



**YAZILI ANLATIM BECERİSİ PUANLAMA ANAHTARININ
GÜVENİRLİĞİNİN KLASİK TEST, GENELLENEBİLİRLİK VE
MADDE TEPKİ KURAMLARINA GÖRE İNCELENMESİ**

Merve Yıldırım Seheryeli

**YÜKSEK LİSANS TEZİ
EĞİTİM BİLİMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

MAYIS, 2018

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

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

YAZARIN

Adı: Merve

Soyadı: Yıldırım Seheryeli

Bölümü: Eğitimde Ölçme ve Değerlendirme

İmza:

Teslim tarihi:

TEZİN

Türkçe Adı: Yazılı Anlatım Becerisi Puanlama Anahtarının Güvenirliğinin Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarına Göre İncelenmesi

İngilizce Adı: An Examination of the Reliability Estimates of A Scoring Rubric of A Writing Skill Examination Using The Classical Test Theory, Generalizability Theory and The Item Response Theory Models

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduđumu, yararlandıđım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiđimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduđunu beyan ederim.

Yazar Adı Soyadı: Merve Yıldırım Seheryeli

İmza:

Jüri onay sayfası

Merve Yıldırım Seheryeli tarafından hazırlanan “Yazılı Anlatım Becerisi Puanlama Anahtarının Güvenirliğinin Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarına Göre İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Şeref Tan

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)

Başkan: Prof. Dr. Selahattin Gelbal

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi)

Üye: Doç. Dr. Nilüfer Kahraman

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)

Tez Savunma Tarihi: 31.05.2018

Bu tezin Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Selma YEL

Eğitim Bilimleri Enstitüsü Müdürü



Kardeřimle beni nice fedakârlıkla büyüten

Anneme ve Babama

TEŞEKKÜR

Yüksek lisans eğitimim sırasında, hem aldığım derslerde hem de tez yazım sürecimde engin bilgilerinden faydalanmamı sağlayan, akademisyenliğinin yanı sıra öğrencilerine karşı gösterdiği hakkaniyetini de örnek aldığım değerli danışmanım Prof. Dr. Şeref TAN'a en içten teşekkürlerimi sunuyorum.

Kocaman bir ailede bana da kucak açan adaleti, merhameti ve babalığını tek bir gün bile esirgemeyen, sorduğum her soruyu bana en güzel ders niteliğinde anlatan değerli öğretmenim Prof. Dr. Şener BÜYÜKÖZTÜRK'e; zorlandığım her alanda yoluma ışık tutan; tecrübelerini, hayat felsefesini istekle örnek aldığım, Kalyoncu ailemizin birleştiricisi ve güçlendiricisi değerli öğretmenim Prof. Dr. Yaşar ÖZBAY'a;

Çalışma hayatımla birlikte akademik bilgileri ile de beni destekleyen, tez yazım sürecimde değerli görüş ve önerilerini esirgemeyen öğretmenim Dr. Öğr. Üyesi Ufuk AKBAŞ'a; yine önerileri ile tez yazım sürecimi şekillendirmeme yardımcı olan Doç. Dr. Neşe GÜLER'e; yüksek lisans eğitimim boyunca bilgi ve tecrübelerimi geliştirmemi sağlayan öğretmenlerim Prof. Dr. Mehtap ÇAKAN, Doç. Dr. İsmail KARAKAYA, Doç. Dr. Nilüfer KAHRAMAN'a; tez jürimde bulunarak katkılarını esirgemeyen kıymetli öğretmenim Prof. Dr. Selahattin GELBAL'a;

Tüm bu süreçlerde yanımda olan Arş. Gör. Büşranur DURMAZ, Arş. Gör. Esra Eker, Arş. Gör. Nurgül ÇAKMAK ve Arş. Gör. Fatma ÇİLOĞLAN başta olmak üzere tüm akademik ailem Hasan Kalyoncu Üniversitesi Eğitim Fakültesi öğretim elemanlarına;

Hayatımın her anında bana inanan, güvenen, beni yüreklendiren ve her konuda bana rol model olan canım ailem, yerini dolduramayacağım babam Abdulvahap YILDIRIM; gücü ile bana güç katan annem Reşide YILDIRIM ve en kıymetli emanetim, biricik kardeşim Safa YILDIRIM'a; yorulduğumda bile bana inancıyla koşmamı sağlayan eşim Sinan SEHERYELİ'ye; desteklerini esirgemeyen anne yarım Dilek KAYMAZ ve Melek TUNÇ'a; kız kardeşlerim Bengisu TUNÇ ve Tuana TUNÇ'a teşekkür ediyorum.

**YAZILI ANLATIM BECERİSİ PUANLAMA ANAHTARININ
GÜVENİRLİĞİNİN KLASİK TEST, GENELLENEBİLİRLİK VE
MADDE TEPKİ KURAMLARINA GÖRE İNCELENMESİ
(Yüksek Lisans Tezi)**

**Merve Yıldırım Seheryeli
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Mayıs, 2018**

ÖZ

Bu araştırmanın amacı, tesadüfî hatayı en aza indirmek için çalışma alanlarına göre farklılaşan ölçme kuramlarından Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarının yazılı anlatım becerisi analitik puanlama anahtarı kullanılarak elde edilen ölçümlere ilişkin güvenirlik belirlemedeki yaklaşımlarının incelenmesidir. Betimsel olarak tasarlanan bu çalışmada uygun örnekleme yöntemi ile seçilen beş farklı sınıf seviyesinde toplam 523 öğrenci tarafından yazılan hikâyeler, puanlama anahtarındaki 11 ölçüte göre ve bir Türkçe öğretmeni, beş sınıf öğretmeni, bir ölçme değerlendirme uzmanından oluşan yedi puanlayıcı tarafından puanlanmıştır. Toplanan verilerin güvenirliğinin belirlenmesi için Klasik Test Kuramı'nda SPSS 22 programında puanlayıcıların puanlamaları arasındaki ilişki için Pearson Korelasyon, puanlayıcılar arasındaki uyum için Eta korelasyon, puanlayıcıların puanlamalarındaki iç tutarlılık için Cronbach Alpha katsayılarına bakılmıştır. Genellenebilirlik Kuramı'nda Edu-G 6.1e programı kullanılarak G ve Phi katsayıları hesaplanmıştır. Bu katsayılar her sınıf seviyesinde ve öğrencilerin tamamı analiz edilerek elde edilmiştir. Madde Tepki Kuramı'nda Multilog 7.03 programı kullanılarak Samejima'nın Derecelendirilmiş Tepki modeline göre madde ve test bilgi fonksiyonları ile marjinal güvenirlik katsayıları tek boyutluluk ve yerel bağımsızlık sayıltılarını sağlayan ilk beş puanlayıcı analiz edilerek hesaplanmıştır. Elde edilen güvenirlik kestirimleri arasındaki farkın manidarlığı için ise Microsoft Excel 2010 programında Fisher Z' istatistiği hesaplanmıştır. Klasik Test Kuramı bulgularında Pearson Korelasyon katsayıları incelendiğinde ilk beş puanlayıcı arasındaki katsayılar 0,90'ın üzerinde, altı ve yedinci puanlayıcıların diğer puanlayıcılarla arasındaki korelasyonların daha düşük seviyede (0,60-0,85 arasında) olduğu; Eta korelasyon katsayıları incelendiğinde puanlayıcıların öğrencileri her sınıf seviyesine ve tüm öğrencilere göre

