2. Telefonunuzu sıvılardan uzak tutunuz. Sıvı temasından koruyunuz.
3. Telefonunuzda herhangi bir arıza olması durumunda telefonunuzu kendiniz tamir etmeyiniz.
4. Telefonunuza, güvenilir olmayan sitelerden uygulama, yazılım vb. ürünler indirmeyiniz.

Zehra bozulan telefonunun garanti kapsamında tamir edilmesini istemektedir. Telefonu aldığı firmaya telefonunun garanti kapsamında yapılması için dilekçe yazmıştır.

Bir buçuk yıl önce firmanızdan aldığım cep telefonum şarj olmuyor. Faturam olmadığı için kimlik bilgilerimi gönderiyorum. Ürünümün değiştirilmesini ya da garanti kapsamında (ücretsiz) tamir edilmesini istiyorum.

3-4 Giriş Kartı

Üç salondan meydana gelen bir spor tesisinin, yaş gruplarına göre aylık kullanım bilgileri tabloda belirtilmiştir.

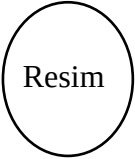
Salonlar	Çocuk (14-18 yaş arası)	Genç (19-23 yaş arası)	Yetişkin (24 yaş ve üzeri)
A- Spor Aletler Salonu	Koşu bandı* Eliptik Bisiklet* Dikey bisiklet** Dambıl**	Benç pres* Koşu bandı** Yatay bisiklet** Kelebek**	Yatay bisiklet* Kelebek* Dikey bisiklet** İstasyon**
B- Havuz (Üyelerin 1 ay içinde en fazla yararlanabilecekleri gün sayısı verilmiştir.)	7 gün* 3 gün**	14 gün* 5 gün**	20 gün* 8 gün**
C- Branş Sporları Salonu (Üyeler sadece birini seçebilir.)	Jimnastik Basketbol	Masa tenisi Salon futbolu	Hentbol Voleybol

* Hafta içi

** Hafta sonu

Bu spor tesisinde kullanıcılara tesis kullanım tercihlerine göre giriş kartı düzenlenmektedir. Giriş kartında kişinin adı, yaşı, havuzdan yararlanılan gün sayısı (hafta içi), havuzdan yararlanılan gün sayısı (hafta sonu), seçilen spor branşı, kullanılacak spor aletleri (hafta içi), kullanılacak spor aletleri (hafta sonu) ile ilgili bilgiler yer almaktadır.

Örnek giriş kartı;

GİRİŞ KARTI	
 Eymen Yaş: 16	Havuzdan yararlanılan gün sayısı (Hafta içi): 5 gün
	Havuzdan yararlanılan gün sayısı (Hafta sonu): 2 gün
	Seçilen spor branşı: Jimnastik
	Kullanılacak spor aletleri (Hafta içi): Koşu bandı Eliptik bisiklet
	Kullanılacak spor aletleri (Hafta sonu): Dikey bisiklet Dambıl

GİRİŞ KARTI	
Resim Kemal Yaş: 20	Havuzdan yararlanılan gün sayısı (Hafta içi): 20 gün
	Havuzdan yararlanılan gün sayısı (Hafta sonu): 3 gün
	Seçilen spor branşı: Masa tenisi
	Kullanılabilecek spor aletleri (Hafta içi): Benç pres
	Kullanılabilecek spor aletleri (Hafta sonu): Koşu bandı
	Yatay bisiklet
	Kelebek

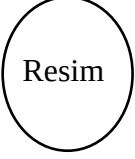
3- Kemal için hazırlanan giriş kartı tesise giriş için uygun mudur? Kutucuğu işaretleyiniz.

Uygun ☐ Uygun değil ☐

Cevabınızı gerekçeleriyle açıklayınız.

[illegible]

4- 28 yařındaki Mete iin giriř kartı hazırlayınız.

 Resim	GİRİŞ KARTI
Mete	

Öğrenci/Öğretmen destek sistemi (Ods, 2022) soru kodu 1337’den uyarlanmıştır.

5- İçindekiler Bölümü Hazırlama

Bir kitabın veya derginin başında bulunan, konu başlıklarının liste şeklinde sayfa numaralarıyla gösterildiği bölüme “İçindekiler” denir. İçindekiler bölümünde bütün başlıklar aynı biçim ve konumda dizilebildiği gibi her bir başlık düzeyi farklı da dizilebilmektedir. Aşağıda bir kitabın “İçindekiler” bölümünün düzenlenmesiyle ilgili kurallar verilmiştir.

- 1. düzey bölüm başlıklarında (genel başlıklar) sözcükler, **kalin-koyu** ve her birinin ilk harfi büyük olacak şekilde yazılmalıdır. Genel başlıklar sırasıyla 1., 2. şeklinde numaralandırılmalıdır.
- 2. düzey bölüm başlıklarında (alt başlıklar) sözcüklerin ilk harfleri büyük olmalı, sözcükler ince- açık yazılmalı ve genel başlığa göre sağa girintili olmalıdır. Alt başlıklar ait olduğu genel başlığın numarasıyla sıralanır. 1. 1., 2. 2. gibi.
- 3. düzey bölüm başlıklarında (alt başlıkların ayrıntıları) ilk sözcük büyük harfle, diğer sözcükler küçük harfle başlamalı; sözcükler ince-açık ve *eğik* yazılmalıdır. Alt başlığa göre sağa girintili olması gereken 3. düzey bölüm başlıkları, ait olduğu alt başlığın numarasıyla sıralanır: 1.1.1. gibi.

(Kalin- koyu yazı tipi: **Kalin**, İnce- açık yazı tipi: İnce açık, Eğik (italik) yazı tipi: *Eğik*)

5-

İÇİNDEKİLER

1. Geçmiş Zaman Edipleri (Kalin-Koyu)	1
1.1. Hamdullah Suphi Kimdir? (İnce-Açık)	9
1.1.1. Hayatı ve Edebi Kişiliği (İnce-Açık-Eğik)	18
1.1.2. Eserleri (İnce-Açık-Eğik)	24
1.2. Halit Ziya Kimdir? (İnce-Açık)	26
1.2.1. Hayatı ve Edebi Kişiliği (İnce-Açık-Eğik)	35
1.2.2. Eserleri (İnce-Açık-Eğik)	48
2. Ediplerimize Dair Hatıralar (Kalin-Koyu)	50

Oluşturulan “İçindekiler” bölümü metinde bahsedilen kurala uygun olup olmadığını aşağıdaki kutucuklara işaretleyiniz.

Uygun ☐ Uygun Değil ☐

Cevabınızı gerekçeleri ile açıklayınız.

.....

.....

.....

.....

.....

Oğrenci/Oğretmen destek sistemi (Ods, 2022) soru kodu 1419’dan uyarlanmıştır.

EK 2. Türkçe Dersi Dereceli Puanlama Anahtarı

1- Garanti Belgesi ve Kullanım Kılavuzu

<i>En Doğru Yanıt</i>	Puan
Öğrenci olumsuz seçeneğini işaretler ve gerekçesinde Zehra'nın faturasının olmadığını ya da kimlik bilgilerinin göndermesinin garanti şartlarına uygun olmadığını ifade eder. Örnek Yanıt: - Olumsuz, çünkü Zehra'nın telefonunun faturası yok. - Olumsuz, çünkü Zehra'nın kimlik bilgileri fatura yerine geçmemektedir. - Olumsuz, çünkü firma kimlik bilgileri değil cihazın faturasını istemektedir.	4
<i>Uzak Doğru Yanıtlar</i>	
Öğrenci olumsuz seçeneğini işaretler ve gerekçesinde garanti şartlarının karşılanmadığını ifade eder. Faturanın olmadığından bahsetmez. Örnek Yanıt: - Olumsuz, çünkü cihaz garanti şartlarını taşıyor.	3
Öğrenci olumsuz seçeneğini işaretler. Ancak gerekçesinde kullanım kılavuzu ile ilgili hatalardan (sıvı teması, kendi tamir etmeye çalışma vb.) bahseder. Örnek Yanıt: - Olumsuz, çünkü Zehra'nın telefonu suya düşmüş olabilir. - Olumsuz, çünkü Zehra telefonunu kullanım kılavuzuna uygun olarak kullanmamıştır.	2
Öğrenci olumsuz seçeneğini işaretler. Ancak gerekçe yazmaz ya da gerekçesi doğru değildir. Örnek Yanıt: - Olumsuz, kimlik bilgilerini göndermesi telefon sahibinin Zehra olduğu anlamına gelmez. - Olumsuz, çünkü telefonun garanti süresi dolmuştur. - Olumsuz, çünkü telefonun neden şarj olmadığı ile ilgili açıklama yapılmamıştır.	
<i>Yanlış/İlişkisiz Yanıtlar ve Boş</i>	
Öğrenci olumlu seçeneğini işaretler ve cevabına yönelik gerekçe yazar. Örnek Yanıt: - Olumlu, çünkü telefonun garantisi devam ediyor. Öğrenci kendi içinde doğru ancak konuyla alakasız açıklama yapar. Örnek Yanıt: - Kimlik bilgilerimizi tanımadığımız kişilerle paylaşmamalıyız.	1

2- Garanti Belgesi ve Kullanım Kılavuzu

<i>En Doğru Yanıt</i>	Puan
Öğrenci garanti şartlarını taşıyan (garanti süresi, fatura, üretim hatası, kullanım kılavuzuna uygun kullanım) bir dilekçe yazar. Örnek Yanıt: - Bir yıl önce firmanızdan aldığım cep telefonumun sıvı teması olmamasına rağmen hoparlöründen ses gelmiyor. Ürünün faturası kutunun altındadır. Ürünün değiştirilmesini ya da garanti kapsamında tamir edilmesini istiyorum.	4
<i>Uzak Doğru Yanıtlar</i>	
Öğrencinin dilekçesinde garanti şartları yeterince yer almayan ancak kullanım kılavuzuna uygun kullanıldığı belirtilen ve firmadan olumlu yanıt alabilecek cevaplar. Örnek Yanıt: - Firmanızdan aldığım cep telefonunu hasarlı prizlerde kullanmamama, sıvı teması olmamasına rağmen yaptığım görüşmelerde sesim karşı tarafa gitmiyor. Faturasını kutu içinde yer almaktadır. Ürünün değiştirilmesini ya da garanti kapsamında tamir edilmesini istiyorum. Öğrencinin dilekçesinde kullanım kılavuzu yeterince yer almayan ancak garanti şartları belirtilen ve firmadan olumlu yanıt alabilecek cevaplar. Örnek Yanıt: - Bir yıl önce firmanızdan aldığım cep telefonumun hoparlöründen ses gelmiyor. Ürünün faturası kutunun altındadır. Ürünün değiştirilmesini ya da garanti kapsamında tamir edilmesini istiyorum.	3
Öğrenci garanti şartlarından bahsetmez sadece kullanım kılavuzuna uygun kullandığını belirtir. Örnek Yanıt: - Firmanızdan aldığım cep telefonuna sıvı teması olmamasına rağmen hoparlörden ses gelmemektedir. Ürünün değiştirilmesini ya da garanti kapsamında tamir edilmesini istiyorum.	2
<i>Yanlış/İlişkisiz Yanıtlar ve Boş</i>	
Öğrenci garanti şartlarından ve kullanım kılavuzundan ilişkisiz cevaplar yazar. Örnek Yanıt: - Firmanızdan aldığım telefona internetten uygulama indirdim. Uygulamayı açtıktan sonra telefonum kapandı ve daha açılmıyor. Ürünün değiştirilmesini ya da garanti kapsamında (ücretsiz) tamir edilmesini istiyorum. Öğrenci kendi içinde doğru ancak konuyla alakasız açıklama yapar. Örnek Yanıt: - Kimlik bilgilerimizi tanımadığımız kişilerle paylaşmamalıyız.	1

3- Giriş Kartı

En Doğru Yanıt	Puan
Öğrenci uygun değil seçeneğini işaretler ve açıklamada hafta içi havuzdan yararlanılan gün sayısının, hafta içi kullanılabilecek spor aletlerinin yanlış olduğunu ifade eder. Örnek Yanıt: - Uygun değil, çünkü hafta içi havuzdan yararlanılacak gün sayısı 14 günden fazladır ve koşu bandı hafta sonu kullanılmalıdır. - Uygun değil, çünkü hafta içi havuzdan yararlanılacak gün sayısı ve hafta içi kullanılabilecek spor aletleri yanlış yazılmıştır.	4
Uzak Doğru Yanıtlar	
Öğrenci uygun değil seçeneğini işaretler ve gerekçeyi detaylı ifade etmez. Örnek Yanıt: - Uygun değil, çünkü bazı başlıklar uygun aralıklarda değildir. Öğrenci uygun değil seçeneğini işaretler ve gerekçesinde sadece 1 tane eksiği söyler. Örnek Yanıt: - Uygun değil, çünkü havuzdan yararlanılan gün sayısı 14 günün altında olmalıdır. - Uygun değil, çünkü hafta içi kullanılabilecek spor aletleri yanlış yazılmıştır. Öğrenci uygundur seçeneğini işaretler ve gerekçeyi detaylı ifade eder. Örnek Yanıt: - Uygundur, çünkü hafta içi havuzdan yararlanılacak gün sayısı 14 günden fazladır ve koşu bandı hafta sonu kullanılmalıdır. - Uygundur, çünkü hafta içi havuzdan yararlanılacak gün sayısı ve hafta içi kullanılabilecek spor aletleri yanlış yazılmıştır.	3
Öğrenci uygun değil seçeneğini işaretler. Ancak gerekçesi doğru değildir ya da yoktur. Örnek Yanıt: - Uygun değil, çünkü seçilen spor branşı yanlış yazılmıştır. Öğrenci uygun seçeneğini işaretler. Ancak gerekçeden bir tanesi doğrudur. Örnek Yanıt: - Uygundur, çünkü havuzdan yararlanılan gün sayısı 14 günün altında olmalıdır. - Uygundur, çünkü hafta içi kullanılabilecek spor aletleri yanlış yazılmıştır.	2
Yanlış/İlişkisiz Yanıtlar ve Boş	
Öğrenci uygun seçeneğini işaretler ve cevabına yönelik gerekçe yazar. Örnek Yanıt: - Uygun. Başlıklar açıklamaya göre düzenlenmiştir. Öğrenci uygun seçeneğini işaretler ve gerekçe yazmaz.	1

4- Giriş Kartı

En Doğru Yanıt		Puan
Öğrenci kartta yer alması gereken tüm başlıkları yazar ve bu başlıkları kurala uygun (salona, yaşa ve hafta içine-sonuna uygun) bir şekilde yerleştirir. Örnek Yanıt: Resim Havuzdan yararlanılan gün sayısı (Hafta içi): 20 gün ve altı Havuzdan yararlanılan gün sayısı (Hafta sonu): 8 gün ve altı Seçilen spor branşı: Hentbol veya Voleybol Kullanılabilecek spor aletleri (Hafta içi): Yatay bisiklet veya İ Mete Yaş: 28 İstasyon Kullanılabilecek spor aletleri (Hafta sonu): Yatay bisiklet ve Kelebek Dikey bisiklet veya Dikey bisiklet ve İstasyon		4
Uzak Doğru Yanıtlar		
Öğrenci kartta yer alması gereken başlıklardan ve başlıklara gelecek kavramlardan en fazla 2 tanesini eksik ya da yanlış yazmıştır. Ancak yazdığı başlıkların ve kavramların tamamı doğrudur. Örnek Yanıt: Resim Havuzdan yararlanılan gün sayısı (Hafta içi): 20 gün ve altı Havuzdan yararlanılan gün sayısı (Hafta sonu): 8 gün ve altı Seçilen spor branşı: YANLIŞ İ Mete EKSİK İstasyon Kullanılabilecek spor aletleri (Hafta içi): Yatay bisiklet veya Kullanılabilecek spor aletleri (Hafta sonu): Yatay bisiklet ve Kelebek Dikey bisiklet veya Dikey bisiklet ve İstasyon		3
Öğrenci kartta yer alması gereken başlıkları tam olarak yazmıştır ancak başlıklara ait bilgiler hatalıdır. Örnek Yanıt: Resim Havuzdan yararlanılan gün sayısı (Hafta içi): 21 gün ve üstü Havuzdan yararlanılan gün sayısı (Hafta sonu): 8 gün ve üstü Seçilen spor branşı: Masa tenisi Kullanılabilecek spor aletleri (Hafta içi): Dikey bisiklet Mete Yaş: 28 Kullanılabilecek spor aletleri (Hafta sonu): BOŞ		2
Yanlış/İlişkisiz Yanıtlar ve Boş		
Öğrencinin kriterleri göz önünde bulundurmadan yazdığı cevaplar. Öğrenci kartta yer alması gereken başlıklardan ve başlıklara gelecek kavramları yanlış yazmıştır.		1

5- İçindekiler Bölümü Hazırlama

<i>En Doğru Yanıt</i>	Puan
Öğrenci uygun değil seçeneğini işaretler ve açıklamada üçüncü düzey açıklamada her sözcüğün büyük harfle başladığını bu durumun yanlış olduğunu ifade eder. Örnek Yanıt: - Uygun değil, çünkü hayatı ve edebi kişiliği başlıklarının ilk harfleri büyük yazılmıştır. - Uygun değil, çünkü hayatı ve edebi kişiliği başlıklarında her sözcüğün baş harfi büyük yazılmamalı, sadece cümlelerin ilk harfi büyük olmalıydı.	4
<i>Uzak Doğru Yanıtlar</i>	
Öğrenci uygun değil seçeneğini işaretler ve açıklamada üçüncü düzey başlıkta hata olduğunu ifade eder. Ancak hatanın ne olduğu ile ilgili açıklama yapmaz. Örnek Yanıt: - Uygun değil, çünkü üçüncü düzey başlıkta (hayatı ve eserleri) hata var.	3
Öğrenci uygun değil seçeneğini işaretler. Ancak gerekçesi yoktur ya da doğru değildir. Örnek Yanıt: - Uygun değil, çünkü birinci/ikinci düzey başlıkta hata var. - Uygun değil, çünkü üçüncü düzey başlık numaralandırılması yanlış. Öğrenci uygun seçeneğini işaretler. Ancak gerekçesi doğrudur. Örnek Yanıt: - Uygun, çünkü hayatı ve edebi kişiliği başlıklarının ilk harfleri büyük yazılmıştır.	2
<i>Yanlış/İlişkisiz Yanıtlar ve Boş</i>	
Öğrenci uygun seçeneğini işaretler ve cevabına yönelik gerekçe yazar. Örnek Yanıt: - Uygun. Başlıklar açıklamaya göre düzenlenmiştir. Öğrenci uygun seçeneğini işaretler ve gerekçe yazmaz.	1

EK 3. Türkçe Dersi Puanlayıcı Değerlendirme Formu

Türkçe Dersi Puanlayıcı Değerlendirme Formu					
Puanlayıcı kodu/numarası					
Kâğıt numarası	Sorular				
	1 Garanti Belgesi	2 Garanti Belgesi	3 Giriş Kartı	4 Giriş Kartı	5 İçindekiler Bölümü
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					

EK 4. Matematik Dersi Açık Uçlu Sorular

KAN DEĞERLERİ

Aşağıda bazı kan değerlerinin sağlıklı bir bireyde olması gereken en yüksek ve en düşük referans aralıkları verilmiştir (Uç değerler dahildir).

Kan Değeri	Referans Aralığı	
	En Düşük	En Yüksek
Ferritin	9,9	199,9
Trigliserid	0,5	149,9
Glukoz	70,5	99,8
Kalsiyum	8,6	10,6

Eğer bir kan sonucu referans aralığı içerisinde ise sonuç 'Normal', olması gereken en yüksek değerden daha yüksekse 'Yüksek', olması gereken en düşük değerden daha düşükse 'Düşük' olarak sınıflandırılır.

1) Ayşe kan değerlerini ölçtürdüğünde sonuçlar aşağıdaki gibi çıkmıştır.

Kan Değeri	Değer
Ferritin	240,5
Trigliserid	120,9
Glukoz	99,9
Kalsiyum	8,5

Ayşe'nin kan değeri sonuçlarını referans aralıklarına göre Normal, Yüksek ve Düşük olarak aşağıdaki tabloya göre örnekteki gibi sınıflandırınız.

Kan Değeri	Sonuç
Ferritin	Yüksek
Trigliserid	
Glukoz	
Kalsiyum	

2) Kan değerlerini ölçtüren Ahmet'in belirtilen Referans Aralıklarına göre kan değerleri durumu aşağıdaki gibi çıkmıştır.

Kan Değeri	Sonuç
Ferritin	Normal
Trigliserid	Yüksek
Glukoz	Normal
Kalsiyum	Düşük

Ahmet'in yukarıda belirtilen sonuçlara göre kan değerleri nasıl çıkmış olabilir? Referans Aralıklarını ve sonuçları göz önünde bulundurarak aşağıdaki tabloyu 'Ferritin' değerinde verilen örnekteki gibi doldurunuz.

Kan Değeri	Sonuç
Ferritin	26,8
Trigliserid	
Glukoz	
Kalsiyum	

BAZAL METABOLİZMA HIZI (BMH)

İnsan vücudu yalnızca hareket halinde değil, dinlenme halinde de enerji, yani kalori harcar. Bunun sebebi, vücudun dinlenme halinde de temel yaşamsal işlevleri devam ettirmesidir. Bazal metabolizma hızı, kişinin vücudunun temel yaşamsal fonksiyonları yerine getirmek için ihtiyaç duyduğu minimum kalori miktarına karşılık gelir ve kişiden kişiye ve cinsiyetlere göre farklılık gösterir.

Aşağıda kadınlar ve erkekler için bazal metabolizma hızı (BMH) hesaplama formülü verilmiştir.

Kadınlar için BMH;

$$655 + (9,5 \times \text{kilo}) + (1,8 \times \text{boy}) - (4,8 \times \text{yaş})$$

Erkekler için BMH;

$$66 + (13,8 \times \text{kilo}) + (5,1 \times \text{boy}) - (6,8 \times \text{yaş})$$

3) Buna göre aşağıda kilo, boy ve yaş bilgileri verilen Ayşe'nin bazal metabolizma hızını hesaplayarak yandaki boşluğa yazınız. İşlemlerinizi aşağıda gösteriniz.

Kişi	Kilo	Boy	Yaş	BMH
Ayşe	70	160	25	

İşlem:.....
.....
.....
.....
.....
.....
.....
.....

4) 5 yıl önce boyu ve kilosu aynı olan Ayşe'nin bazal metabolizma hızı şimdiki haline göre nasıldır? (Artış veya azalış miktarı şeklinde) Gerekçeleriyle açıklayınız.

.....
.....
.....
.....
.....
.....
.....
.....

ROBOT YARIŞMASI

Bir robot yarışmasında kişilerin hareket eden birer robot tasarımı yapmaları isteniyor ve bu tasarım daha sonra puanlanıyor.

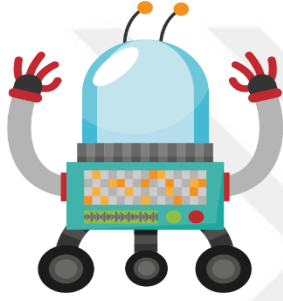
Tasarım puanlanırken;

$$A + (\text{Kol sayısı} \times 50) + (\text{Anten sayısı} \times 35,5) - (\text{Tekerlek sayısı} \times 110,5) + K$$

A katsayısı robot belirtilen sürede tamamlandıysa +100 puan olarak etki etmektedir.

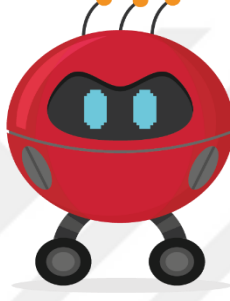
K katsayısı robot hareket edebiliyorsa +100 puan olarak, hareket edemiyorsa -100 puan olarak hesaplanmaktadır.

Aşağıda belirtilen sürelerde tamamlanmış 3 tane robot tasarımı verilmiştir.



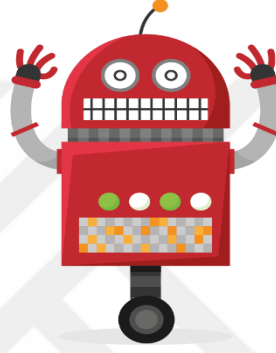
A

Hareket edebiliyor.



B

Hareket edebiliyor.



C

Hareket edemiyor.

5) Robotların puanını hesaplayınız.

A puanı:.....

.....
.....
.....
.....

B puanı:.....

.....
.....
.....
.....

C puanı:.....

.....
.....
.....
.....

6) En yüksek puanı alan robotun birinci olduđu bir yarışma için C robotunun en yüksek puanı alabilmesi için verilen hesaplamanın herhangi bir yerinde (işaretlerde veya katsayılarda) düzeltme yapmanız isteniyor.

Formülde veya katsayılarda nasıl bir düzeltme yaparsınız? Düzelttiğiniz formül için robotların puanlarını hesaplayınız ve C robotunun birinciliğini destekleyiniz.

Formülün yeni hali:

.....

.....

.....

.....

.....

.....

.....

A puanı:

.....

.....

.....

.....

.....

.....

.....

B puanı:

.....

.....

.....

.....

.....

.....

.....

C puanı:

.....

.....

.....

.....

.....

.....

.....

EK 5. Matematik Dersi Dereceli Puanlama Anahtarı**KAN DEĞERLERİ- 1**

Tam Doğru Yanıt	Puan
Tabloda Trigliserid değeri karşısına ‘Normal’, Glukoz karşısına ‘Yüksek’, Kalsiyum karşısına ‘Düşük’ yazan yanıtlar.	4
Kısmen Doğru Yanıtlar	
Tabloda Trigliserid değeri karşısına ‘Normal’, Glukoz değeri karşısına ‘Yüksek’ yazan ancak Kalsiyum karşısına ‘Normal’ veya ‘Yüksek’ yazan yanıtlar.	3
Tabloda Glukoz karşısına ‘Yüksek’, Kalsiyum karşısına ‘Düşük’ yazan ancak Trigliserid değeri karşısına ‘Yüksek’ veya ‘Düşük’ yazan yanıtlar.	3
Tabloda Trigliserid değeri karşısına ‘Normal’, Kalsiyum karşısına ‘Düşük’ yazan ancak Glukoz karşısına ‘Normal’ veya ‘Düşük’ yazan yanıtlar.	3
Tabloda Trigliserid değeri karşısına ‘Normal’ yazan ancak Glukoz karşısına ‘Normal’ veya ‘Düşük’, Kalsiyum karşısına ise ‘Normal’ veya ‘Yüksek’ yazan yanıtlar.	2
Tabloda Glukoz karşısına ‘Yüksek’ yazan ancak Kalsiyum karşısına ‘Normal’ veya ‘Yüksek’, Trigliserid değeri karşısına ise ‘Yüksek’ veya ‘Düşük’ yazan yanıtlar.	2
Tabloda Kalsiyum karşısına ‘Düşük’ yazan ancak Trigliserid değeri karşısına ‘Yüksek’ veya ‘Düşük’ yazan, Glukoz karşısına ise ‘Normal’ veya ‘Düşük’ yazan yanıtlar.	2
Yanlış Yanıtlar	
Tabloda Trigliserid değeri karşısına ‘Yüksek’ veya ‘Düşük’, Glukoz karşısına ‘Normal’ veya ‘Düşük’ ve Kalsiyum karşısına ‘Normal’ veya ‘Yüksek’ yazan yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

KAN DEĞERLERİ- 2

Tam Doğru Yanıt	Puan
Tabloda Trigliserid değeri karşısına 149,9'dan büyük değerler, Glukoz karşısına 70.5 ve 99.8 arasında değerler (70.5 ve 99.8 dahil) ve Kalsiyum karşısına 8.6'dan küçük değerler yazan yanıtlar.	4
Kısmen Doğru Yanıtlar	
Tabloda Trigliserid değeri karşısına 149,9'dan büyük değerler, Glukoz değeri karşısına 70.5 ve 99.8 arasında değerler (70.5 ve 99.8 dahil) yazan ancak Kalsiyum karşısına 8.6'dan büyük değerler yazan yanıtlar.	3
Tabloda Glukoz karşısına '70.5 ve 99.8 arasında değerler (70.5 ve 99.8 dahil) ve Kalsiyum karşısına 8.6'dan küçük değerler yazan ancak Trigliserid değeri karşısına 149,9'dan küçük değerler yazan yanıtlar.	3
Tabloda Trigliserid değeri karşısına 149,9'dan büyük değerler, Kalsiyum karşısına 8.6'dan küçük değerler yazan ancak Glukoz karşısına 70.5'tan küçük veya 99.8'den büyük değerler yazan yanıtlar.	3
Tabloda Trigliserid değeri karşısına 149,9'dan büyük değerler yazan ancak Glukoz karşısına 70.5'tan küçük veya 99.8'den büyük değerler, Kalsiyum karşısına ise 8.6'dan büyük değerler yazan yanıtlar.	2
Tabloda Glukoz karşısına 70.5 ve 99.8 arasında değerler (70.5 ve 99.8 dahil) yazan ancak Kalsiyum karşısına 8.6'dan büyük değerler, Trigliserid değeri karşısına ise 149,9'dan küçük değerler yazan yanıtlar.	2
Tabloda Kalsiyum karşısına 8.6'dan küçük değerler yazan ancak Trigliserid değeri karşısına 149,9'dan küçük değerler yazan, Glukoz karşısına ise 70.5'tan küçük veya 99.8'den büyük değerler yazan yanıtlar.	2
Yanlış Yanıtlar	
Tabloda Trigliserid değeri karşısına 149,9'dan küçük değerler, Glukoz karşısına 70.5'tan küçük veya 99.8'den büyük değerler ve Kalsiyum karşısına 8.6'dan büyük değerler yazan yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

BAZAL METABOLİZMA HIZI-3

Tam Doğru Yanıt	Puan
Bazal metabolizma hızını 1488 bulmuş ve işlem adımlarını da tam doğru olarak göstermiş yanıtlar. İşlem adımları: $655 + (9.5 \times 70) + (1.8 \times 160) - (4.8 \times 25) = 655 + 665 + 288 - 120 = 1488$	4
Kısmen Doğru Yanıtlar	
Çarpma işlemlerinden herhangi birinde işlem hatası yapmış ve bu yüzden sonucu yanlış bulmuş yanıtlar.	3
(4.8 x 25) sonucunu hesaplamış ancak işlem adımında çıkarma yerine toplama işlemi yaparak sonucu 1728 bulmuş yanıtlar.	3
Ayşe'nin bazal metabolizma hızını erkek bireylerin formülüne göre tam doğru hesaplayarak 1678 bulmuş yanıtlar.	3
Çarpma işlemlerinden herhangi ikisinde işlem hatası yapmış, toplama işlemlerini doğru yapmış ve bu yüzden sonucu yanlış bulmuş yanıtlar.	2
Ayşe'nin bazal metabolizma hızını erkek bireylerin formülüne göre hesaplayıp işlem adımlarının yalnızca bir tanesinde hata yaparak sonucu yanlış bulmuş yanıtlar.	2
Yanlış Yanıtlar	
Ayşe'nin bazal metabolizma hızını erkek bireylerin formülüne göre hesaplayıp işlem adımlarında birden fazla şekilde hata yaparak sonucu yanlış bulmuş yanıtlar.	1
Çarpma işlemlerinin tamamında işlem hatası yapmış ve bu yüzden sonucu yanlış bulmuş yanıtlar.	1
Çarpma işlemlerinin 2 tanesinde ve toplama işlemlerinin de bir veya birkaçında işlem hatası yapmış yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

BAZAL METABOLİZMA HIZI- 4

Tam Doğru Yanıt	Puan
Yaş 5 azaldığında $-(4.8 \times 5) = -24$ olduğundan, 5 yıl önceki BHM'nin şimdiki BM'ye göre 24 puan daha az azalacağını bu yüzden 24 puan daha yüksek olması gerektiğini vurgulayan yanıtlar.	4
Ayşe'nin 5 yıl önceki BMH'sini işlem adımları gerçekleştirerek 24 puan daha yüksek bularak bunu vurgulayan yanıtlar.	4
Kısmen Doğru Yanıtlar	
Yaş 5 azaldığında $(4.8 \times 5) = 24$ olduğundan, 5 yıl önceki BHM'nin şimdiki BM'ye göre 24 puan AZALACAĞINI yazan yanıtlar.	3
Ayşe'nin 5 yıl önceki BMH'sini işlem adımları gerçekleştirerek hesaplayan ancak son adımda 24 puan daha AZ olduğunu vurgulayan yanıtlar.	3
Gerekli hesaplamaları gerçekleştirerek 24 bulan ancak FAZLA olması ifadesini kullanmayan yanıtlar.	3
Yalnızca öncekine göre daha fazla olacağını yazan ancak herhangi bir hesaplama yapmayan veya 24 değerini belirtmeyen yanıtlar.	2
Öncekine göre daha fazla olacağını yazan ancak işlem hatası nedeniyle 24 dışında değer bulan yanıtlar.	2
Yanlış Yanıtlar	
Yalnızca öncekine göre daha AZ olacağını yazan ancak herhangi bir hesaplama yapmayan yanıtlar.	1
Öncekine göre daha AZ olacağını yazan ve hesaplama sonucu 24 dışında değer bulan yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