puanlamada yüksek uyum gösterdikleri; Cronbach Alfa katsayılarında ise tüm sınıf seviyelerinde ve puanlayıcılarda katsayılar 0,85-0,95 arasında olup altıncı puanlayıcının puanlamalarının iç tutarlığının daha düşük olduğu (0,80-0,90 arasında) bulunmuştur. G ve Phi katsayıları incelendiğinde ise G katsayılarının 0,97'nin üzerinde Phi katsayılarının 0,95'in üzerinde olduğu görülmüştür. Genellenebilirlik Kuramı bulgularına göre, öğrenciler arasında beklenen farklılaşma ortaya çıkmış, ölçütlerin güçlük düzeyleri bir öğrenciden diğerine değişmemiş, puanlayıcılar arasındaki puanlama tutarlılığı mükemmel seviyede (toplam hata varyansını açıklama oranı 0,00) bulunmuştur. Karar çalışmaları katsayıları yükseltmek amacı yerine 0,80'in üzerinde kalacak şekilde ölçüt ve puanlayıcı sayıları düşürülerek gerçekleştirilmiştir. Beş ölçüt ve iki puanlayıcının olduğu karar çalışmalarında en düşük katsayılar 3. sınıflarda G katsayısı 0,83 ve Phi katsayısı 0,79 olarak elde edilmiştir. Madde Tepki kuramı bulgularına göre kestirilen a parametrelerinin (1,24 ile 6,44 arasında) yüksek düzeyde ayırt ediciliğe sahip olduğu görülmüştür. 1 ve 3. puanlayıcılara göre en fazla ayırt edici olan madde 8. ölçüt iken 2, 4 ve 5. puanlayıcılara göre en fazla ayırt edici olan 9. ölçüttür. Ayırt ediciliği en az olan madde ise tüm puanlayıcılara göre 11. ölçüt olarak bulunmuştur. b parametrelerinde ise tüm puanlayıcılar için bireylerin 0,50 ihtimalle 0 kategorisinden daha yüksek puan almaları için en düşük -2,35 yetenek düzeyinde, 1 kategorisinden daha yüksek puan almaları için en düşük -0,80 yetenek düzeyinde, 2 kategorisinden daha yüksek puan almaları için en düşük 0,41 yetenek düzeyinde bulunmaları gerekmektedir. θ yetenek düzeylerinde ise 1. puanlayıcıya göre 3 ve 4. maddeleri; 2, 3, 4 ve 5. puanlayıcılara göre 4. maddeyi gruba göre en düşük yetenek düzeyindeki bireylerin, tüm puanlayıcılara göre 11. maddeyi gruba göre en yüksek yetenek düzeyine sahip bireylerin doğru yanıtlaması beklenmektedir. Test bilgi fonksiyonlarında ise bilgi miktarı arttıkça standart hata azalmaktadır. -1,00 ile 1,50 arasında yeteneğe sahip olan bireyler için ölçme sonuçları daha az hata ile kestirilmiştir. Bu aralıktaki θ değerlerinden uzaklaşıldıkça ölçme sonuçlarının içerdiği hata miktarı da artmaktadır. Marjinal güvenilirlik katsayıları incelendiğinde ise güvenirliliğin oldukça yüksek (0,93 civarında) olduğu görülmüştür. Tüm güvenilirlik kestirimlerinin arasındaki farkın manidarlığı için, sınıf düzeylerine ve tüm öğrencilere ait KTK'de Eta korelasyon katsayıları, Cronbach Alpha katsayılarının ortancaları ile GK'de G ve Phi katsayıları için Fisher Z' testi sonuçlarında α ile G, α ile Phi, G ve Eta, Phi ve Eta katsayıları arasında 0,05 düzeyinde anlamlı bir fark bulunmuştur. Beş puanlayıcıya ait KTK'deki Cronbach Alpha katsayıları ile MTK'deki marjinal güvenilirlik katsayıları arasında 0,05 anlamlılık düzeyinde anlamlı bir fark olmadığı tespit edilmiştir. Sonuç olarak hata varyansı kaynaklarını açıklamada ve güvenilirlik belirlemede Genellenebilirlik Kuramı, madde bazında hata kestirimlerinde ve yetenek düzeyi belirlemede Madde Tepki Kuramı, Klasik Test Kuramı'ndan daha ayrıntılı bilgiler ortaya koymuştur. Puanlayıcılar arası güvenirlilikte ise KTK ile GK'ye göre kestirilen parametreler arasında anlamlı bir fark ($p < 0,05$) bulunurken; KTK ile MTK'ye göre kestirilen parametreler arasında anlamlı bir fark ($p < 0,05$) saptanmamıştır.

Bilim kodu:

Anahtar Kelimeler: Klasik test kuramı, genellenebilirlik kuramı, madde tepki kuramı, güvenilirlik, hata varyansı kaynakları, puanlayıcılar arası uyum.

Sayfa Adedi: xviii + 164

Danışman: Prof. Dr. Şeref Tan



**AN EXAMINATION OF THE RELIABILITY ESTIMATES OF A
SCRORING RUBRIC OF A WRITING SKILL EXAMINATION
USING THE CLASSICAL TEST THEORY, GENERALIZABILITY
THEORY AND THE ITEM RESPONSE THEORY MODELS
(M.S Thesis)**

Merve Yıldırım Seheryeli

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

May, 2018

ABSTRACT

The aim of this research is to evaluate the performance of a writing scale rubric and to compare the reliability of the resulting test scores using the Classical Test Theory, Generalizability Theory and the Item Response Theory, each taking a different approach to minimizing random errors. In this study, stories written by 523 students at five different grade levels who were selected by convenient sampling were scored by seven raters including a Turkish Language teacher, five classroom teachers, and a measurement expert using a 11-criteria scoring rubric. Pearson Correlation was calculated to find interrater correlations to determine the reliability of the collected data based on the Classical Test Theory, Eta correlation was used to determine rater reliability and finally Cronbach Alpha coefficient was used to determine the overall internal consistency using SPSS 22 program. G and Phi coefficients were calculated by using Edu-G 6.1e program for Generalizability Theory. These coefficients were computed for each grade level. As for Item Response Theory, based on Samejima's Scaled Response Model, the Multilog 7.03 program was used to estimate item parameters, test information functions and marginal reliability coefficients using ratings data of the first five raters which satisfied the assumptions of unidimensionality and local independence. Fisher Z 'statistic was calculated using the Microsoft Excel 2010 program to test the significance of the differences observed between the resulting reliability estimates. When the Pearson Correlation coefficients of the Classical Test Theory were examined, it was found that the coefficients between the first five scorers were above 0.90 and that the correlations between the sixth and seventh raters were lower (between 0.60-0.85) when compared with the other raters. Eta correlation

coefficients showed that the raters had a high level of consistency in scoring according to each grade level and the students; Cronbach's alpha coefficients of raters were between 0,85 and 0,95 within grade levels; the ratings of the sixth rater being the lowest (between 0,80 and 0,90). As for G and Phi coefficients, G coefficients were over 0.97 while the Phi coefficients were over 0.95. According to the findings of Generalizability Theory, the expected differentiation occurred among the students, the difficulty levels of the criteria were not changed from one student to the other, the scoring consistency between the raters was perfect (the explanation ratio of the total error variance was 0.00). Decision studies were carried out by reducing the number of raters and criteria in order to keep the coefficients above 0.80 instead of the increasing the coefficients. The lowest coefficients in the decision studies with five criteria and two raters were obtained in Grade 3 with a G coefficient of 0.83 and a Phi coefficient of 0.79. It was seen that the a parameters estimated from the Item Response Theory findings (between 1.24 and 6.44) had a high level of discrimination. Item 8 was the most distinctive item with respect to raters 1 and 3, while item 9 was the most distinctive according to raters 2, 4, and 5. The item with the least distinctiveness was the 11th criterion according to all the raters. For b-parameters, for all raters should be in at least -2.35 skill level to get a higher score than category 0; at least -0.80 skill level to score higher than category 1; at least 0.41 skill level to score higher than category 2. It was expected that the individuals with the lowest ability level would give the correct answers in 0 levels to 3rd and 4th items according to the 1st rater; the 4th item according to the 2nd, 3rd, 4th and 5th raters. The individuals with the highest ability level were expected to answer item 11 correctly according to all the raters. In test information functions, the standard error decreases as the amount of information increases. For individuals who had the ability between -1.00 and 1.50, the measurement results were estimated with fewer errors. As the distance from the θ values in this range decreases, the error amount of the measurement results increases. The reliability was very high in terms of the marginal reliability coefficients (around 0.93). For the purpose of examining the significant difference between the reliability estimates, a significant difference was found at the level of 0,05 between the coefficients of α and G, α and Phi, G and Eta, Phi and Eta as a result of Fisher Z' test with G and Phi in GT and Eta correlation and Cronbach Alpha coefficients in CTT, It was determined that there was no significant difference between the Cronbach Alpha coefficients of the five scorers in the CTT and the marginal reliability coefficients in the IRT at the significance level of 0.05. As a result, while Generalizability Theory can be used to explain sources of error variance and to determine reliability, Item Response Theory and Classical Test Theory are better in terms of error estimates in the item level and determining the ability level. There was a significant difference ($p < 0,05$) between the estimated parameters according to CTT and GT in the interrater reliability; whereas there was no significant difference ($p < 0,05$) between the estimated parameters according to CTT and IRT.

Science Code:

Key Words: Classical test theory, generalizability theory, item response theory, reliability, source of variance, interrater reliability.

Page Number: xviii + 164

Supervisor: Prof. Dr. Şeref TAN



İÇİNDEKİLER

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
ÖZ.....	vi
ABSTRACT.....	ix
İÇİNDEKİLER	xii
TABLolar LİSTESİ.....	xv
ŞEKİLLER LİSTESİ	xvii
SİMGELER VE KISALTMALAR LİSTESİ.....	xviii
BÖLÜM I.....	1
GİRİŞ	1
1.1. Problem Durumu	1
1.1.1. Ölçmede Hata Türleri.....	5
1.1.2. Ölçümlerde/Ölçme Araçlarında Bulunması Gereken Nitelikler	7
1.2. Araştırmanın Amacı	10
1.3. Araştırmanın Önemi.....	11
1.4. Problem Cümlesi.....	11
1.5. Araştırma Soruları.....	12
1.6. Tanımlar	13
1.7. Sınırlılıklar.....	13
BÖLÜM II	15
İLGİLİ ARAŞTIRMALAR ve KURAMSAL ÇERÇEVE.....	15
2.1. İlgili Araştırmalar	15

2.1.1. Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar	15
2.1.2. Klasik Test ve Genellenebilirlik Kuramlarını Karşılaştıran Araştırmalar	16
2.1.3. Klasik Test ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar	20
2.1.4. Genellenebilirlik ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar	30
2.2. Kuramsal Çerçeve.....	32
2.2.1. Klasik Test Kuramı.....	32
2.2.2. Genellenebilirlik Kuramı	38
2.2.3. Madde Tepki Kuramı	49
BÖLÜM III.....	67
YÖNTEM.....	67
3.1. Araştırma Deseni	67
3.2. Çalışma Grubu	67
3.3. Veri Toplama Araçları	69
3.3.1. 3-7. Sınıflar İçin Kullanılan (Hikâyeler) Formlar.....	69
3.3.2. Yazılı Anlatım (Hikâye Yazma) Becerisi Değerlendirme Formu	69
3.4. Veri Analizi.....	70
BÖLÜM IV	73
BULGULAR VE YORUM.....	73
4.1. Birinci Araştırma Sorusuna Yönelik Bulgular	73
4.2. İkinci Araştırma Sorusuna Yönelik Bulgular	79
4.3. Üçüncü Araştırma Sorusuna Yönelik Bulgular	88
4.4. Dördüncü Araştırma Sorusuna Yönelik Bulgular	96
BÖLÜM V.....	101
SONUÇ, TARTIŞMA ve ÖNERİLER.....	101
5.1. Sonuç ve Tartışma.....	101

5.1.1. Birinci Alt Problem.....	101
5.1.2. İkinci Alt Problem.....	104
5.1.3. Üçüncü Alt Problem.....	106
5.1.4. Dördüncü Alt Problem	109
5.2. Öneriler	110
KAYNAKLAR	112
EKLER.....	120
EK 1.Yazılı Anlatım Becerisi Puanlama Anahtarı	121
EK 2a. Form 1: 3. sınıf hikâye yazma	123
EK 2b. Form 2: 4. sınıf hikâye yazma	124
EK 2c. Form 3: 5. sınıf hikâye yazma	125
EK 2d. Form 4: 6. sınıf hikâye yazma	126
EK 2e. Form 5: 7. sınıf hikâye yazma	127
EK 3. Verilerin dağılımına ilişkin histogram grafikleri	128
EK 4. Her bir puanlayıcıya ait tüm veriler için yamaç-birikinti grafikleri	130
EK 5. GRM için MULTİLOG komut dosyası örneği.....	134
EK 6. Yazılı anlatım becerileri puanlama anahtarının madde parametreleri kestirimi MULTİLOG çıktısı örneği – 1. Puanlayıcı.....	135
EK 7. Yazılı anlatım becerileri puanlama anahtarının madde kategorilerine ait beklenen ve gözlenen oranlar ve aralarındaki farklar	145
EK 8. Öğrencilerin ilk beş puanlayıcıya göre kestirilen yazılı anlatım becerisi düzeyleri (Theta Yetenek Düzeyleri).....	148
EK 9. Maddelere ait karakteristik eğrileri ve bilgi fonksiyonları örneği-1. puanlayıcı	160

TABLÖLAR LİSTESİ

Tablo 1. Puanlama Güvenirliğindeki Farklı Durumlarda Kullanılabilecek İstatistikler	36
Tablo 2. Sıklıkla Kullanılan Güvenirlik Belirleme Yöntemleri ve Hata Kaynakları	39
Tablo 3. İki Yüzeyli Çaprazlanmış ve Yuvalanmış Desene Göre Puan Tablosu	46
Tablo 4. İki yüzeyli ($b \times m \times p$) Çaprazlanmış Desene İlişkin Değişkenlik Kaynağı, Serbestlik Dereceleri, Değişkenlik türü, Kestirilen Varyans Bileşenleri ve Kareler Ortalamaları	46
Tablo 5. İki Yüzeyli ($b \times (m : p)$) Yuvalanmış Desene İlişkin Değişkenlik Kaynağı, Serbestlik Dereceleri, Değişkenlik türü, Kestirilen Varyans Bileşenleri ve Kareler Ortalamaları	47
Tablo 6. Çok Kategorili Puanlanan Maddeler İçin Samejima'nın Teorik Çerçevesi.....	62
Tablo 7. Örneklem: Öğrenci Sayıları	68
Tablo 8. Puanlayıcılar İle İlgili Bilgiler	68
Tablo 9. Yedi Puanlayıcının Sınıf Seviyelerine Göre Öğrencilere Verdikleri Toplam Puanlara Ait Betimsel İstatistikler.....	74
Tablo 10. Yedi Puanlayıcının Sınıf Seviyelerine Göre Öğrencilere Verdikleri Toplam Puanlar Arasındaki Pearson Korelasyon Katsayıları	76
Tablo 11. Yedi Puanlayıcının Test Puanlarının Tutarlıklarına Ait Eta Korelasyon Katsayıları.....	77
Tablo 12. Her Bir Puanlayıcının Sınıf Seviyelerine ve Toplam Öğrenci Sayısına Göre Cronbach α İç Tutarlılık Katsayıları	78

Tablo 13. Öğrencilerinin <i>G</i> Çalışması Sonucunda Kestirilen Varyansları ve Toplam Varyansı Açıklama Yüzdeleri	80
Tablo 14. Öğrencilerin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin <i>K</i> Çalışması Sonucu Elde Edilen <i>G</i> ve <i>Phi</i> Katsayıları	83
Tablo 15. Tüm Öğrenciler İçin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin Beş Puanlayıcının Olduğu Karar Çalışmaları İle İlk Beş Puanlayıcının Sabit Tutulduğu Karar Çalışmalarının Sonuçları.....	87
Tablo 16. Tüm Öğrenciler İçin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin 6 Puanlayıcının Sabit Tutulduğu Karar Çalışmaları İle 6. Puanlayıcı Dışındaki 6 Puanlayıcının Sabit Tutulduğu Karar Çalışmalarının Sonuçları	88
Tablo 17. Yedi Puanlayıcıya Ait Verilerin Tek Boyutluluk Sayıltısı İçin TBA Sonuçları	89
Tablo 18. Yazılı Anlatım Becerileri Puanlama Anahtarının Maddelerine Ait Ayırt Edicilik İle Adım Fonksiyonları Parametreleri Ve Standart Hataları	92
Tablo 19. Yazılı Anlatım Becerileri Puanlama Anahtarındaki Maddelerin Yetenek Parametreleri.....	93
Tablo 20. Beş Puanlayıcıya Ait Marjinal Güvenirlik Katsayıları	96
Tablo 21. Sınıf Seviyeleri ve Toplamdaki Öğrenciler İçin α (Ortancaları), <i>Eta</i> , <i>G</i> ve <i>Phi</i> Katsayıları.....	97
Tablo 22. Dört Katsayıya Ait Fisher <i>Z'</i> Testi Sonuçları.....	98
Tablo 23. Puanlayıcılar İçin α ve Marjinal Güvenirlik Katsayıları	99
Tablo 24. İki Katsayıya Ait Fisher <i>Z'</i> Testi Sonuçları	99

ŞEKİLLER LİSTESİ

Şekil 1. Bilimde kuram ile deneysel yapı arasındaki ilişki	4
Şekil 2. Tek yüzeyli çaprazlanmış ve yuvalanmış desen gösterimi	45
Şekil 3. Madde karakteristik eğrisi	50
Şekil 4. b parametreleri farklı a = 1,2 olan üç maddeye ait MKE.....	52
Şekil 5. b parametreleri farklı a = 1 olan üç maddeye ait MKE.....	53
Şekil 6. b parametreleri aynı a parametreleri farklı olan üç maddeye ait MKE.....	54
Şekil 7. $\theta = 1$ yetenek düzeyinde mükemmel ayırt ediciliğe sahip maddeye ait MKE.....	55
Şekil 8. Madde bilgi fonksiyonu	57
Şekil 9. Test bilgi fonksiyonu.....	58
Şekil 10. İki kategorili ve dört kategorili puanlanan maddelere ait madde bilgi fonksiyonları	59
Şekil 11. Dört kategorili bir maddedeki adımlar	60
Şekil 12. Dört kategorili puanlanan bir maddeye ait adım fonksiyonları	61
Şekil 13. G ve Phi katsayılarındaki değişimler	86
Şekil 14. Beş puanlayıcıya ait test bilgi fonksiyonları	95

SİMGELER VE KISALTMALAR LİSTESİ

KTK	Klasik Test Kuramı
GK	Genellenebilirlik Kuramı
MTK	Madde Tepki Kuramı
ÇBMTK	Çok Boyutlu MTK
ÇDKRM	Çok Değişkenlik Kaynaklı Rasch Modeli
α	Cronbach Alfa Katsayısı
$b \times o \times p$	birey $\times$ ölçüt $\times$ puanlayıcı
MKE	Madde Karakteristik Eğrisi
DTM	Derecelendirilmiş Tepki Modeli
TBA	Temel Bileşenler Analizi
SH	Standart Hata

BÖLÜM I

GİRİŞ

Bu bölümde araştırmanın çıkış noktası olan problem durumu, problem cümlesi ve araştırma soruları; araştırmanın amacı, önemi ve sınırlılıkları ile birlikte bazı tanımlar verilmiştir.

1.1. Problem Durumu

İnsanların varoluşunu, gelişimini ve bu gelişim sürecindeki davranışları anlamlandırma çalışmaları yüzyıllardır araştırmalara konu olmuştur. Bilinmeyeni ortaya çıkarmak, kesin olmayanı kanıtlamak üzerine yapılan bu çalışmalar araştırmacılara ışık tutarak bilimin gelişmesine katkı sağlamıştır. Büyüköztürk (2017)'e göre bilim dünyayı tanımak, anlamak ve gerçeği bulmaktır; amacı ise kontrol etmek, yordamak, açıklamak ve anlamaktır (s. 6-14).

Bilimin gelişmesine katkı sağlayan araştırmaların konusu olan insan ve onun davranışlarını incelemek ilk olarak psikoloji bilimine aittir. Önceleri sadece görülebilen ya da işitilebilen vücut hareketleri davranış olarak kabul edilse de günümüzde doğrudan ya da dolaylı yollarla ölçülebilen her şey davranış olarak nitelendirilmektedir. Diğer bir deyişle davranış; gözlenebilen vücut hareketleri ile birlikte hisleri, ilgi ve tutumları, zihinsel süreçleri de içine alan gözlenemeyen tüm içsel olayları kapsamaktadır (Morgan, 2013, s. 6). O hâlde

bireylerin tüm bu davranışlarını anlamlandırma, bireyi tanıma ve birey hakkında karar verme sürecinde ölçme ve değerlendirme temel ihtiyaçlardan biri olmaktadır.