ROBOT YARIŞMASI- 5

Tam Doğru Yanıt	Puan
Her 3 robotun puanını da tam doğru hesaplayarak A robotunun puanını 39.5, B robotunun puanını 85.5 ve C robotunun puanını 25 olarak hesaplayan yanıtlar.	4
Kısmen Doğru Yanıtlar	
A robotunun puanını 39.5, B robotunun puanını 85.5 olarak hesaplayan ancak C robotunun puanını 25 dışında bulan yanıtlar.	3
A robotunun puanını 39.5, C robotunun puanını 25 olarak hesaplayan ancak B robotunun puanını 85.5 dışında bulan yanıtlar.	3
B robotunun puanını 85.5 ve C robotunun puanını 25 olarak hesaplayan ancak A robotunun puanını 39.5 dışında bulan yanıtlar.	3
A robotunun puanını 39.5 bulan ancak B robotunun puanını 85.5 dışında ve C robotunun puanını 25 dışında bulan yanıtlar.	2
B robotunun puanını 85.5 bulan ancak A robotunun puanını 39.5 dışında ve C robotunun puanını 25 dışında bulan yanıtlar.	2
C robotunun puanını 25 bulan ancak B robotunun puanını 85.5 dışında ve A robotunun puanını 39.5 dışında bulan yanıtlar.	2
Yanlış Yanıtlar	
A robotunun puanını 39.5 dışında, B robotunun puanını 85.5 dışında ve C robotunun puanını 25 dışında bulan yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

ROBOT YARIŞMASI- 6

Tam Doğru Yanıt	Puan
Düzeltilme yaptığı formülde C robotunu avantajlı hale geçirerek robotların puanını yeni formüle göre doğru hesaplayan ve puan üstünlüğüne göre C robotunun daha avantajlı olduğunu ispatlayan yanıtlar. Örnek yanıt: K katsayısını robot hareket edebiliyorsa +50, edemiyorsa -50 olarak değiştiririz.	4
Kısmen Doğru Yanıtlar	
Formülde C robotunu avantajlı konuma getirecek doğru formülü yazan ancak robotların puanlarını hesaplamayıp yalnızca C robotunun puanını hesaplayarak üstün olduğunu belirten yanıtlar.	3
Formülde C robotunu avantajlı konuma getirecek doğru formülü yazan ancak herhangi bir hesaplama yapmayan yanıtlar.	2
Yanlış Yanıtlar	
Formülde C robotunu avantajlı konuma getirecek doğru yanıt yazamayan yanıtlar.	1
Boş	1
Diğer Yanıtlar	
Soruyla ilgili verilen alakasız yanıtlar.	1

EK 6. Matematik Dersi Puanlayıcı Değerlendirme Formu

Matematik Dersi Puanlayıcı Değerlendirme Formu					
Puanlayıcı kodu/numarası					
Kâğıt numarası	Sorular				
	1 Kan değerleri	2 Kan değerleri	3 Bazal Metabolizma Hızı	4 Bazal Metabolizma Hızı	5 Robot Yarışması
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					

EK 7. Ön Test Türkçe Puanlayıcı Grubu *Puanlayıcı X Birey* Etkileşimleri

Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Yanlılık Büyüklüğü	SH	t	sd	p	infit	outfit
P1	12	20,00	18,96	0,65	1,36	0,48	4	0,66	0,10	0,10
P1	13	20,00	17,31	1,49	1,36	1,10	4	0,33	0,10	0,10
P1	26	20,00	19,41	0,15	1,36	0,11	4	0,92	0,10	0,10
P1	37	20,00	18,37	1,05	1,36	0,77	4	0,48	0,10	0,10
P2	26	20,00	19,54	-0,07	1,36	-0,05	4	0,96	0,10	0,10
P3	12	20,00	19,44	0,11	1,36	0,08	4	0,94	0,10	0,10
P3	26	20,00	19,67	-0,39	1,36	-0,29	4	0,79	0,10	0,10
P3	28	20,00	19,40	0,17	1,36	0,13	4	0,91	0,10	0,10
P4	12	20,00	18,61	0,91	1,36	0,67	4	0,54	0,10	0,10
P5	12	20,00	19,22	0,41	1,36	0,30	4	0,78	0,10	0,10
P5	26	20,00	19,55	-0,09	1,36	-0,07	4	0,95	0,10	0,10
P5	37	20,00	18,77	0,80	1,36	0,59	4	0,59	0,10	0,10
P6	12	20,00	19,39	0,19	1,36	0,14	4	0,90	0,10	0,10
P6	38	20,00	18,02	1,21	1,36	0,89	4	0,42	0,10	0,10
P8	13	20,00	17,66	1,36	1,36	1,00	4	0,37	0,10	0,10
P8	22	20,00	17,38	1,46	1,36	1,08	4	0,34	0,10	0,10
P8	23	20,00	18,91	0,70	1,36	0,51	4	0,63	0,10	0,10
P8	25	20,00	17,66	1,36	1,36	1,00	4	0,37	0,10	0,10
P8	26	20,00	19,48	0,03	1,36	0,02	4	0,98	0,10	0,10
P8	28	20,00	19,04	0,59	1,36	0,43	4	0,69	0,10	0,10
P8	37	20,00	18,59	0,92	1,36	0,68	4	0,53	0,10	0,10
P8	38	20,00	17,11	1,55	1,36	1,14	4	0,32	0,10	0,10
P9	12	20,00	19,17	0,45	1,36	0,33	4	0,75	0,10	0,10
P9	13	20,00	17,84	1,29	1,36	0,95	4	0,40	0,10	0,10
P9	25	20,00	17,84	1,29	1,36	0,95	4	0,40	0,10	0,10
P9	26	20,00	19,52	-0,04	1,36	-0,03	4	0,98	0,10	0,10
P9	28	20,00	19,11	0,52	1,36	0,38	4	0,72	0,10	0,10
P9	37	20,00	18,70	0,85	1,36	0,63	4	0,57	0,10	0,10
P9	39	20,00	17,07	1,57	1,36	1,16	4	0,31	0,10	0,10
P10	12	20,00	19,18	0,45	1,36	0,33	4	0,76	0,10	0,10
P10	25	20,00	17,86	1,28	1,36	0,94	4	0,40	0,10	0,10
P10	26	20,00	19,53	-0,05	1,36	-0,04	4	0,97	0,10	0,10
P10	28	20,00	19,12	0,51	1,36	0,37	4	0,73	0,10	0,10
P10	37	20,00	18,71	0,84	1,36	0,62	4	0,57	0,10	0,10
P11	23	20,00	18,69	0,86	1,36	0,63	4	0,56	0,10	0,10
P12	26	20,00	19,38	0,19	1,36	0,14	4	0,89	0,10	0,10
P12	37	20,00	18,29	1,08	1,36	0,80	4	0,47	0,10	0,10
P13	23	20,00	18,54	0,95	1,36	0,70	4	0,52	0,10	0,10
P13	25	20,00	16,92	1,61	1,36	1,19	4	0,30	0,10	0,10
P13	37	20,00	18,11	1,17	1,36	0,86	4	0,44	0,10	0,10
P6	28	16,00	19,35	-1,63	0,51	-3,17	4	0,03	1,00	0,80
P6	37	15,00	19,05	-1,54	0,49	-3,16	4	0,03	0,90	1,00
P9	16	6,00	10,78	-1,53	0,89	-1,72	4	0,16	0,40	0,20
P8	16	6,00	10,44	-1,46	0,89	-1,64	4	0,18	0,60	0,50

P6	22	13,00	18,22	-1,45	0,46	-3,14	4	0,03	2,90	3,50
P4	1	6,00	10,38	-1,45	0,89	-1,63	4	0,18	0,60	0,50
P1	16	6,00	9,86	-1,34	0,89	-1,50	4	0,21	0,40	0,20
P10	23	17,00	19,01	-0,97	0,56	-1,73	4	0,16	0,20	0,40
P8	35	10,00	14,50	-0,96	0,47	-2,07	4	0,11	0,80	0,70
P6	26	19,00	19,64	-0,93	0,94	-1,00	4	0,37	1,00	2,60
P12	12	17,00	18,92	-0,90	0,56	-1,60	4	0,18	3,10	4,50
P6	25	16,00	18,42	-0,85	0,51	-1,66	4	0,17	0,80	1,10
P10	13	15,00	17,86	-0,84	0,49	-1,73	4	0,16	0,20	0,30
P7	26	18,00	19,24	-0,84	0,66	-1,27	4	0,27	1,10	2,00
P4	2	10,00	13,93	-0,84	0,47	-1,79	4	0,15	0,40	0,60
P6	40	15,00	17,82	-0,82	0,49	-1,69	4	0,17	0,80	0,90
P13	12	17,00	18,80	-0,81	0,56	-1,44	4	0,22	3,10	4,50
P8	11	12,00	15,56	-0,80	0,46	-1,74	4	0,16	0,80	0,80
P13	38	13,00	16,26	-0,77	0,46	-1,66	4	0,17	0,70	0,80
P9	1	9,00	12,55	-0,76	0,49	-1,56	4	0,19	0,80	1,10
P9	30	15,00	17,65	-0,76	0,49	-1,56	4	0,19	0,30	0,40
P12	16	7,00	9,69	-0,75	0,63	-1,19	4	0,30	0,20	0,20
P8	31	8,00	11,18	-0,74	0,54	-1,38	4	0,24	1,30	0,80
P14	36	8,00	11,10	-0,72	0,54	-1,35	4	0,25	0,40	0,30
P14	26	18,00	19,13	-0,72	0,66	-1,09	4	0,34	1,10	2,00
P3	37	18,00	19,13	-0,72	0,66	-1,09	4	0,34	0,10	0,10
P7	12	17,00	18,65	-0,71	0,56	-1,26	4	0,28	1,90	2,70
P6	20	14,00	16,64	-0,66	0,47	-1,39	4	0,24	1,70	1,60
P3	13	17,00	18,55	-0,64	0,56	-1,15	4	0,32	0,20	0,40
P1	19	8,00	10,73	-0,64	0,54	-1,20	4	0,30	2,00	1,50
P4	38	13,00	15,76	-0,64	0,46	-1,37	4	0,24	0,40	0,40
P1	35	11,00	13,95	-0,63	0,46	-1,37	4	0,24	1,50	1,40
P11	25	15,00	17,20	-0,60	0,49	-1,23	4	0,29	1,80	2,20
P7	37	16,00	17,88	-0,60	0,51	-1,17	4	0,31	3,10	3,90
P1	39	14,00	16,40	-0,59	0,47	-1,25	4	0,28	0,90	0,90
P2	7	10,00	12,74	-0,58	0,47	-1,24	4	0,28	0,40	0,40
P12	31	8,00	10,41	-0,57	0,54	-1,07	4	0,34	0,40	0,30
P3	27	14,00	16,31	-0,56	0,47	-1,19	4	0,30	0,30	0,40
P4	6	8,00	10,31	-0,55	0,54	-1,03	4	0,36	0,30	0,60
P4	40	13,00	15,41	-0,55	0,46	-1,19	4	0,30	1,20	1,20
P11	36	10,00	12,60	-0,55	0,47	-1,17	4	0,31	0,40	0,40
P5	14	14,00	16,24	-0,54	0,47	-1,15	4	0,31	1,50	1,80
P10	18	12,00	14,50	-0,54	0,46	-1,19	4	0,30	0,70	0,70
P7	22	14,00	16,22	-0,54	0,47	-1,14	4	0,32	1,80	1,80
P6	14	15,00	16,99	-0,53	0,49	-1,09	4	0,34	1,10	1,70
P12	1	9,00	11,43	-0,53	0,49	-1,08	4	0,34	0,60	1,00
P11	30	15,00	16,96	-0,52	0,49	-1,07	4	0,34	0,60	0,80
P14	13	14,00	16,14	-0,52	0,47	-1,09	4	0,34	3,60	3,70
P14	25	14,00	16,14	-0,52	0,47	-1,09	4	0,34	1,80	1,90
P4	23	17,00	18,30	-0,51	0,56	-0,91	4	0,42	0,20	0,40
P11	37	17,00	18,29	-0,50	0,56	-0,90	4	0,42	3,10	4,90
P3	20	15,00	16,89	-0,50	0,49	-1,03	4	0,36	0,50	0,50

P14	7	8,00	10,02	-0,49	0,54	-0,92	4	0,41	0,70	0,60
P8	29	17,00	18,26	-0,49	0,56	-0,87	4	0,43	1,20	0,80
P5	27	13,00	15,16	-0,49	0,46	-1,05	4	0,35	0,50	0,50
P13	31	8,00	9,99	-0,48	0,54	-0,91	4	0,42	0,40	0,30
P6	11	15,00	16,82	-0,48	0,49	-0,98	4	0,38	2,10	3,80
P11	28	18,00	18,84	-0,47	0,66	-0,71	4	0,51	1,10	2,00
P8	1	10,00	12,20	-0,46	0,47	-1,00	4	0,38	1,00	1,10
P2	27	13,00	15,05	-0,46	0,46	-1,00	4	0,38	0,50	0,50
P6	35	14,00	15,90	-0,45	0,47	-0,96	4	0,39	1,40	1,60
P10	11	14,00	15,88	-0,45	0,47	-0,95	4	0,40	0,20	0,30
P3	18	14,00	15,88	-0,45	0,47	-0,95	4	0,40	0,30	0,40
P1	38	15,00	16,71	-0,45	0,49	-0,91	4	0,41	1,10	1,70
P7	11	12,00	14,06	-0,44	0,46	-0,97	4	0,39	0,80	0,80
P13	8	10,00	12,10	-0,44	0,47	-0,95	4	0,40	1,50	1,20
P9	11	14,00	15,85	-0,44	0,47	-0,93	4	0,40	1,50	1,70
P13	1	9,00	11,00	-0,44	0,49	-0,90	4	0,42	0,60	1,00
P9	27	13,00	14,95	-0,44	0,46	-0,94	4	0,40	0,50	0,50
P11	26	19,00	19,38	-0,43	0,94	-0,46	4	0,67	1,00	1,60
P14	23	17,00	18,11	-0,42	0,56	-0,74	4	0,50	2,00	3,90
P7	13	15,00	16,58	-0,41	0,49	-0,84	4	0,45	2,70	3,10
P10	16	9,00	10,82	-0,40	0,49	-0,82	4	0,46	0,60	1,00
P10	38	16,00	17,35	-0,40	0,51	-0,78	4	0,48	0,40	0,70
P2	37	18,00	18,74	-0,40	0,66	-0,60	4	0,58	0,10	0,10
P3	9	16,00	17,33	-0,40	0,51	-0,77	4	0,48	0,60	0,50
P1	25	16,00	17,31	-0,39	0,51	-0,76	4	0,49	1,00	2,20
P4	19	8,00	9,54	-0,38	0,54	-0,72	4	0,51	0,10	0,10
P11	35	12,00	13,77	-0,38	0,46	-0,83	4	0,45	0,90	0,80
P6	31	11,00	12,81	-0,38	0,46	-0,83	4	0,45	1,70	1,60
P4	25	15,00	16,47	-0,38	0,49	-0,78	4	0,48	1,10	1,70
P5	18	13,00	14,67	-0,37	0,46	-0,80	4	0,47	0,50	0,50
P3	10	14,00	15,57	-0,37	0,47	-0,78	4	0,48	0,30	0,40
P3	39	17,00	18,00	-0,36	0,56	-0,65	4	0,55	0,20	0,40
P8	27	13,00	14,64	-0,36	0,46	-0,79	4	0,48	0,50	0,50
P7	40	14,00	15,53	-0,36	0,47	-0,76	4	0,49	1,30	1,10
P1	8	11,00	12,70	-0,36	0,46	-0,78	4	0,48	1,80	1,60
P7	36	10,00	11,68	-0,36	0,47	-0,76	4	0,49	0,40	0,40
P11	13	16,00	17,20	-0,35	0,51	-0,68	4	0,53	2,50	2,70
P2	18	13,00	14,57	-0,35	0,46	-0,75	4	0,49	0,50	0,50
P3	23	19,00	19,32	-0,34	0,94	-0,37	4	0,73	1,00	1,70
P13	26	19,00	19,32	-0,34	0,94	-0,36	4	0,73	1,00	1,60
P1	18	12,00	13,58	-0,34	0,46	-0,74	4	0,50	0,50	0,40
P13	11	13,00	14,51	-0,33	0,46	-0,72	4	0,51	0,60	0,80
P1	9	14,00	15,43	-0,33	0,47	-0,71	4	0,52	0,80	0,80
P11	29	17,00	17,90	-0,32	0,56	-0,58	4	0,59	0,60	0,60
P13	27	12,00	13,52	-0,32	0,46	-0,71	4	0,52	0,30	0,30
P14	2	12,00	13,51	-0,32	0,46	-0,71	4	0,52	0,40	0,40
P11	40	15,00	16,27	-0,32	0,49	-0,66	4	0,55	1,40	1,30
P6	9	16,00	17,11	-0,32	0,51	-0,62	4	0,57	0,60	0,50

P12	8	11,00	12,52	-0,32	0,46	-0,70	4	0,52	1,80	1,60
P5	34	12,00	13,49	-0,32	0,46	-0,69	4	0,53	0,70	0,70
P5	4	14,00	15,36	-0,32	0,47	-0,67	4	0,54	0,70	0,70
P5	33	13,00	14,40	-0,31	0,46	-0,67	4	0,54	0,50	0,50
P3	2	16,00	17,06	-0,30	0,51	-0,59	4	0,58	0,30	0,40
P11	1	10,00	11,43	-0,30	0,47	-0,65	4	0,55	0,80	0,90
P4	9	13,00	14,33	-0,29	0,46	-0,63	4	0,56	1,50	1,40
P5	10	13,00	14,33	-0,29	0,46	-0,63	4	0,56	0,50	0,50
P2	34	12,00	13,38	-0,29	0,46	-0,64	4	0,56	0,70	0,70
P2	4	14,00	15,26	-0,29	0,47	-0,62	4	0,57	0,70	0,70
P13	35	12,00	13,37	-0,29	0,46	-0,64	4	0,56	1,00	0,90
P6	39	17,00	17,82	-0,29	0,56	-0,52	4	0,63	0,20	0,40
P4	32	11,00	12,38	-0,29	0,46	-0,63	4	0,56	0,20	0,20
P11	6	10,00	11,35	-0,29	0,47	-0,61	4	0,57	0,80	0,70
P7	14	13,00	14,30	-0,29	0,46	-0,62	4	0,57	2,20	2,30
P4	37	17,00	17,81	-0,28	0,56	-0,51	4	0,64	0,30	0,70
P2	33	13,00	14,29	-0,28	0,46	-0,61	4	0,57	0,50	0,50
P3	25	18,00	18,55	-0,28	0,66	-0,42	4	0,70	0,10	0,10
P14	5	11,00	12,32	-0,28	0,46	-0,60	4	0,58	0,20	0,20
P14	35	11,00	12,32	-0,28	0,46	-0,60	4	0,58	0,60	0,60
P2	10	13,00	14,22	-0,27	0,46	-0,58	4	0,59	0,50	0,50
P6	23	19,00	19,26	-0,27	0,94	-0,29	4	0,79	1,00	1,70
P11	14	14,00	15,12	-0,26	0,47	-0,55	4	0,61	1,70	1,70
P4	14	13,00	14,17	-0,26	0,46	-0,56	4	0,61	1,20	1,20
P10	10	13,00	14,15	-0,25	0,46	-0,54	4	0,62	0,50	0,50
P11	22	16,00	16,89	-0,25	0,51	-0,49	4	0,65	0,80	1,10
P12	22	16,00	16,89	-0,25	0,51	-0,49	4	0,65	2,00	2,10
P1	27	13,00	14,09	-0,24	0,46	-0,52	4	0,63	0,50	0,50
P8	6	11,00	12,13	-0,24	0,46	-0,52	4	0,63	1,60	1,50
P13	29	17,00	17,68	-0,24	0,56	-0,42	4	0,70	0,60	0,60
P5	5	14,00	15,02	-0,23	0,47	-0,50	4	0,65	0,70	0,70
P3	6	13,00	14,06	-0,23	0,46	-0,50	4	0,64	0,60	0,60
P10	30	17,00	17,67	-0,23	0,56	-0,41	4	0,70	0,70	1,00
P6	2	16,00	16,82	-0,23	0,51	-0,45	4	0,68	1,40	1,80
P4	33	11,00	12,07	-0,22	0,46	-0,49	4	0,65	0,80	0,70
P14	12	18,00	18,45	-0,22	0,66	-0,33	4	0,76	1,10	2,00
P3	32	15,00	15,88	-0,22	0,49	-0,45	4	0,68	0,50	0,50
P4	26	19,00	19,21	-0,21	0,94	-0,22	4	0,83	1,00	1,60
P9	2	15,00	15,85	-0,21	0,49	-0,43	4	0,69	1,10	1,10
P2	5	14,00	14,91	-0,21	0,47	-0,44	4	0,68	0,70	0,70
P5	20	15,00	15,84	-0,21	0,49	-0,42	4	0,69	0,50	0,50
P11	2	14,00	14,89	-0,21	0,47	-0,43	4	0,69	0,10	0,10
P10	29	18,00	18,42	-0,20	0,66	-0,31	4	0,78	0,70	0,40
P6	13	18,00	18,42	-0,20	0,66	-0,31	4	0,78	0,10	0,10
P12	27	13,00	13,92	-0,20	0,46	-0,44	4	0,69	0,50	0,50
P8	4	14,00	14,85	-0,20	0,47	-0,41	4	0,70	0,70	0,70
P10	5	14,00	14,84	-0,19	0,47	-0,41	4	0,70	0,70	0,70
P4	27	12,00	12,91	-0,19	0,46	-0,42	4	0,70	0,50	0,40

P14	37	17,00	17,57	-0,19	0,56	-0,34	4	0,75	3,10	4,90
P2	12	19,00	19,20	-0,19	0,94	-0,21	4	0,85	1,00	1,60
P7	35	12,00	12,89	-0,19	0,46	-0,41	4	0,70	0,90	0,80
P9	5	14,00	14,81	-0,19	0,47	-0,39	4	0,72	0,70	0,70
P13	30	16,00	16,67	-0,19	0,51	-0,36	4	0,74	0,20	0,30
P5	8	13,00	13,84	-0,18	0,46	-0,40	4	0,71	0,50	0,50
P3	34	14,00	14,80	-0,18	0,47	-0,38	4	0,72	0,80	0,80
P2	20	15,00	15,74	-0,18	0,49	-0,37	4	0,73	0,50	0,50
P5	7	12,00	12,85	-0,18	0,46	-0,39	4	0,71	0,70	0,70
P12	35	13,00	13,77	-0,17	0,46	-0,36	4	0,73	1,20	1,10
P10	20	15,00	15,67	-0,17	0,49	-0,34	4	0,75	0,50	0,50
P13	4	13,00	13,74	-0,16	0,46	-0,35	4	0,74	0,50	0,50
P5	31	11,00	11,77	-0,16	0,46	-0,35	4	0,74	0,40	0,40
P4	5	12,00	12,76	-0,16	0,46	-0,35	4	0,74	0,50	0,40
P14	28	18,00	18,34	-0,16	0,66	-0,24	4	0,82	1,10	2,00
P2	8	13,00	13,73	-0,16	0,46	-0,34	4	0,75	0,50	0,50
P7	25	16,00	16,58	-0,16	0,51	-0,31	4	0,77	0,40	0,70
P11	38	16,00	16,57	-0,16	0,51	-0,31	4	0,77	0,40	0,70
P12	38	16,00	16,57	-0,16	0,51	-0,31	4	0,77	1,00	2,20
P10	36	13,00	13,72	-0,16	0,46	-0,34	4	0,75	0,50	0,50
P6	6	13,00	13,72	-0,16	0,46	-0,34	4	0,75	1,30	1,50
P9	20	15,00	15,64	-0,16	0,49	-0,32	4	0,76	0,50	0,50
P5	28	19,00	19,16	-0,16	0,94	-0,17	4	0,88	1,00	1,70
P7	19	9,00	9,67	-0,15	0,49	-0,31	4	0,77	1,20	1,00
P5	32	14,00	14,67	-0,15	0,47	-0,32	4	0,76	0,70	0,70
P8	30	17,00	17,45	-0,15	0,56	-0,27	4	0,80	1,20	0,80
P10	19	11,00	11,72	-0,15	0,46	-0,33	4	0,76	1,20	1,00
P9	36	13,00	13,69	-0,15	0,46	-0,32	4	0,76	0,50	0,50
P1	7	11,00	11,68	-0,14	0,46	-0,31	4	0,77	0,40	0,40
P9	19	11,00	11,68	-0,14	0,46	-0,31	4	0,77	1,20	1,00
P10	7	12,00	12,66	-0,14	0,46	-0,30	4	0,78	0,70	0,70
P2	31	11,00	11,65	-0,14	0,46	-0,30	4	0,78	0,40	0,40
P7	29	17,00	17,40	-0,13	0,56	-0,24	4	0,82	0,60	0,60
P3	4	16,00	16,48	-0,13	0,51	-0,26	4	0,81	0,60	0,50
P2	28	19,00	19,14	-0,13	0,94	-0,14	4	0,89	1,00	1,70
P2	32	14,00	14,57	-0,13	0,47	-0,27	4	0,80	0,70	0,70
P6	30	18,00	18,27	-0,13	0,66	-0,19	4	0,86	0,80	1,40
P8	5	14,00	14,50	-0,11	0,47	-0,24	4	0,82	0,70	0,70
P10	32	14,00	14,50	-0,11	0,47	-0,24	4	0,82	0,70	0,70
P14	11	13,00	13,51	-0,11	0,46	-0,24	4	0,82	0,60	0,80
P7	1	10,00	10,51	-0,11	0,47	-0,24	4	0,83	0,80	0,90
P9	18	14,00	14,46	-0,10	0,47	-0,22	4	0,84	0,70	0,70
P9	32	14,00	14,46	-0,10	0,47	-0,22	4	0,84	0,70	0,70
P5	9	16,00	16,37	-0,10	0,51	-0,20	4	0,85	0,60	0,50
P4	39	15,00	15,41	-0,10	0,49	-0,21	4	0,85	1,10	1,70
P7	6	10,00	10,44	-0,09	0,47	-0,20	4	0,85	0,80	0,70
P14	38	15,00	15,39	-0,09	0,49	-0,19	4	0,86	1,10	1,70
P8	12	19,00	19,10	-0,09	0,94	-0,10	4	0,92	1,00	1,60

P12	18	13,00	13,41	-0,09	0,46	-0,19	4	0,86	0,50	0,50
P7	30	16,00	16,31	-0,08	0,51	-0,16	4	0,88	0,20	0,30
P2	9	16,00	16,28	-0,08	0,51	-0,15	4	0,89	0,60	0,50
P3	14	17,00	17,22	-0,07	0,56	-0,13	4	0,90	0,20	0,40
P11	39	16,00	16,27	-0,07	0,51	-0,14	4	0,90	1,10	1,30
P1	34	12,00	12,34	-0,07	0,46	-0,15	4	0,88	0,70	0,70
P1	4	14,00	14,31	-0,07	0,47	-0,15	4	0,89	0,70	0,70
P5	39	17,00	17,22	-0,07	0,56	-0,13	4	0,91	0,20	0,40
P13	16	9,00	9,29	-0,07	0,49	-0,14	4	0,89	0,60	1,00
P8	19	11,00	11,33	-0,07	0,46	-0,15	4	0,89	1,20	1,00
P11	9	15,00	15,27	-0,07	0,49	-0,13	4	0,90	1,10	0,90
P10	34	13,00	13,30	-0,06	0,46	-0,14	4	0,90	0,50	0,50
P1	33	13,00	13,29	-0,06	0,46	-0,13	4	0,90	0,70	0,60
P8	7	12,00	12,28	-0,06	0,46	-0,13	4	0,91	0,70	0,70
P6	4	16,00	16,21	-0,06	0,51	-0,11	4	0,92	0,90	1,30
P9	34	13,00	13,26	-0,06	0,46	-0,12	4	0,91	0,50	0,50
P7	34	11,00	11,24	-0,05	0,46	-0,11	4	0,92	0,60	0,50
P3	5	16,00	16,19	-0,05	0,51	-0,10	4	0,93	0,60	0,50
P10	33	14,00	14,22	-0,05	0,47	-0,10	4	0,92	0,70	0,70
P5	23	19,00	19,05	-0,05	0,94	-0,05	4	0,96	1,00	1,70
P1	10	13,00	13,21	-0,05	0,46	-0,10	4	0,93	0,50	0,50
P2	39	17,00	17,14	-0,05	0,56	-0,08	4	0,94	0,20	0,40
P3	7	14,00	14,20	-0,04	0,47	-0,09	4	0,93	0,30	0,40
P9	33	14,00	14,18	-0,04	0,47	-0,09	4	0,94	0,70	0,70
P2	14	16,00	16,15	-0,04	0,51	-0,08	4	0,94	0,90	1,30
P11	34	12,00	12,16	-0,03	0,46	-0,07	4	0,95	0,70	0,70
P12	34	12,00	12,16	-0,03	0,46	-0,07	4	0,95	0,70	0,70
P12	4	14,00	14,15	-0,03	0,47	-0,07	4	0,95	0,70	0,70
P8	18	14,00	14,14	-0,03	0,47	-0,07	4	0,95	0,70	0,70
P8	32	14,00	14,14	-0,03	0,47	-0,07	4	0,95	0,70	0,70
P4	4	13,00	13,14	-0,03	0,46	-0,07	4	0,95	0,10	0,10
P10	39	17,00	17,09	-0,03	0,56	-0,05	4	0,96	0,20	0,20
P7	10	12,00	12,13	-0,03	0,46	-0,06	4	0,95	0,70	0,60
P9	10	14,00	14,11	-0,02	0,47	-0,05	4	0,96	0,70	0,70
P11	33	13,00	13,11	-0,02	0,46	-0,05	4	0,96	0,50	0,50
P12	33	13,00	13,11	-0,02	0,46	-0,05	4	0,96	0,50	0,50
P2	23	19,00	19,03	-0,02	0,94	-0,02	4	0,98	1,00	1,70
P4	34	11,00	11,10	-0,02	0,46	-0,05	4	0,97	0,60	0,50
P7	2	14,00	14,06	-0,01	0,47	-0,03	4	0,98	0,50	0,60
P14	29	17,00	17,04	-0,01	0,56	-0,02	4	0,98	0,60	0,60
P12	10	13,00	13,04	-0,01	0,46	-0,02	4	0,99	0,50	0,50
P14	39	15,00	15,03	-0,01	0,49	-0,02	4	0,99	2,10	2,20
P14	40	15,00	15,03	-0,01	0,49	-0,02	4	0,99	2,10	2,20
P5	16	11,00	11,02	0,00	0,46	-0,01	4	0,99	1,80	2,40
P1	22	17,00	17,01	0,00	0,56	0,00	4	1,00	2,00	3,90
P3	40	18,00	18,00	0,00	0,66	0,00	4	1,00	0,10	0,10
P13	18	13,00	12,99	0,00	0,46	0,00	4	1,00	0,50	0,50
P13	32	13,00	12,99	0,00	0,46	0,00	4	1,00	0,50	0,50