İnsan ve davranışlarını inceleyen diğer bir alan da eğitim bilimleridir. Eğitimin amacı bireyleri; küreselleşen dünyanın ihtiyaçlarına yanıt verecek şekilde yetiştirmektir. Bu ihtiyaçları karşılayabilecek nitelikteki kişilerin sorunları çözebilen, sorgulayan, olaylara eleştirel gözle bakabilen, araştırma-geliştirme becerisine sahip, üst düzey düşünebilen ve üretebilen bireyler olması gerekmektedir. Bu amaç doğrultusunda değişen eğitim politikaları beraberinde ölçme ve değerlendirme yaklaşımlarında da değişiklikler getirmiştir. Bu değişiklikler, bilginin varlığını sorgulamak yerine ne derece var olduğunu ortaya koyacak, üst düzey düşünme becerilerini belirleyebilecek değerlendirme çalışmalarının kullanılmasını arttırmıştır. Bu nedenle bireylerin yalnız (çoktan seçmeli, boşluk doldurma, eşleştirme, vb. madde tipi barındıran) testlerle değil gerçek yaşam durumlarındaki düşünme becerilerini, performanslarını ortaya koyabilecekleri değerlendirme yaklaşımları kullanılması daha önemli hâle gelmiştir (Kutlu, Doğan & Karakaya, 2014, s. 3-4).

Kutlu, Doğan & Karakaya (2014)'ya göre üst düzey düşünme becerileri; bireylerin bilişsel, duyuşsal ve psikomotor özelliklerinden birkaçını aynı anda ortaya koymalarını gerektiren yetenekleri sergilemeleri olarak belirtilse de (s. 15-16) alanyazında üst düzey düşünme becerileri ile ilgili ortak bir tanım bulunmamakla birlikte genellikle taksonomilerin üst basamakları bu beceriyi tanımlamak için kullanılmıştır. Tüm taksonomilere referans olan 1956 Bloom (ve 2010 Revize Bloom) Taksonomisine göre üst düzey düşünme becerileri bilgi basamağının üstünde olan kavrama, uygulama, analiz, sentez, değerlendirme olarak belirtilmiş olsa da farklı sınıflamalarda üst düzey düşünme becerileri olarak farklı tanımlamalar yapılmıştır. Zoller (1993) üst düzey bilişsel beceriler olarak soru sorma, eleştirel ve sistematik düşünme, problem çözme, analiz, değerlendirme, yeni bilgiler sentezleme ve karar verme yeteneklerini belirtmiştir.

SOLO (1982) Taksonomisinde ise yapı öncesi, tekli ve çoklu yapısal basamaklarının üstünde olan ilişkilendirme ve soyutlama üst düzey beceriler olarak ele alınmaktadır. Dettmer Taksonomisinde bilmek ve kavramak fazlarının üstünde olan uygulamak, analiz etmek, değerlendirmek, sentezlemek, düşünmek, yaratmak fazları üst düzey düşünme becerilerini tanımlamaktadır (Arı, 2013). Marzano ve Kendall (2007) ise geri getirme ve anlama basamaklarının üstünde olan analiz, bilgiyi genelleme, üst bilişsel sistem ve öz-sistem basamakları ile üst düzey düşünme becerilerini tanımlamışlardır (s. 66). Haladyna (1997) da bu süreçleri anlama dışındaki problem çözme, eleştirel düşünme ve yaratıcılık olarak belirtmiştir. Alınan uzman görüşleri doğrultusunda bu çalışmada bahsedilen üst düzey düşünme becerileri de Haladyna'nın görüşleri ile paralellik gösterecektir.

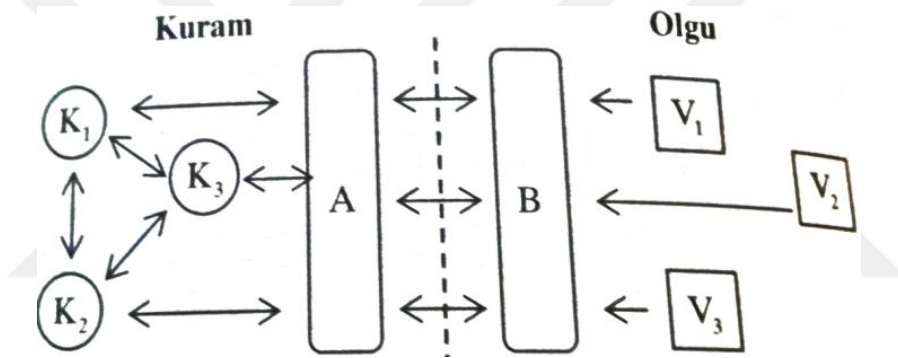
Bireylerin üst düzey becerilerini ortaya koymalarına yardımcı olan birçok araç bulunmasına rağmen soyut durumdaki düşünceleri somutlaştırarak incelenebilir hâle dönüştüren en önemli araç yazma ya da yazılı anlatım becerisidir. Yazma, "düşünme üzerine düşünme" olarak tanımlanmakta ve bireylerin bilgiyi düzenleyerek düşüncelerini genişletmelerini sağlamaktadır (Karatay, 2015, s. 21-22).

Öğretim sürecinde ortak bir yazma yaklaşımının olmaması, öğrencilerinin yazma becerilerinin gelişim sürecinin takip edilmesini de güçleştirmektedir. Bu durumun bir sonucu olarak da öğretmenler tarafından öğrencilerin eksiklerinin belirlenip geri bildirim verilmesi mümkün olmamakta, öğrencilerin de kendi eksiklerinin farkında olmalarının önüne geçilmektedir. Gelişmiş ülkelerde yazılı anlatım becerilerinin detaylı olarak ölçülmesi için çalışmalar yapılırken ülkemize henüz öğrencilerin bu yöndeki eksikliklerinin belirlenmesi üzerine ortak bir çalışma yapılmamaktadır (Karatay, 2015, s. 38-39). Bu nedenle, yazılı anlatım (hikâye yazma) becerilerinin değerlendirilmesi bu çalışmanın da konusunu oluşturmaktadır.

Öğrencilerin tüm bu becerileri hakkında doğru karar verebilmek için doğru bir değerlendirme, doğru bir değerlendirme için de geçerli ve güvenilir bir ölçme yapmak

gerekmektedir. Ölçme, gözlemlerin ve deneylerin herkes tarafından kabul edilebilir olmasını ayrıca elde edilen sonuçların kesin, açık ve doğru yansıtılmasını sağlar.

Torgerson'dan aktaran Baykul (2010) bilimi, kuramsal yapı ile deneysel verilerin birlikte içinde bulunduğu bir sistem olarak anlatmıştır. Bu sistemin elemanları arasındaki ilişkiler, bu ilişkilerin genellenmesi ve genellemelerin kanunlaşması ölçme yoluyla sağlanmaktadır. Bu sistemin (bilimin) kuramsal tarafında; K_1 , K_2 ve K_3 kuramsal yapının elemanları olmak üzere aşağıda Şekil 1'de oklarla gösterilen ilişkilerden A yargısına varılmış olsun. Benzer şekilde sistemin olgusal tarafında V_1 , V_2 ve V_3 verilerinden çıkarımlar yapılarak B bulgusuna ulaşılmış olsun. Aşağıdaki şekilde bu kuramsal ve deneysel yapılar arasındaki ilişki gösterilmiştir:



Şekil 1. Bilimde kuram ile deneysel yapı arasındaki ilişki (Torgerson, 1958, s. 10)

Şekil 1'de görüldüğü gibi A vargıları ile B bulguları arasındaki ilişki de dâhil olmak üzere ok işareti ile gösterilen tüm ilişkilerin doğruluğu yapılan ölçmelerin hatasızlığı ile doğru orantılıdır. O hâlde, bilime ve teoriye katkı sağlayabilmek için geçerli ve güvenilir ölçmelerin yapılması gerekir (Baykul, 2010, s. 95-97).

Crocker ve Algina'ya (2006) göre ölçme, bireye ait birtakım özelliklere sayısal değerler verme işidir (s. 3-5). Yaşar (2014) ise Magnusson'un ölçmeye bakışını "Geçerli görgül yollarla test edilebilecek kurallar çerçevesinde nesnelere belli özelliklere sahip oluş derecelerine göre sayılar ve semboller vermektir." şeklinde aktarmıştır. Değerlendirme ise bu ölçme sonuçlarını daha önce belirlenen bir ölçüt ya da ölçütler grubu ile kıyaslayarak birey ya da nesnelerin ölçülen özellikleri hakkında karar verme sürecidir (Turgut &

Baykul, 1992, s. 17-29; Baykul, 2010, s. 97-98). Popham (2000)'a göre ise eğitimde ölçme ve değerlendirme; öğrencilerin bilgi, beceri ya da duygusal durumu hakkında çıkarımlar yapmak için eğitimcilerin, özel olarak yaratılan veya doğal olarak oluşan uyaranlara öğrenciler tarafından verilen cevapları kullandıkları bir süreçtir (s. 3-4).

Yukarıda bahsedilen açıklamalar kapsamında bireyin davranışlarının ölçülmesi için iyi tanımlanmış bir psikolojik yapı (operasyonel tanım) ve bu yapıyı ortaya koyacak bir ölçme süreci gerekmektedir. Yapılan ölçme işlemlerindeki mükemmellik arayışı sonucu alanımızı sorguladığımızda karşılaştığımız 5 temel sorunu Crocker ve Algina (2006) aşağıdaki gibi belirtmiştir:

1. Bir özelliği ölçmek için birden fazla yaklaşım bulunabilir. Bu durum ölçmeyi yapan kişilere göre farklılaşmaktadır.
2. Yapıları ölçmek için belirlenen davranış örnekleri, tüm davranışların bir alt kümesidir.
3. Yapılan tüm ölçmelerde hata yapma ihtimali vardır.
4. Ölçmede iyi tanımlanmış bir ölçme birimi yoktur.
5. Psikolojik yapıyı yalnızca operasyonel olarak tanımlamak yeterli değildir. Bu yapının diğer yapılarla ya da ölçülebilir fenomenlerle deneysel olarak ilişkisi de gösterilmelidir.

(Crocker & Algina, 2006, s. 5-7)

Bu ölçme sorunları incelendiğinde hata kaynaklarının ölçülen yapıdan, ölçmeyi yapan kişiye, ölçme aracından ölçme işleminin kendisine kadar çeşitlendiği görülmektedir. Bu hata kaynakları ve ölçme işlemine karışan hatalar ne kadar bilinirse yapılan ölçme kesinliğe o kadar yaklaşır; fakat hiçbir zaman kesinliğinden söz etmemiz mümkün olamamaktadır.

1.1.1. Ölçmede Hata Türleri

Yukarıdaki bölümde bahsedilen bu hatalar; sabit, sistematik ve tesadüfi hata türleri olarak sınıflandırılmaktadır.

Sabit Hatalar

Tekrarlı ölçümlerde sonucu deęiřtirmese bile ölçme sonuçlarını sabit bir şekilde etkileyen hatalardır. Hata sebebi belli olduęu için giderilmesi mümkündür. Örneęin; bir sınavdan en yüksek 95 puan alan kişilerin bulunduęu bir sınıfta öğretmen herkesin puanını 5 puan arttırması sabit bir hatadır. Hata kaynaęı olarak her öğrencinin aldığı puan aynıdır. Sabit hatalar mutlak deęerlendirme yapılacaęı durumlarda bireyler hakkında verilen kararların doğruluęunu (güvenirlilik) etkilerken baęlı deęerlendirme yapılacaęı durumlarda göreceli sıralamayı deęiřtirmedięinden ölçme sonuçlarını da deęiřtirmez.

Sistematiik Hatalar

Hem tekrarlı ölçümlerde sonucu deęiřtiren hem de aynı ölçümde farklı kişilere üstünlük saęlayan hatalardır. Yanlılık olarak da nitelendirilebilen bu hata türünde de hata sebebi belli olduęu için giderilmesi mümkündür. Örneęin; bir önceki hata (her öğrencinin puanını 5 puan arttırmak) yerine öğretmenin her öğrencinin puanını %5 arttırdığını kabul edelim. Bu durumda tüm sınıf aynı puanı almadığı sürece öğrencilerin aldığı ek puanlar farklılaşacaktır. Ayrıca bu puan artışı yüksek puan alanlara üstünlük saęlamaktadır. O hâlde herkesin puanını %5 arttırmak sistematiik bir hatadır. Sistematiik hatalar bu nedenle ölçümlerin daha çok geçerliğini etkiler.

Tesadüfi Hatalar

Ölçme sonuçlarını tesadüfi olarak etkiler. Bireyin ölçülen özellięi deęiřmemesine rağmen ölçme sonuçlarında farklılıklar gözlenebilir. Bu farklılıklar hata kaynaęı, ne derece ve hangi yönde etki ettięi bilinemedięi için giderilemez. Örneęin; öğrencilerin sınav kâğıtlarını deęerlendiren bir öğretmenin yorgun olması sebebiyle öğrencilerin yanıtlarını dikkatsizce okuması tesadüfi hatadır. Tesadüfi hatalar ölçümlere pozitif ya da negatif

yönde karışabildiği için tekrarlı ölçümlerin sayısı arttıkça hata puanlarının ortalaması sıfıra yaklaşır (Özbek, 2014, s. 46; Tan, 2014, s.135-139). Bu sebeple “Hata teorileri (test teorileri/kuramları) tesadüfi hatalar üzerine kurulmuştur.” (Baykul, 2010, s. 118).

Görüldüğü gibi bazı hata kaynakları bilindiği için giderilebilirken bazıları bilinemediği için hatalar giderilememektedir. Buna rağmen hataların en aza indirgenebilmesi ve ölçümlerin ideal kestirimlerine ulaşabilmek için ölçümlerin ve ölçme araçlarının bazı özelliklere sahip olması gerekmektedir. Bu özellikler: Geçerlik, güvenirlik ve kullanışlılıktır.

1.1.2. Ölçümlerde/Ölçme Araçlarında Bulunması Gereken Nitelikler

Bu kısımda geçerlik, güvenirlik ve kullanışlılık kavramlarından bahsedilmiştir.

Geçerlik

Geçerlik; ölçülmek istenen özelliğin tam ve doğru olarak ölçülebilme, diğer bir ifadeyle ölçme aracının geliştirilme amacına hizmet etme derecesidir (Tekin, 2015, s. 42). Köse (2014)’ye göre geçerlik, bir ölçme aracından elde edilen puanlara göre yapılan yorumların çıkarımların, test puanlarının dayanak olduğu çeşitli kararlar için uygun olma derecesidir (s.66-68). Geçerlik katsayısı hatasız ölçümler yapıldığı varsayıldığında +1 olur. Fakat böyle ideal bir durum mümkün değildir. Aynı zamanda geçerlik yapılan ölçmenin amacına göre de farklılaşabilmektedir. Bir ölçme aracı başka bir amaçla kullanıldığında uygun geçerlikte olmayabilir. Bu durumda tek bir geçerlikten bahsetmek yerine geçerlik türlerine değinmek gerekmektedir (Baykul, 2010, s. 224; Crocker & Algina, 2006, s. 217; Tekin, 2015, s. 43).

Kapsam geçerliliği, testteki davranışların ölçülmek istenen davranışlar evrenini ne derece yansıttığıdır. Testi yapan kişinin teste girenlerin sonuçlarından testin kendisinde var olan büyük alan maddelerine çıkarım yapmak istemesi durumunda kullanılır (Crocker & Algina, 2006, s. 218; Tekin, 2015, s. 44-48). Kapsam geçerliğini belirlemek için uzman

görüşüne başvurma, aynı kapsamı ölçtüğü bilinen bir testle arasındaki korelasyon ve tekrarlı test geliştirme gibi yöntemler kullanılabilir (Baykul, 2010, s. 226-230).

Ölçüt geçerliği, testi yapan kişinin testi alanların sonuçlarından pratik önemi ve bazı gerçek davranışsal çeşitliliği olan performansı hakkında çıkarım (yordama) yapmak istemesi durumunda kullanılır (Crocker & Algina, 2006, s. 224; Tekin, 2015, s. 48-50).

Yordayıcının seçilen bir ölçütle ilişki düzeyi bize ölçüt geçerliğini verir. Uygunluk ve yordama geçerliği olmak üzere iki türü vardır (Baykul, 2010, s. 231).

Yapı geçerliliği, ölçülmek istenen yapının ölçme aracı kullanılarak ortaya konulma derecesidir (Lord & Novick, 1968, s. 278-279). Testi yapanlar test sonuçlarından performansa yönelik belirli psikolojik yapılar altında gruplaştırmalar yapabilecekleri çıkarımlarda bulunmak isterler (Crocker & Algina, 2006, s. 230). Yapı geçerliğini belirlemek için test maddelerinin uzmanlar tarafından incelenmesi, cevaplayıcıların cevaplama süreçlerinin çalışılması, puanların kararlılığı, diğer testlerle korelasyon, pratik ölçütlerle korelasyon, grup farklılıkları, test puanlarını geliştirici çalışmalar ve doğrulayıcı faktör analizi gibi yöntemler kullanılmaktadır.

Görünüş geçerliği, ölçme aracının görünüş itibarı ile neyi ölçtüğüdür. (Tan, 2014, s. 198).

Günümüzde geçerlik tartışmaları sürmesine rağmen birçok uzman tek bir konuda görüş birliğine varmıştır: Geçerlik testin bir özelliği değil testten alınan puanlara ait bir özelliktir.

Bu açıdan düşünüldüğünde geçerlik türlerini geçerleme çalışmaları için kanıt toplama süreçleri olarak nitelendirmek daha doğru olmaktadır (Kelecioğlu & Göçer Şahin, 2014).

Geçerlik bu araştırmanın asıl konusunu oluşturmadığından ayrıntılı olarak ele alınmamıştır.

Güvenirlilik

Güvenirlilik; ölçümlerin tesadüfi hatalardan arınık olma derecesidir. Yine güvenirliliğin; bir ölçme aracının aynı özelliği ölçme derecesi, ölçülen özelliği ne derece hassas ölçtüğü ve özelliğin zaman içindeki kararlılığı olmak üzere üç farklı anlamı vardır (Özbek, 2014, s.