P9	23	19,00	19,00	0,00	0,94	0,00	4	1,00	1,00	1,70
P10	27	15,00	14,98	0,00	0,49	0,01	4	0,99	0,20	0,30
P12	30	17,00	16,96	0,01	0,56	0,02	4	0,98	0,20	0,20
P14	1	10,00	9,95	0,01	0,47	0,02	4	0,98	0,80	0,90
P1	5	14,00	13,95	0,01	0,47	0,03	4	0,98	0,70	0,70
P14	32	12,00	11,93	0,01	0,46	0,03	4	0,98	0,50	0,40
P5	13	18,00	17,96	0,02	0,66	0,03	4	0,98	0,10	0,10
P5	25	18,00	17,96	0,02	0,66	0,03	4	0,98	0,10	0,10
P5	19	12,00	11,91	0,02	0,46	0,04	4	0,97	0,70	0,60
P13	40	16,00	15,93	0,02	0,51	0,04	4	0,97	0,60	0,60
P5	36	14,00	13,91	0,02	0,47	0,04	4	0,97	0,70	0,70
P13	9	15,00	14,90	0,02	0,49	0,05	4	0,96	0,50	0,40
P11	11	15,00	14,89	0,03	0,49	0,05	4	0,96	0,30	0,40
P12	2	15,00	14,89	0,03	0,49	0,05	4	0,96	0,80	0,90
P14	6	10,00	9,88	0,03	0,47	0,06	4	0,96	0,80	0,70
P8	33	14,00	13,85	0,03	0,47	0,07	4	0,95	0,70	0,70
P7	16	9,00	8,86	0,03	0,49	0,07	4	0,95	1,30	1,40
P10	35	15,00	14,84	0,04	0,49	0,08	4	0,94	0,20	0,30
P14	30	16,00	15,85	0,04	0,51	0,07	4	0,94	0,20	0,30
P7	20	14,00	13,82	0,04	0,47	0,08	4	0,94	0,30	0,30
P1	20	15,00	14,83	0,04	0,49	0,08	4	0,94	0,50	0,50
P12	29	18,00	17,90	0,04	0,66	0,06	4	0,95	0,70	0,80
P2	13	18,00	17,90	0,04	0,66	0,06	4	0,95	0,10	0,10
P2	25	18,00	17,90	0,04	0,66	0,06	4	0,95	0,10	0,10
P2	19	12,00	11,79	0,04	0,46	0,09	4	0,93	0,70	0,60
P2	36	14,00	13,80	0,04	0,47	0,09	4	0,93	0,70	0,70
P1	36	13,00	12,78	0,05	0,46	0,10	4	0,92	0,50	0,50
P8	10	14,00	13,78	0,05	0,47	0,10	4	0,92	0,70	0,70
P12	5	14,00	13,77	0,05	0,47	0,11	4	0,92	0,70	0,70
P8	39	17,00	16,83	0,05	0,56	0,09	4	0,93	0,60	0,60
P14	14	14,00	13,76	0,05	0,47	0,11	4	0,91	1,70	1,70
P13	14	15,00	14,74	0,06	0,49	0,12	4	0,91	1,10	1,70
P14	4	13,00	12,71	0,06	0,46	0,13	4	0,90	0,10	0,10
P5	6	13,00	12,71	0,06	0,46	0,13	4	0,90	0,60	0,60
P14	22	16,00	15,76	0,06	0,51	0,12	4	0,91	3,00	3,70
P3	29	19,00	18,93	0,06	0,94	0,07	4	0,95	0,60	0,40
P13	33	13,00	12,69	0,07	0,46	0,14	4	0,89	0,50	0,50
P11	16	10,00	9,69	0,07	0,47	0,15	4	0,89	0,80	0,90
P11	12	19,00	18,92	0,07	0,94	0,07	4	0,95	1,00	1,60
P14	34	11,00	10,66	0,07	0,46	0,16	4	0,88	1,30	1,10
P10	8	14,00	13,65	0,08	0,47	0,16	4	0,88	0,70	0,70
P12	20	15,00	14,67	0,08	0,49	0,16	4	0,88	0,50	0,50
P9	7	13,00	12,62	0,08	0,46	0,17	4	0,87	0,50	0,50
P7	8	12,00	11,61	0,08	0,46	0,18	4	0,87	0,50	0,40
P13	10	13,00	12,62	0,08	0,46	0,18	4	0,87	0,50	0,50
P9	8	14,00	13,62	0,09	0,47	0,18	4	0,87	0,70	0,70
P2	6	13,00	12,59	0,09	0,46	0,19	4	0,86	0,60	0,60
P1	31	11,00	10,59	0,09	0,46	0,19	4	0,86	1,30	1,10

P10	1	13,00	12,59	0,09	0,46	0,19	4	0,86	1,70	2,00
P1	28	19,00	18,89	0,09	0,94	0,10	4	0,93	1,00	1,70
P14	10	12,00	11,55	0,09	0,46	0,20	4	0,85	1,10	1,00
P1	32	14,00	13,58	0,09	0,47	0,20	4	0,85	0,70	0,70
P11	19	11,00	10,55	0,09	0,46	0,21	4	0,85	1,20	1,00
P3	33	16,00	15,63	0,10	0,51	0,19	4	0,86	0,60	0,50
P1	6	12,00	11,53	0,10	0,46	0,21	4	0,84	1,20	1,10
P9	31	12,00	11,53	0,10	0,46	0,21	4	0,84	0,70	0,60
P11	8	13,00	12,52	0,10	0,46	0,22	4	0,84	0,50	0,50
P11	7	12,00	11,50	0,10	0,46	0,23	4	0,83	0,70	0,70
P12	7	12,00	11,50	0,10	0,46	0,23	4	0,83	0,70	0,70
P6	32	16,00	15,58	0,11	0,51	0,21	4	0,85	1,00	2,20
P4	8	12,00	11,47	0,11	0,46	0,24	4	0,82	1,10	1,00
P13	2	15,00	14,51	0,11	0,49	0,24	4	0,83	0,80	0,90
P7	39	16,00	15,53	0,12	0,51	0,23	4	0,83	1,10	1,30
P13	22	17,00	16,58	0,12	0,56	0,22	4	0,83	1,90	2,70
P7	9	15,00	14,46	0,13	0,49	0,26	4	0,81	1,40	1,30
P12	28	19,00	18,84	0,13	0,94	0,14	4	0,90	1,00	1,70
P12	32	14,00	13,41	0,13	0,47	0,28	4	0,80	0,70	0,70
P4	18	13,00	12,38	0,13	0,46	0,29	4	0,79	0,50	0,50
P6	29	19,00	18,83	0,14	0,94	0,15	4	0,89	0,60	0,40
P13	5	14,00	13,37	0,14	0,47	0,30	4	0,78	0,70	0,70
P8	36	14,00	13,35	0,14	0,47	0,30	4	0,78	0,70	0,70
P3	19	14,00	13,30	0,15	0,47	0,33	4	0,76	0,30	0,30
P4	13	17,00	16,47	0,16	0,56	0,28	4	0,79	0,30	0,70
P8	8	14,00	13,28	0,16	0,47	0,33	4	0,76	0,70	0,70
P14	20	14,00	13,27	0,16	0,47	0,34	4	0,75	0,40	0,40
P9	22	18,00	17,59	0,17	0,66	0,25	4	0,81	1,90	3,10
P13	20	15,00	14,28	0,17	0,49	0,34	4	0,75	0,90	0,70
P7	33	13,00	12,21	0,17	0,46	0,36	4	0,74	0,50	0,50
P13	36	13,00	12,17	0,18	0,46	0,38	4	0,72	0,50	0,50
P13	19	11,00	10,13	0,18	0,46	0,40	4	0,71	1,20	1,00
P12	9	16,00	15,27	0,18	0,51	0,36	4	0,74	0,60	0,50
P3	31	14,00	13,16	0,18	0,47	0,39	4	0,72	0,30	0,30
P1	23	19,00	18,74	0,20	0,94	0,21	4	0,84	1,00	1,70
P14	8	12,00	11,03	0,20	0,46	0,44	4	0,68	0,50	0,40
P10	4	16,00	15,19	0,20	0,51	0,40	4	0,71	0,30	0,30
P3	36	16,00	15,18	0,21	0,51	0,40	4	0,71	0,60	0,50
P11	10	14,00	13,04	0,21	0,47	0,45	4	0,68	0,70	0,70
P9	4	16,00	15,16	0,21	0,51	0,41	4	0,70	0,30	0,30
P12	39	17,00	16,27	0,21	0,56	0,38	4	0,72	0,20	0,40
P12	40	17,00	16,27	0,21	0,56	0,38	4	0,72	0,20	0,20
P13	28	19,00	18,71	0,22	0,94	0,23	4	0,83	1,00	1,60
P3	8	16,00	15,12	0,22	0,51	0,43	4	0,69	0,60	0,50
P12	14	16,00	15,12	0,22	0,51	0,43	4	0,69	2,60	3,80
P13	6	12,00	10,92	0,22	0,46	0,49	4	0,65	1,20	1,10
P14	18	13,00	11,93	0,23	0,46	0,49	4	0,65	0,50	0,50
P10	9	17,00	16,22	0,23	0,56	0,40	4	0,71	1,20	0,80

P6	19	14,00	12,95	0,23	0,47	0,48	4	0,65	2,10	2,90
P9	9	17,00	16,19	0,23	0,56	0,42	4	0,70	1,20	0,80
P8	34	14,00	12,92	0,23	0,47	0,50	4	0,65	0,70	0,70
P12	23	19,00	18,69	0,24	0,94	0,25	4	0,81	1,00	1,70
P4	22	17,00	16,11	0,26	0,56	0,46	4	0,67	1,90	2,80
P9	38	18,00	17,33	0,26	0,66	0,39	4	0,72	2,10	5,50
P6	7	15,00	13,86	0,26	0,49	0,54	4	0,62	0,20	0,30
P4	29	18,00	17,32	0,26	0,66	0,40	4	0,71	0,10	0,10
P13	34	13,00	11,73	0,27	0,46	0,58	4	0,59	0,50	0,50
P9	14	17,00	16,05	0,27	0,56	0,48	4	0,65	0,20	0,40
P12	11	16,00	14,89	0,27	0,51	0,54	4	0,62	1,00	2,20
P5	2	17,00	16,04	0,28	0,56	0,49	4	0,65	0,20	0,40
P6	27	17,00	16,03	0,28	0,56	0,50	4	0,65	0,70	1,20
P14	33	13,00	11,63	0,29	0,46	0,62	4	0,57	0,50	0,50
P2	1	14,00	12,66	0,29	0,47	0,61	4	0,57	1,50	1,80
P9	35	16,00	14,81	0,29	0,51	0,57	4	0,60	0,30	0,30
P5	40	18,00	17,22	0,30	0,66	0,45	4	0,68	0,10	0,10
P7	7	12,00	10,58	0,30	0,46	0,65	4	0,55	0,70	0,70
P4	20	15,00	13,69	0,30	0,49	0,61	4	0,57	0,50	0,50
P2	2	17,00	15,94	0,30	0,56	0,54	4	0,62	0,20	0,40
P12	13	18,00	17,20	0,30	0,66	0,46	4	0,67	0,10	0,10
P12	25	18,00	17,20	0,30	0,66	0,46	4	0,67	0,10	0,10
P12	19	12,00	10,55	0,30	0,46	0,66	4	0,54	0,70	0,60
P13	39	17,00	15,93	0,30	0,56	0,54	4	0,62	0,60	0,60
P12	36	14,00	12,60	0,30	0,47	0,64	4	0,55	0,70	0,70
P4	36	13,00	11,55	0,31	0,46	0,66	4	0,54	0,50	0,50
P8	9	17,00	15,92	0,31	0,56	0,55	4	0,61	1,20	0,80
P6	5	17,00	15,90	0,31	0,56	0,56	4	0,61	0,50	0,40
P7	31	11,00	9,53	0,32	0,46	0,69	4	0,53	0,60	0,50
P7	38	17,00	15,87	0,32	0,56	0,57	4	0,60	0,20	0,40
P7	28	19,00	18,55	0,32	0,94	0,34	4	0,75	1,00	1,60
P2	40	18,00	17,14	0,32	0,66	0,49	4	0,65	0,10	0,10
P10	6	14,00	12,52	0,32	0,47	0,68	4	0,53	2,20	2,00
P7	32	14,00	12,51	0,32	0,47	0,68	4	0,53	0,50	0,60
P4	7	12,00	10,45	0,33	0,46	0,71	4	0,52	0,70	0,70
P9	6	14,00	12,48	0,33	0,47	0,70	4	0,52	2,20	2,00
P11	31	12,00	10,41	0,33	0,46	0,73	4	0,51	0,70	0,60
P4	31	11,00	9,40	0,35	0,46	0,75	4	0,49	0,40	0,40
P9	40	18,00	17,07	0,35	0,66	0,52	4	0,63	0,70	0,40
P12	6	13,00	11,35	0,35	0,46	0,75	4	0,50	0,60	0,60
P3	11	18,00	17,06	0,35	0,66	0,53	4	0,63	0,10	0,10
P4	28	19,00	18,50	0,35	0,94	0,37	4	0,73	1,00	1,70
P5	29	19,00	18,49	0,36	0,94	0,38	4	0,72	0,60	0,40
P11	32	15,00	13,41	0,36	0,49	0,74	4	0,50	0,20	0,20
P14	16	10,00	8,39	0,38	0,47	0,82	4	0,46	0,80	0,90
P2	29	19,00	18,45	0,38	0,94	0,41	4	0,70	0,60	0,40
P7	4	15,00	13,28	0,39	0,49	0,80	4	0,47	0,30	0,40
P13	13	18,00	16,92	0,39	0,66	0,59	4	0,59	0,10	0,10

P8	2	17,00	15,56	0,40	0,56	0,71	4	0,52	0,60	0,60
P3	30	19,00	18,42	0,40	0,94	0,43	4	0,69	0,60	0,40
P13	7	13,00	11,07	0,41	0,46	0,88	4	0,43	0,50	0,50
P9	29	19,00	18,40	0,41	0,94	0,44	4	0,69	0,40	0,20
P14	19	11,00	9,14	0,41	0,46	0,89	4	0,42	1,20	1,00
P8	40	18,00	16,83	0,42	0,66	0,63	4	0,56	0,70	0,40
P3	22	19,00	18,37	0,42	0,94	0,45	4	0,67	1,00	1,70
P7	23	19,00	18,36	0,43	0,94	0,46	4	0,67	1,00	2,60
P4	10	14,00	12,00	0,43	0,47	0,91	4	0,41	0,20	0,30
P14	31	11,00	9,01	0,44	0,46	0,96	4	0,39	1,20	1,00
P2	16	13,00	10,90	0,44	0,46	0,95	4	0,39	1,90	1,80
P11	4	16,00	14,15	0,45	0,51	0,87	4	0,43	0,30	0,30
P8	20	17,00	15,35	0,45	0,56	0,80	4	0,47	1,20	0,80
P6	33	17,00	15,33	0,46	0,56	0,81	4	0,46	0,20	0,40
P1	14	17,00	15,28	0,47	0,56	0,84	4	0,45	1,90	2,80
P7	5	15,00	12,89	0,47	0,49	0,97	4	0,39	0,20	0,20
P6	10	17,00	15,26	0,47	0,56	0,84	4	0,45	0,70	1,20
P14	9	16,00	13,92	0,50	0,51	0,97	4	0,39	1,00	0,80
P1	1	14,00	11,61	0,51	0,47	1,08	4	0,34	1,50	1,80
P3	38	19,00	18,19	0,52	0,94	0,55	4	0,61	1,00	1,70
P10	31	14,00	11,57	0,52	0,47	1,10	4	0,33	0,10	0,10
P11	5	16,00	13,77	0,53	0,51	1,03	4	0,36	0,30	0,30
P1	40	18,00	16,40	0,54	0,66	0,82	4	0,46	0,70	0,80
P7	18	15,00	12,51	0,55	0,49	1,13	4	0,32	0,30	0,40
P6	8	17,00	14,80	0,58	0,56	1,04	4	0,36	1,20	0,80
P4	30	18,00	16,20	0,60	0,66	0,90	4	0,42	0,80	1,40
P1	29	19,00	17,99	0,60	0,94	0,65	4	0,55	0,40	0,20
P11	18	16,00	13,41	0,61	0,51	1,19	4	0,30	0,30	0,30
P11	20	17,00	14,67	0,61	0,56	1,09	4	0,34	0,20	0,20
P10	14	18,00	16,09	0,63	0,66	0,95	4	0,40	0,10	0,10
P5	11	18,00	16,04	0,64	0,66	0,97	4	0,39	0,10	0,10
P6	34	17,00	14,47	0,66	0,56	1,18	4	0,31	0,20	0,40
P2	11	18,00	15,94	0,67	0,66	1,01	4	0,37	0,10	0,10
P10	2	18,00	15,88	0,68	0,66	1,03	4	0,36	2,00	3,40
P5	30	19,00	17,78	0,69	0,94	0,74	4	0,50	0,60	0,40
P8	14	18,00	15,78	0,71	0,66	1,07	4	0,34	0,10	0,10
P5	22	19,00	17,72	0,72	0,94	0,77	4	0,49	1,00	1,70
P2	30	19,00	17,72	0,72	0,94	0,77	4	0,49	0,60	0,40
P2	22	19,00	17,65	0,74	0,94	0,79	4	0,47	1,00	1,70
P5	1	16,00	12,78	0,75	0,51	1,45	4	0,22	2,60	3,80
P10	22	19,00	17,61	0,76	0,94	0,81	4	0,46	1,00	1,60
P5	38	19,00	17,47	0,81	0,94	0,87	4	0,44	1,00	1,70
P3	16	16,00	12,43	0,82	0,51	1,60	4	0,19	0,30	0,40
P2	38	19,00	17,40	0,83	0,94	0,89	4	0,42	1,00	1,70
P1	2	18,00	15,05	0,89	0,66	1,34	4	0,25	0,70	0,80
P1	11	18,00	15,05	0,89	0,66	1,34	4	0,25	0,10	0,10
P5	35	18,00	15,02	0,90	0,66	1,36	4	0,25	0,10	0,10
P4	16	13,00	8,75	0,92	0,46	1,99	4	0,12	1,90	1,80

P2	35	18,00	14,91	0,92	0,66	1,39	4	0,24	0,10	0,10
P6	36	18,00	14,87	0,93	0,66	1,41	4	0,23	0,10	0,10
P10	40	19,00	17,09	0,94	0,94	1,00	4	0,37	0,40	0,20
P1	30	19,00	17,08	0,94	0,94	1,01	4	0,37	1,00	2,60
P7	27	17,00	13,05	0,97	0,56	1,74	4	0,16	1,30	1,60
P14	27	17,00	12,48	1,10	0,56	1,95	4	0,12	1,30	1,60
P3	1	18,00	14,13	1,10	0,66	1,67	4	0,17	1,10	2,00
P4	11	18,00	13,93	1,15	0,66	1,73	4	0,16	0,10	0,10
P3	35	19,00	16,19	1,20	0,94	1,29	4	0,27	0,40	0,20
P6	18	19,00	15,58	1,36	0,94	1,45	4	0,22	0,40	0,20
P4	35	18,00	12,76	1,40	0,66	2,12	4	0,10	0,10	0,10
P6	16	18,00	12,07	1,55	0,66	2,34	4	0,08	0,50	0,70
P11	27	19,00	13,92	1,75	0,94	1,87	4	0,13	1,00	1,70
P6	1	19,00	13,79	1,78	0,94	1,90	4	0,13	1,00	2,60



EK 8. Son Test Türkçe Puanlayıcı Grubu *Puanlayıcı X Birey* Etkileşimleri

Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Yanlılık Büyüklüğü	SH	t	Sd	p	infit	outfit
P1	26	20,00	19,62	-0,27	1,41	-0,19	4	0,86	0,20	0,10
P2	12	20,00	19,02	0,66	1,41	0,47	4	0,66	0,20	0,10
P2	26	20,00	19,61	-0,24	1,41	-0,17	4	0,87	0,20	0,10
P3	12	20,00	19,09	0,59	1,41	0,42	4	0,70	0,20	0,10
P3	26	20,00	19,63	-0,31	1,41	-0,22	4	0,84	0,20	0,10
P4	25	20,00	18,11	1,31	1,41	0,93	4	0,40	0,20	0,10
P4	37	20,00	18,57	1,03	1,41	0,73	4	0,51	0,20	0,10
P5	38	20,00	17,21	1,73	1,41	1,23	4	0,29	0,20	0,10
P6	12	20,00	19,09	0,59	1,41	0,42	4	0,70	0,20	0,10
P6	26	20,00	19,63	-0,31	1,41	-0,22	4	0,84	0,20	0,10
P8	13	20,00	18,57	1,03	1,41	0,73	4	0,50	0,20	0,10
P8	22	20,00	17,56	1,58	1,41	1,12	4	0,33	0,20	0,10
P8	23	20,00	18,37	1,16	1,41	0,82	4	0,46	0,20	0,10
P8	25	20,00	18,70	0,94	1,41	0,67	4	0,54	0,20	0,10
P8	26	20,00	19,61	-0,25	1,41	-0,18	4	0,87	0,20	0,10
P8	28	20,00	19,29	0,35	1,41	0,25	4	0,82	0,20	0,10
P8	37	20,00	19,03	0,65	1,41	0,46	4	0,67	0,20	0,10
P9	12	20,00	19,10	0,58	1,41	0,41	4	0,70	0,20	0,10
P9	13	20,00	18,67	0,96	1,41	0,68	4	0,53	0,20	0,10
P9	26	20,00	19,64	-0,32	1,41	-0,23	4	0,83	0,20	0,10
P9	28	20,00	19,34	0,27	1,41	0,20	4	0,85	0,20	0,10
P9	37	20,00	19,10	0,58	1,41	0,41	4	0,70	0,20	0,10
P9	39	20,00	17,40	1,65	1,41	1,17	4	0,31	0,20	0,10
P10	12	20,00	19,07	0,61	1,41	0,43	4	0,69	0,20	0,10
P10	25	20,00	18,76	0,89	1,41	0,64	4	0,56	0,20	0,10
P10	26	20,00	19,63	-0,29	1,41	-0,21	4	0,85	0,20	0,10
P10	28	20,00	19,32	0,31	1,41	0,22	4	0,84	0,20	0,10
P10	37	20,00	19,07	0,61	1,41	0,43	4	0,69	0,20	0,10
P11	23	20,00	17,85	1,45	1,41	1,03	4	0,36	0,20	0,10
P12	12	20,00	19,22	0,43	1,41	0,31	4	0,77	0,20	0,10
P12	26	20,00	19,69	-0,47	1,41	-0,33	4	0,76	0,20	0,10
P12	37	20,00	19,22	0,43	1,41	0,31	4	0,77	0,20	0,10
P12	38	20,00	18,08	1,32	1,41	0,94	4	0,40	0,20	0,10
P13	25	20,00	18,35	1,17	1,41	0,83	4	0,45	0,20	0,10
P13	26	20,00	19,50	-0,01	1,41	-0,01	4	0,99	0,20	0,10
P13	28	20,00	19,09	0,59	1,41	0,42	4	0,70	0,20	0,10
P13	37	20,00	18,76	0,89	1,41	0,63	4	0,56	0,20	0,10
P14	28	20,00	19,06	0,62	1,41	0,44	4	0,68	0,20	0,10
P9	16	6,00	10,31	-1,63	0,91	-1,79	4	0,15	0,20	0,10
P8	16	6,00	10,00	-1,55	0,91	-1,71	4	0,16	0,90	1,10
P5	26	18,00	19,53	-1,44	0,72	-2,00	4	0,12	0,40	0,60
P5	12	16,00	18,83	-1,32	0,57	-2,34	4	0,08	1,50	1,90
P11	25	15,00	18,27	-1,24	0,53	-2,34	4	0,08	1,90	1,70
P5	23	15,00	18,05	-1,12	0,53	-2,11	4	0,10	1,50	3,00

P14	25	16,00	18,30	-0,96	0,57	-1,69	4	0,17	3,30	2,70
P9	30	15,00	17,66	-0,93	0,53	-1,75	4	0,15	0,90	1,10
P8	35	12,00	15,65	-0,92	0,48	-1,94	4	0,12	0,80	0,70
P7	12	17,00	18,75	-0,91	0,62	-1,46	4	0,22	2,40	3,10
P7	37	17,00	18,75	-0,91	0,62	-1,46	4	0,22	3,70	3,50
P8	11	14,00	16,95	-0,90	0,50	-1,79	4	0,15	1,60	2,40
P14	37	17,00	18,72	-0,89	0,62	-1,43	4	0,23	3,70	3,50
P7	13	16,00	18,17	-0,88	0,57	-1,55	4	0,20	2,30	2,30
P11	37	17,00	18,70	-0,87	0,62	-1,40	4	0,23	3,70	3,50
P11	13	16,00	18,10	-0,84	0,57	-1,49	4	0,21	2,70	2,20
P5	28	18,00	19,14	-0,84	0,72	-1,17	4	0,31	1,20	1,10
P14	36	12,00	15,34	-0,83	0,48	-1,74	4	0,16	1,20	1,00
P10	13	17,00	18,63	-0,82	0,62	-1,32	4	0,26	0,30	0,30
P4	38	14,00	16,70	-0,80	0,50	-1,60	4	0,18	1,00	2,00
P4	12	17,00	18,57	-0,78	0,62	-1,26	4	0,28	3,80	4,00
P14	39	14,00	16,51	-0,74	0,50	-1,47	4	0,22	1,40	1,20
P11	28	18,00	19,05	-0,73	0,72	-1,02	4	0,36	1,40	2,80
P5	4	13,00	15,78	-0,73	0,48	-1,52	4	0,20	0,70	0,60
P9	11	15,00	17,14	-0,71	0,53	-1,33	4	0,25	2,20	3,60
P14	16	7,00	9,00	-0,69	0,67	-1,03	4	0,36	1,30	3,70
P10	23	17,00	18,44	-0,69	0,62	-1,11	4	0,33	0,50	0,50
P1	27	12,00	14,83	-0,69	0,48	-1,44	4	0,22	1,10	0,90
P7	26	19,00	19,50	-0,68	0,99	-0,69	4	0,53	1,10	1,40
P13	38	15,00	17,06	-0,67	0,53	-1,28	4	0,27	0,40	0,60
P14	26	19,00	19,49	-0,66	0,99	-0,67	4	0,54	1,10	1,40
P8	31	11,00	13,85	-0,66	0,48	-1,37	4	0,24	1,60	1,30
P11	40	15,00	17,02	-0,66	0,53	-1,25	4	0,28	1,00	0,90
P14	35	12,00	14,72	-0,66	0,48	-1,38	4	0,24	0,80	0,70
P10	38	16,00	17,72	-0,66	0,57	-1,16	4	0,31	0,70	1,10
P11	26	19,00	19,48	-0,64	0,99	-0,65	4	0,55	1,10	1,40
P7	22	15,00	16,96	-0,63	0,53	-1,20	4	0,30	1,30	1,40
P10	16	8,00	10,18	-0,61	0,58	-1,06	4	0,35	0,60	2,10
P8	29	17,00	18,30	-0,61	0,62	-0,98	4	0,38	0,90	0,60
P5	34	11,00	13,61	-0,60	0,48	-1,25	4	0,28	0,60	0,60
P11	36	13,00	15,28	-0,58	0,48	-1,21	4	0,29	0,70	0,60
P10	18	14,00	16,05	-0,58	0,50	-1,16	4	0,31	0,50	0,50
P8	1	10,00	12,50	-0,58	0,50	-1,16	4	0,31	1,70	2,10
P11	30	15,00	16,78	-0,57	0,53	-1,07	4	0,34	0,70	0,60
P4	26	19,00	19,43	-0,55	0,99	-0,56	4	0,61	1,10	1,40
P12	28	19,00	19,43	-0,55	0,99	-0,56	4	0,61	1,10	1,20
P12	14	15,00	16,70	-0,54	0,53	-1,02	4	0,37	1,20	1,30
P5	37	18,00	18,83	-0,53	0,72	-0,74	4	0,50	2,30	2,80
P4	31	10,00	12,24	-0,52	0,50	-1,05	4	0,35	0,60	0,50
P4	11	14,00	15,86	-0,52	0,50	-1,04	4	0,36	1,10	2,00
P4	8	13,00	15,04	-0,51	0,48	-1,07	4	0,35	2,50	3,80
P9	27	13,00	15,02	-0,51	0,48	-1,06	4	0,35	1,10	0,90
P14	13	17,00	18,13	-0,51	0,62	-0,82	4	0,46	3,70	3,50
P3	27	13,00	14,98	-0,50	0,48	-1,04	4	0,36	1,10	0,90

P6	27	13,00	14,98	-0,50	0,48	-1,04	4	0,36	1,10	0,90
P12	18	15,00	16,58	-0,50	0,53	-0,94	4	0,40	0,30	0,30
P13	12	18,00	18,76	-0,48	0,72	-0,66	4	0,54	2,30	2,80
P8	6	11,00	13,07	-0,47	0,48	-0,98	4	0,38	1,80	1,60
P5	33	14,00	15,70	-0,47	0,50	-0,94	4	0,40	0,60	0,60
P13	35	13,00	14,82	-0,46	0,48	-0,94	4	0,40	1,30	1,20
P1	19	12,00	13,93	-0,45	0,48	-0,95	4	0,40	1,20	1,00
P9	2	15,00	16,42	-0,44	0,53	-0,83	4	0,45	0,70	0,60
P8	27	13,00	14,75	-0,44	0,48	-0,90	4	0,42	1,10	0,90
P12	10	15,00	16,40	-0,43	0,53	-0,82	4	0,46	0,30	0,30
P12	23	18,00	18,70	-0,43	0,72	-0,60	4	0,58	1,20	1,10
P11	6	10,00	11,82	-0,43	0,50	-0,86	4	0,44	0,90	0,80
P2	27	13,00	14,71	-0,43	0,48	-0,88	4	0,43	1,10	0,90
P5	9	14,00	15,55	-0,42	0,50	-0,85	4	0,44	0,10	0,10
P2	25	18,00	18,69	-0,42	0,72	-0,59	4	0,59	0,40	0,60
P12	7	14,00	15,53	-0,42	0,50	-0,83	4	0,45	0,50	0,50
P9	1	11,00	12,82	-0,42	0,48	-0,86	4	0,44	2,10	2,60
P13	8	14,00	15,51	-0,41	0,50	-0,82	4	0,46	1,40	2,20
P11	35	13,00	14,65	-0,41	0,48	-0,85	4	0,44	0,90	0,80
P13	40	16,00	17,14	-0,41	0,57	-0,72	4	0,51	0,20	0,30
P5	22	16,00	17,13	-0,40	0,57	-0,71	4	0,52	0,20	0,20
P7	40	16,00	17,12	-0,40	0,57	-0,70	4	0,52	0,20	0,30
P3	28	19,00	19,33	-0,40	0,99	-0,40	4	0,71	1,10	1,20
P6	28	19,00	19,33	-0,40	0,99	-0,40	4	0,71	1,10	1,20
P13	11	15,00	16,28	-0,39	0,53	-0,74	4	0,50	0,40	0,70
P7	11	15,00	16,26	-0,38	0,53	-0,72	4	0,51	1,00	1,20
P7	2	14,00	15,40	-0,38	0,50	-0,76	4	0,49	0,80	0,70
P7	36	14,00	15,40	-0,38	0,50	-0,76	4	0,49	0,60	0,50
P2	5	15,00	16,23	-0,38	0,53	-0,71	4	0,52	0,40	0,40
P2	33	15,00	16,23	-0,38	0,53	-0,71	4	0,52	0,40	0,40
P10	11	16,00	17,06	-0,37	0,57	-0,66	4	0,55	0,50	1,00
P13	29	17,00	17,86	-0,37	0,62	-0,60	4	0,58	0,40	0,40
P14	11	15,00	16,20	-0,36	0,53	-0,69	4	0,53	1,30	2,80
P14	2	14,00	15,34	-0,36	0,50	-0,72	4	0,51	1,20	1,00
P1	28	19,00	19,30	-0,35	0,99	-0,36	4	0,74	1,10	1,20
P11	2	14,00	15,28	-0,34	0,50	-0,69	4	0,53	0,30	0,30
P3	18	15,00	16,12	-0,34	0,53	-0,64	4	0,56	0,30	0,30
P6	18	15,00	16,12	-0,34	0,53	-0,64	4	0,56	0,30	0,30
P13	16	8,00	9,10	-0,33	0,58	-0,57	4	0,60	0,60	2,10
P12	4	16,00	16,95	-0,33	0,57	-0,58	4	0,59	0,60	0,50
P11	38	16,00	16,94	-0,33	0,57	-0,58	4	0,59	0,70	1,10
P12	20	15,00	16,09	-0,33	0,53	-0,62	4	0,57	0,30	0,30
P11	29	17,00	17,76	-0,32	0,62	-0,52	4	0,63	0,40	0,40
P2	28	19,00	19,28	-0,32	0,99	-0,32	4	0,76	1,10	1,20
P4	35	13,00	14,30	-0,32	0,48	-0,66	4	0,55	1,30	1,20
P14	14	14,00	15,18	-0,32	0,50	-0,63	4	0,56	1,50	1,50
P2	34	13,00	14,30	-0,32	0,48	-0,65	4	0,55	0,20	0,20
P1	40	17,00	17,75	-0,31	0,62	-0,51	4	0,64	0,30	0,30

P13	30	16,00	16,91	-0,31	0,57	-0,55	4	0,61	0,20	0,30
P14	1	10,00	11,32	-0,31	0,50	-0,63	4	0,56	2,20	2,50
P2	32	15,00	16,03	-0,31	0,53	-0,58	4	0,59	0,40	0,40
P12	5	16,00	16,89	-0,31	0,57	-0,54	4	0,62	0,60	0,50
P12	8	16,00	16,89	-0,31	0,57	-0,54	4	0,62	0,60	0,50
P12	33	16,00	16,89	-0,31	0,57	-0,54	4	0,62	0,60	0,50
P9	22	17,00	17,73	-0,31	0,62	-0,50	4	0,65	2,20	2,00
P12	34	14,00	15,15	-0,31	0,50	-0,61	4	0,57	0,50	0,50
P11	22	16,00	16,86	-0,30	0,57	-0,52	4	0,63	1,20	1,10
P11	14	14,00	15,12	-0,30	0,50	-0,60	4	0,58	1,50	1,50
P1	18	15,00	15,99	-0,30	0,53	-0,56	4	0,61	0,30	0,30
P5	10	14,00	15,10	-0,29	0,50	-0,59	4	0,59	0,10	0,10
P4	27	12,00	13,26	-0,29	0,48	-0,61	4	0,58	0,70	0,60
P9	23	18,00	18,49	-0,28	0,72	-0,39	4	0,71	1,20	1,10
P3	10	15,00	15,92	-0,27	0,53	-0,52	4	0,63	0,30	0,30
P6	10	15,00	15,92	-0,27	0,53	-0,52	4	0,63	0,30	0,30
P3	23	18,00	18,47	-0,27	0,72	-0,38	4	0,72	1,20	1,10
P6	23	18,00	18,47	-0,27	0,72	-0,38	4	0,72	1,20	1,10
P4	40	16,00	16,78	-0,27	0,57	-0,47	4	0,66	0,20	0,30
P2	18	15,00	15,89	-0,26	0,53	-0,50	4	0,64	0,30	0,30
P12	9	16,00	16,77	-0,26	0,57	-0,46	4	0,67	0,60	0,50
P3	7	14,00	14,98	-0,26	0,50	-0,52	4	0,63	0,50	0,50
P6	7	14,00	14,98	-0,26	0,50	-0,52	4	0,63	0,50	0,50
P5	7	13,00	14,06	-0,26	0,48	-0,53	4	0,62	0,60	0,60
P10	10	15,00	15,85	-0,25	0,53	-0,48	4	0,66	0,40	0,40
P5	25	18,00	18,44	-0,25	0,72	-0,35	4	0,74	0,40	0,60
P10	30	17,00	17,59	-0,24	0,62	-0,39	4	0,71	0,60	1,40
P10	19	13,00	14,02	-0,24	0,48	-0,51	4	0,64	1,30	1,00
P12	32	16,00	16,70	-0,24	0,57	-0,42	4	0,69	0,60	0,50
P10	7	14,00	14,91	-0,24	0,50	-0,48	4	0,66	0,50	0,50
P12	6	13,00	13,99	-0,24	0,48	-0,49	4	0,65	0,30	0,30
P4	29	17,00	17,56	-0,23	0,62	-0,37	4	0,73	0,40	0,40
P1	10	15,00	15,78	-0,23	0,53	-0,44	4	0,68	0,30	0,30
P1	23	18,00	18,41	-0,23	0,72	-0,32	4	0,77	1,20	1,10
P1	7	14,00	14,83	-0,22	0,50	-0,43	4	0,69	0,50	0,50
P10	29	18,00	18,38	-0,21	0,72	-0,29	4	0,78	0,50	0,30
P4	22	16,00	16,61	-0,21	0,57	-0,36	4	0,73	2,40	2,40
P5	8	15,00	15,70	-0,21	0,53	-0,39	4	0,72	0,30	0,30
P7	35	14,00	14,79	-0,21	0,50	-0,41	4	0,70	0,30	0,30
P4	1	10,00	10,85	-0,20	0,50	-0,41	4	0,70	1,10	1,50
P8	30	17,00	17,50	-0,20	0,62	-0,33	4	0,76	0,90	0,60
P2	23	18,00	18,35	-0,20	0,72	-0,28	4	0,80	1,20	1,10
P13	27	13,00	13,83	-0,20	0,48	-0,41	4	0,70	1,10	0,90
P4	16	8,00	8,62	-0,20	0,58	-0,34	4	0,75	0,60	2,10
P8	7	14,00	14,75	-0,20	0,50	-0,39	4	0,72	0,50	0,50
P13	31	12,00	12,84	-0,19	0,48	-0,40	4	0,71	0,20	0,30
P2	30	17,00	17,47	-0,19	0,62	-0,31	4	0,77	1,40	0,90
P2	31	13,00	13,80	-0,19	0,48	-0,40	4	0,71	0,20	0,20

P7	39	16,00	16,57	-0,19	0,57	-0,34	4	0,75	0,90	1,00
P5	20	14,00	14,73	-0,19	0,50	-0,38	4	0,72	0,40	0,60
P7	25	18,00	18,33	-0,18	0,72	-0,26	4	0,81	0,40	0,50
P12	19	14,00	14,71	-0,18	0,50	-0,37	4	0,73	0,70	0,60
P12	31	14,00	14,71	-0,18	0,50	-0,37	4	0,73	0,60	0,50
P9	20	15,00	15,62	-0,18	0,53	-0,34	4	0,75	0,30	0,30
P4	30	16,00	16,53	-0,18	0,57	-0,31	4	0,77	0,20	0,30
P3	4	16,00	16,52	-0,17	0,57	-0,31	4	0,77	0,60	0,50
P6	4	16,00	16,52	-0,17	0,57	-0,31	4	0,77	0,60	0,50
P3	20	15,00	15,59	-0,17	0,53	-0,32	4	0,76	0,30	0,30
P6	20	15,00	15,59	-0,17	0,53	-0,32	4	0,76	0,30	0,30
P11	39	16,00	16,46	-0,15	0,57	-0,27	4	0,80	0,90	1,00
P3	39	17,00	17,38	-0,15	0,62	-0,25	4	0,82	0,50	0,50
P6	39	17,00	17,38	-0,15	0,62	-0,25	4	0,82	0,50	0,50
P3	5	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P3	8	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P3	33	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P6	5	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P6	8	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P6	33	16,00	16,45	-0,15	0,57	-0,27	4	0,80	0,60	0,50
P3	34	14,00	14,58	-0,15	0,50	-0,30	4	0,78	0,50	0,50
P6	34	14,00	14,58	-0,15	0,50	-0,30	4	0,78	0,50	0,50
P10	20	15,00	15,52	-0,15	0,53	-0,28	4	0,79	0,30	0,30
P9	36	16,00	16,42	-0,14	0,57	-0,24	4	0,82	0,60	0,50
P5	14	15,00	15,48	-0,14	0,53	-0,26	4	0,81	1,20	1,30
P1	1	12,00	12,59	-0,13	0,48	-0,28	4	0,79	3,50	6,00
P1	4	16,00	16,40	-0,13	0,57	-0,23	4	0,83	0,60	0,50
P10	39	17,00	17,33	-0,13	0,62	-0,21	4	0,84	0,30	0,30
P10	5	16,00	16,39	-0,13	0,57	-0,23	4	0,83	0,60	0,50
P10	33	16,00	16,39	-0,13	0,57	-0,23	4	0,83	0,60	0,50
P1	20	15,00	15,44	-0,13	0,53	-0,24	4	0,82	0,30	0,30
P7	20	14,00	14,48	-0,12	0,50	-0,25	4	0,82	0,20	0,40
P8	4	16,00	16,33	-0,11	0,57	-0,19	4	0,86	0,60	0,50
P1	39	17,00	17,28	-0,11	0,62	-0,18	4	0,87	0,50	0,50
P1	5	16,00	16,33	-0,11	0,57	-0,19	4	0,86	0,60	0,50
P1	33	16,00	16,33	-0,11	0,57	-0,19	4	0,86	0,60	0,50
P1	34	14,00	14,42	-0,11	0,50	-0,21	4	0,84	0,50	0,50
P3	9	16,00	16,32	-0,11	0,57	-0,19	4	0,86	0,60	0,50
P6	9	16,00	16,32	-0,11	0,57	-0,19	4	0,86	0,60	0,50
P14	20	14,00	14,41	-0,10	0,50	-0,21	4	0,85	0,10	0,10
P13	9	15,00	15,36	-0,10	0,53	-0,19	4	0,86	0,40	0,40
P13	1	11,00	11,45	-0,10	0,48	-0,21	4	0,84	0,70	1,10
P2	4	16,00	16,30	-0,10	0,57	-0,18	4	0,87	0,20	0,30
P2	20	15,00	15,34	-0,10	0,53	-0,18	4	0,86	0,30	0,30
P9	32	16,00	16,29	-0,09	0,57	-0,17	4	0,88	0,60	0,50
P5	18	15,00	15,33	-0,09	0,53	-0,18	4	0,87	0,30	0,30
P3	37	19,00	19,09	-0,09	0,99	-0,09	4	0,93	0,20	0,20
P6	37	19,00	19,09	-0,09	0,99	-0,09	4	0,93	0,20	0,20

P8	39	17,00	17,22	-0,09	0,62	-0,14	4	0,89	0,30	0,40
P8	5	16,00	16,27	-0,09	0,57	-0,15	4	0,89	0,60	0,50
P1	2	16,00	16,26	-0,09	0,57	-0,15	4	0,89	0,50	0,40
P7	28	19,00	19,08	-0,08	0,99	-0,08	4	0,94	1,10	1,40
P3	14	16,00	16,25	-0,08	0,57	-0,15	4	0,89	1,10	1,00
P3	32	16,00	16,25	-0,08	0,57	-0,15	4	0,89	0,60	0,50
P6	14	16,00	16,25	-0,08	0,57	-0,15	4	0,89	1,10	1,00
P6	32	16,00	16,25	-0,08	0,57	-0,15	4	0,89	0,60	0,50
P13	14	15,00	15,28	-0,08	0,53	-0,15	4	0,89	1,50	3,00
P13	32	15,00	15,28	-0,08	0,53	-0,15	4	0,89	0,40	0,40
P3	6	13,00	13,34	-0,08	0,48	-0,17	4	0,88	0,30	0,30
P6	6	13,00	13,34	-0,08	0,48	-0,17	4	0,88	0,30	0,30
P7	34	13,00	13,33	-0,08	0,48	-0,16	4	0,88	0,60	0,60
P2	8	16,00	16,23	-0,08	0,57	-0,14	4	0,90	0,60	0,50
P7	14	15,00	15,25	-0,07	0,53	-0,13	4	0,90	2,30	2,20
P7	32	15,00	15,25	-0,07	0,53	-0,13	4	0,90	0,60	0,60
P5	6	12,00	12,28	-0,06	0,48	-0,13	4	0,90	0,30	0,30
P1	9	16,00	16,19	-0,06	0,57	-0,11	4	0,92	0,60	0,50
P10	32	16,00	16,19	-0,06	0,57	-0,11	4	0,92	0,60	0,50
P14	34	13,00	13,25	-0,06	0,48	-0,12	4	0,91	1,10	0,90
P11	9	15,00	15,20	-0,06	0,53	-0,11	4	0,92	0,70	0,90
P11	1	11,00	11,24	-0,06	0,48	-0,12	4	0,91	1,30	1,50
P2	2	16,00	16,17	-0,05	0,57	-0,09	4	0,93	1,10	1,00
P1	37	19,00	19,05	-0,05	0,99	-0,05	4	0,96	0,20	0,20
P1	32	16,00	16,12	-0,04	0,57	-0,07	4	0,95	0,60	0,50
P1	6	13,00	13,16	-0,04	0,48	-0,08	4	0,94	0,30	0,30
P9	19	14,00	14,15	-0,04	0,50	-0,07	4	0,95	1,70	1,40
P9	31	14,00	14,15	-0,04	0,50	-0,07	4	0,95	0,70	0,60
P13	18	15,00	15,13	-0,04	0,53	-0,07	4	0,95	0,30	0,30
P4	4	15,00	15,12	-0,03	0,53	-0,06	4	0,95	0,40	0,40
P2	9	16,00	16,10	-0,03	0,57	-0,05	4	0,96	0,60	0,50
P8	12	19,00	19,03	-0,03	0,99	-0,03	4	0,98	1,10	1,40
P3	19	14,00	14,10	-0,03	0,50	-0,05	4	0,96	0,70	0,60
P3	31	14,00	14,10	-0,03	0,50	-0,05	4	0,96	0,60	0,50
P6	19	14,00	14,10	-0,03	0,50	-0,05	4	0,96	0,70	0,60
P6	31	14,00	14,10	-0,03	0,50	-0,05	4	0,96	0,60	0,50
P1	16	10,00	10,09	-0,02	0,50	-0,04	4	0,97	0,80	1,20
P2	37	19,00	19,02	-0,02	0,99	-0,02	4	0,99	0,20	0,20
P7	38	17,00	17,04	-0,02	0,62	-0,02	4	0,98	0,50	0,50
P12	22	18,00	18,03	-0,01	0,72	-0,02	4	0,99	1,20	1,10
P2	14	16,00	16,03	-0,01	0,57	-0,02	4	0,99	1,10	1,00
P14	38	17,00	16,99	0,00	0,62	0,01	4	1,00	0,50	0,50
P2	16	10,00	9,96	0,01	0,50	0,02	4	0,99	1,60	4,20
P4	2	15,00	14,96	0,01	0,53	0,02	4	0,98	0,90	0,90
P5	29	18,00	17,97	0,01	0,72	0,02	4	0,99	1,20	1,10
P11	16	9,00	8,93	0,02	0,53	0,04	4	0,97	1,20	2,00
P10	27	15,00	14,91	0,03	0,53	0,05	4	0,96	0,60	0,50
P13	10	15,00	14,90	0,03	0,53	0,05	4	0,96	0,30	0,30

P14	22	17,00	16,91	0,03	0,62	0,05	4	0,96	3,80	4,00
P12	25	19,00	18,96	0,03	0,99	0,03	4	0,97	0,20	0,20
P4	9	15,00	14,87	0,03	0,53	0,07	4	0,95	0,40	0,40
P13	19	13,00	12,84	0,04	0,48	0,07	4	0,94	1,30	1,00
P7	10	15,00	14,86	0,04	0,53	0,07	4	0,95	0,70	0,60
P8	19	14,00	13,85	0,04	0,50	0,08	4	0,94	1,70	1,40
P7	23	18,00	17,92	0,04	0,72	0,06	4	0,96	1,40	2,80
P4	28	19,00	18,95	0,04	0,99	0,05	4	0,97	1,10	1,40
P7	30	17,00	16,88	0,05	0,62	0,07	4	0,95	0,30	0,30
P7	19	13,00	12,80	0,05	0,48	0,09	4	0,93	0,70	0,70
P2	19	14,00	13,80	0,05	0,50	0,10	4	0,93	0,70	0,60
P3	1	13,00	12,77	0,05	0,48	0,11	4	0,92	3,30	4,90
P6	1	13,00	12,77	0,05	0,48	0,11	4	0,92	3,30	4,90
P4	14	15,00	14,79	0,06	0,53	0,11	4	0,92	1,50	3,00
P4	32	15,00	14,79	0,06	0,53	0,11	4	0,92	0,40	0,40
P12	2	17,00	16,83	0,07	0,62	0,11	4	0,92	0,50	0,50
P12	36	17,00	16,83	0,07	0,62	0,11	4	0,92	0,90	0,60
P10	35	16,00	15,78	0,07	0,57	0,12	4	0,91	0,20	0,20
P14	7	14,00	13,71	0,07	0,50	0,14	4	0,89	0,50	0,50
P9	40	18,00	17,86	0,07	0,72	0,10	4	0,93	0,50	0,30
P10	1	13,00	12,68	0,07	0,48	0,15	4	0,89	2,50	3,90
P2	7	15,00	14,71	0,08	0,53	0,15	4	0,89	0,90	0,80
P7	29	18,00	17,84	0,08	0,72	0,11	4	0,92	0,50	0,30
P3	40	18,00	17,84	0,08	0,72	0,11	4	0,91	0,40	0,50
P6	40	18,00	17,84	0,08	0,72	0,11	4	0,91	0,40	0,50
P11	19	13,00	12,64	0,08	0,48	0,17	4	0,87	1,30	1,00
P5	39	17,00	16,75	0,09	0,62	0,15	4	0,89	0,50	0,50
P5	5	16,00	15,70	0,09	0,57	0,17	4	0,88	0,60	0,50
P14	29	18,00	17,80	0,10	0,72	0,14	4	0,90	0,50	0,30
P2	10	16,00	15,68	0,10	0,57	0,18	4	0,87	0,20	0,20
P4	18	15,00	14,63	0,10	0,53	0,19	4	0,86	0,30	0,30
P9	38	18,00	17,79	0,10	0,72	0,14	4	0,89	2,60	9,00
P9	34	15,00	14,62	0,10	0,53	0,20	4	0,85	0,30	0,30
P3	38	18,00	17,77	0,11	0,72	0,16	4	0,88	1,20	1,10
P6	38	18,00	17,77	0,11	0,72	0,16	4	0,88	1,20	1,10
P2	1	13,00	12,45	0,13	0,48	0,26	4	0,81	3,30	4,90
P13	4	16,00	15,59	0,13	0,57	0,23	4	0,83	0,60	0,50
P12	13	19,00	18,86	0,13	0,99	0,13	4	0,90	0,20	0,20
P12	39	18,00	17,73	0,13	0,72	0,18	4	0,86	0,40	0,50
P4	6	12,00	11,42	0,13	0,48	0,28	4	0,80	0,80	0,70
P13	20	15,00	14,52	0,13	0,53	0,25	4	0,82	0,40	0,40
P10	34	15,00	14,50	0,14	0,53	0,26	4	0,81	0,30	0,30
P3	22	18,00	17,71	0,14	0,72	0,20	4	0,85	1,20	1,10
P6	22	18,00	17,71	0,14	0,72	0,20	4	0,85	1,20	1,10
P8	40	18,00	17,70	0,15	0,72	0,20	4	0,85	0,50	0,30
P12	27	16,00	15,53	0,15	0,57	0,26	4	0,81	0,20	0,20
P9	6	14,00	13,39	0,15	0,50	0,30	4	0,78	1,70	1,40
P13	39	17,00	16,59	0,15	0,62	0,24	4	0,82	0,40	0,40

P13	5	16,00	15,51	0,15	0,57	0,27	4	0,80	0,60	0,50
P13	33	16,00	15,51	0,15	0,57	0,27	4	0,80	0,60	0,50
P1	38	18,00	17,68	0,16	0,72	0,22	4	0,84	1,20	1,10
P14	4	16,00	15,50	0,16	0,57	0,28	4	0,80	0,60	0,50
P2	40	18,00	17,68	0,16	0,72	0,22	4	0,84	0,40	0,50
P7	5	16,00	15,48	0,16	0,57	0,28	4	0,79	0,20	0,30
P7	33	16,00	15,48	0,16	0,57	0,28	4	0,79	0,60	0,50
P5	32	16,00	15,48	0,16	0,57	0,29	4	0,79	0,60	0,50
P10	22	18,00	17,66	0,17	0,72	0,23	4	0,83	1,30	1,10
P4	10	15,00	14,38	0,17	0,53	0,31	4	0,77	0,30	0,30
P4	19	13,00	12,24	0,17	0,48	0,36	4	0,74	1,30	1,00
P13	2	16,00	15,44	0,17	0,57	0,31	4	0,77	0,50	0,40
P13	36	16,00	15,44	0,17	0,57	0,31	4	0,77	0,60	0,50
P3	16	11,00	10,26	0,17	0,48	0,36	4	0,74	1,50	3,00
P6	16	11,00	10,26	0,17	0,48	0,36	4	0,74	1,50	3,00
P14	5	16,00	15,42	0,18	0,57	0,32	4	0,77	0,60	0,50
P9	25	19,00	18,80	0,18	0,99	0,18	4	0,86	1,10	1,30
P10	6	14,00	13,25	0,18	0,50	0,36	4	0,73	1,70	1,40
P1	22	18,00	17,61	0,19	0,72	0,26	4	0,81	1,20	1,10
P2	38	18,00	17,61	0,19	0,72	0,26	4	0,81	1,20	1,10
P9	5	17,00	16,48	0,19	0,62	0,30	4	0,78	0,90	0,60
P9	8	17,00	16,48	0,19	0,62	0,30	4	0,78	0,90	0,60
P9	33	17,00	16,48	0,19	0,62	0,30	4	0,78	0,90	0,60
P3	25	19,00	18,78	0,19	0,99	0,19	4	0,86	0,20	0,20
P6	25	19,00	18,78	0,19	0,99	0,19	4	0,86	0,20	0,20
P10	4	17,00	16,46	0,20	0,62	0,32	4	0,77	0,30	0,40
P11	5	16,00	15,36	0,20	0,57	0,35	4	0,74	0,90	0,90
P11	8	16,00	15,36	0,20	0,57	0,35	4	0,74	0,60	0,50
P11	33	16,00	15,36	0,20	0,57	0,35	4	0,74	0,60	0,50
P11	34	14,00	13,17	0,20	0,50	0,40	4	0,71	0,50	0,50
P7	9	16,00	15,33	0,21	0,57	0,36	4	0,73	0,90	1,00
P2	22	18,00	17,54	0,22	0,72	0,30	4	0,78	1,20	1,10
P5	19	14,00	13,09	0,22	0,50	0,44	4	0,68	0,70	0,60
P5	31	14,00	13,09	0,22	0,50	0,44	4	0,68	0,60	0,50
P10	8	17,00	16,39	0,22	0,62	0,36	4	0,74	0,90	0,60
P13	6	13,00	12,03	0,22	0,48	0,46	4	0,67	0,30	0,30
P3	2	17,00	16,39	0,22	0,62	0,36	4	0,74	0,50	0,50
P3	36	17,00	16,39	0,22	0,62	0,36	4	0,74	0,90	0,60
P6	2	17,00	16,39	0,22	0,62	0,36	4	0,74	0,50	0,50
P6	36	17,00	16,39	0,22	0,62	0,36	4	0,74	0,90	0,60
P14	9	16,00	15,26	0,22	0,57	0,40	4	0,71	0,60	0,50
P7	6	13,00	11,99	0,23	0,48	0,48	4	0,66	0,40	0,40
P9	9	17,00	16,35	0,23	0,62	0,38	4	0,72	0,90	0,60
P1	25	19,00	18,73	0,23	0,99	0,24	4	0,82	0,20	0,20
P2	6	14,00	13,02	0,23	0,50	0,47	4	0,66	0,10	0,10
P14	12	19,00	18,72	0,24	0,99	0,24	4	0,82	1,10	1,40
P1	8	17,00	16,33	0,24	0,62	0,39	4	0,72	0,90	0,60
P10	36	17,00	16,32	0,24	0,62	0,39	4	0,71	0,40	0,40

P14	32	16,00	15,18	0,25	0,57	0,44	4	0,69	0,60	0,50
P14	6	13,00	11,90	0,25	0,48	0,52	4	0,63	1,20	1,00
P9	14	17,00	16,29	0,26	0,62	0,41	4	0,70	0,50	0,50
P11	12	19,00	18,70	0,26	0,99	0,26	4	0,81	1,10	1,40
P8	8	17,00	16,27	0,26	0,62	0,42	4	0,69	0,90	0,60
P8	33	17,00	16,27	0,26	0,62	0,42	4	0,69	0,90	0,60
P1	36	17,00	16,26	0,27	0,62	0,43	4	0,69	0,90	0,60
P11	32	16,00	15,12	0,27	0,57	0,47	4	0,66	0,20	0,30
P10	9	17,00	16,26	0,27	0,62	0,43	4	0,69	0,90	0,60
P4	20	15,00	13,98	0,27	0,53	0,51	4	0,64	0,40	0,40
P1	31	15,00	13,93	0,28	0,53	0,53	4	0,62	0,90	0,80
P7	31	14,00	12,80	0,29	0,50	0,57	4	0,60	0,60	0,50
P8	2	17,00	16,20	0,29	0,62	0,46	4	0,67	0,30	0,40
P8	36	17,00	16,20	0,29	0,62	0,46	4	0,67	0,90	0,60
P3	13	19,00	18,66	0,29	0,99	0,29	4	0,79	0,20	0,20
P6	13	19,00	18,66	0,29	0,99	0,29	4	0,79	0,20	0,20
P4	39	17,00	16,19	0,29	0,62	0,46	4	0,67	0,40	0,40
P4	5	16,00	15,04	0,29	0,57	0,51	4	0,64	0,60	0,50
P4	33	16,00	15,04	0,29	0,57	0,51	4	0,64	0,60	0,50
P14	18	16,00	15,03	0,29	0,57	0,51	4	0,63	0,60	0,50
P5	35	16,00	15,03	0,29	0,57	0,51	4	0,63	1,20	1,10
P9	7	16,00	15,02	0,29	0,57	0,52	4	0,63	0,60	0,50
P2	36	17,00	16,17	0,30	0,62	0,48	4	0,66	0,90	0,60
P12	29	19,00	18,64	0,30	0,99	0,30	4	0,78	1,10	1,20
P9	18	17,00	16,15	0,30	0,62	0,48	4	0,65	0,90	0,60
P11	11	17,00	16,14	0,30	0,62	0,49	4	0,65	0,60	1,40
P14	19	14,00	12,72	0,30	0,50	0,61	4	0,58	1,70	1,40
P14	31	14,00	12,72	0,30	0,50	0,61	4	0,58	1,70	1,40
P13	7	15,00	13,83	0,31	0,53	0,58	4	0,59	0,30	0,30
P8	9	17,00	16,13	0,31	0,62	0,50	4	0,64	0,90	0,60
P1	14	17,00	16,12	0,31	0,62	0,50	4	0,64	0,50	0,50
P4	36	16,00	14,96	0,31	0,57	0,55	4	0,61	0,60	0,50
P7	7	15,00	13,79	0,32	0,53	0,60	4	0,58	0,60	0,50
P7	27	15,00	13,79	0,32	0,53	0,60	4	0,58	2,80	2,40
P11	31	14,00	12,64	0,32	0,50	0,65	4	0,55	0,70	0,60
P1	13	19,00	18,60	0,33	0,99	0,33	4	0,76	0,20	0,20
P8	32	17,00	16,06	0,33	0,62	0,53	4	0,62	0,90	0,60
P11	7	15,00	13,64	0,35	0,53	0,67	4	0,54	0,90	0,80
P14	10	16,00	14,80	0,36	0,57	0,63	4	0,56	0,60	0,50
P2	13	19,00	18,55	0,36	0,99	0,37	4	0,73	0,20	0,20
P2	39	18,00	17,20	0,36	0,72	0,50	4	0,64	1,20	1,10
P9	10	17,00	15,95	0,37	0,62	0,59	4	0,59	0,90	0,60
P11	10	16,00	14,73	0,37	0,57	0,66	4	0,55	0,60	0,50
P8	18	17,00	15,92	0,38	0,62	0,61	4	0,58	0,90	0,60
P9	35	17,00	15,89	0,39	0,62	0,62	4	0,57	0,30	0,40
P12	1	15,00	13,45	0,40	0,53	0,75	4	0,49	2,50	2,50
P14	40	18,00	17,07	0,41	0,72	0,57	4	0,60	0,50	0,30
P13	34	15,00	13,37	0,42	0,53	0,79	4	0,47	0,30	0,30

P5	30	18,00	17,06	0,42	0,72	0,58	4	0,59	1,20	1,10
P8	10	17,00	15,72	0,44	0,62	0,71	4	0,52	0,90	0,60
P4	7	15,00	13,26	0,44	0,53	0,84	4	0,45	0,30	0,30
P9	29	19,00	18,43	0,45	0,99	0,45	4	0,68	0,20	0,20
P13	22	18,00	16,98	0,45	0,72	0,62	4	0,57	1,30	1,10
P3	29	19,00	18,41	0,46	0,99	0,46	4	0,67	1,10	1,20
P6	29	19,00	18,41	0,46	0,99	0,46	4	0,67	1,10	1,20
P5	2	17,00	15,63	0,47	0,62	0,75	4	0,49	0,50	0,50
P5	36	17,00	15,63	0,47	0,62	0,75	4	0,49	0,90	0,60
P2	35	17,00	15,61	0,47	0,62	0,76	4	0,49	0,30	0,40
P8	34	16,00	14,34	0,48	0,57	0,84	4	0,45	0,60	0,50
P7	16	11,00	9,06	0,49	0,48	1,01	4	0,37	0,70	0,90
P7	4	17,00	15,56	0,49	0,62	0,79	4	0,48	0,30	0,40
P1	29	19,00	18,34	0,50	0,99	0,50	4	0,64	1,10	1,20
P14	30	18,00	16,83	0,50	0,72	0,70	4	0,52	0,50	0,30
P7	8	17,00	15,48	0,51	0,62	0,82	4	0,46	0,40	0,40
P11	4	17,00	15,43	0,53	0,62	0,85	4	0,44	0,30	0,40
P14	8	17,00	15,42	0,53	0,62	0,85	4	0,44	0,90	0,60
P14	33	17,00	15,42	0,53	0,62	0,85	4	0,44	0,90	0,60
P2	29	19,00	18,29	0,53	0,99	0,54	4	0,62	1,10	1,20
P5	13	19,00	18,28	0,53	0,99	0,54	4	0,62	0,20	0,20
P5	1	14,00	11,70	0,54	0,50	1,07	4	0,34	1,30	1,50
P8	20	17,00	15,37	0,54	0,62	0,88	4	0,43	0,90	0,60
P5	27	16,00	14,06	0,55	0,57	0,97	4	0,39	0,20	0,20
P4	34	15,00	12,78	0,55	0,53	1,05	4	0,35	0,30	0,30
P10	31	16,00	14,02	0,56	0,57	0,99	4	0,38	0,20	0,30
P13	13	19,00	18,18	0,59	0,99	0,60	4	0,58	0,20	0,20
P7	1	14,00	11,41	0,60	0,50	1,21	4	0,29	0,30	0,40
P9	4	18,00	16,55	0,61	0,72	0,84	4	0,45	0,50	0,30
P12	40	19,00	18,14	0,61	0,99	0,62	4	0,57	1,10	1,20
P7	18	17,00	15,10	0,62	0,62	1,00	4	0,37	0,60	1,40
P5	11	18,00	16,45	0,64	0,72	0,89	4	0,42	0,40	0,50
P12	35	18,00	16,33	0,68	0,72	0,95	4	0,40	0,40	0,50
P10	2	18,00	16,32	0,68	0,72	0,95	4	0,39	2,30	2,50
P12	30	19,00	17,97	0,70	0,99	0,71	4	0,52	1,10	1,20
P13	23	19,00	17,94	0,72	0,99	0,73	4	0,51	1,10	1,30
P4	13	19,00	17,92	0,73	0,99	0,74	4	0,50	0,20	0,20
P10	14	18,00	16,19	0,73	0,72	1,02	4	0,37	0,40	0,50
P14	23	19,00	17,88	0,75	0,99	0,75	4	0,49	1,10	1,20
P8	14	18,00	16,06	0,77	0,72	1,07	4	0,34	0,40	0,50
P10	40	19,00	17,79	0,79	0,99	0,80	4	0,47	0,20	0,20
P11	20	17,00	14,34	0,83	0,62	1,34	4	0,25	0,30	0,30
P3	35	18,00	15,85	0,84	0,72	1,17	4	0,31	0,40	0,50
P6	35	18,00	15,85	0,84	0,72	1,17	4	0,31	0,40	0,50
P4	23	19,00	17,65	0,86	0,99	0,87	4	0,44	1,10	1,30
P3	30	19,00	17,64	0,86	0,99	0,87	4	0,43	1,10	1,20
P6	30	19,00	17,64	0,86	0,99	0,87	4	0,43	1,10	1,20
P8	38	19,00	17,63	0,86	0,99	0,88	4	0,43	1,10	1,30

P1	35	18,00	15,72	0,88	0,72	1,23	4	0,29	0,40	0,50
P1	30	19,00	17,54	0,90	0,99	0,91	4	0,41	1,10	1,20
P12	11	19,00	17,50	0,93	0,99	0,94	4	0,40	0,20	0,20
P5	40	19,00	17,28	1,01	0,99	1,03	4	0,36	1,10	1,20
P3	11	19,00	17,12	1,08	0,99	1,10	4	0,33	0,20	0,20
P6	11	19,00	17,12	1,08	0,99	1,10	4	0,33	0,20	0,20
P11	18	18,00	14,96	1,10	0,72	1,53	4	0,20	0,50	0,30
P1	11	19,00	17,01	1,13	0,99	1,14	4	0,32	0,20	0,20
P2	11	19,00	16,92	1,16	0,99	1,17	4	0,31	0,20	0,20
P12	16	16,00	10,93	1,28	0,57	2,25	4	0,09	1,70	1,70
P5	16	15,00	9,31	1,38	0,53	2,61	4	0,06	0,70	0,90
P14	27	19,00	13,71	2,11	0,99	2,14	4	0,10	1,10	1,20
P11	27	19,00	13,64	2,13	0,99	2,16	4	0,10	1,10	1,20
P1	26	20,00	19,62	-0,27	1,41	-0,19	4	0,86	0,20	0,10
P2	12	20,00	19,02	0,66	1,41	0,47	4	0,66	0,20	0,10