44-45). Eğitimde ve psikolojide ölçülmek istenen özellikler (zekâ, başarı, tutum, yetenek, vb.) çoğunlukla gizil (örtük, doğrudan ölçülemeyen) özelliklerdir. Bu nedenle ölçme işlemlerini yapabilmek için ölçmeyi yapan kişi ya da uzmanlar tarafından bu özellikler gözlenebilir hâle dönüştürülerek davranışlar tanımlanır. Bu tanımların varlığı ya da var olma derecesi sorgulanarak ölçme işlemi yapılır. Bireylerde var olan özellikler hakkında doğru kararlar verebilmek için yapılan ölçmenin doğruluğuna güvenmek gerekir. Bu sebeple ölçme işlemlerinin güvenilirliği ile bahsedilen ölçmenin hatalardan arınlığıdır. Ölçme işlemine karışan hatalar azaldıkça ölçme işleminin güvenilirliği dolayısıyla da bireyde ölçülen özellik hakkında verdiğimiz kararın doğruluğu da artmaktadır (Köse, 2014, s. 86-87). Güvenirlik, Kuramsal Çerçeve başlığı altında daha ayrıntılı olarak ele alınacaktır.

Kullanışlılık

Kullanışlılık ise ölçme aracının kullanımındaki kolaylık olarak düşünülebilir. Bu kolaylık zaman, ölçme aracının uzunluğu, ekonomik olması ve uygulama kolaylığı olarak farklılaşabilmektedir (Tan, 2014, s. 206; Baykul, 2010, s. 223).

Tüm bu açıklamalar ışığında ölçme işleminin bilimsel çalışmaları ve bu çalışmaların uygulamaya konmasını hızlandırdığı ve bu gelişme sonucunda da bilim dalları için özelleşen ölçme teorilerinin geliştirildiği görülmüştür (Baykul, 2010, s. 98). Bu teoriler (kuramlar) kullanım amacına, sınırlıklarına, ölçme sonuçlarının ne şekilde kullanılacağına göre farklılaşmıştır: Klasik Test Kuramı (KTK), Genellenebilirlik Kuramı (GK) ve Madde Tepki Kuramı (MTK) gibi.

Klasik Test Kuramı'na göre, bir kişide var olan özellik ölçülmek istendiğinde testten alınan puan kişinin gözlenen puanı olmaktadır. Gözlenen puan, bir testteki maddelerin her birinden alınan puanların ağırlıklı toplamıdır. Ayrıca gözlenen puan, kişinin gerçek puanı

ve hata puanının toplamı ile hesaplanır. Klasik Test Kuramı'nda bu hata puanı (bireyden, ölçmeyi yapan kişiden, ortamdan, puanlayıcıdan vb. kaynaklanan tesadüfi hataların toplamı) tek bir puan olarak karşımıza çıkmaktadır. Klasik Test Kuramı'nın ve varyans analizinin bir evrimi olan Genellenebilirlik Kuramı'nda ise bu hata kaynakları ölçme süreçlerine dâhil edilmiş, böylelikle hatalar kontrol altına alınmaya çalışılmıştır. Genellenebilirlik Kuramı'nın en büyük üstünlüğü bu hata kaynaklarından gelen varyansları ayrıştırarak ya da aynı anda açıklayabilmesidir. Madde Tepki Kuramı'nda ise kişilerde var olan gizil özelliğin (yeteneğin) derecesine göre kişileri farklılaştırabilen maddeler kullanılarak, maddeler ve kişiler arasındaki ilişkiler gruplardan bağımsız olarak ortaya çıkarılabilmektedir (Atılğan, 2005; Baykul, 2010, s. 107-118; Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 111-216). Daha ayrıntılı bilgiler Kuramsal Çerçeve başlığı altında verilecektir.

Sonuç olarak bu çalışmada, araştırmacı tarafından geliştirilen yazılı anlatım becerisi (hikâye) puanlama anahtarının güvenilirliği Klasik Test Kuramı, Genellenebilirlik Kuramı ve Madde Tepki Kuramı'na göre belirlenerek ilişkili olup olmadıkları incelenmiştir. Puanlama anahtarının güvenilirliğinin üç kurama göre de kestirilmesinin, puanların yorumlanmasındaki zorlukları azaltacağı ve kişiler hakkında verilen kararların uygunluğunu arttırması beklenmektedir.

1.2. Araştırmanın Amacı

Bu araştırmanın amacı, üst düzey düşünme becerilerinden biri olan yazılı anlatım becerisi temel alınarak, tesadüfi hatayı en aza indirmek için çalışma alanlarına göre farklılaşan hata teorilerinin güvenilirlik belirlemedeki yaklaşımlarının incelenmesidir.

1.3. Araştırmanın Önemi

Bu araştırma ile öğrencilerin üst düzey düşünme becerilerinden biri olan yazılı anlatım becerisini değerlendirmek için KTK, GK ve MTK'ye dayalı güvenilirlik kestirme yöntemlerinin karşılaştırılması, birbirlerine göre üstünlükleri ve farkları, sınırlıkları ve sayıltıları belirlenerek alana kuramsal bir katkı sağlanması hedeflenmektedir.

Aynı ölçme aracı kullanılarak yapılan güvenilirlik belirlemeleri ile kuramların kendi içinde ya da birbirleri arasındaki tutarlılıkları da görülebilecektir. Böylelikle bu çalışma ile kuramsal çalışmaların gelişimine katkı sağlayarak tartışmalara ışık tutulabileceği de umulmaktadır. Yine bu çalışmanın, araştırmacıların ya da test geliştiricilerin amacına ve verilerine göre hangi kuramı seçmeleri gerektiği ile ilgili sorulara bir çözüm önerisi getirebilmesi açısından bilimsel bir kaynak niteliği taşıması hedefler arasındadır.

Bu çalışmanın, yukarıdaki teorik hedeflerinin yanı sıra yazılı anlatım becerilerinin ölçülmesinde hatayı en aza indirgeyebilmek ve en etkili değişkenlik kaynaklarını belirleyebilmek, öğrencilerin yazılı anlatım performansı hakkında karar verilirken puanlamanın güvenilirliğini arttırmak için hem ekonomik hem de etkili olacak şekilde ne gibi çalışmalar yapılabileceğini göstermek gibi pratikte de uygulanabilir hedefleri bulunmaktadır.

Bu çalışma incelenen üç kuram için sayıltıların sağlanması ve bu kuramların uygulanabilirliğine yönelik olarak araştırmacıların karşılaşılabileceği güçlükler ve çözüm yollarına yönelik bulgu ve yorumların yapılması bakımından da önem taşımaktadır.

1.4. Problem Cümlesi

Üst düzey düşünme becerilerinden biri olan yazılı anlatım (hikâye yazma) becerisini puanlamak için hazırlanan puanlama anahtarı kullanılarak öğrencilerin performanslarının birden fazla kişi tarafından puanlanması ile elde edilen puanların güvenilirlikleri Klasik

Test, Genellenebilirlik ve Madde Tepki Kuramlarının güvenilirlik kestirme yöntemlerine göre nasıl farklılaşmaktadır?

1.5. Araştırma Soruları

Araştırmanın amacına yönelik olarak aşağıdaki sorulara yanıt aranacaktır:

1. KTK'ye göre; yazılı anlatım becerisinin ölçülmesinde, yedi farklı puanlayıcının aynı öğrencilerin hikâyelerini (formlarını) puanlaması sonucu elde edilen puanlar arasındaki tutarlılık derecesi nedir?
 - 1.1. Puanlayıcıların verdikleri toplam puanlar arasındaki Pearson korelasyon katsayısı ve Eta korelasyon katsayısı ne düzeydedir?
 - 1.2. Her bir puanlayıcının verdiği puanların sınıf düzeyleri ve tüm öğrencilerin puanları için Cronbach α katsayıları ne düzeydedir?
2. GK'ye göre yazılı anlatım becerisinin ölçülmesinde kestirilen parametre değerleri nelerdir?
 - 2.1. Öğrenci (b), ölçüt (ö) ve puanlayıcı (p) değişkenlik kaynaklarının tümüyle çaprazlandığı $b \times ö \times p$ deseninin G çalışması sonucunda kestirilen hata varyansları ve toplam varyansı açıklama yüzdeleri nasıldır?
 - 2.2. $b \times ö \times p$ deseninde puanlayıcı ve ölçüt sayılarının arttırılıp azaltılmasıyla yapılan karar çalışması (K) sonucunda elde edilen G ve Phi katsayıları nasıldır?
3. Çok kategorili MTK modellerinden biri olan Samejima'nın Derecelendirilmiş Tepki Modeli (DTM)'ne göre yazılı anlatım becerisinin ölçülmesinde kestirilen madde ve test istatistikleri nelerdir?
4. Bu üç teoride elde edilen güvenilirlik kestirimleri arasında manidar bir farklılık var mıdır?

1.6. Tanımlar

Hata: "Yanlışlık" anlamında kullanılmamaktadır. Ölçmedeki hata ölçme sonuçlarının ölçme aracı, ölçmeyi yapan kişi, ortam, zaman vb. kaynaklardan etkilenmesi sebebiyle ölçme sürecinin doğal bir parçasıdır (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 1-3)

Derecelendirilmiş Tepki Modeli: Sıralı cevap kategorilerine sahip çok kategorili (0, 1, 2 vb.) puanlanan maddelerde, seçilen bir kategorinin ve bu kategoriden yüksek kategorilerin seçilme olasılığına dayanan iki parametrelili bir MTK modelidir.

Derecelendirilmiş Tepki Modeli'nde Madde Bazında Verilen Yetenek Parametresi: Her bir madde için verilen yetenek parametreleri, o maddenin verdiği bilgi miktarının en fazla olduğu yetenek düzeyini göstermektedir.