EK 9. Ön Test Matematik Puanlayıcı Grubu *Puanlayıcı X Birey* Etkileşimleri

Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Yanlılık Büyüklüğü	SH	t	sd	p	infit	outfit
P1	25	24	21,34	1,58	1,38	1,14	5	0,31	0,10	0,10
P3	8	24	22,17	1,22	1,38	0,88	5	0,42	0,10	0,10
P3	21	24	21,03	1,69	1,38	1,22	5	0,28	0,10	0,10
P6	8	24	21,09	1,67	1,38	1,20	5	0,28	0,10	0,10
P19	8	24	22,41	1,09	1,38	0,79	5	0,47	0,10	0,10
P19	21	24	21,4	1,56	1,38	1,12	5	0,31	0,10	0,10
P19	25	24	22,27	1,16	1,38	0,84	5	0,44	0,10	0,10
P8	2	9	10,67	-1,47	1,12	-1,31	5	0,25	0,30	0,10
P7	2	9	10,63	-1,44	1,12	-1,28	5	0,26	0,30	0,10
P10	32	10	12,27	-1,23	0,87	-1,42	5	0,22	0,10	0,10
P9	16	12	15,05	-1,11	0,67	-1,65	5	0,16	0,60	0,30
P1	16	12	15,02	-1,10	0,67	-1,63	5	0,16	0,60	0,30
P2	8	18	21,69	-1,09	0,50	-2,18	5	0,08	0,70	0,40
P10	2	10	11,72	-0,98	0,87	-1,13	5	0,31	0,10	0,10
P1	32	10	11,58	-0,92	0,87	-1,05	5	0,34	0,10	0,10
P11	14	18	20,99	-0,83	0,50	-1,66	5	0,16	0,90	0,70
P11	12	15	18,04	-0,81	0,54	-1,49	5	0,20	2,80	3,30
P4	8	18	20,9	-0,80	0,50	-1,60	5	0,17	1,20	0,90
P11	11	18	20,85	-0,79	0,50	-1,57	5	0,18	0,70	9,00
P11	40	13	15,38	-0,78	0,63	-1,25	5	0,27	3,20	8,10
P18	2	10	11,26	-0,76	0,87	-0,87	5	0,42	0,10	0,10
P18	10	13	15,29	-0,76	0,63	-1,21	5	0,28	1,30	1,10
P19	39	16	18,92	-0,75	0,52	-1,45	5	0,21	1,30	0,90
P14	22	12	13,87	-0,74	0,67	-1,10	5	0,32	0,60	0,30
P12	22	12	13,82	-0,72	0,67	-1,07	5	0,33	0,60	0,30
P6	37	18	20,64	-0,72	0,50	-1,43	5	0,21	0,90	1,40
P10	1	12	13,78	-0,71	0,67	-1,05	5	0,34	0,60	0,30
P8	39	14	16,42	-0,70	0,58	-1,22	5	0,28	0,20	0,20
P8	29	15	17,57	-0,69	0,54	-1,27	5	0,26	1,10	0,90
P7	39	14	16,32	-0,68	0,58	-1,17	5	0,29	0,20	0,20
P16	19	19	21,27	-0,68	0,51	-1,32	5	0,24	1,50	0,90
P9	2	10	11,1	-0,68	0,87	-0,78	5	0,47	0,10	0,10
P3	26	16	18,6	-0,67	0,52	-1,29	5	0,25	1,30	2,30
P1	2	10	11,08	-0,67	0,87	-0,77	5	0,48	0,10	0,10
P10	12	16	18,59	-0,66	0,52	-1,28	5	0,26	0,60	0,40
P7	29	15	17,47	-0,66	0,54	-1,22	5	0,28	1,10	0,90
P15	32	10	11,06	-0,65	0,87	-0,75	5	0,49	0,10	0,10
P20	32	10	11,06	-0,65	0,87	-0,75	5	0,49	0,10	0,10
P13	25	19	21,21	-0,65	0,51	-1,28	5	0,26	0,80	0,50
P14	25	19	21,21	-0,65	0,51	-1,28	5	0,26	0,70	0,50
P5	25	19	21,16	-0,64	0,51	-1,25	5	0,27	0,80	0,50
P10	10	14	16,1	-0,62	0,58	-1,07	5	0,33	1,20	0,80
P19	33	17	19,41	-0,62	0,51	-1,22	5	0,28	0,40	0,30
P11	38	15	17,27	-0,61	0,54	-1,13	5	0,31	0,80	9,00

P8	13	18	20,28	-0,61	0,50	-1,21	5	0,28	0,40	0,30
P19	17	16	18,37	-0,61	0,52	-1,17	5	0,29	0,60	0,40
P9	13	19	21,02	-0,59	0,51	-1,15	5	0,30	0,40	0,30
P7	13	18	20,2	-0,58	0,50	-1,16	5	0,30	0,40	0,30
P1	19	18	20,18	-0,58	0,50	-1,15	5	0,30	0,10	0,10
P19	38	16	18,21	-0,57	0,52	-1,10	5	0,32	0,80	0,60
P8	22	12	13,38	-0,57	0,67	-0,84	5	0,44	0,60	0,30
P1	23	16	18,15	-0,55	0,52	-1,07	5	0,33	0,40	0,30
P16	17	16	18,12	-0,54	0,52	-1,05	5	0,34	2,70	2,40
P7	22	12	13,31	-0,54	0,67	-0,80	5	0,46	0,60	0,30
P3	40	14	15,75	-0,52	0,58	-0,91	5	0,41	0,50	0,40
P15	22	12	13,26	-0,52	0,67	-0,78	5	0,47	0,60	0,30
P20	22	12	13,26	-0,52	0,67	-0,78	5	0,47	0,60	0,30
P7	27	14	15,75	-0,52	0,58	-0,91	5	0,41	0,70	9,00
P17	21	18	19,99	-0,52	0,50	-1,04	5	0,35	1,10	0,70
P15	17	14	15,73	-0,52	0,58	-0,90	5	0,41	0,80	0,50
P20	17	14	15,73	-0,52	0,58	-0,90	5	0,41	0,80	0,50
P13	21	18	19,96	-0,51	0,50	-1,02	5	0,35	1,10	0,70
P18	7	17	19,02	-0,51	0,51	-1,02	5	0,36	1,10	0,80
P16	35	20	21,55	-0,51	0,53	-0,96	5	0,38	1,40	1,50
P15	27	14	15,68	-0,51	0,58	-0,88	5	0,42	0,80	0,70
P20	27	14	15,68	-0,51	0,58	-0,88	5	0,42	0,80	0,70
P5	21	18	19,9	-0,50	0,50	-0,99	5	0,37	1,10	0,70
P12	21	18	19,9	-0,50	0,50	-0,99	5	0,37	1,10	0,70
P10	7	18	19,85	-0,48	0,50	-0,96	5	0,38	1,90	1,40
P3	33	17	18,9	-0,48	0,51	-0,96	5	0,38	0,40	0,30
P6	36	19	20,7	-0,48	0,51	-0,94	5	0,39	0,70	0,50
P2	31	15	16,76	-0,48	0,54	-0,89	5	0,42	2,50	3,10
P11	9	15	16,75	-0,48	0,54	-0,88	5	0,42	2,80	2,40
P10	35	20	21,46	-0,48	0,53	-0,89	5	0,41	1,40	1,50
P11	23	17	18,85	-0,47	0,51	-0,93	5	0,40	1,10	0,80
P6	29	16	17,8	-0,46	0,52	-0,90	5	0,41	0,40	0,30
P19	37	21	22,14	-0,46	0,58	-0,79	5	0,46	1,50	0,80
P19	2	11	11,93	-0,46	0,73	-0,63	5	0,56	0,40	0,30
P8	16	13	14,29	-0,45	0,63	-0,73	5	0,50	4,30	2,80
P12	8	20	21,37	-0,44	0,53	-0,83	5	0,45	1,30	0,80
P18	12	16	17,71	-0,44	0,52	-0,85	5	0,43	0,60	0,40
P3	38	16	17,68	-0,43	0,52	-0,84	5	0,44	0,80	0,60
P1	38	15	16,58	-0,43	0,54	-0,80	5	0,46	1,00	0,60
P17	17	15	16,57	-0,43	0,54	-0,80	5	0,46	0,60	0,50
P7	16	13	14,21	-0,43	0,63	-0,69	5	0,52	4,30	2,80
P9	7	17	18,69	-0,43	0,51	-0,85	5	0,43	1,10	0,80
P13	17	15	16,54	-0,42	0,54	-0,78	5	0,47	0,60	0,50
P4	14	18	19,63	-0,42	0,50	-0,84	5	0,44	0,90	1,40
P2	11	19	20,48	-0,41	0,51	-0,81	5	0,46	0,60	1,50
P2	12	16	17,57	-0,41	0,52	-0,79	5	0,47	0,60	0,40
P5	17	15	16,48	-0,41	0,54	-0,75	5	0,49	0,60	0,50
P12	17	15	16,48	-0,41	0,54	-0,75	5	0,49	2,20	1,40

P15	26	15	16,46	-0,40	0,54	-0,74	5	0,49	1,60	1,80
P20	26	15	16,46	-0,40	0,54	-0,74	5	0,49	1,60	1,80
P18	13	20	21,25	-0,40	0,53	-0,75	5	0,49	0,40	0,30
P7	36	19	20,43	-0,40	0,51	-0,78	5	0,47	0,50	0,40
P17	4	16	17,53	-0,40	0,52	-0,76	5	0,48	0,40	0,30
P1	1	12	12,92	-0,39	0,67	-0,58	5	0,59	0,60	0,30
P17	25	20	21,23	-0,39	0,53	-0,73	5	0,50	0,80	0,50
P8	9	14	15,26	-0,39	0,58	-0,67	5	0,53	0,50	0,40
P13	4	16	17,5	-0,39	0,52	-0,75	5	0,49	0,40	0,30
P14	4	16	17,5	-0,39	0,52	-0,75	5	0,49	0,90	0,90
P6	23	16	17,49	-0,39	0,52	-0,75	5	0,49	0,40	0,30
P19	1	13	14,07	-0,38	0,63	-0,61	5	0,57	0,30	0,20
P3	28	18	19,49	-0,38	0,50	-0,76	5	0,48	0,50	0,40
P7	12	15	16,37	-0,38	0,54	-0,70	5	0,52	3,20	4,40
P11	24	18	19,44	-0,37	0,50	-0,74	5	0,49	2,50	2,10
P5	4	16	17,43	-0,37	0,52	-0,72	5	0,51	0,40	0,30
P12	4	16	17,43	-0,37	0,52	-0,72	5	0,51	0,40	0,30
P14	31	15	16,34	-0,37	0,54	-0,68	5	0,52	2,30	2,30
P1	22	13	14,02	-0,37	0,63	-0,59	5	0,58	0,80	0,70
P2	13	20	21,16	-0,37	0,53	-0,68	5	0,52	0,50	0,40
P15	29	16	17,41	-0,36	0,52	-0,70	5	0,51	0,60	0,70
P20	29	16	17,41	-0,36	0,52	-0,70	5	0,51	0,60	0,70
P12	25	20	21,16	-0,36	0,53	-0,68	5	0,53	0,80	0,50
P7	9	14	15,18	-0,36	0,58	-0,63	5	0,56	0,50	0,40
P9	11	19	20,31	-0,36	0,51	-0,71	5	0,51	0,60	1,50
P16	16	15	16,29	-0,36	0,54	-0,66	5	0,54	1,00	0,60
P9	12	16	17,38	-0,36	0,52	-0,69	5	0,52	0,60	0,40
P3	39	17	18,4	-0,36	0,51	-0,70	5	0,51	1,20	0,80
P12	31	15	16,28	-0,35	0,54	-0,65	5	0,54	2,30	2,30
P2	32	11	11,71	-0,35	0,73	-0,48	5	0,65	0,20	0,10
P3	6	19	20,28	-0,35	0,51	-0,69	5	0,52	1,50	0,90
P15	9	14	15,12	-0,35	0,58	-0,60	5	0,57	0,50	0,40
P20	9	14	15,12	-0,35	0,58	-0,60	5	0,57	0,50	0,40
P4	13	19	20,25	-0,34	0,51	-0,68	5	0,53	0,40	0,30
P16	10	15	16,23	-0,34	0,54	-0,63	5	0,56	1,20	0,80
P14	26	16	17,3	-0,34	0,52	-0,65	5	0,54	0,60	0,70
P18	40	14	15,08	-0,34	0,58	-0,58	5	0,59	0,50	0,40
P9	6	18	19,3	-0,33	0,50	-0,67	5	0,53	0,80	0,70
P4	21	18	19,29	-0,33	0,50	-0,66	5	0,54	0,70	0,40
P19	15	15	16,19	-0,33	0,54	-0,61	5	0,57	1,10	1,10
P17	22	13	13,89	-0,33	0,63	-0,52	5	0,62	0,60	0,40
P3	2	11	11,64	-0,32	0,73	-0,44	5	0,68	0,20	0,10
P15	15	13	13,89	-0,32	0,63	-0,52	5	0,63	0,40	0,50
P20	15	13	13,89	-0,32	0,63	-0,52	5	0,63	0,40	0,50
P10	16	15	16,16	-0,32	0,54	-0,59	5	0,58	1,00	0,60
P16	1	13	13,88	-0,32	0,63	-0,51	5	0,63	0,80	0,70
P18	15	14	15,03	-0,32	0,58	-0,56	5	0,60	0,60	0,50
P11	31	16	17,22	-0,32	0,52	-0,61	5	0,57	2,40	2,20

P17	28	17	18,24	-0,31	0,51	-0,62	5	0,56	0,60	0,90
P9	10	14	15	-0,31	0,58	-0,54	5	0,61	1,20	0,80
P13	28	17	18,21	-0,31	0,51	-0,61	5	0,57	0,60	0,90
P14	28	17	18,21	-0,31	0,51	-0,61	5	0,57	0,60	0,90
P9	32	11	11,6	-0,30	0,73	-0,42	5	0,69	0,20	0,10
P1	10	14	14,97	-0,30	0,58	-0,53	5	0,62	1,20	0,80
P11	26	17	18,2	-0,30	0,51	-0,60	5	0,58	1,50	1,50
P1	30	18	19,17	-0,30	0,50	-0,60	5	0,58	0,40	0,30
P11	25	21	21,78	-0,30	0,58	-0,51	5	0,63	0,30	0,40
P15	21	18	19,17	-0,30	0,50	-0,59	5	0,58	1,10	0,70
P20	21	18	19,17	-0,30	0,50	-0,59	5	0,58	1,10	0,70
P19	31	17	18,16	-0,29	0,51	-0,58	5	0,59	0,90	0,60
P5	28	17	18,14	-0,29	0,51	-0,57	5	0,59	0,60	0,90
P12	28	17	18,14	-0,29	0,51	-0,57	5	0,59	0,60	0,90
P2	15	14	14,92	-0,29	0,58	-0,50	5	0,64	0,60	0,50
P3	22	14	14,92	-0,29	0,58	-0,50	5	0,64	1,00	0,90
P3	16	15	16,03	-0,29	0,54	-0,53	5	0,62	1,00	0,60
P15	30	17	18,13	-0,29	0,51	-0,57	5	0,60	0,60	0,50
P20	30	17	18,13	-0,29	0,51	-0,57	5	0,60	0,60	0,50
P18	24	18	19,12	-0,29	0,50	-0,57	5	0,59	0,70	0,50
P13	39	16	17,09	-0,29	0,52	-0,55	5	0,61	0,40	0,60
P17	6	18	19,11	-0,28	0,50	-0,57	5	0,60	1,60	1,00
P16	40	15	16,01	-0,28	0,54	-0,52	5	0,62	0,40	0,30
P17	19	19	20,04	-0,28	0,51	-0,55	5	0,60	1,50	0,90
P6	24	17	18,11	-0,28	0,51	-0,56	5	0,60	0,30	0,20
P3	13	21	21,74	-0,28	0,58	-0,48	5	0,65	1,50	0,80
P13	6	18	19,08	-0,28	0,50	-0,55	5	0,61	1,60	1,00
P13	19	19	20,01	-0,27	0,51	-0,54	5	0,61	1,50	0,90
P14	19	19	20,01	-0,27	0,51	-0,54	5	0,61	1,50	0,90
P3	10	15	15,97	-0,27	0,54	-0,50	5	0,64	0,60	0,50
P5	39	16	17,03	-0,27	0,52	-0,52	5	0,63	0,40	0,60
P6	38	15	15,95	-0,27	0,54	-0,49	5	0,64	1,00	0,60
P16	15	15	15,95	-0,27	0,54	-0,49	5	0,64	1,10	1,10
P6	25	20	20,87	-0,26	0,53	-0,50	5	0,64	0,80	0,50
P19	35	21	21,7	-0,26	0,58	-0,45	5	0,67	1,50	0,80
P6	35	19	19,98	-0,26	0,51	-0,52	5	0,63	0,40	0,30
P19	28	19	19,97	-0,26	0,51	-0,51	5	0,63	0,80	0,50
P5	6	18	19,02	-0,26	0,50	-0,52	5	0,63	1,60	1,00
P14	21	19	19,96	-0,26	0,51	-0,51	5	0,63	0,70	0,50
P5	19	19	19,95	-0,26	0,51	-0,50	5	0,64	1,50	0,90
P12	19	19	19,95	-0,26	0,51	-0,50	5	0,64	1,50	0,90
P10	24	19	19,94	-0,25	0,51	-0,50	5	0,64	1,50	0,90
P6	7	17	18,01	-0,25	0,51	-0,50	5	0,64	1,10	0,80
P17	32	11	11,5	-0,25	0,73	-0,35	5	0,74	0,60	1,60
P2	24	18	18,99	-0,25	0,50	-0,50	5	0,64	0,70	0,50
P9	40	14	14,8	-0,25	0,58	-0,44	5	0,68	0,50	0,40
P18	35	20	20,83	-0,25	0,53	-0,47	5	0,66	0,40	0,30
P8	17	15	15,88	-0,25	0,54	-0,46	5	0,67	0,20	0,20

P13	32	11	11,48	-0,25	0,73	-0,34	5	0,75	0,60	1,60
P14	32	11	11,48	-0,25	0,73	-0,34	5	0,75	0,60	1,60
P11	36	21	21,66	-0,24	0,58	-0,42	5	0,69	1,00	2,90
P1	40	14	14,77	-0,24	0,58	-0,42	5	0,69	0,50	0,40
P15	8	20	20,8	-0,24	0,53	-0,45	5	0,67	0,80	0,50
P20	8	20	20,8	-0,24	0,53	-0,45	5	0,67	0,80	0,50
P9	15	14	14,75	-0,24	0,58	-0,41	5	0,70	0,60	0,50
P8	27	15	15,84	-0,24	0,54	-0,44	5	0,68	0,30	0,30
P5	30	18	18,91	-0,23	0,50	-0,46	5	0,66	0,40	0,30
P10	15	15	15,82	-0,23	0,54	-0,43	5	0,69	0,60	0,60
P5	32	11	11,45	-0,23	0,73	-0,31	5	0,77	0,60	1,60
P12	32	11	11,45	-0,23	0,73	-0,31	5	0,77	0,60	1,60
P11	30	19	19,83	-0,22	0,51	-0,43	5	0,68	1,20	0,80
P11	2	11	11,43	-0,22	0,73	-0,30	5	0,77	0,40	0,30
P18	14	20	20,73	-0,22	0,53	-0,41	5	0,70	0,40	0,30
P2	35	20	20,72	-0,22	0,53	-0,41	5	0,70	0,40	0,30
P4	24	17	17,85	-0,21	0,51	-0,42	5	0,69	0,60	0,50
P19	6	20	20,71	-0,21	0,53	-0,40	5	0,70	1,00	0,60
P10	38	17	17,82	-0,21	0,51	-0,41	5	0,70	1,40	1,00
P18	29	18	18,82	-0,21	0,50	-0,41	5	0,70	0,70	0,50
P8	35	19	19,77	-0,21	0,51	-0,40	5	0,70	0,90	0,90
P4	4	16	16,78	-0,21	0,52	-0,40	5	0,71	0,40	0,30
P9	24	18	18,8	-0,20	0,50	-0,40	5	0,70	0,70	0,50
P4	38	15	15,71	-0,20	0,54	-0,37	5	0,73	1,00	0,60
P4	25	20	20,66	-0,20	0,53	-0,37	5	0,73	0,80	0,50
P3	15	15	15,7	-0,20	0,54	-0,36	5	0,73	1,10	1,10
P10	31	17	17,77	-0,20	0,51	-0,39	5	0,72	2,10	1,60
P1	24	18	18,76	-0,19	0,50	-0,39	5	0,72	0,70	0,50
P3	7	19	19,72	-0,19	0,51	-0,38	5	0,72	1,50	0,90
P6	28	17	17,75	-0,19	0,51	-0,38	5	0,72	0,30	0,20
P4	31	15	15,66	-0,19	0,54	-0,35	5	0,74	3,10	3,50
P19	11	21	21,51	-0,19	0,58	-0,32	5	0,76	1,00	2,90
P6	11	19	19,7	-0,19	0,51	-0,37	5	0,73	0,60	1,50
P2	14	20	20,63	-0,19	0,53	-0,35	5	0,74	0,40	0,30
P16	12	18	18,73	-0,19	0,50	-0,37	5	0,73	0,60	0,50
P15	24	17	17,71	-0,18	0,51	-0,36	5	0,73	0,20	0,20
P20	24	17	17,71	-0,18	0,51	-0,36	5	0,73	0,20	0,20
P7	35	19	19,68	-0,18	0,51	-0,35	5	0,74	0,90	0,90
P1	27	16	16,68	-0,18	0,52	-0,35	5	0,74	0,40	0,30
P18	11	20	20,59	-0,18	0,53	-0,33	5	0,76	0,40	0,30
P11	13	21	21,48	-0,17	0,58	-0,30	5	0,78	0,40	0,20
P8	1	12	12,38	-0,17	0,67	-0,25	5	0,81	0,50	0,90
P15	25	20	20,56	-0,17	0,53	-0,31	5	0,77	0,80	0,50
P20	25	20	20,56	-0,17	0,53	-0,31	5	0,77	0,80	0,50
P15	35	19	19,62	-0,16	0,51	-0,32	5	0,76	0,40	0,30
P20	35	19	19,62	-0,16	0,51	-0,32	5	0,76	0,40	0,30
P18	6	19	19,62	-0,16	0,51	-0,32	5	0,76	0,30	0,20
P17	33	17	17,63	-0,16	0,51	-0,32	5	0,76	0,50	0,40

P3	31	17	17,63	-0,16	0,51	-0,32	5	0,76	0,70	0,60
P18	36	21	21,44	-0,16	0,58	-0,27	5	0,80	0,40	0,20
P19	19	21	21,44	-0,16	0,58	-0,27	5	0,80	0,60	0,50
P17	8	21	21,44	-0,16	0,58	-0,27	5	0,80	1,50	0,80
P16	32	12	12,35	-0,16	0,67	-0,23	5	0,83	1,20	1,30
P15	31	15	15,54	-0,16	0,54	-0,29	5	0,79	0,40	0,30
P20	31	15	15,54	-0,16	0,54	-0,29	5	0,79	0,40	0,30
P15	7	17	17,61	-0,16	0,51	-0,31	5	0,77	0,30	0,20
P20	7	17	17,61	-0,16	0,51	-0,31	5	0,77	0,30	0,20
P4	26	16	16,58	-0,15	0,52	-0,30	5	0,78	0,60	0,70
P13	33	17	17,6	-0,15	0,51	-0,30	5	0,78	0,50	0,40
P14	33	17	17,6	-0,15	0,51	-0,30	5	0,78	0,50	0,40
P8	36	20	20,51	-0,15	0,53	-0,28	5	0,79	0,80	0,50
P9	31	16	16,56	-0,15	0,52	-0,29	5	0,79	2,40	2,20
P13	8	21	21,41	-0,15	0,58	-0,26	5	0,81	1,50	0,80
P14	8	21	21,41	-0,15	0,58	-0,26	5	0,81	1,50	0,80
P11	1	13	13,39	-0,15	0,63	-0,24	5	0,82	0,30	0,20
P18	19	20	20,49	-0,15	0,53	-0,27	5	0,80	1,00	0,60
P7	1	12	12,32	-0,14	0,67	-0,21	5	0,84	0,50	0,90
P14	17	16	16,54	-0,14	0,52	-0,28	5	0,79	2,00	1,20
P6	21	19	19,54	-0,14	0,51	-0,28	5	0,79	0,90	0,60
P18	37	21	21,4	-0,14	0,58	-0,24	5	0,82	0,40	0,20
P1	31	16	16,53	-0,14	0,52	-0,27	5	0,80	0,90	0,70
P6	13	20	20,47	-0,14	0,53	-0,26	5	0,80	0,80	0,50
P17	27	16	16,52	-0,14	0,52	-0,27	5	0,80	0,40	0,30
P7	26	16	16,52	-0,14	0,52	-0,27	5	0,80	1,50	2,50
P6	10	14	14,42	-0,14	0,58	-0,24	5	0,82	1,20	0,80
P10	39	18	18,54	-0,14	0,50	-0,27	5	0,80	0,80	0,50
P9	14	20	20,46	-0,14	0,53	-0,26	5	0,81	0,40	0,30
P18	30	19	19,52	-0,14	0,51	-0,27	5	0,80	0,80	0,50
P4	29	17	17,54	-0,14	0,51	-0,27	5	0,80	0,60	0,50
P5	33	17	17,53	-0,14	0,51	-0,27	5	0,80	0,50	0,40
P12	33	17	17,53	-0,14	0,51	-0,27	5	0,80	0,50	0,40
P4	22	13	13,35	-0,13	0,63	-0,22	5	0,84	0,30	0,20
P6	30	18	18,53	-0,13	0,50	-0,27	5	0,80	0,40	0,30
P5	8	21	21,37	-0,13	0,58	-0,23	5	0,83	1,50	0,80
P3	35	21	21,37	-0,13	0,58	-0,23	5	0,83	1,50	0,80
P13	27	16	16,49	-0,13	0,52	-0,25	5	0,81	0,40	0,30
P14	27	16	16,49	-0,13	0,52	-0,25	5	0,81	0,40	0,30
P2	6	19	19,49	-0,13	0,51	-0,25	5	0,81	0,20	0,10
P11	33	18	18,5	-0,13	0,50	-0,25	5	0,81	1,80	9,00
P17	7	18	18,5	-0,13	0,50	-0,25	5	0,81	1,60	1,00
P2	36	21	21,35	-0,13	0,58	-0,22	5	0,84	0,40	0,20
P16	11	21	21,35	-0,13	0,58	-0,21	5	0,84	1,00	2,90
P19	10	16	16,47	-0,12	0,52	-0,24	5	0,82	0,50	0,40
P4	28	17	17,49	-0,12	0,51	-0,24	5	0,82	0,60	0,50
P4	11	19	19,46	-0,12	0,51	-0,24	5	0,82	0,60	1,50
P13	7	18	18,47	-0,12	0,50	-0,23	5	0,82	1,60	1,00

P1	29	18	18,46	-0,12	0,50	-0,23	5	0,83	0,40	0,30
P4	12	16	16,44	-0,12	0,52	-0,22	5	0,83	0,20	0,20
P5	27	16	16,43	-0,11	0,52	-0,22	5	0,84	0,40	0,30
P12	27	16	16,43	-0,11	0,52	-0,22	5	0,84	0,40	0,30
P2	19	20	20,38	-0,11	0,53	-0,21	5	0,84	1,00	0,60
P16	9	17	17,44	-0,11	0,51	-0,22	5	0,84	0,80	0,50
P15	36	20	20,38	-0,11	0,53	-0,21	5	0,84	0,40	0,30
P20	36	20	20,38	-0,11	0,53	-0,21	5	0,84	0,40	0,30
P2	37	21	21,31	-0,11	0,58	-0,19	5	0,86	0,40	0,20
P19	22	15	15,37	-0,11	0,54	-0,20	5	0,85	1,80	1,60
P2	30	19	19,39	-0,10	0,51	-0,20	5	0,85	0,80	0,50
P2	2	11	11,2	-0,10	0,73	-0,14	5	0,89	0,20	0,10
P5	7	18	18,4	-0,10	0,50	-0,20	5	0,85	1,60	1,00
P11	4	18	18,4	-0,10	0,50	-0,20	5	0,85	2,50	2,10
P10	23	19	19,38	-0,10	0,51	-0,19	5	0,85	1,50	0,90
P18	16	15	15,34	-0,10	0,54	-0,18	5	0,86	1,00	0,60
P15	37	20	20,32	-0,09	0,53	-0,17	5	0,87	0,40	0,30
P20	37	20	20,32	-0,09	0,53	-0,17	5	0,87	0,40	0,30
P4	19	19	19,35	-0,09	0,51	-0,18	5	0,87	0,30	0,20
P15	28	17	17,35	-0,09	0,51	-0,18	5	0,87	0,60	0,50
P20	28	17	17,35	-0,09	0,51	-0,18	5	0,87	0,60	0,50
P10	11	21	21,26	-0,09	0,58	-0,15	5	0,88	1,00	2,90
P1	12	17	17,34	-0,09	0,51	-0,17	5	0,87	0,40	0,30
P19	36	22	22,17	-0,08	0,69	-0,12	5	0,91	0,90	0,50
P17	26	17	17,33	-0,08	0,51	-0,17	5	0,87	0,80	1,10
P16	21	21	21,23	-0,08	0,58	-0,14	5	0,90	2,10	1,10
P19	27	18	18,31	-0,08	0,50	-0,16	5	0,88	0,10	0,10
P8	10	14	14,24	-0,08	0,58	-0,14	5	0,90	1,20	0,80
P14	14	20	20,27	-0,08	0,53	-0,15	5	0,89	1,50	2,50
P9	36	21	21,22	-0,08	0,58	-0,13	5	0,90	0,40	0,20
P10	9	17	17,3	-0,08	0,51	-0,15	5	0,89	1,20	0,80
P13	26	17	17,3	-0,08	0,51	-0,15	5	0,89	0,80	1,10
P8	30	18	18,3	-0,08	0,50	-0,15	5	0,89	0,40	0,30
P17	29	18	18,29	-0,07	0,50	-0,15	5	0,89	0,60	0,50
P8	32	11	11,13	-0,07	0,73	-0,10	5	0,93	0,60	0,80
P8	23	17	17,26	-0,07	0,51	-0,13	5	0,90	0,20	0,20
P13	29	18	18,26	-0,07	0,50	-0,13	5	0,90	0,60	0,50
P19	40	16	16,24	-0,07	0,52	-0,13	5	0,90	0,50	0,30
P2	16	15	15,22	-0,06	0,54	-0,12	5	0,91	1,00	0,60
P18	1	13	13,16	-0,06	0,63	-0,10	5	0,92	0,80	0,70
P6	15	14	14,19	-0,06	0,58	-0,11	5	0,92	0,60	0,50
P5	14	20	20,21	-0,06	0,53	-0,12	5	0,91	1,50	2,50
P12	14	20	20,21	-0,06	0,53	-0,12	5	0,91	1,50	2,50
P9	19	20	20,21	-0,06	0,53	-0,12	5	0,91	1,00	0,60
P15	6	18	18,24	-0,06	0,50	-0,12	5	0,91	0,10	0,10
P20	6	18	18,24	-0,06	0,50	-0,12	5	0,91	0,10	0,10
P4	23	17	17,23	-0,06	0,51	-0,12	5	0,91	0,30	0,20
P5	26	17	17,23	-0,06	0,51	-0,12	5	0,91	0,80	1,10

P12	26	17	17,23	-0,06	0,51	-0,12	5	0,91	0,80	1,10
P9	37	21	21,17	-0,06	0,58	-0,10	5	0,93	0,40	0,20
P19	7	20	20,19	-0,06	0,53	-0,10	5	0,92	1,00	0,60
P7	10	14	14,16	-0,05	0,58	-0,09	5	0,93	1,20	0,80
P9	30	19	19,2	-0,05	0,51	-0,10	5	0,92	0,80	0,50
P7	30	18	18,2	-0,05	0,50	-0,10	5	0,92	0,40	0,30
P2	10	15	15,17	-0,05	0,54	-0,09	5	0,93	0,60	0,50
P5	29	18	18,2	-0,05	0,50	-0,10	5	0,93	0,60	0,50
P16	22	15	15,16	-0,05	0,54	-0,08	5	0,94	3,10	3,30
P16	33	19	19,17	-0,05	0,51	-0,09	5	0,93	1,70	1,10
P7	23	17	17,17	-0,04	0,51	-0,08	5	0,94	0,20	0,20
P1	21	20	20,13	-0,04	0,53	-0,07	5	0,95	0,40	0,30
P15	10	14	14,11	-0,04	0,58	-0,07	5	0,95	1,20	0,80
P20	10	14	14,11	-0,04	0,58	-0,07	5	0,95	1,20	0,80
P6	33	17	17,14	-0,04	0,51	-0,07	5	0,95	0,50	0,40
P2	1	13	13,07	-0,03	0,63	-0,05	5	0,96	0,80	0,70
P6	17	16	16,1	-0,03	0,52	-0,05	5	0,96	0,20	0,20
P16	36	22	22,05	-0,02	0,69	-0,03	5	0,98	0,90	0,50
P1	9	16	16,08	-0,02	0,52	-0,04	5	0,97	0,80	0,60
P16	27	18	18,07	-0,02	0,50	-0,03	5	0,97	1,70	1,90
P18	4	18	18,06	-0,02	0,50	-0,03	5	0,98	1,90	1,40
P19	13	22	22,03	-0,01	0,69	-0,02	5	0,98	0,90	0,50
P4	40	14	14,03	-0,01	0,58	-0,02	5	0,99	0,50	0,40
P6	4	17	17,04	-0,01	0,51	-0,02	5	0,98	2,10	2,20
P12	39	17	17,03	-0,01	0,51	-0,01	5	0,99	0,90	1,10
P8	15	14	14,01	0,00	0,58	-0,01	5	0,99	1,20	1,40
P16	37	22	22,01	0,00	0,69	-0,01	5	1,00	0,90	0,50
P17	30	19	19,01	0,00	0,51	-0,01	5	1,00	0,50	0,40
P14	37	21	21	0,00	0,58	0,00	5	1,00	1,50	0,80
P17	23	18	17,99	0,00	0,50	0,01	5	1,00	0,10	0,10
P4	15	14	13,99	0,00	0,58	0,01	5	0,99	0,60	0,50
P13	30	19	18,98	0,01	0,51	0,01	5	0,99	0,50	0,40
P14	30	19	18,98	0,01	0,51	0,01	5	0,99	0,50	0,40
P16	7	20	19,98	0,01	0,53	0,01	5	0,99	1,30	1,40
P19	12	19	18,97	0,01	0,51	0,01	5	0,99	1,30	0,80
P13	23	18	17,96	0,01	0,50	0,02	5	0,98	0,10	0,10
P14	23	18	17,96	0,01	0,50	0,02	5	0,98	0,10	0,10
P10	36	22	21,97	0,01	0,69	0,02	5	0,98	0,90	0,50
P18	38	17	16,94	0,02	0,51	0,03	5	0,98	1,40	1,00
P5	37	21	20,95	0,02	0,58	0,03	5	0,98	1,50	0,80
P12	37	21	20,95	0,02	0,58	0,03	5	0,98	1,50	0,80
P11	8	22	21,96	0,02	0,69	0,03	5	0,98	1,30	4,80
P17	9	16	15,92	0,02	0,52	0,04	5	0,97	0,80	0,60
P9	1	13	12,95	0,02	0,63	0,03	5	0,97	0,80	0,70
P11	6	20	19,92	0,02	0,53	0,04	5	0,97	1,40	9,00
P12	30	19	18,91	0,02	0,51	0,04	5	0,97	0,50	0,40
P5	23	18	17,89	0,03	0,50	0,06	5	0,96	0,10	0,10
P12	23	18	17,89	0,03	0,50	0,06	5	0,96	0,10	0,10

P2	7	19	18,89	0,03	0,51	0,06	5	0,96	0,80	0,80
P13	9	16	15,89	0,03	0,52	0,06	5	0,96	0,80	0,60
P14	9	16	15,89	0,03	0,52	0,06	5	0,96	0,80	0,60
P4	33	17	16,88	0,03	0,51	0,06	5	0,95	0,50	0,40
P10	37	22	21,93	0,03	0,69	0,05	5	0,97	0,90	0,50
P16	26	19	18,88	0,03	0,51	0,06	5	0,95	2,20	6,30
P10	40	16	15,88	0,03	0,52	0,06	5	0,95	1,10	1,00
P17	16	15	14,88	0,04	0,54	0,07	5	0,95	0,50	0,40
P6	14	20	19,87	0,04	0,53	0,07	5	0,95	0,80	0,50
P6	26	17	16,84	0,04	0,51	0,08	5	0,94	0,40	0,80
P13	22	14	13,87	0,04	0,58	0,08	5	0,94	0,20	0,20
P13	16	15	14,85	0,04	0,54	0,08	5	0,94	0,50	0,40
P5	9	16	15,83	0,05	0,52	0,09	5	0,93	0,80	0,60
P12	9	16	15,83	0,05	0,52	0,09	5	0,93	0,80	0,60
P16	13	22	21,9	0,05	0,69	0,07	5	0,95	0,90	0,50
P2	38	17	16,81	0,05	0,51	0,10	5	0,93	1,40	1,00
P3	36	22	21,89	0,05	0,69	0,07	5	0,95	0,90	0,50
P3	4	19	18,8	0,05	0,51	0,10	5	0,92	1,50	0,90
P4	27	16	15,81	0,05	0,52	0,10	5	0,92	0,40	0,30
P3	27	18	17,79	0,05	0,50	0,11	5	0,92	0,10	0,10
P7	17	16	15,79	0,06	0,52	0,11	5	0,92	0,80	0,60
P16	29	20	19,79	0,06	0,53	0,11	5	0,92	0,40	1,80
P18	28	19	18,77	0,06	0,51	0,12	5	0,91	1,50	0,90
P5	22	14	13,82	0,06	0,58	0,10	5	0,92	0,20	0,20
P5	16	15	14,8	0,06	0,54	0,11	5	0,92	0,50	0,40
P15	33	17	16,75	0,06	0,51	0,13	5	0,90	0,50	0,40
P20	33	17	16,75	0,06	0,51	0,13	5	0,90	0,50	0,40
P3	37	22	21,86	0,07	0,69	0,10	5	0,93	0,90	0,50
P9	4	18	17,73	0,07	0,50	0,14	5	0,90	1,90	1,40
P8	38	16	15,74	0,07	0,52	0,14	5	0,90	0,60	0,40
P17	1	13	12,82	0,07	0,63	0,11	5	0,91	0,80	0,70
P4	35	20	19,74	0,07	0,53	0,14	5	0,90	0,40	0,30
P6	12	17	16,69	0,08	0,51	0,16	5	0,88	0,40	0,30
P13	1	13	12,8	0,08	0,63	0,13	5	0,90	0,80	0,70
P14	1	13	12,8	0,08	0,63	0,13	5	0,90	0,80	0,70
P2	29	19	18,69	0,08	0,51	0,16	5	0,88	0,80	0,50
P16	39	19	18,68	0,08	0,51	0,16	5	0,88	0,40	0,30
P10	13	22	21,82	0,08	0,69	0,12	5	0,91	0,90	0,50
P18	39	18	17,66	0,09	0,50	0,17	5	0,87	0,80	0,50
P1	7	19	18,66	0,09	0,51	0,17	5	0,87	1,50	0,90
P15	4	17	16,65	0,09	0,51	0,18	5	0,87	0,60	0,50
P20	4	17	16,65	0,09	0,51	0,18	5	0,87	0,60	0,50
P11	21	21	20,72	0,09	0,58	0,16	5	0,88	1,10	3,10
P6	6	19	18,63	0,10	0,51	0,19	5	0,86	0,30	0,20
P16	2	12	11,79	0,10	0,67	0,14	5	0,89	0,50	0,90
P7	38	16	15,65	0,10	0,52	0,18	5	0,86	0,60	0,40
P8	14	20	19,66	0,10	0,53	0,18	5	0,86	0,80	0,50
P5	1	13	12,76	0,10	0,63	0,15	5	0,88	0,80	0,70

P12	1	13	12,76	0,10	0,63	0,15	5	0,88	0,80	0,70
P9	38	17	16,62	0,10	0,51	0,20	5	0,85	1,40	1,00
P8	25	21	20,69	0,10	0,58	0,18	5	0,87	0,40	0,20
P17	24	19	18,6	0,10	0,51	0,20	5	0,85	1,00	0,70
P18	8	22	21,77	0,11	0,69	0,15	5	0,88	1,30	0,70
P10	28	20	19,62	0,11	0,53	0,20	5	0,85	1,60	0,90
P3	1	14	13,68	0,11	0,58	0,19	5	0,86	0,40	0,40
P17	40	15	14,63	0,11	0,54	0,20	5	0,85	0,40	0,30
P13	24	19	18,57	0,11	0,51	0,22	5	0,84	1,00	0,70
P14	24	19	18,57	0,11	0,51	0,22	5	0,84	1,00	0,70
P13	40	15	14,61	0,12	0,54	0,22	5	0,84	0,40	0,30
P14	40	15	14,61	0,12	0,54	0,22	5	0,84	0,40	0,30
P2	39	18	17,52	0,12	0,50	0,24	5	0,82	0,80	0,50
P7	14	20	19,57	0,12	0,53	0,23	5	0,83	0,80	0,50
P19	16	17	16,53	0,12	0,51	0,24	5	0,82	1,20	0,80
P6	2	11	10,78	0,12	0,73	0,17	5	0,87	0,20	0,10
P18	23	19	18,52	0,12	0,51	0,24	5	0,82	1,50	0,90
P17	15	15	14,58	0,13	0,54	0,23	5	0,83	1,10	1,10
P7	25	21	20,61	0,13	0,58	0,22	5	0,84	0,40	0,20
P1	26	18	17,49	0,13	0,50	0,25	5	0,81	0,60	1,20
P5	24	19	18,5	0,13	0,51	0,25	5	0,81	1,00	0,70
P12	24	19	18,5	0,13	0,51	0,25	5	0,81	1,00	0,70
P9	29	19	18,49	0,13	0,51	0,26	5	0,81	0,80	0,50
P13	15	15	14,56	0,13	0,54	0,25	5	0,81	1,10	1,10
P14	15	15	14,56	0,13	0,54	0,25	5	0,81	1,10	1,10
P5	40	15	14,55	0,14	0,54	0,25	5	0,81	0,40	0,30
P12	40	15	14,55	0,14	0,54	0,25	5	0,81	0,40	0,30
P16	23	20	19,51	0,14	0,53	0,25	5	0,81	1,00	0,60
P15	14	20	19,5	0,14	0,53	0,26	5	0,81	0,80	0,50
P20	14	20	19,5	0,14	0,53	0,26	5	0,81	0,80	0,50
P14	7	19	18,47	0,14	0,51	0,27	5	0,80	1,20	1,00
P8	12	17	16,47	0,14	0,51	0,27	5	0,80	0,70	1,00
P3	12	19	18,45	0,14	0,51	0,28	5	0,79	1,30	0,80
P9	35	21	20,57	0,14	0,58	0,24	5	0,82	0,40	0,20
P8	11	20	19,49	0,14	0,53	0,27	5	0,80	0,80	0,50
P6	9	16	15,47	0,15	0,52	0,28	5	0,79	0,80	0,60
P18	9	17	16,43	0,15	0,51	0,29	5	0,78	0,80	0,50
P17	38	17	16,42	0,15	0,51	0,30	5	0,78	1,20	0,80
P12	15	15	14,5	0,15	0,54	0,28	5	0,79	1,10	1,10
P8	6	19	18,4	0,15	0,51	0,30	5	0,78	1,50	0,90
P12	7	19	18,4	0,15	0,51	0,30	5	0,77	1,20	1,00
P11	27	18	17,38	0,16	0,50	0,31	5	0,77	1,90	1,40
P14	38	17	16,39	0,16	0,51	0,31	5	0,77	1,20	0,80
P2	23	19	18,39	0,16	0,51	0,31	5	0,77	1,50	0,90
P4	39	17	16,39	0,16	0,51	0,31	5	0,77	1,00	0,60
P4	6	19	18,37	0,16	0,51	0,32	5	0,76	1,20	1,00
P17	31	17	16,37	0,16	0,51	0,32	5	0,76	1,50	1,50
P19	14	22	21,63	0,16	0,69	0,24	5	0,82	0,90	0,50

P4	36	21	20,49	0,17	0,58	0,28	5	0,79	0,40	0,20
P7	11	20	19,4	0,17	0,53	0,31	5	0,77	0,80	0,50
P11	37	22	21,61	0,17	0,69	0,25	5	0,81	0,70	4,10
P9	39	18	17,32	0,17	0,50	0,34	5	0,75	0,80	0,50
P13	31	17	16,34	0,17	0,51	0,34	5	0,75	1,50	1,50
P12	38	17	16,33	0,17	0,51	0,34	5	0,74	1,20	0,80
P8	37	21	20,46	0,18	0,58	0,30	5	0,78	0,40	0,20
P7	6	19	18,31	0,18	0,51	0,35	5	0,74	1,50	0,90
P1	39	18	17,29	0,18	0,50	0,36	5	0,74	0,80	0,50
P2	9	17	16,3	0,18	0,51	0,36	5	0,73	0,80	0,50
P15	11	20	19,33	0,18	0,53	0,34	5	0,75	0,50	2,00
P20	11	20	19,33	0,18	0,53	0,34	5	0,75	0,50	2,00
P4	37	21	20,43	0,18	0,58	0,31	5	0,77	0,40	0,20
P16	30	21	20,42	0,18	0,58	0,32	5	0,76	0,10	0,10
P8	21	20	19,32	0,19	0,53	0,35	5	0,74	0,40	0,30
P5	31	17	16,28	0,19	0,51	0,37	5	0,73	1,50	1,50
P19	4	20	19,31	0,19	0,53	0,35	5	0,74	1,00	0,60
P4	2	11	10,66	0,19	0,73	0,26	5	0,81	0,60	1,60
P14	29	19	18,26	0,19	0,51	0,37	5	0,73	0,50	0,40
P9	8	22	21,56	0,19	0,69	0,28	5	0,79	1,30	0,70
P6	1	13	12,52	0,20	0,63	0,31	5	0,77	0,80	0,70
P1	8	22	21,54	0,20	0,69	0,29	5	0,79	1,30	0,70
P1	6	20	19,27	0,20	0,53	0,37	5	0,72	1,00	0,60
P14	35	21	20,37	0,20	0,58	0,34	5	0,75	0,80	0,70
P7	37	21	20,37	0,20	0,58	0,34	5	0,75	0,40	0,20
P12	29	19	18,2	0,21	0,51	0,40	5	0,70	0,50	0,40
P3	23	20	19,25	0,21	0,53	0,39	5	0,71	1,20	0,80
P17	12	18	17,18	0,21	0,50	0,41	5	0,70	0,60	0,50
P9	23	19	18,19	0,21	0,51	0,41	5	0,70	1,50	0,90
P7	21	20	19,23	0,21	0,53	0,40	5	0,71	0,40	0,30
P3	9	18	17,16	0,21	0,50	0,42	5	0,69	1,90	1,30
P18	33	19	18,16	0,21	0,51	0,42	5	0,69	1,70	1,10
P12	35	21	20,32	0,22	0,58	0,37	5	0,73	0,80	0,70
P17	39	18	17,13	0,22	0,50	0,44	5	0,68	0,10	0,10
P17	14	21	20,3	0,22	0,58	0,38	5	0,72	1,50	0,80
P1	11	21	20,29	0,23	0,58	0,39	5	0,71	1,00	2,90
P16	14	22	21,47	0,23	0,69	0,33	5	0,76	0,90	0,50
P4	16	15	14,26	0,23	0,54	0,42	5	0,69	1,00	0,60
P15	23	18	17,1	0,23	0,50	0,45	5	0,67	0,10	0,10
P20	23	18	17,1	0,23	0,50	0,45	5	0,67	0,10	0,10
P14	39	18	17,09	0,23	0,50	0,46	5	0,67	0,70	0,40
P13	14	21	20,27	0,23	0,58	0,40	5	0,71	1,50	0,80
P9	9	17	16,11	0,23	0,51	0,46	5	0,67	0,80	0,50
P6	40	15	14,24	0,23	0,54	0,43	5	0,68	0,40	0,30
P11	28	20	19,1	0,25	0,53	0,46	5	0,66	1,00	0,60
P2	22	15	14,2	0,25	0,54	0,46	5	0,67	3,10	3,30
P2	33	19	18,03	0,25	0,51	0,48	5	0,65	1,70	1,10
P6	27	17	16,05	0,25	0,51	0,49	5	0,65	0,30	0,20

P14	6	20	19,08	0,25	0,53	0,47	5	0,66	0,60	0,40
P16	4	20	19,07	0,25	0,53	0,47	5	0,66	1,60	0,90
P3	30	21	20,19	0,25	0,58	0,44	5	0,68	1,00	0,50
P10	33	20	19,04	0,26	0,53	0,49	5	0,65	1,30	0,80
P10	14	22	21,38	0,26	0,69	0,38	5	0,72	0,90	0,50
P4	1	13	12,36	0,26	0,63	0,42	5	0,69	0,80	0,70
P16	38	19	17,96	0,26	0,51	0,52	5	0,63	1,40	0,90
P12	6	20	19,02	0,27	0,53	0,50	5	0,64	0,60	0,40
P17	11	21	20,14	0,27	0,58	0,46	5	0,67	1,00	2,90
P10	22	16	15,03	0,27	0,52	0,52	5	0,62	3,00	2,60
P2	4	19	17,93	0,27	0,51	0,54	5	0,62	1,50	0,90
P10	27	19	17,93	0,27	0,51	0,54	5	0,61	1,50	0,90
P8	33	18	16,91	0,28	0,50	0,55	5	0,61	0,30	0,20
P13	11	21	20,11	0,28	0,58	0,47	5	0,66	0,40	0,20
P14	11	21	20,11	0,28	0,58	0,47	5	0,66	0,40	0,20
P18	31	18	16,89	0,28	0,50	0,56	5	0,60	1,60	1,10
P8	24	19	17,88	0,29	0,51	0,56	5	0,60	0,20	0,10
P10	4	20	18,94	0,29	0,53	0,54	5	0,61	1,60	0,90
P6	31	17	15,91	0,29	0,51	0,57	5	0,59	0,10	0,10
P2	40	16	14,97	0,29	0,52	0,56	5	0,60	1,10	1,00
P16	24	21	20,07	0,29	0,58	0,50	5	0,64	0,60	0,50
P18	26	19	17,86	0,29	0,51	0,57	5	0,59	1,50	2,00
P5	11	21	20,06	0,29	0,58	0,50	5	0,64	0,40	0,20
P12	11	21	20,06	0,29	0,58	0,50	5	0,64	0,40	0,20
P8	40	15	14,06	0,29	0,54	0,54	5	0,61	0,20	0,20
P3	17	19	17,84	0,30	0,51	0,58	5	0,59	2,30	1,70
P3	14	22	21,28	0,30	0,69	0,43	5	0,68	0,90	0,50
P9	22	15	14,04	0,30	0,54	0,55	5	0,61	3,10	3,30
P9	33	19	17,83	0,30	0,51	0,58	5	0,58	1,70	1,10
P7	33	18	16,82	0,30	0,50	0,60	5	0,58	0,30	0,20
P8	4	18	16,81	0,30	0,50	0,60	5	0,58	1,70	1,90
P1	33	19	17,8	0,31	0,51	0,60	5	0,57	1,70	1,10
P7	24	19	17,78	0,31	0,51	0,61	5	0,57	0,20	0,10
P7	40	15	13,98	0,32	0,54	0,59	5	0,58	0,20	0,20
P4	7	19	17,74	0,32	0,51	0,63	5	0,56	0,20	0,10
P1	17	18	16,73	0,32	0,50	0,64	5	0,55	0,80	0,70
P14	16	16	14,85	0,32	0,52	0,63	5	0,56	0,90	0,60
P2	26	19	17,73	0,32	0,51	0,64	5	0,55	1,50	2,00
P7	4	18	16,72	0,33	0,50	0,65	5	0,55	1,70	1,90
P1	36	22	21,19	0,33	0,69	0,48	5	0,65	0,90	0,50
P17	10	16	14,83	0,33	0,52	0,64	5	0,55	0,50	0,40
P1	4	19	17,7	0,33	0,51	0,65	5	0,54	1,50	0,90
P7	15	15	13,94	0,33	0,54	0,61	5	0,57	1,10	1,10
P15	40	15	13,93	0,33	0,54	0,62	5	0,56	0,40	0,30
P20	40	15	13,93	0,33	0,54	0,62	5	0,56	0,40	0,30
P19	9	19	17,68	0,34	0,51	0,66	5	0,54	1,70	1,10
P3	32	13	12,19	0,34	0,63	0,54	5	0,61	0,30	0,20
P10	19	22	21,17	0,34	0,69	0,49	5	0,65	1,00	1,00

P13	10	16	14,8	0,34	0,52	0,65	5	0,54	0,50	0,40
P14	10	16	14,8	0,34	0,52	0,65	5	0,54	0,50	0,40
P12	16	16	14,8	0,34	0,52	0,66	5	0,54	0,90	0,60
P3	11	22	21,16	0,34	0,69	0,50	5	0,64	0,90	0,50
P8	31	17	15,69	0,35	0,51	0,68	5	0,53	0,80	0,70
P6	39	18	16,64	0,35	0,50	0,69	5	0,52	0,80	0,50
P1	37	22	21,14	0,35	0,69	0,50	5	0,64	0,90	0,50
P10	21	22	21,13	0,35	0,69	0,51	5	0,63	1,30	0,70
P8	26	18	16,62	0,35	0,50	0,70	5	0,51	0,30	1,00
P5	10	16	14,75	0,36	0,52	0,69	5	0,52	0,50	0,40
P12	10	16	14,75	0,36	0,52	0,69	5	0,52	0,50	0,40
P2	28	20	18,64	0,36	0,53	0,68	5	0,53	1,60	0,90
P11	35	22	21,08	0,37	0,69	0,54	5	0,61	0,80	0,80
P7	31	17	15,6	0,37	0,51	0,73	5	0,50	0,80	0,70
P17	36	22	21,07	0,37	0,69	0,54	5	0,61	0,90	0,50
P3	19	22	21,07	0,37	0,69	0,54	5	0,61	1,00	1,00
P15	38	17	15,59	0,37	0,51	0,74	5	0,49	1,20	0,80
P20	38	17	15,59	0,37	0,51	0,74	5	0,49	1,20	0,80
P9	26	19	17,53	0,37	0,51	0,73	5	0,50	1,50	2,00
P8	28	19	17,52	0,38	0,51	0,74	5	0,49	0,40	0,30
P13	36	22	21,05	0,38	0,69	0,55	5	0,61	0,90	0,50
P14	36	22	21,05	0,38	0,69	0,55	5	0,61	0,90	0,50
P16	28	21	19,74	0,38	0,58	0,65	5	0,54	1,00	0,50
P19	23	21	19,74	0,38	0,58	0,66	5	0,54	0,70	0,60
P17	37	22	21,02	0,39	0,69	0,57	5	0,60	0,90	0,50
P5	36	22	21	0,40	0,69	0,58	5	0,59	0,90	0,50
P12	36	22	21	0,40	0,69	0,58	5	0,59	0,90	0,50
P13	37	22	21	0,40	0,69	0,58	5	0,59	0,90	0,50
P11	17	19	17,43	0,40	0,51	0,78	5	0,47	2,80	2,60
P1	13	22	20,99	0,40	0,69	0,58	5	0,59	0,90	0,50
P7	28	19	17,42	0,40	0,51	0,79	5	0,47	0,40	0,30
P13	38	18	16,39	0,41	0,50	0,82	5	0,45	0,60	0,50
P9	28	20	18,44	0,41	0,53	0,78	5	0,47	1,60	0,90
P6	19	21	19,59	0,42	0,58	0,72	5	0,50	0,70	0,60
P8	8	22	20,92	0,42	0,69	0,61	5	0,57	0,20	0,10
P1	28	20	18,41	0,42	0,53	0,79	5	0,46	1,60	0,90
P5	38	18	16,33	0,43	0,50	0,85	5	0,43	0,60	0,50
P4	32	12	11,12	0,43	0,67	0,63	5	0,55	0,50	0,90
P5	15	16	14,5	0,43	0,52	0,83	5	0,44	0,70	0,80
P15	12	18	16,31	0,43	0,50	0,86	5	0,43	0,60	0,50
P20	12	18	16,31	0,43	0,50	0,86	5	0,43	0,60	0,50
P7	32	12	11,09	0,44	0,67	0,66	5	0,54	0,40	0,40
P15	39	18	16,26	0,44	0,50	0,89	5	0,42	0,80	0,50
P20	39	18	16,26	0,44	0,50	0,89	5	0,42	0,80	0,50
P7	8	22	20,85	0,45	0,69	0,65	5	0,54	0,20	0,10
P11	15	17	15,32	0,45	0,51	0,89	5	0,42	1,20	0,80
P13	13	22	20,84	0,45	0,69	0,65	5	0,54	0,40	0,30
P14	13	22	20,84	0,45	0,69	0,65	5	0,54	0,60	0,80

P4	30	20	18,26	0,46	0,53	0,86	5	0,43	0,10	0,10
P5	13	22	20,79	0,47	0,69	0,68	5	0,53	0,40	0,30
P12	13	22	20,79	0,47	0,69	0,68	5	0,53	0,60	0,80
P13	12	19	17,14	0,47	0,51	0,92	5	0,40	0,40	0,30
P14	12	19	17,14	0,47	0,51	0,92	5	0,40	0,40	0,30
P6	22	15	13,54	0,47	0,54	0,87	5	0,42	1,00	0,60
P4	9	17	15,24	0,47	0,51	0,94	5	0,39	0,80	0,50
P11	19	22	20,76	0,48	0,69	0,69	5	0,52	0,90	9,00
P8	19	21	19,38	0,48	0,58	0,82	5	0,45	0,70	0,60
P18	17	19	17,1	0,48	0,51	0,95	5	0,39	1,20	0,80
P5	12	19	17,08	0,49	0,51	0,96	5	0,38	0,40	0,30
P12	12	19	17,08	0,49	0,51	0,96	5	0,38	0,40	0,30
P17	2	12	11	0,49	0,67	0,72	5	0,50	0,50	0,90
P16	8	23	22,3	0,49	0,97	0,51	5	0,63	1,00	1,60
P13	2	12	10,99	0,50	0,67	0,74	5	0,49	0,50	0,90
P14	2	12	10,99	0,50	0,67	0,74	5	0,49	0,50	0,90
P18	27	19	17,04	0,50	0,51	0,97	5	0,37	1,50	0,90
P7	19	21	19,28	0,51	0,58	0,87	5	0,43	0,70	0,60
P5	2	12	10,96	0,51	0,67	0,76	5	0,48	0,50	0,90
P12	2	12	10,96	0,51	0,67	0,76	5	0,48	0,50	0,90
P2	17	19	16,96	0,52	0,51	1,01	5	0,36	2,80	2,60
P4	10	16	14,22	0,52	0,52	1,01	5	0,36	0,50	0,40
P15	19	21	19,22	0,52	0,58	0,89	5	0,41	0,70	0,60
P20	19	21	19,22	0,52	0,58	0,89	5	0,41	0,70	0,60
P11	39	20	17,99	0,53	0,53	0,99	5	0,37	1,20	0,70
P10	8	23	22,24	0,53	0,97	0,55	5	0,61	1,00	1,60
P2	27	19	16,91	0,53	0,51	1,04	5	0,35	1,50	0,90
P15	16	16	14,16	0,54	0,52	1,04	5	0,35	0,90	0,60
P20	16	16	14,16	0,54	0,52	1,04	5	0,35	0,90	0,60
P11	29	21	19,15	0,54	0,58	0,93	5	0,40	0,90	2,80
P1	35	22	20,54	0,55	0,69	0,79	5	0,46	0,90	0,50
P19	26	21	19,12	0,55	0,58	0,94	5	0,39	1,80	3,50
P4	17	18	15,85	0,55	0,50	1,10	5	0,32	0,90	0,70
P16	6	22	20,51	0,55	0,69	0,80	5	0,46	0,40	0,30
P19	32	14	12,5	0,56	0,58	0,98	5	0,37	1,70	2,10
P9	17	19	16,77	0,57	0,51	1,11	5	0,32	2,80	2,60
P16	25	23	22,16	0,57	0,97	0,59	5	0,58	0,60	0,30
P18	21	22	20,44	0,58	0,69	0,84	5	0,44	1,30	0,70
P1	14	22	20,44	0,58	0,69	0,84	5	0,44	0,90	0,50
P9	27	19	16,72	0,58	0,51	1,14	5	0,31	1,50	0,90
P8	7	20	17,78	0,58	0,53	1,09	5	0,32	1,00	0,60
P17	35	22	20,4	0,59	0,69	0,85	5	0,43	0,90	0,50
P10	6	22	20,4	0,59	0,69	0,86	5	0,43	1,80	1,40
P13	35	22	20,37	0,60	0,69	0,87	5	0,43	0,90	0,50
P10	25	23	22,09	0,61	0,97	0,63	5	0,56	0,50	0,20
P7	7	20	17,68	0,61	0,53	1,14	5	0,31	1,00	0,60
P2	21	22	20,33	0,61	0,69	0,88	5	0,42	1,30	0,70
P11	16	18	15,64	0,61	0,50	1,22	5	0,28	1,90	1,30

P5	35	22	20,32	0,61	0,69	0,89	5	0,41	0,90	0,50
P10	30	22	20,31	0,62	0,69	0,90	5	0,41	0,60	0,40
P19	24	22	20,28	0,62	0,69	0,91	5	0,41	1,00	1,00
P1	15	17	14,72	0,63	0,51	1,24	5	0,27	1,90	1,70
P11	10	18	15,59	0,63	0,50	1,25	5	0,27	0,40	0,30
P3	25	23	22,01	0,64	0,97	0,66	5	0,54	0,50	0,20
P10	26	21	18,74	0,65	0,58	1,11	5	0,32	1,80	3,50
P15	1	14	12,29	0,66	0,58	1,14	5	0,31	0,50	0,40
P20	1	14	12,29	0,66	0,58	1,14	5	0,31	0,50	0,40
P9	21	22	20,16	0,66	0,69	0,96	5	0,38	1,30	0,70
P15	13	22	20,14	0,67	0,69	0,97	5	0,38	0,20	0,10
P20	13	22	20,14	0,67	0,69	0,97	5	0,38	0,20	0,10
P6	16	17	14,47	0,70	0,51	1,39	5	0,22	1,40	1,00
P3	24	22	19,82	0,76	0,69	1,10	5	0,32	1,00	1,00
P6	32	13	11,25	0,78	0,63	1,25	5	0,27	1,40	0,90
P10	29	22	19,66	0,80	0,69	1,16	5	0,30	0,60	0,40
P11	32	14	11,96	0,80	0,58	1,39	5	0,22	2,00	1,40
P18	25	23	21,58	0,83	0,97	0,86	5	0,43	0,60	0,30
P3	29	22	19,53	0,84	0,69	1,21	5	0,28	0,60	0,40
P10	17	21	17,98	0,84	0,58	1,44	5	0,21	3,40	4,70
P2	25	23	21,5	0,86	0,97	0,89	5	0,41	0,60	0,30
P18	32	14	11,78	0,89	0,58	1,53	5	0,19	0,40	0,30
P11	7	22	19,34	0,89	0,69	1,29	5	0,25	1,00	1,00
P9	25	23	21,37	0,91	0,97	0,95	5	0,39	0,60	0,30
P18	22	18	14,3	1,01	0,50	2,01	5	0,10	2,30	1,60
P17	13	23	20,87	1,09	0,97	1,13	5	0,31	0,50	0,20
P15	2	13	10,6	1,14	0,63	1,82	5	0,13	0,70	0,40
P20	2	13	10,6	1,14	0,63	1,82	5	0,13	0,70	0,40
P19	30	23	20,63	1,17	0,97	1,21	5	0,28	0,90	0,70
P11	22	19	14,57	1,18	0,51	2,31	5	0,07	2,00	1,30
P16	31	22	17,91	1,25	0,69	1,82	5	0,13	1,80	1,40
P19	29	23	20,01	1,35	0,97	1,40	5	0,22	0,90	0,70