1.7. Sınırlılıklar

1. Araştırma üst düzey becerilerin ölçülmesinde sadece yazılı anlatım becerisinin ölçülmesi için hazırlanan puanlama anahtarıyla sınırlıdır.
2. Araştırma yazılı anlatım becerisini değerlendirmede sadece hikâye yazımıyla sınırlıdır.
3. Araştırma, Gaziantep ve Karabük illerinde bulunan birer okulun 3, 4, 5, 6 ve 7. sınıf öğrencileri ile sınırlandırılmıştır.
4. Araştırma, bir Türkçe öğretmeni, beş sınıf öğretmeni ve bir ölçme ve değerlendirme uzmanından oluşan yedi puanlayıcı ile sınırlıdır. MTK analizlerinde tek boyutluluk varsayımı karşılanmadığı için 6 ve 7. puanlayıcılar analize dâhil edilememiştir.



BÖLÜM II

İLGİLİ ARAŞTIRMALAR ve KURAMSAL ÇERÇEVE

Bu bölümde KTK, GK ve MTK'ye göre güvenirlik belirleme yöntemlerini karşılaştırmayı konu alan araştırmalara yer verilmiştir.

2.1. İlgili Araştırmalar

Alanyazın incelendiğinde üç kuramı karşılaştıran çalışmalara göre kuramları ikili karşılaştıran çalışmaların çoğunlukta olduğu görülmüştür.

2.1.1. Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar

Güler (2008)'in araştırmasında matematik performansının ölçülmesinde kullanılan maddelerin birden fazla puanlayıcı tarafından puanlanması sonucu oluşan değişkenlik kaynaklarının arttığı ölçme durumlarında, KTK, GK ve çok değişkenlik kaynaklı Rasch ölçme modelinin karşılaştırılması amaçlanmıştır. Veriler, 2007 yılında bahar döneminde farklı yerleşim yerlerinde ve okul türlerinde öğrenim gören 8 ve 9. sınıf öğrencilerinin oluşturduğu 203 kişilik gruba 1999 TIMSS'te yer alan 24 açık uçlu madde uygulanarak toplanmış; bütüncül puanlama anahtarı kullanılarak dört kişi tarafından puanlama yapılmıştır. Araştırmanın sonunda matematik performansının ölçülmesinde yer alan

maddelerin oldukça tutarlı sonuçlar verdiği, puanlamalar arasında anlamlı bir farklılığın olmadığı, puanlayıcılar arası güvenilirliğin yüksek olduğu, genellenebilirlik ve güvenilirliği arttırmak için puanlayıcı sayısı yerine madde sayısını arttırmanın kullanışlılığı arttıracağı sonuçlarına ulaşılmıştır. Kuramlardan hangisinin ya da hangilerinin bir arada kullanılacağına karar verilmesinin amacımıza göre değişebileceği belirtilmiş olsa da en az iki kuramın birlikte kullanılması daha güvenilir sonuçlar vermesi açısından önerilmiştir.

Brennan (2011)'ın araştırmasında, KTK ile GK teorik açıdan ele alınarak karşılaştırılmış, iki kuramın da MTK ile güvenilirlik kestiriminde kıyaslamalarını yapmıştır. Araştırmacılara sunduğu öneriler şöyledir: KTK'da bütün hata kaynakları tek bir hata terimi ile gösterilir. GK'de ise pek çok hata kaynağı kestirilebilmektedir. Araştırmacı bu hata kaynaklarından hangilerinin ölçme yüzeylerini etkili biçimde tanımladığını bilmelidir. Kuramların varsayım kıyaslamalarında ise KTK ve GK'nin aksine MTK'nin varsayımlarının güçlü olduğu belirtilmiştir. KTK'nin basit ve yaygın kullanımlı olması; GK'nin farklı hata kaynaklarını tek bir analizle kestirmeye olanak sağlaması, sabit ve sonsuz yüzeylere uygulanabilirliği; MTK'nin matematiksel karmaşık yapıda olan ölçme durumlarına uygun olması bu üç kuramın güçlü yönleri iken KTK'nin hata varyansını ayrıştıramaması, GK'nin kavramsal zorluğu, MTK'nin ise yalnızca sabit yüzeylere uygulanabilir olması ise bu kuramların zayıf yönleridir.

2.1.2. Klasik Test ve Genellenebilirlik Kuramlarını Karşılaştıran Araştırmalar

Yelboğa (2007)'nin araştırmasında bir iş performansı ölçeğinin güvenilirliği ile elde edilen verilerin KTK ve GK'ye göre incelenmesi amaçlanmıştır. Veriler 2005-2006 eğitim öğretim yılında hizmet sektörünün 11 farklı biriminde çalışan 176 kişiden araştırmacı tarafından geliştirilen iş performansı ölçeği kullanılarak elde edilmiştir. KTK'de test tekrar test ve Cronbach α iç tutarlılık güvenilirlik katsayısı, puanlayıcılar arası güvenilirlik için

Kendall'ın uyum katsayısı, GK'de ise G ve Phi katsayıları incelenmiş ve bu analizlerin tutarlı; iki kuramın benzer sonuçlar verdiği gözlenmiştir. Farklı ölçekler ve farklı gruplarla çalışmalar yapılması, farklı madde ve puanlayıcı sayılarının kullanılması durumunda GK'nin kullanılmasının daha ayrıntılı sonuçlar vereceği önerilmektedir.

Deliceoğlu (2009)'nun araştırmasında futbolcuların teknik yeteneklerini ölçmek için Futbol Yetilerine İlişkin Dereceleme Ölçeğinin güvenirlik katsayılarının KTK ve GK'ye göre belirlenmesi ve karşılaştırılması amaçlanmıştır. Veriler 2007-2008 yıllarında Ankara'daki profesyonel spor kulüplerinin alt yapılarındaki profesyonelliğe aday ve antrenman yılları beş ve üzeri olan 72 futbolcudan elde edilmiştir. Video kamera çekimleri ile kayıt altına alınan performanslar B lisans futbol antrenörlük belgesine sahip dört uzman tarafından Futbol Yetilerine İlişkin Dereceleme Ölçeği kullanılarak bir hafta aralıkla iki kez puanlanmıştır. İki puanlama arasındaki tutarlılık için Pearson momentler çarpımı ilişkisi katsayısı, maddelerin iç tutarlılığı için Cronbach α katsayısı, puanlayıcılar arasındaki tutarlılık için Kendall uyum katsayısı, genellenebilirlik kuramına göre G ve Phi katsayıları incelenmiştir. Phi katsayısı ve Kendall'ın W katsayıları dışındaki değerler beklenen düzeyde iken bu iki katsayı beklenen değerden düşük bulunmuştur. Ayrıca hata kaynaklarının birden fazla olduğu durumlarda GK'nin KTK'ye alternatif olduğu ve kullanılan ölçeğin uygun güvenirlik düzeyinde olduğu sonuçlarına ulaşılmıştır. GK ve KTK'nin benzer sonuçlar vermesi sebebiyle farklı madde ve puanlayıcı sayıları ile alternatif analizler yapılması; farklı ölçeklerden elde edilen veriler ile GK'deki güvenirlik katsayıları arasındaki tutarlılığa bakılması önerilmiştir.

Öztürk (2011)'ün araştırmasında voleybolcuların becerilerinin teknik açıdan tespit edilebilmesi için kullanılan voleybol becerilerine ilişkin gözlem formu ile KTK ve GK'ye dayalı belirlenen güvenirlik katsayılarının düzeylerinin belirlenip karşılaştırılması amaçlanmıştır. Veriler 2009-2010 yıllarında Türkiye Voleybol Federasyonu tarafından

belirlenen 10 ilden hakemlerin seçtiği 13-14 yaş grubundaki 102 erkek voleybolcudan elde edilmiştir. Video kayıt ile kaydedilen performanslar gözlem formundaki 25 madde ile dört puanlayıcı tarafından değerlendirilmiştir. Araştırma sonunda KTK'ye göre gözlem formunun iç tutarlılığının sağlandığı fakat puanlamalar arasındaki tutarlılığın uygun düzeyde olmadığı görülmüştür. Ayrıca GK'ye göre puanlayıcı ve madde sayılarının artırılarak yüksek genellenebilirlik ve güvenilirliğe ulaşılmasını sağlayacağı belirtilmiştir. Seçilen puanlayıcıların uzmanlık alanlarının dikkate alınması, voleybol becerilerinin değerlendirilmesinde gözlem formu yerine analitik rubrik kullanılması, farklı geçerlik ve güvenilirlik katsayılarına sahip ölçekler kullanılarak ölçmelerin psikometrik niteliklerindeki farklılaşmaların araştırılabileceği önerilmiştir.

Güler (2011)'in araştırmasında 125 öğrencinin 18 maddeye verdikleri yanıtların dört farklı puanlayıcı tarafından puanlandığı varsayımı ile rastgele veriler üretilerek GK ve KTK'nin benzer ve farklı yönlerinin ortaya çıkarılması amaçlanmıştır. GK'nin değişkenlik kaynağının sadece maddeler olduğu tek değişken kaynaklı çaprazlanmış desendeki G katsayısı ile KTK'nin Cronbach α değeri başka bir deyişle güvenilirlik katsayıları yaklaşık olarak aynı bulunmuştur. Değişkenlik kaynağının birden fazla olduğu durumlarda GK'nin kullanılmasının aksi durumda ise KTK'nin kullanılmasının daha ekonomik (anlaşılması ve hesaplanması daha kolay) olduğu belirtilmiştir. Verilerin rastgele üretilmesi sebebiyle ölçme hatalarının yüksek ve güvenirliliğin düşük olduğu durumda da KTK ve GK'nin paralel sonuçlar vererek güvenirliliğin düşük olduğu sonucuna ulaşılmıştır. Araştırmanın sonunda öğrenci, madde ve puanlayıcı sayılarının artırılıp azaltılarak yeni çalışmalar yapılabileceği önerilmiştir.