EK 10. Son Test Matematik Puanlayıcı Grubu *Puanlayıcı X Birey* Etkileşimleri

Puanlayıcı	Birey	Gözlenen Puan	Beklenen Puan	Yanlılık Büyüklüğü	SH	t	sd	p	infit	outfit
P1	25	24,00	21,83	1,38	1,38	1,00	5	0,36	0,20	0,10
P3	8	24,00	21,87	1,36	1,38	0,99	5	0,37	0,20	0,10
P3	21	24,00	21,26	1,62	1,38	1,17	5	0,29	0,20	0,10
P14	8	24,00	21,53	1,51	1,38	1,10	5	0,32	0,20	0,10
P16	8	24,00	21,51	1,52	1,38	1,10	5	0,32	0,20	0,10
P19	8	24,00	22,22	1,19	1,38	0,87	5	0,43	0,20	0,10
P19	21	24,00	21,68	1,44	1,38	1,05	5	0,34	0,20	0,10
P19	25	24,00	22,57	0,98	1,38	0,71	5	0,51	0,20	0,10
P9	32	9,00	12,18	-2,84	1,52	-1,86	5	0,12	0,10	0,10
P8	2	9,00	10,63	-2,01	1,52	-1,32	5	0,24	0,10	0,10
P18	16	12,00	16,35	-1,63	0,69	-2,35	5	0,07	0,70	0,30
P9	16	12,00	16,26	-1,60	0,69	-2,31	5	0,07	0,70	0,30
P10	32	10,00	12,54	-1,46	0,95	-1,55	5	0,18	0,10	0,10
P1	32	10,00	12,40	-1,40	0,95	-1,48	5	0,20	0,10	0,10
P8	16	12,00	15,59	-1,39	0,69	-2,01	5	0,10	4,60	2,20
P18	32	10,00	12,24	-1,33	0,95	-1,40	5	0,22	0,10	0,10
P11	19	16,00	20,39	-1,29	0,56	-2,32	5	0,07	4,30	5,20
P1	22	12,00	15,15	-1,25	0,69	-1,80	5	0,13	0,70	0,30
P11	26	14,00	17,56	-1,13	0,61	-1,86	5	0,12	0,70	1,00
P11	27	14,00	17,51	-1,12	0,61	-1,84	5	0,13	0,70	1,00
P10	31	14,00	17,48	-1,11	0,61	-1,83	5	0,13	0,40	0,50
P2	17	13,00	16,14	-1,11	0,65	-1,71	5	0,15	0,80	0,50
P11	30	15,00	18,57	-1,07	0,58	-1,85	5	0,12	2,30	1,80
P17	22	12,00	14,61	-1,07	0,69	-1,54	5	0,18	0,70	0,30
P12	17	13,00	15,64	-0,96	0,65	-1,47	5	0,20	0,80	0,50
P8	22	12,00	14,26	-0,94	0,69	-1,36	5	0,23	0,70	0,30
P3	32	11,00	12,92	-0,94	0,76	-1,24	5	0,27	0,30	0,10
P5	25	18,00	21,04	-0,92	0,53	-1,73	5	0,14	0,70	0,50
P11	4	15,00	17,86	-0,87	0,58	-1,50	5	0,19	3,90	3,40
P7	8	18,00	20,83	-0,85	0,53	-1,59	5	0,17	1,90	1,40
P8	29	15,00	17,76	-0,84	0,58	-1,45	5	0,21	1,50	1,80
P10	2	10,00	11,26	-0,84	0,95	-0,89	5	0,42	0,10	0,10
P2	8	18,00	20,78	-0,83	0,53	-1,56	5	0,18	1,90	1,40
P18	10	13,00	15,24	-0,83	0,65	-1,27	5	0,26	0,80	0,50
P4	25	19,00	21,43	-0,78	0,54	-1,46	5	0,20	0,80	0,50
P3	26	16,00	18,68	-0,78	0,56	-1,41	5	0,22	1,40	2,70
P7	23	15,00	17,55	-0,78	0,58	-1,35	5	0,24	0,80	0,50
P1	2	10,00	11,15	-0,77	0,95	-0,82	5	0,45	0,10	0,10
P3	40	14,00	16,25	-0,75	0,61	-1,23	5	0,27	0,40	0,30
P15	4	15,00	17,39	-0,73	0,58	-1,27	5	0,26	0,70	0,70
P20	4	15,00	17,39	-0,73	0,58	-1,27	5	0,26	0,70	0,70
P7	17	14,00	16,20	-0,73	0,61	-1,20	5	0,28	1,60	1,00
P17	17	14,00	16,20	-0,73	0,61	-1,20	5	0,28	1,10	0,90
P3	10	14,00	16,19	-0,73	0,61	-1,20	5	0,28	1,60	1,00

P19	39	16,00	18,51	-0,73	0,56	-1,32	5	0,25	1,60	1,00
P19	33	17,00	19,53	-0,72	0,54	-1,34	5	0,24	0,40	0,30
P5	8	18,00	20,42	-0,71	0,53	-1,34	5	0,24	0,30	0,20
P18	2	10,00	11,02	-0,70	0,95	-0,74	5	0,49	0,10	0,10
P18	6	17,00	19,39	-0,68	0,54	-1,27	5	0,26	0,60	0,40
P12	8	18,00	20,31	-0,68	0,53	-1,27	5	0,26	2,70	1,80
P9	2	10,00	10,97	-0,68	0,95	-0,71	5	0,51	0,10	0,10
P7	24	16,00	18,20	-0,64	0,56	-1,16	5	0,30	0,10	0,10
P19	17	16,00	18,19	-0,64	0,56	-1,15	5	0,30	0,70	0,50
P1	23	16,00	18,18	-0,64	0,56	-1,15	5	0,30	0,40	0,30
P3	22	14,00	15,86	-0,63	0,61	-1,04	5	0,35	1,30	1,20
P2	24	16,00	18,14	-0,63	0,56	-1,13	5	0,31	0,10	0,10
P4	22	13,00	14,64	-0,62	0,65	-0,96	5	0,38	1,00	1,20
P11	7	17,00	19,13	-0,61	0,54	-1,13	5	0,31	1,40	1,50
P4	2	10,00	10,85	-0,60	0,95	-0,64	5	0,55	0,10	0,10
P14	38	14,00	15,77	-0,60	0,61	-0,99	5	0,37	0,40	0,30
P14	40	14,00	15,77	-0,60	0,61	-0,99	5	0,37	0,40	0,30
P17	15	13,00	14,56	-0,60	0,65	-0,92	5	0,40	0,40	1,00
P1	28	17,00	19,09	-0,60	0,54	-1,11	5	0,32	0,50	0,40
P7	14	18,00	20,05	-0,59	0,53	-1,11	5	0,32	1,80	1,50
P14	10	14,00	15,72	-0,58	0,61	-0,96	5	0,38	1,60	1,00
P14	12	16,00	17,96	-0,58	0,56	-1,04	5	0,35	0,40	0,40
P16	16	15,00	16,84	-0,57	0,58	-0,99	5	0,37	0,90	0,50
P10	10	14,00	15,66	-0,57	0,61	-0,93	5	0,40	1,60	1,00
P10	16	15,00	16,81	-0,57	0,58	-0,98	5	0,37	0,90	0,50
P10	12	16,00	17,90	-0,56	0,56	-1,00	5	0,36	0,40	0,40
P5	26	15,00	16,78	-0,56	0,58	-0,96	5	0,38	0,30	0,40
P3	33	17,00	18,94	-0,55	0,54	-1,02	5	0,35	0,40	0,30
P11	36	18,00	19,88	-0,54	0,53	-1,02	5	0,35	2,60	3,10
P3	28	18,00	19,87	-0,54	0,53	-1,01	5	0,36	0,30	0,20
P8	1	11,00	12,04	-0,54	0,76	-0,71	5	0,51	0,30	0,10
P10	6	18,00	19,85	-0,53	0,53	-1,00	5	0,36	0,60	0,50
P13	37	16,00	17,74	-0,51	0,56	-0,92	5	0,40	0,70	0,80
P1	10	14,00	15,46	-0,50	0,61	-0,82	5	0,45	1,60	1,00
P1	21	19,00	20,61	-0,49	0,54	-0,92	5	0,40	0,90	0,50
P16	7	18,00	19,69	-0,48	0,53	-0,91	5	0,40	0,50	0,40
P11	37	18,00	19,68	-0,48	0,53	-0,91	5	0,41	1,10	1,70
P19	2	11,00	11,92	-0,48	0,76	-0,64	5	0,55	0,40	0,30
P19	31	17,00	18,67	-0,48	0,54	-0,88	5	0,42	1,20	0,90
P5	19	18,00	19,65	-0,47	0,53	-0,89	5	0,41	1,90	1,40
P12	24	16,00	17,60	-0,47	0,56	-0,85	5	0,43	0,10	0,10
P4	23	16,00	17,58	-0,47	0,56	-0,84	5	0,44	0,40	0,30
P5	21	18,00	19,60	-0,46	0,53	-0,86	5	0,43	0,30	0,20
P7	4	16,00	17,55	-0,46	0,56	-0,82	5	0,45	0,40	0,40
P19	22	15,00	16,42	-0,45	0,58	-0,78	5	0,47	2,30	1,80
P18	40	14,00	15,29	-0,45	0,61	-0,73	5	0,50	0,40	0,30
P2	23	16,00	17,48	-0,44	0,56	-0,79	5	0,47	0,40	0,30
P12	19	18,00	19,53	-0,44	0,53	-0,82	5	0,45	1,50	1,90

P11	29	17,00	18,52	-0,43	0,54	-0,80	5	0,46	1,10	0,70
P19	15	15,00	16,36	-0,43	0,58	-0,75	5	0,49	1,40	2,00
P13	19	17,00	18,50	-0,43	0,54	-0,80	5	0,46	2,20	2,20
P19	28	19,00	20,42	-0,43	0,54	-0,79	5	0,46	0,50	0,30
P16	36	19,00	20,41	-0,42	0,54	-0,79	5	0,47	0,90	0,50
P12	14	18,00	19,48	-0,42	0,53	-0,79	5	0,46	0,90	0,50
P9	40	14,00	15,20	-0,42	0,61	-0,69	5	0,52	0,40	0,30
P3	6	19,00	20,38	-0,42	0,54	-0,78	5	0,47	1,20	0,80
P15	23	16,00	17,39	-0,41	0,56	-0,74	5	0,49	0,40	0,30
P20	23	16,00	17,39	-0,41	0,56	-0,74	5	0,49	0,40	0,30
P14	1	12,00	12,90	-0,41	0,69	-0,59	5	0,58	0,70	0,30
P9	10	14,00	15,15	-0,40	0,61	-0,66	5	0,54	1,60	1,00
P16	9	15,00	16,26	-0,40	0,58	-0,70	5	0,52	0,90	0,50
P4	35	18,00	19,41	-0,40	0,53	-0,75	5	0,48	1,00	0,70
P11	14	19,00	20,34	-0,40	0,54	-0,75	5	0,49	1,20	0,80
P11	21	19,00	20,34	-0,40	0,54	-0,75	5	0,49	0,50	0,30
P8	32	11,00	11,75	-0,40	0,76	-0,53	5	0,62	0,60	1,00
P16	33	17,00	18,39	-0,40	0,54	-0,73	5	0,50	0,30	0,30
P6	23	16,00	17,33	-0,39	0,56	-0,71	5	0,51	0,40	0,30
P10	9	15,00	16,23	-0,39	0,58	-0,68	5	0,53	0,90	0,50
P7	13	19,00	20,31	-0,39	0,54	-0,73	5	0,50	0,90	0,50
P4	17	15,00	16,23	-0,39	0,58	-0,68	5	0,53	1,70	1,00
P5	32	11,00	11,74	-0,39	0,76	-0,52	5	0,63	0,40	0,30
P10	1	12,00	12,86	-0,39	0,69	-0,56	5	0,60	0,70	0,30
P5	31	15,00	16,19	-0,38	0,58	-0,66	5	0,54	0,40	0,40
P4	26	16,00	17,28	-0,38	0,56	-0,68	5	0,52	0,60	0,70
P11	6	18,00	19,33	-0,38	0,53	-0,71	5	0,51	0,40	0,30
P11	11	18,00	19,33	-0,38	0,53	-0,71	5	0,51	0,20	0,20
P1	31	16,00	17,26	-0,38	0,56	-0,68	5	0,53	1,30	1,20
P8	9	14,00	15,07	-0,37	0,61	-0,61	5	0,57	0,40	0,30
P15	28	17,00	18,30	-0,37	0,54	-0,69	5	0,52	0,20	0,20
P20	28	17,00	18,30	-0,37	0,54	-0,69	5	0,52	0,20	0,20
P9	6	18,00	19,30	-0,37	0,53	-0,70	5	0,52	0,50	0,40
P16	37	19,00	20,22	-0,36	0,54	-0,68	5	0,53	0,90	0,50
P14	13	20,00	21,08	-0,36	0,55	-0,65	5	0,55	0,40	0,30
P11	8	20,00	21,08	-0,36	0,55	-0,65	5	0,55	0,30	0,40
P3	7	19,00	20,20	-0,36	0,54	-0,67	5	0,53	1,20	0,80
P4	24	17,00	18,23	-0,35	0,54	-0,65	5	0,54	0,20	0,20
P4	29	17,00	18,23	-0,35	0,54	-0,65	5	0,54	0,50	0,40
P12	32	11,00	11,66	-0,35	0,76	-0,47	5	0,66	0,70	2,20
P6	28	17,00	18,23	-0,35	0,54	-0,65	5	0,54	0,20	0,20
P19	27	18,00	19,23	-0,35	0,53	-0,66	5	0,54	0,10	0,10
P15	37	18,00	19,22	-0,35	0,53	-0,65	5	0,54	1,10	1,70
P20	37	18,00	19,22	-0,35	0,53	-0,65	5	0,54	1,10	1,70
P12	31	15,00	16,08	-0,34	0,58	-0,60	5	0,58	0,90	0,50
P18	7	18,00	19,19	-0,34	0,53	-0,64	5	0,55	0,50	0,40
P15	17	15,00	16,05	-0,34	0,58	-0,58	5	0,59	0,20	0,20
P20	17	15,00	16,05	-0,34	0,58	-0,58	5	0,59	0,20	0,20

P1	39	16,00	17,11	-0,33	0,56	-0,60	5	0,58	2,60	1,50
P6	37	18,00	19,16	-0,33	0,53	-0,62	5	0,56	1,10	1,70
P4	12	16,00	17,08	-0,32	0,56	-0,58	5	0,59	0,40	0,40
P17	10	14,00	14,91	-0,32	0,61	-0,53	5	0,62	1,60	1,00
P17	16	15,00	16,00	-0,32	0,58	-0,55	5	0,60	0,40	0,40
P6	17	15,00	15,99	-0,32	0,58	-0,55	5	0,61	0,20	0,20
P9	7	18,00	19,10	-0,31	0,53	-0,59	5	0,58	0,50	0,40
P18	15	14,00	14,88	-0,31	0,61	-0,51	5	0,63	0,70	0,80
P3	2	11,00	11,58	-0,31	0,76	-0,41	5	0,70	0,30	0,10
P7	21	19,00	20,05	-0,31	0,54	-0,58	5	0,59	0,70	0,40
P17	21	19,00	20,05	-0,31	0,54	-0,58	5	0,59	0,70	0,40
P12	25	20,00	20,94	-0,31	0,55	-0,55	5	0,60	0,80	0,50
P3	31	17,00	18,06	-0,31	0,54	-0,57	5	0,60	0,60	0,40
P12	39	15,00	15,93	-0,30	0,58	-0,52	5	0,63	0,20	0,20
P15	24	17,00	18,04	-0,30	0,54	-0,55	5	0,60	0,20	0,20
P15	29	17,00	18,04	-0,30	0,54	-0,55	5	0,60	0,40	0,30
P20	24	17,00	18,04	-0,30	0,54	-0,55	5	0,60	0,20	0,20
P20	29	17,00	18,04	-0,30	0,54	-0,55	5	0,60	0,40	0,30
P4	6	18,00	19,05	-0,30	0,53	-0,56	5	0,60	0,60	0,50
P14	19	20,00	20,90	-0,29	0,55	-0,53	5	0,62	0,70	0,60
P2	14	19,00	19,99	-0,29	0,54	-0,54	5	0,61	1,20	0,80
P2	21	19,00	19,99	-0,29	0,54	-0,54	5	0,61	0,70	0,40
P19	6	20,00	20,89	-0,29	0,55	-0,52	5	0,62	0,70	0,60
P17	11	18,00	19,02	-0,29	0,53	-0,54	5	0,61	1,10	1,70
P12	4	16,00	16,95	-0,29	0,56	-0,51	5	0,63	0,40	0,30
P11	31	16,00	16,95	-0,29	0,56	-0,51	5	0,63	0,80	0,50
P9	15	14,00	14,80	-0,28	0,61	-0,47	5	0,66	0,70	0,80
P19	19	21,00	21,72	-0,28	0,60	-0,47	5	0,66	0,90	0,90
P12	36	18,00	18,99	-0,28	0,53	-0,53	5	0,62	0,40	0,30
P6	24	17,00	17,98	-0,28	0,54	-0,52	5	0,62	0,20	0,20
P6	29	17,00	17,98	-0,28	0,54	-0,52	5	0,62	0,40	0,30
P19	16	17,00	17,97	-0,28	0,54	-0,52	5	0,63	1,50	1,10
P14	14	20,00	20,86	-0,28	0,55	-0,50	5	0,64	0,40	0,30
P9	31	16,00	16,92	-0,28	0,56	-0,50	5	0,64	2,60	2,00
P19	1	13,00	13,68	-0,28	0,65	-0,42	5	0,69	0,30	0,20
P13	36	17,00	17,95	-0,27	0,54	-0,51	5	0,63	1,10	1,50
P14	11	19,00	19,91	-0,27	0,54	-0,50	5	0,64	0,60	1,90
P13	17	14,00	14,75	-0,27	0,61	-0,44	5	0,68	0,80	9,00
P13	38	13,00	13,66	-0,27	0,65	-0,41	5	0,70	0,60	0,30
P16	12	17,00	17,93	-0,27	0,54	-0,49	5	0,64	0,40	0,30
P15	21	19,00	19,90	-0,26	0,54	-0,49	5	0,64	0,70	0,40
P20	21	19,00	19,90	-0,26	0,54	-0,49	5	0,64	0,70	0,40
P8	13	19,00	19,90	-0,26	0,54	-0,49	5	0,64	0,80	0,50
P3	15	15,00	15,81	-0,26	0,58	-0,45	5	0,67	1,40	2,00
P3	39	17,00	17,91	-0,26	0,54	-0,48	5	0,65	1,30	0,90
P16	6	19,00	19,88	-0,26	0,54	-0,48	5	0,65	0,20	0,20
P16	11	19,00	19,88	-0,26	0,54	-0,48	5	0,65	0,60	1,90
P6	10	14,00	14,72	-0,26	0,61	-0,42	5	0,69	1,60	1,00

P5	35	18,00	18,91	-0,26	0,53	-0,48	5	0,65	0,40	0,30
P5	37	18,00	18,91	-0,26	0,53	-0,48	5	0,65	0,40	0,30
P8	17	15,00	15,79	-0,26	0,58	-0,44	5	0,68	0,50	0,50
P18	1	12,00	12,55	-0,25	0,69	-0,37	5	0,73	0,70	0,30
P1	30	18,00	18,89	-0,25	0,53	-0,47	5	0,66	0,40	0,30
P13	28	16,00	16,83	-0,25	0,56	-0,45	5	0,67	0,90	1,10
P6	21	19,00	19,84	-0,25	0,54	-0,46	5	0,66	0,70	0,40
P5	17	15,00	15,76	-0,25	0,58	-0,43	5	0,69	0,40	0,30
P19	40	16,00	16,81	-0,25	0,56	-0,44	5	0,68	0,50	0,40
P16	38	15,00	15,74	-0,24	0,58	-0,41	5	0,70	0,90	0,50
P16	40	15,00	15,74	-0,24	0,58	-0,41	5	0,70	0,50	0,50
P1	24	18,00	18,84	-0,24	0,53	-0,45	5	0,67	0,80	0,60
P1	29	18,00	18,84	-0,24	0,53	-0,45	5	0,67	0,40	0,30
P1	27	17,00	17,82	-0,24	0,54	-0,44	5	0,68	0,20	0,20
P11	1	12,00	12,51	-0,24	0,69	-0,34	5	0,75	0,70	0,30
P19	7	20,00	20,72	-0,23	0,55	-0,42	5	0,69	0,70	0,60
P10	38	15,00	15,71	-0,23	0,58	-0,40	5	0,71	0,90	0,50
P10	40	15,00	15,71	-0,23	0,58	-0,40	5	0,71	0,90	0,50
P19	10	16,00	16,76	-0,23	0,56	-0,41	5	0,70	0,80	0,60
P9	1	12,00	12,49	-0,23	0,69	-0,33	5	0,76	0,70	0,30
P16	10	15,00	15,69	-0,22	0,58	-0,39	5	0,72	0,80	0,50
P5	30	17,00	17,78	-0,22	0,54	-0,41	5	0,70	0,30	0,30
P15	8	20,00	20,70	-0,22	0,55	-0,40	5	0,70	1,80	0,90
P20	8	20,00	20,70	-0,22	0,55	-0,40	5	0,70	1,80	0,90
P12	35	18,00	18,78	-0,22	0,53	-0,42	5	0,70	0,40	0,30
P12	37	18,00	18,78	-0,22	0,53	-0,42	5	0,70	0,40	0,30
P12	13	19,00	19,75	-0,22	0,54	-0,41	5	0,70	0,90	0,50
P4	15	14,00	14,59	-0,21	0,61	-0,35	5	0,74	0,70	0,80
P8	40	14,00	14,59	-0,21	0,61	-0,35	5	0,74	0,40	0,30
P14	7	19,00	19,72	-0,21	0,54	-0,39	5	0,71	1,20	0,80
P18	13	20,00	20,64	-0,20	0,55	-0,37	5	0,73	0,40	0,30
P7	15	14,00	14,56	-0,20	0,61	-0,33	5	0,75	2,30	1,90
P11	35	19,00	19,68	-0,20	0,54	-0,37	5	0,73	2,00	1,20
P12	26	16,00	16,66	-0,20	0,56	-0,36	5	0,73	0,60	0,70
P8	10	14,00	14,54	-0,20	0,61	-0,32	5	0,76	1,60	1,00
P1	12	17,00	17,67	-0,19	0,54	-0,36	5	0,73	0,40	0,30
P13	30	16,00	16,63	-0,19	0,56	-0,35	5	0,74	0,10	0,10
P12	30	17,00	17,65	-0,19	0,54	-0,35	5	0,74	0,30	0,30
P12	27	16,00	16,61	-0,18	0,56	-0,33	5	0,75	0,10	0,10
P8	21	19,00	19,63	-0,18	0,54	-0,34	5	0,75	0,70	0,40
P1	16	16,00	16,59	-0,18	0,56	-0,32	5	0,76	0,90	0,60
P3	27	18,00	18,63	-0,18	0,53	-0,34	5	0,75	0,10	0,10
P3	13	21,00	21,47	-0,18	0,60	-0,30	5	0,78	2,20	1,10
P13	29	16,00	16,58	-0,18	0,56	-0,32	5	0,76	0,10	0,10
P5	12	16,00	16,58	-0,18	0,56	-0,32	5	0,76	0,90	0,60
P9	13	20,00	20,56	-0,18	0,55	-0,32	5	0,76	0,40	0,30
P5	14	19,00	19,60	-0,17	0,54	-0,33	5	0,76	0,10	0,10
P7	36	19,00	19,58	-0,17	0,54	-0,32	5	0,76	0,90	0,50

P1	40	15,00	15,51	-0,17	0,58	-0,29	5	0,78	0,50	0,50
P14	2	11,00	11,30	-0,16	0,76	-0,22	5	0,84	0,30	0,10
P4	9	15,00	15,48	-0,16	0,58	-0,27	5	0,80	0,90	0,50
P16	2	11,00	11,28	-0,15	0,76	-0,20	5	0,85	0,30	0,10
P4	1	12,00	12,33	-0,15	0,69	-0,22	5	0,83	0,70	0,30
P5	11	18,00	18,54	-0,15	0,53	-0,29	5	0,78	0,30	1,20
P2	36	19,00	19,52	-0,15	0,54	-0,28	5	0,79	0,90	0,50
P17	25	21,00	21,40	-0,15	0,60	-0,25	5	0,81	0,50	0,30
P7	39	16,00	16,50	-0,15	0,56	-0,27	5	0,80	1,60	1,00
P7	9	15,00	15,45	-0,15	0,58	-0,26	5	0,81	0,90	0,50
P17	9	15,00	15,45	-0,15	0,58	-0,26	5	0,81	0,90	0,50
P7	1	12,00	12,31	-0,14	0,69	-0,21	5	0,84	0,70	0,30
P17	1	12,00	12,31	-0,14	0,69	-0,21	5	0,84	0,70	0,30
P17	33	17,00	17,50	-0,14	0,54	-0,27	5	0,80	0,30	0,30
P8	8	20,00	20,45	-0,14	0,55	-0,26	5	0,81	1,20	0,70
P6	15	14,00	14,39	-0,14	0,61	-0,23	5	0,83	0,70	0,80
P2	4	17,00	17,48	-0,14	0,54	-0,26	5	0,81	0,50	0,40
P18	19	20,00	20,44	-0,14	0,55	-0,25	5	0,81	0,70	0,60
P14	36	20,00	20,44	-0,14	0,55	-0,25	5	0,81	0,40	0,30
P13	15	13,00	13,32	-0,13	0,65	-0,20	5	0,85	0,40	1,00
P2	9	15,00	15,40	-0,13	0,58	-0,23	5	0,83	0,90	0,50
P1	7	19,00	19,44	-0,13	0,54	-0,24	5	0,82	1,20	0,80
P13	14	18,00	18,45	-0,13	0,53	-0,24	5	0,82	0,10	0,10
P13	21	18,00	18,45	-0,13	0,53	-0,24	5	0,82	0,30	0,20
P2	1	12,00	12,27	-0,13	0,69	-0,18	5	0,86	0,70	0,30
P2	33	17,00	17,43	-0,13	0,54	-0,23	5	0,83	0,30	0,30
P15	36	19,00	19,43	-0,12	0,54	-0,23	5	0,83	0,90	0,50
P20	36	19,00	19,43	-0,12	0,54	-0,23	5	0,83	0,90	0,50
P16	4	18,00	18,44	-0,12	0,53	-0,23	5	0,82	1,50	2,00
P18	14	20,00	20,39	-0,12	0,55	-0,22	5	0,83	0,40	0,30
P18	12	17,00	17,42	-0,12	0,54	-0,22	5	0,83	0,40	0,30
P14	33	18,00	18,42	-0,12	0,53	-0,22	5	0,83	2,20	1,30
P12	11	18,00	18,41	-0,12	0,53	-0,22	5	0,83	0,20	0,20
P14	15	15,00	15,35	-0,11	0,58	-0,20	5	0,85	1,60	1,60
P14	39	17,00	17,39	-0,11	0,54	-0,21	5	0,84	1,10	0,70
P18	11	19,00	19,39	-0,11	0,54	-0,21	5	0,84	0,60	1,90
P10	22	15,00	15,34	-0,11	0,58	-0,20	5	0,85	3,50	3,10
P9	19	20,00	20,36	-0,11	0,55	-0,20	5	0,85	0,70	0,60
P2	28	18,00	18,39	-0,11	0,53	-0,21	5	0,84	0,30	0,20
P3	16	17,00	17,38	-0,11	0,54	-0,20	5	0,85	1,50	1,10
P7	35	19,00	19,38	-0,11	0,54	-0,20	5	0,85	0,90	0,50
P6	36	19,00	19,37	-0,11	0,54	-0,20	5	0,85	0,90	0,50
P15	25	21,00	21,29	-0,11	0,60	-0,18	5	0,87	0,50	0,30
P20	25	21,00	21,29	-0,11	0,60	-0,18	5	0,87	0,50	0,30
P16	15	15,00	15,32	-0,10	0,58	-0,18	5	0,86	1,40	2,00
P3	1	13,00	13,25	-0,10	0,65	-0,16	5	0,88	0,30	0,20
P16	39	17,00	17,36	-0,10	0,54	-0,19	5	0,85	1,10	0,70
P15	9	15,00	15,31	-0,10	0,58	-0,18	5	0,86	0,90	0,50

P20	9	15,00	15,31	-0,10	0,58	-0,18	5	0,86	0,90	0,50
P16	28	19,00	19,34	-0,10	0,54	-0,18	5	0,86	1,20	0,80
P15	33	17,00	17,34	-0,10	0,54	-0,18	5	0,86	0,30	0,30
P20	33	17,00	17,34	-0,10	0,54	-0,18	5	0,86	0,30	0,30
P13	8	19,00	19,33	-0,10	0,54	-0,18	5	0,86	0,40	0,40
P9	14	20,00	20,31	-0,10	0,55	-0,17	5	0,87	0,40	0,30
P10	15	15,00	15,29	-0,10	0,58	-0,17	5	0,88	0,70	0,80
P18	38	15,00	15,29	-0,09	0,58	-0,16	5	0,88	0,90	0,50
P9	12	17,00	17,32	-0,09	0,54	-0,17	5	0,87	0,40	0,30
P2	35	19,00	19,31	-0,09	0,54	-0,17	5	0,87	0,90	0,50
P2	37	19,00	19,31	-0,09	0,54	-0,17	5	0,87	0,90	0,50
P14	9	16,00	16,29	-0,09	0,56	-0,16	5	0,88	0,90	0,60
P6	25	21,00	21,24	-0,09	0,60	-0,15	5	0,89	0,50	0,30
P9	11	19,00	19,30	-0,09	0,54	-0,16	5	0,88	0,60	1,90
P6	9	15,00	15,26	-0,09	0,58	-0,15	5	0,89	0,90	0,50
P6	1	12,00	12,17	-0,08	0,69	-0,12	5	0,91	0,70	0,30
P4	30	18,00	18,28	-0,08	0,53	-0,15	5	0,89	0,40	0,30
P6	33	17,00	17,28	-0,08	0,54	-0,15	5	0,89	0,30	0,30
P2	13	20,00	20,25	-0,08	0,55	-0,14	5	0,89	1,20	0,70
P8	15	14,00	14,21	-0,08	0,61	-0,13	5	0,90	1,50	2,70
P14	35	20,00	20,25	-0,08	0,55	-0,14	5	0,90	0,40	0,30
P14	37	20,00	20,25	-0,08	0,55	-0,14	5	0,90	0,40	0,30
P3	38	16,00	16,25	-0,08	0,56	-0,14	5	0,90	0,90	0,60
P17	30	18,00	18,25	-0,07	0,53	-0,13	5	0,90	0,40	0,30
P17	26	17,00	17,24	-0,07	0,54	-0,13	5	0,90	0,50	0,90
P9	38	15,00	15,20	-0,07	0,58	-0,12	5	0,91	0,90	0,50
P16	35	20,00	20,22	-0,07	0,55	-0,12	5	0,91	0,40	0,30
P19	35	21,00	21,18	-0,07	0,60	-0,11	5	0,92	2,20	1,10
P4	27	17,00	17,23	-0,07	0,54	-0,12	5	0,91	0,20	0,20
P15	35	19,00	19,22	-0,06	0,54	-0,12	5	0,91	0,90	0,50
P20	35	19,00	19,22	-0,06	0,54	-0,12	5	0,91	0,90	0,50
P13	32	11,00	11,10	-0,06	0,76	-0,08	5	0,94	0,40	0,30
P7	29	18,00	18,20	-0,06	0,53	-0,11	5	0,92	0,40	0,30
P17	27	17,00	17,19	-0,06	0,54	-0,10	5	0,92	0,20	0,20
P2	30	18,00	18,19	-0,05	0,53	-0,10	5	0,92	1,00	0,70
P13	7	17,00	17,17	-0,05	0,54	-0,09	5	0,93	0,80	1,20
P15	13	20,00	20,16	-0,05	0,55	-0,09	5	0,93	1,20	0,70
P20	13	20,00	20,16	-0,05	0,55	-0,09	5	0,93	1,20	0,70
P14	30	19,00	19,17	-0,05	0,54	-0,09	5	0,93	0,50	0,30
P13	31	15,00	15,15	-0,05	0,58	-0,09	5	0,94	0,40	0,40
P12	22	14,00	14,13	-0,05	0,61	-0,08	5	0,94	0,50	0,70
P6	35	19,00	19,16	-0,05	0,54	-0,09	5	0,94	0,90	0,50
P2	29	18,00	18,14	-0,04	0,53	-0,07	5	0,94	0,40	0,30
P2	27	17,00	17,13	-0,04	0,54	-0,07	5	0,95	0,20	0,20
P14	24	19,00	19,12	-0,04	0,54	-0,07	5	0,95	1,20	0,80
P14	29	19,00	19,12	-0,04	0,54	-0,07	5	0,95	0,50	0,30
P13	9	14,00	14,09	-0,03	0,61	-0,06	5	0,96	0,40	0,30
P5	36	19,00	19,12	-0,03	0,54	-0,06	5	0,95	0,80	0,50

P10	30	19,00	19,11	-0,03	0,54	-0,06	5	0,95	0,50	0,30
P8	23	17,00	17,11	-0,03	0,54	-0,06	5	0,96	0,40	0,30
P10	26	18,00	18,10	-0,03	0,53	-0,05	5	0,96	1,70	2,00
P16	24	19,00	19,09	-0,03	0,54	-0,05	5	0,96	1,20	0,80
P16	29	19,00	19,09	-0,03	0,54	-0,05	5	0,96	0,20	0,20
P15	30	18,00	18,09	-0,03	0,53	-0,05	5	0,96	0,40	0,30
P20	30	18,00	18,09	-0,03	0,53	-0,05	5	0,96	0,40	0,30
P15	26	17,00	17,09	-0,03	0,54	-0,05	5	0,96	0,50	0,90
P20	26	17,00	17,09	-0,03	0,54	-0,05	5	0,96	0,50	0,90
P8	25	21,00	21,07	-0,02	0,60	-0,04	5	0,97	0,50	0,30
P8	39	16,00	16,08	-0,02	0,56	-0,04	5	0,97	1,60	1,00
P4	21	20,00	20,08	-0,02	0,55	-0,04	5	0,97	0,40	0,30
P19	12	19,00	19,08	-0,02	0,54	-0,04	5	0,97	2,00	1,20
P16	13	21,00	21,06	-0,02	0,60	-0,04	5	0,97	0,50	0,30
P10	29	19,00	19,06	-0,02	0,54	-0,03	5	0,98	0,50	0,30
P5	39	16,00	16,05	-0,02	0,56	-0,03	5	0,98	0,40	0,30
P7	12	17,00	17,04	-0,01	0,54	-0,02	5	0,98	0,40	0,30
P13	25	20,00	20,04	-0,01	0,55	-0,02	5	0,98	0,40	0,50
P15	27	17,00	17,04	-0,01	0,54	-0,02	5	0,98	0,20	0,20
P20	27	17,00	17,04	-0,01	0,54	-0,02	5	0,98	0,20	0,20
P6	30	18,00	18,03	-0,01	0,53	-0,02	5	0,99	0,40	0,30
P1	9	16,00	16,03	-0,01	0,56	-0,02	5	0,99	0,90	0,60
P6	26	17,00	17,03	-0,01	0,54	-0,01	5	0,99	0,50	0,90
P5	33	17,00	17,03	-0,01	0,54	-0,01	5	0,99	0,30	0,30
P17	32	12,00	12,01	-0,01	0,69	-0,01	5	0,99	0,80	0,60
P13	39	15,00	15,01	-0,01	0,58	-0,01	5	0,99	0,70	0,70
P7	11	19,00	19,02	0,00	0,54	-0,01	5	0,99	0,60	1,90
P17	6	19,00	19,02	0,00	0,54	-0,01	5	0,99	1,20	0,80
P14	25	22,00	22,01	0,00	0,70	0,00	5	1,00	0,90	0,40
P3	4	19,00	18,99	0,00	0,54	0,01	5	1,00	1,20	0,80
P4	38	15,00	14,99	0,00	0,58	0,01	5	0,99	0,90	0,50
P4	40	15,00	14,99	0,00	0,58	0,01	5	0,99	0,50	0,50
P2	12	17,00	16,98	0,01	0,54	0,01	5	0,99	0,40	0,30
P6	27	17,00	16,98	0,01	0,54	0,01	5	0,99	0,20	0,20
P13	4	16,00	15,97	0,01	0,56	0,02	5	0,99	0,70	0,50
P2	32	12,00	11,98	0,01	0,69	0,02	5	0,99	2,20	2,50
P2	11	19,00	18,95	0,01	0,54	0,02	5	0,98	0,60	1,90
P7	38	15,00	14,96	0,01	0,58	0,02	5	0,98	0,90	0,50
P17	38	15,00	14,96	0,01	0,58	0,02	5	0,98	0,90	0,50
P17	40	15,00	14,96	0,01	0,58	0,02	5	0,98	0,50	0,50
P12	23	17,00	16,95	0,01	0,54	0,03	5	0,98	0,40	0,30
P10	25	22,00	21,97	0,02	0,70	0,02	5	0,98	0,90	0,40
P18	36	20,00	19,94	0,02	0,55	0,03	5	0,98	0,40	0,30
P8	35	19,00	18,94	0,02	0,54	0,03	5	0,98	1,30	1,50
P12	1	12,00	11,94	0,03	0,69	0,04	5	0,97	0,70	0,30
P14	6	20,00	19,91	0,03	0,55	0,05	5	0,96	1,10	0,60
P12	33	17,00	16,90	0,03	0,54	0,05	5	0,96	0,30	0,30
P15	14	20,00	19,90	0,03	0,55	0,06	5	0,96	1,20	0,70

P20	14	20,00	19,90	0,03	0,55	0,06	5	0,96	1,20	0,70
P2	38	15,00	14,91	0,03	0,58	0,05	5	0,96	0,90	0,50
P15	12	17,00	16,89	0,03	0,54	0,06	5	0,95	0,40	0,30
P20	12	17,00	16,89	0,03	0,54	0,06	5	0,95	0,40	0,30
P19	11	21,00	20,89	0,04	0,60	0,06	5	0,95	1,40	4,00
P5	13	20,00	19,87	0,04	0,55	0,07	5	0,95	0,40	0,30
P15	11	19,00	18,86	0,04	0,54	0,08	5	0,94	0,60	1,90
P20	11	19,00	18,86	0,04	0,54	0,08	5	0,94	0,60	1,90
P18	39	17,00	16,86	0,04	0,54	0,08	5	0,94	1,10	0,70
P16	19	21,00	20,88	0,04	0,60	0,07	5	0,94	0,70	0,80
P9	36	20,00	19,85	0,04	0,55	0,08	5	0,94	0,40	0,30
P6	12	17,00	16,83	0,05	0,54	0,09	5	0,93	0,40	0,30
P4	8	21,00	20,85	0,05	0,60	0,09	5	0,93	0,50	0,30
P10	19	21,00	20,85	0,05	0,60	0,09	5	0,93	0,60	0,40
P17	7	19,00	18,81	0,05	0,54	0,10	5	0,92	1,20	0,80
P19	38	17,00	16,81	0,05	0,54	0,10	5	0,92	0,80	0,50
P8	30	18,00	17,81	0,05	0,53	0,10	5	0,92	0,40	0,30
P8	26	17,00	16,81	0,06	0,54	0,10	5	0,92	0,50	0,90
P6	6	19,00	18,80	0,06	0,54	0,11	5	0,92	1,20	0,80
P6	11	19,00	18,80	0,06	0,54	0,11	5	0,92	0,60	1,90
P15	38	15,00	14,83	0,06	0,58	0,10	5	0,92	0,90	0,50
P15	40	15,00	14,83	0,06	0,58	0,10	5	0,92	0,50	0,50
P20	38	15,00	14,83	0,06	0,58	0,10	5	0,92	0,90	0,50
P20	40	15,00	14,83	0,06	0,58	0,10	5	0,92	0,50	0,50
P16	14	21,00	20,83	0,06	0,60	0,10	5	0,92	0,50	0,30
P16	21	21,00	20,83	0,06	0,60	0,10	5	0,92	1,50	0,70
P17	8	21,00	20,83	0,06	0,60	0,10	5	0,92	2,20	1,10
P19	13	22,00	21,87	0,06	0,70	0,09	5	0,93	1,30	0,70
P18	9	16,00	15,79	0,06	0,56	0,12	5	0,91	0,90	0,60
P9	39	17,00	16,77	0,07	0,54	0,13	5	0,90	1,10	0,70
P8	27	17,00	16,76	0,07	0,54	0,13	5	0,90	0,20	0,20
P13	35	18,00	17,74	0,07	0,53	0,14	5	0,90	0,40	0,30
P15	10	15,00	14,78	0,08	0,58	0,13	5	0,90	0,80	0,50
P20	10	15,00	14,78	0,08	0,58	0,13	5	0,90	0,80	0,50
P6	38	15,00	14,77	0,08	0,58	0,13	5	0,90	0,90	0,50
P6	40	15,00	14,77	0,08	0,58	0,13	5	0,90	0,50	0,50
P18	35	20,00	19,74	0,08	0,55	0,14	5	0,89	0,40	0,30
P18	37	20,00	19,74	0,08	0,55	0,14	5	0,89	0,40	0,30
P5	24	18,00	17,72	0,08	0,53	0,15	5	0,89	0,50	0,40
P5	27	17,00	16,73	0,08	0,54	0,15	5	0,89	0,60	0,40
P9	9	16,00	15,71	0,09	0,56	0,16	5	0,88	0,90	0,60
P4	31	17,00	16,67	0,10	0,54	0,18	5	0,87	0,80	0,50
P13	26	16,00	15,69	0,10	0,56	0,17	5	0,87	0,40	0,60
P15	7	19,00	18,65	0,10	0,54	0,18	5	0,86	1,20	0,80
P20	7	19,00	18,65	0,10	0,54	0,18	5	0,86	1,20	0,80
P7	2	11,00	10,83	0,10	0,76	0,13	5	0,90	0,70	3,10
P17	2	11,00	10,83	0,10	0,76	0,13	5	0,90	0,30	0,10
P10	7	20,00	19,66	0,10	0,55	0,19	5	0,86	1,10	0,60

P3	35	21,00	20,70	0,10	0,60	0,17	5	0,87	2,20	1,10
P9	35	20,00	19,65	0,10	0,55	0,19	5	0,86	0,40	0,30
P9	37	20,00	19,65	0,10	0,55	0,19	5	0,86	0,40	0,30
P1	6	20,00	19,64	0,11	0,55	0,20	5	0,85	0,70	0,60
P1	11	20,00	19,64	0,11	0,55	0,20	5	0,85	2,20	3,40
P13	27	16,00	15,65	0,11	0,56	0,20	5	0,85	0,10	0,10
P8	14	20,00	19,63	0,11	0,55	0,20	5	0,85	1,20	0,70
P4	36	20,00	19,61	0,12	0,55	0,21	5	0,84	0,40	0,30
P6	7	19,00	18,59	0,12	0,54	0,22	5	0,84	1,20	0,80
P1	19	21,00	20,66	0,12	0,60	0,20	5	0,85	0,70	0,80
P2	2	11,00	10,80	0,12	0,76	0,15	5	0,88	0,70	3,10
P8	6	19,00	18,57	0,12	0,54	0,23	5	0,83	1,20	0,80
P13	24	17,00	16,58	0,12	0,54	0,23	5	0,83	0,20	0,20
P6	8	21,00	20,64	0,12	0,60	0,21	5	0,84	2,20	1,10
P1	1	13,00	12,71	0,12	0,65	0,19	5	0,86	1,00	1,20
P17	36	20,00	19,58	0,13	0,55	0,23	5	0,83	0,40	0,30
P19	4	20,00	19,57	0,13	0,55	0,23	5	0,83	0,70	0,60
P13	40	14,00	13,66	0,13	0,61	0,21	5	0,84	0,50	0,70
P14	31	18,00	17,55	0,13	0,53	0,24	5	0,82	2,20	1,30
P17	23	18,00	17,55	0,13	0,53	0,24	5	0,82	0,80	0,60
P5	6	19,00	18,54	0,13	0,54	0,24	5	0,82	2,30	1,50
P9	30	19,00	18,54	0,13	0,54	0,25	5	0,82	0,50	0,30
P7	22	15,00	14,61	0,13	0,58	0,23	5	0,83	0,90	0,50
P4	33	18,00	17,53	0,13	0,53	0,25	5	0,81	0,60	0,50
P5	16	16,00	15,57	0,14	0,56	0,24	5	0,82	0,80	0,50
P11	24	19,00	18,52	0,14	0,54	0,26	5	0,81	0,10	0,10
P11	13	21,00	20,59	0,14	0,60	0,24	5	0,82	0,30	0,20
P7	33	18,00	17,50	0,14	0,53	0,27	5	0,80	0,60	0,50
P19	14	22,00	21,68	0,14	0,70	0,21	5	0,84	1,30	0,70
P13	10	14,00	13,62	0,15	0,61	0,24	5	0,82	0,20	0,20
P9	29	19,00	18,49	0,15	0,54	0,27	5	0,80	0,50	0,30
P4	28	19,00	18,49	0,15	0,54	0,27	5	0,80	0,10	0,10
P13	16	15,00	14,57	0,15	0,58	0,25	5	0,81	0,40	0,40
P3	12	19,00	18,48	0,15	0,54	0,28	5	0,79	2,00	1,20
P5	38	15,00	14,56	0,15	0,58	0,26	5	0,81	0,90	0,50
P17	39	17,00	16,50	0,15	0,54	0,28	5	0,79	1,10	0,70
P2	22	15,00	14,56	0,15	0,58	0,26	5	0,81	0,90	0,50
P15	31	17,00	16,49	0,15	0,54	0,28	5	0,79	0,10	0,10
P20	31	17,00	16,49	0,15	0,54	0,28	5	0,79	0,10	0,10
P14	4	19,00	18,47	0,15	0,54	0,28	5	0,79	1,30	2,30
P14	23	19,00	18,47	0,15	0,54	0,28	5	0,79	1,20	0,80
P18	25	22,00	21,67	0,15	0,70	0,22	5	0,84	0,90	0,40
P13	12	16,00	15,51	0,15	0,56	0,28	5	0,79	0,90	0,60
P1	38	16,00	15,51	0,15	0,56	0,28	5	0,79	0,90	0,60
P7	28	19,00	18,46	0,16	0,54	0,29	5	0,78	0,50	0,30
P17	28	19,00	18,46	0,16	0,54	0,29	5	0,78	0,60	0,40
P12	21	20,00	19,48	0,16	0,55	0,28	5	0,79	0,40	0,30
P16	23	19,00	18,44	0,16	0,54	0,30	5	0,78	1,20	0,80

P6	2	11,00	10,73	0,16	0,76	0,21	5	0,84	0,30	0,10
P2	39	17,00	16,43	0,17	0,54	0,31	5	0,77	1,10	0,70
P6	31	17,00	16,43	0,17	0,54	0,31	5	0,77	0,10	0,10
P11	25	22,00	21,62	0,17	0,70	0,24	5	0,82	0,60	1,20
P15	22	15,00	14,48	0,18	0,58	0,31	5	0,77	0,90	0,50
P20	22	15,00	14,48	0,18	0,58	0,31	5	0,77	0,90	0,50
P4	37	20,00	19,41	0,18	0,55	0,32	5	0,76	1,20	0,70
P13	6	18,00	17,38	0,18	0,53	0,33	5	0,75	0,50	0,40
P13	11	18,00	17,38	0,18	0,53	0,33	5	0,75	0,30	1,20
P9	25	22,00	21,60	0,18	0,70	0,26	5	0,81	0,90	0,40
P8	7	19,00	18,37	0,18	0,54	0,34	5	0,75	1,20	0,80
P14	32	13,00	12,58	0,18	0,65	0,28	5	0,79	2,60	1,30
P10	33	19,00	18,36	0,18	0,54	0,34	5	0,75	1,40	0,80
P11	12	18,00	17,36	0,18	0,53	0,35	5	0,74	1,40	0,90
P12	38	15,00	14,46	0,19	0,58	0,32	5	0,76	0,90	0,50
P7	37	20,00	19,38	0,19	0,55	0,34	5	0,75	1,20	0,70
P17	35	20,00	19,38	0,19	0,55	0,34	5	0,75	0,40	0,30
P17	37	20,00	19,38	0,19	0,55	0,34	5	0,75	0,40	0,30
P14	28	20,00	19,37	0,19	0,55	0,34	5	0,75	1,10	0,60
P16	32	13,00	12,56	0,19	0,65	0,29	5	0,78	1,40	0,80
P10	39	18,00	17,33	0,19	0,53	0,36	5	0,73	1,00	0,70
P6	4	18,00	17,33	0,19	0,53	0,36	5	0,73	1,50	2,00
P15	15	15,00	14,44	0,19	0,58	0,34	5	0,75	1,40	2,00
P20	15	15,00	14,44	0,19	0,58	0,34	5	0,75	1,40	2,00
P15	39	17,00	16,35	0,19	0,54	0,36	5	0,73	1,10	0,70
P20	39	17,00	16,35	0,19	0,54	0,36	5	0,73	1,10	0,70
P6	22	15,00	14,43	0,19	0,58	0,34	5	0,75	0,90	0,50
P10	28	20,00	19,31	0,21	0,55	0,37	5	0,73	1,10	0,60
P3	11	21,00	20,38	0,21	0,60	0,35	5	0,74	1,40	4,00
P6	39	17,00	16,29	0,21	0,54	0,39	5	0,71	1,10	0,70
P7	30	19,00	18,25	0,21	0,54	0,40	5	0,71	0,80	0,50
P7	26	18,00	17,24	0,22	0,53	0,41	5	0,70	0,30	1,20
P4	13	21,00	20,34	0,22	0,60	0,37	5	0,72	0,50	0,30
P12	7	19,00	18,21	0,23	0,54	0,42	5	0,69	1,40	0,80
P10	8	22,00	21,49	0,23	0,70	0,33	5	0,76	0,90	0,40
P17	24	19,00	18,20	0,23	0,54	0,42	5	0,69	0,40	0,40
P17	29	19,00	18,20	0,23	0,54	0,42	5	0,69	0,80	0,50
P7	27	18,00	17,19	0,23	0,53	0,43	5	0,68	0,50	0,40
P8	31	17,00	16,22	0,23	0,54	0,43	5	0,69	1,00	0,80
P17	13	21,00	20,31	0,23	0,60	0,39	5	0,71	0,50	0,30
P1	4	19,00	18,18	0,23	0,54	0,43	5	0,68	1,20	0,80
P2	26	18,00	17,18	0,23	0,53	0,44	5	0,68	0,30	1,20
P14	26	19,00	18,16	0,24	0,54	0,44	5	0,68	1,30	2,30
P11	38	16,00	15,23	0,24	0,56	0,44	5	0,68	0,40	0,60
P13	22	14,00	13,36	0,25	0,61	0,40	5	0,70	0,40	0,50
P16	26	19,00	18,13	0,25	0,54	0,46	5	0,66	1,30	2,30
P1	33	19,00	18,13	0,25	0,54	0,46	5	0,66	1,40	0,80
P14	27	19,00	18,11	0,25	0,54	0,47	5	0,66	1,20	0,80

P16	30	20,00	19,14	0,25	0,55	0,46	5	0,67	0,10	0,10
P8	4	18,00	17,11	0,26	0,53	0,48	5	0,65	1,50	2,00
P19	32	14,00	13,33	0,26	0,61	0,42	5	0,69	2,00	2,00
P7	25	22,00	21,40	0,26	0,70	0,37	5	0,72	0,10	0,10
P16	27	19,00	18,08	0,26	0,54	0,49	5	0,65	1,20	0,80
P5	4	18,00	17,08	0,27	0,53	0,50	5	0,64	0,50	0,40
P5	23	18,00	17,08	0,27	0,53	0,50	5	0,64	0,10	0,10
P5	22	15,00	14,23	0,27	0,58	0,46	5	0,66	0,70	0,60
P8	33	18,00	17,06	0,27	0,53	0,51	5	0,63	0,20	0,20
P10	27	19,00	18,05	0,27	0,54	0,50	5	0,64	1,20	0,80
P12	2	11,00	10,56	0,27	0,76	0,36	5	0,73	0,70	3,10
P17	12	18,00	17,04	0,27	0,53	0,51	5	0,63	1,00	0,70
P10	24	20,00	19,06	0,28	0,55	0,50	5	0,64	0,30	0,40
P2	25	22,00	21,36	0,28	0,70	0,40	5	0,71	0,10	0,10
P4	11	20,00	19,05	0,28	0,55	0,51	5	0,63	0,70	2,60
P8	28	19,00	18,01	0,28	0,54	0,53	5	0,62	0,40	0,40
P10	17	18,00	17,02	0,28	0,53	0,53	5	0,62	3,40	2,90
P18	31	18,00	17,01	0,28	0,53	0,53	5	0,62	1,50	0,90
P19	36	22,00	21,34	0,28	0,70	0,41	5	0,70	1,30	0,70
P5	15	15,00	14,18	0,28	0,58	0,49	5	0,64	0,50	0,50
P4	16	17,00	16,03	0,29	0,54	0,54	5	0,61	1,20	0,80
P6	13	21,00	20,11	0,30	0,60	0,49	5	0,64	0,50	0,30
P7	19	21,00	20,10	0,30	0,60	0,50	5	0,64	1,20	0,90
P3	19	22,00	21,30	0,30	0,70	0,43	5	0,69	1,20	1,50
P3	23	20,00	18,99	0,30	0,55	0,54	5	0,61	1,30	1,00
P18	4	19,00	17,93	0,31	0,54	0,57	5	0,59	1,20	0,80
P18	23	19,00	17,93	0,31	0,54	0,57	5	0,59	1,20	0,80
P5	9	16,00	15,04	0,31	0,56	0,55	5	0,60	0,40	0,40
P13	23	17,00	15,97	0,31	0,54	0,57	5	0,59	0,20	0,20
P17	14	21,00	20,05	0,31	0,60	0,52	5	0,62	0,50	0,30
P3	14	22,00	21,26	0,31	0,70	0,45	5	0,67	1,30	0,70
P2	19	21,00	20,04	0,32	0,60	0,53	5	0,62	1,20	0,90
P18	33	19,00	17,88	0,32	0,54	0,60	5	0,58	1,40	0,80
P12	15	15,00	14,08	0,32	0,58	0,55	5	0,60	1,10	1,00
P13	1	12,00	11,36	0,32	0,69	0,47	5	0,66	0,70	0,30
P13	33	17,00	15,92	0,32	0,54	0,60	5	0,58	0,30	0,30
P14	16	18,00	16,87	0,32	0,53	0,61	5	0,57	2,20	1,30
P12	28	19,00	17,85	0,33	0,54	0,61	5	0,57	0,10	0,10
P9	4	19,00	17,83	0,33	0,54	0,62	5	0,56	1,20	0,80
P9	23	19,00	17,83	0,33	0,54	0,62	5	0,56	1,20	0,80
P7	40	16,00	14,96	0,34	0,56	0,60	5	0,57	0,90	0,60
P15	6	20,00	18,86	0,34	0,55	0,61	5	0,57	0,70	0,60
P20	6	20,00	18,86	0,34	0,55	0,61	5	0,57	0,70	0,60
P11	33	19,00	17,81	0,34	0,54	0,63	5	0,56	0,60	0,40
P4	7	20,00	18,84	0,34	0,55	0,61	5	0,57	1,10	0,60
P18	28	20,00	18,84	0,34	0,55	0,62	5	0,56	1,10	0,60
P15	19	21,00	19,95	0,34	0,60	0,57	5	0,59	0,70	0,80
P20	19	21,00	19,95	0,34	0,60	0,57	5	0,59	0,70	0,80

P12	9	16,00	14,93	0,34	0,56	0,62	5	0,56	0,40	0,30
P18	22	16,00	14,93	0,34	0,56	0,62	5	0,56	3,60	2,70
P19	37	22,00	21,18	0,34	0,70	0,50	5	0,64	1,30	0,70
P11	39	18,00	16,80	0,35	0,53	0,65	5	0,54	1,10	1,70
P9	33	19,00	17,78	0,35	0,54	0,65	5	0,55	1,40	0,80
P7	7	20,00	18,81	0,35	0,55	0,63	5	0,56	1,10	0,60
P3	9	18,00	16,79	0,35	0,53	0,66	5	0,54	2,70	1,80
P7	10	16,00	14,91	0,35	0,56	0,63	5	0,56	0,40	0,40
P2	40	16,00	14,91	0,35	0,56	0,63	5	0,55	0,90	0,60
P8	24	19,00	17,76	0,35	0,54	0,66	5	0,54	0,40	0,40
P11	28	20,00	18,77	0,36	0,55	0,65	5	0,54	1,10	0,70
P6	19	21,00	19,89	0,36	0,60	0,60	5	0,57	0,70	0,80
P5	29	19,00	17,72	0,36	0,54	0,68	5	0,53	0,10	0,10
P18	8	22,00	21,13	0,36	0,70	0,52	5	0,62	0,90	0,40
P6	16	17,00	15,80	0,36	0,54	0,67	5	0,53	1,20	0,80
P2	7	20,00	18,75	0,37	0,55	0,66	5	0,54	1,10	0,60
P9	28	20,00	18,74	0,37	0,55	0,67	5	0,54	1,10	0,60
P13	13	20,00	18,73	0,37	0,55	0,67	5	0,53	0,40	0,30
P9	22	16,00	14,85	0,37	0,56	0,67	5	0,53	3,60	2,70
P10	11	21,00	19,85	0,37	0,60	0,63	5	0,56	1,40	4,00
P6	14	21,00	19,84	0,38	0,60	0,63	5	0,56	0,50	0,30
P11	9	17,00	15,74	0,38	0,54	0,71	5	0,51	1,20	1,30
P9	8	22,00	21,06	0,39	0,70	0,56	5	0,60	0,90	0,40
P17	31	18,00	16,64	0,39	0,53	0,74	5	0,50	0,50	0,30
P18	26	19,00	17,62	0,39	0,54	0,73	5	0,50	1,30	2,30
P10	13	22,00	21,03	0,40	0,70	0,57	5	0,59	1,30	0,70
P12	29	19,00	17,60	0,40	0,54	0,74	5	0,49	0,10	0,10
P18	30	20,00	18,63	0,40	0,55	0,72	5	0,50	0,10	0,10
P4	4	19,00	17,58	0,40	0,54	0,76	5	0,48	1,20	0,80
P18	27	19,00	17,57	0,41	0,54	0,76	5	0,48	1,20	0,80
P2	31	18,00	16,58	0,41	0,53	0,77	5	0,48	0,20	0,20
P18	24	20,00	18,58	0,41	0,55	0,75	5	0,49	0,70	0,60
P18	29	20,00	18,58	0,41	0,55	0,75	5	0,49	0,10	0,10
P17	4	19,00	17,55	0,41	0,54	0,77	5	0,47	1,20	0,80
P8	11	20,00	18,57	0,42	0,55	0,75	5	0,49	1,20	0,70
P18	17	18,00	16,56	0,42	0,53	0,79	5	0,47	3,40	2,90
P9	26	19,00	17,53	0,42	0,54	0,78	5	0,47	1,30	2,30
P16	31	19,00	17,52	0,42	0,54	0,79	5	0,47	1,20	0,80
P8	19	21,00	19,68	0,42	0,60	0,71	5	0,51	0,70	0,80
P3	30	21,00	19,68	0,43	0,60	0,71	5	0,51	0,60	0,40
P4	39	18,00	16,53	0,43	0,53	0,80	5	0,46	1,00	0,70
P9	27	19,00	17,48	0,43	0,54	0,81	5	0,45	1,20	0,80
P9	24	20,00	18,49	0,44	0,55	0,80	5	0,46	0,70	0,60
P5	1	13,00	12,02	0,44	0,65	0,68	5	0,53	0,30	0,20
P7	32	13,00	12,01	0,44	0,65	0,68	5	0,52	1,60	1,40
P9	17	18,00	16,47	0,45	0,53	0,84	5	0,44	3,40	2,90
P16	1	14,00	12,88	0,45	0,61	0,73	5	0,50	0,40	0,30
P3	36	22,00	20,88	0,45	0,70	0,65	5	0,54	1,30	0,70

P19	23	21,00	19,57	0,46	0,60	0,76	5	0,48	0,70	0,80
P8	38	16,00	14,59	0,46	0,56	0,83	5	0,44	0,40	0,40
P14	21	22,00	20,86	0,46	0,70	0,66	5	0,54	0,90	0,40
P12	6	20,00	18,41	0,46	0,55	0,83	5	0,44	0,70	0,60
P1	13	22,00	20,85	0,46	0,70	0,66	5	0,54	1,30	0,70
P19	9	19,00	17,37	0,46	0,54	0,87	5	0,43	2,40	1,50
P10	4	20,00	18,41	0,46	0,55	0,84	5	0,44	1,10	0,60
P10	23	20,00	18,41	0,46	0,55	0,84	5	0,44	0,70	0,60
P5	40	16,00	14,56	0,47	0,56	0,85	5	0,44	0,20	0,20
P10	14	22,00	20,80	0,48	0,70	0,69	5	0,52	1,30	0,70
P10	21	22,00	20,80	0,48	0,70	0,69	5	0,52	0,90	0,40
P5	10	16,00	14,52	0,49	0,56	0,88	5	0,42	0,20	0,20
P2	15	16,00	14,51	0,49	0,56	0,88	5	0,42	1,70	1,80
P15	32	13,00	11,92	0,49	0,65	0,75	5	0,49	1,40	0,80
P20	32	13,00	11,92	0,49	0,65	0,75	5	0,49	1,40	0,80
P11	16	18,00	16,29	0,50	0,53	0,93	5	0,39	0,60	0,50
P3	25	23,00	22,29	0,50	0,96	0,52	5	0,62	0,70	0,30
P12	40	16,00	14,46	0,51	0,56	0,91	5	0,40	0,20	0,20
P6	32	13,00	11,88	0,51	0,65	0,78	5	0,47	1,40	0,80
P3	37	22,00	20,70	0,52	0,70	0,74	5	0,49	1,30	0,70
P19	26	21,00	19,28	0,54	0,60	0,91	5	0,41	2,10	4,60
P1	14	22,00	20,61	0,54	0,70	0,78	5	0,47	1,30	0,70
P11	10	17,00	15,18	0,56	0,54	1,04	5	0,35	0,30	0,30
P13	2	11,00	10,15	0,57	0,76	0,75	5	0,49	0,30	0,10
P8	36	21,00	19,15	0,58	0,60	0,97	5	0,38	0,50	0,30
P5	28	20,00	17,98	0,59	0,55	1,06	5	0,34	0,10	0,10
P7	16	18,00	16,00	0,59	0,53	1,10	5	0,32	1,50	0,90
P1	15	17,00	15,10	0,59	0,54	1,09	5	0,33	2,50	2,70
P2	16	18,00	15,94	0,61	0,53	1,14	5	0,31	1,50	0,90
P1	26	20,00	17,88	0,61	0,55	1,11	5	0,32	1,40	3,10
P18	21	22,00	20,39	0,62	0,70	0,88	5	0,42	0,90	0,40
P11	23	20,00	17,86	0,62	0,55	1,12	5	0,31	0,10	0,10
P10	36	22,00	20,38	0,62	0,70	0,89	5	0,41	1,30	0,70
P1	17	19,00	16,80	0,63	0,54	1,18	5	0,29	1,50	1,50
P2	6	21,00	18,95	0,64	0,60	1,06	5	0,34	0,60	0,40
P8	37	21,00	18,94	0,64	0,60	1,07	5	0,33	0,50	0,30
P4	10	17,00	14,94	0,64	0,54	1,19	5	0,29	0,50	0,40
P9	21	22,00	20,31	0,64	0,70	0,92	5	0,40	0,90	0,40
P16	25	23,00	21,99	0,66	0,96	0,68	5	0,52	0,40	0,20
P11	22	17,00	14,88	0,66	0,54	1,23	5	0,27	3,10	2,30
P15	2	12,00	10,76	0,66	0,69	0,96	5	0,38	0,60	1,80
P20	2	12,00	10,76	0,66	0,69	0,96	5	0,38	0,60	1,80
P2	10	17,00	14,86	0,67	0,54	1,24	5	0,27	0,50	0,40
P11	15	17,00	14,83	0,68	0,54	1,26	5	0,26	4,00	3,40
P19	24	22,00	20,19	0,68	0,70	0,98	5	0,37	1,20	1,50
P10	35	22,00	20,19	0,68	0,70	0,98	5	0,37	1,30	0,70
P10	37	22,00	20,19	0,68	0,70	0,98	5	0,37	1,30	0,70
P1	36	22,00	20,18	0,68	0,70	0,98	5	0,37	1,30	0,70

P8	12	19,00	16,61	0,69	0,54	1,28	5	0,26	0,80	0,50
P4	19	22,00	20,13	0,70	0,70	1,00	5	0,36	0,40	0,30
P17	19	22,00	20,10	0,71	0,70	1,02	5	0,36	0,40	0,30
P4	14	22,00	20,08	0,72	0,70	1,03	5	0,35	1,30	0,70
P11	17	19,00	16,50	0,72	0,54	1,35	5	0,24	0,40	9,00
P12	12	19,00	16,46	0,73	0,54	1,36	5	0,23	0,90	0,50
P1	35	22,00	19,98	0,74	0,70	1,07	5	0,33	1,30	0,70
P1	37	22,00	19,98	0,74	0,70	1,07	5	0,33	1,30	0,70
P15	1	14,00	12,21	0,75	0,61	1,22	5	0,28	0,40	0,30
P20	1	14,00	12,21	0,75	0,61	1,22	5	0,28	0,40	0,30
P5	2	12,00	10,61	0,75	0,69	1,09	5	0,33	0,60	1,80
P12	16	18,00	15,45	0,76	0,53	1,43	5	0,21	1,50	0,90
P14	22	18,00	15,40	0,78	0,53	1,46	5	0,20	2,20	1,30
P16	22	18,00	15,37	0,79	0,53	1,48	5	0,20	2,20	1,30
P5	7	21,00	18,34	0,81	0,60	1,36	5	0,23	0,70	0,80
P4	32	14,00	12,03	0,83	0,61	1,36	5	0,23	2,00	1,40
P3	24	22,00	19,63	0,85	0,70	1,22	5	0,28	1,20	1,50
P3	29	22,00	19,63	0,85	0,70	1,22	5	0,28	0,40	0,30
P16	17	20,00	17,05	0,85	0,55	1,54	5	0,18	1,40	1,00
P15	16	19,00	15,85	0,92	0,54	1,71	5	0,15	1,40	0,80
P20	16	19,00	15,85	0,92	0,54	1,71	5	0,15	1,40	0,80
P1	8	23,00	21,32	0,94	0,96	0,98	5	0,37	0,40	0,20
P7	31	20,00	16,64	0,97	0,55	1,75	5	0,14	0,30	0,20
P3	17	21,00	17,59	1,02	0,60	1,71	5	0,15	0,80	0,90
P7	6	22,00	19,02	1,03	0,70	1,48	5	0,20	0,40	0,30
P11	40	19,00	15,23	1,12	0,54	2,08	5	0,09	1,90	1,80
P14	17	21,00	17,08	1,17	0,60	1,96	5	0,11	0,80	0,90
P19	30	23,00	20,24	1,32	0,96	1,37	5	0,23	0,90	0,80
P19	29	23,00	20,19	1,33	0,96	1,38	5	0,23	0,90	0,80
P11	2	14,00	10,98	1,37	0,61	2,25	5	0,07	0,20	0,20
P12	10	19,00	14,41	1,40	0,54	2,60	5	0,05	0,90	0,50
P11	32	19,00	12,20	2,29	0,54	4,28	5	0,01	1,40	0,80



GAZİLİ OLMAK AYRICALIKTIR..