Büyükkıdık (2012)'in araştırmasında aynı gruba uygulanan problem çözme ve çözüm yollarını sunma becerisini ölçen iki performans görevinin puanlanmasında kullanılan puanlama anahtarlarının güvenirliliği KTK ve GK ile elde edilen bilgiler ışığında

karşılaştırılması amaçlanmıştır. Veriler 2011-2012 eğitim ve öğretim yılında Kütahya'da öğrenim gören 6, 7 ve 8. sınıf öğrencilerinin oluşturduğu 132 kişiden toplanmıştır. Bu veriler farklı bölgelerde görev yapan dört matematik öğretmeni tarafından puanlanmıştır. Puanların güvenilirliği; KTK'da sınıf içi ilişki katsayısı ve puanlayıcılar arası ilişki katsayısı ile, GK'de tümüyle çaprazlanmış desenler kullanılarak hesaplanmıştır. Araştırma sonucunda her iki kurama göre de güvenilirlikler yüksek bulunmuştur. Buna rağmen GK'nin kullanılmasının KTK'ye göre daha detaylı bilgiler verdiği tespit edilmiştir. Ayrıca kullanılan puanlama anahtarı türünün ve GK'de seçilen modelin de güvenilirliği etkilediği görülmüştür. KTK, GK ve MTK'deki güvenilirlik katsayılarını karşılaştırmayı içeren bir araştırma yapılması, matematiksel muhakeme becerilerinin ölçülmesi için çoktan seçmeli testler yerine puanlama anahtarlarının kullanılması, KTK'deki madde güçlük düzeyleri ile GK'deki değişkenlik kaynaklarına ait varyans bileşenlerinin de karşılaştırılabileceği önerilmiştir.

Bağcı (2015)'in araştırmasında öğrencilerin matematiksel muhakeme becerilerinin ölçülmesinde kullanılan aracın KTK ve GK'nin yuvalanmış ve çaprazlanmış desenler sonucunda elde edilen katsayıların karşılaştırılması amaçlanmıştır. Veriler üç bağımsız puanlayıcı ile 2014-2015 eğitim öğretim yılında 187 yedinci sınıf öğrencisinden elde edilmiştir. Araştırma sonucunda iki kurama göre elde edilen katsayılar da kabul edilebilir düzeyde bulunmuştur. Yuvalanmış desende elde edilen G ve Phi katsayılarının çaprazlanmış desenden elde edilen katsayılara göre daha yüksek olduğu görülmüştür. Cronbach alfa katsayılarının ise G katsayıları ile paralel sonuçlar verdiği belirlenmiştir. GK'de yapılan analizler KTK'ye göre daha detaylı bilgiler ortaya koymuştur.

Şalgam (2016)'ın araştırmasında kısa cevaplı matematik yazılı sorularına aynı öğrencilerin verdikleri cevapların güvenilirlik katsayılarının KTK'deki test tekrar test yöntemi ve GK'ye göre incelenmesi hedeflenmiştir. Veriler 2014-2015 eğitim öğretim yılında Çorum ili

Sungurlu ilçesindeki farklı ortaokullarda eğitim gören 7. sınıf öğrencilerinden iki hafta süre ile yazılı soruları iki kez uygulanarak toplanmıştır. Farklı zamanlarda uygulandığı için iki uygulamaya da katılan 99 öğrenci araştırmaya dâhil edilmiştir. Öğrencilerden toplanan yanıtlar üç puanlayıcı tarafından puanlanmıştır. Araştırma sonucunda uygulanan yazılının güvenilirliği Pearson momentler çarpımı korelasyon katsayısı ile hesaplanmış ve yüksek güvenilirlikte olduğu bulunmuştur. Tümüyle çaprazlanmış desen kullanılarak yapılan G çalışması sonucunda maddelerin güçlük düzeylerinin farklı olduğu, öğrenciler arasındaki farklılıklar yeterince ortaya çıkmadığı görülmüştür. Ayrıca yazılının güvenirliliğinin madde sayısı ile doğru orantılı olduğu görülürken puanlayıcı sayısının güvenirliliği etkilemediği tespit edilmiştir. Son olarak da KTK ve GK'de güvenirlilik katsayıları arasında manidar bir fark olmadığı bulunmuştur. Farklı madde sayıları, farklı ölçme araçları, farklı desenler ve değişkenlik kaynakları, bu iki kurama ek olarak MTK'ye göre de analizler yapılabileceği önerilmiştir.

2.1.3. Klasik Test ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar

Koch (1983) araştırmasında, Likert tipi bir tutum ölçeğinin madde ve birey parametrelerini KTK ve MTK'ye göre kestirip karşılaştırmıştır. Gerçek veriler ile yapılan çalışmanın bulgularına göre geleneksel yolla ve DTM'ye göre kestirilen madde ve birey parametreleri arasında yüksek düzeyde ilişkiler bulunmuştur. KTK'ye göre kestirilen madde ortalamaları ve Samejima'nın DTM'sine göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde, madde ayıricılık güçleri ve a parametreleri arasında ise pozitif yönde yüksek düzeyde ilişkiler tespit edilmiştir. Elde edilen bulgular doğrultusunda derecelendirilmiş tepki modelinin tutum ölçüleri için kullanılabilir olduğu ve dahası tutumun her bir düzeyinde ölçmenin kesinliği ile ilgili bilgi fonksiyonları vermesi açısından daha avantajlı olduğu belirtilmiştir.

Gelbal (1994)'ın arařtırmasında KTK ile MTK'nin Rasch modelinin benzer yönlerinin ortaya konması amaçlanmıřtır. Veriler UNICEF, MEB ve Hacettepe Üniversitesi Eđitim Fakóltesince yapılan bir arařtırma kapsamında 5. sınıf öđrencilerinden Türkçe ve matematik testleri ile toplanmıřtır. Arařtırmanın sonucunda KTK ile kestirilen madde güçlük indeksleri ile Rasch modelinde kestirilen b parametresinin benzer olduđu görölmüřtür. Örneklem büyüklüğünün aksine test puanlarının normal dağılım özelliđi gösterip göstermemesi bu iki kurama göre elde edilen kestirimlerin paralelliđini etkilememektedir. Arařtırma sonunda parametre kestirimlerinde büyük örneklem seçilmesi, KTK ile kestirilen parametrelerin Rasch modeli ile kestirilen yetenek ölçüleri yerine kullanılması öneriler arasındadır.

Keleciođlu (2001)'nın arařtırmasında 1994 yılında MEB tarafından Anadolu Liseleri Giriř Sınavı için hazırlanan otuzar sorudan oluřan matematik ve Türkçe alt test maddelerinin örtük özellikler kuramı ile kestirilen parametreleri ve geçiř formülleri ile elde edilen parametreleri boş bırakılan maddelerin yanlış ya da atlanmış olarak analize dâhil edilmesine göre nasıl farklılařtığının incelenmesi amaçlanmıřtır. Veriler 24.701 öđrenciden elde edilmiřtir. İki alt test için de KTK'ye göre ve örtük özellikler kuramının iki parametrelili modeline göre incelemeler yapılmıřtır. Boş bırakılan maddelerin yanlış ya da atlanmış olarak ele alınması ile KTK'ye göre elde edilen p ve r parametreleri geçiř formülleri uygulanarak örtük özellikler kuramındaki b ve a parametrelerine dönüřtürölmüş ve aradaki korelasyon incelenmiřtir. Arařtırma, KTK'deki p ve r parametrelerinin kullanılarak örtük özellikler kuramındaki b ve a parametrelerinin kestirimde normal dağılım özelliđi gösterip göstermemesinin sonuçları etkilediđini ortaya koymuřtur. Arařtırma sonunda, verilerin dağılımı normale yaklařtıkça geçiř formüllerinden elde edilen deđerlerin kestirilen parametre deđerlerine yaklařtığı tespit edilmiřtir.

Özdemir (2002)'in araştırmasında çoktan seçmeli test maddelerinin iki (0-1) ve çok (1, 2, 3, 4) kategorili puanlanmaları durumunda KTK ve MTK'ye göre testin geçerlik ve güvenirliğinin incelenmesi amaçlanmıştır. Veriler 2001-2002 yıllarında 4, 5, 6 ve 7. sınıflarda öğrenim gören 1.608 öğrenciden Türkçe okuduğunu anlama becerisini ölçmeyi amaçlayan dört seçenekli ve 20 maddelik çoktan seçmeli bir test kullanılarak elde edilmiştir. İki kategorili maddelerde KTK'ye göre elde edilen güvenirlik katsayısı MTK'ye göre elde edilen katsayıdan düşüktür; KTK'de iki kategorili puanlamanın güvenirlik katsayısı çok kategorili puanlamadan düşüktür; MTK'de çok kategorili puanlamanın güvenirlik katsayısı iki kategorili puanlamadan düşüktür. Geçerlik sonuçları incelendiğinde ise KTK için 4-5 ve 5-7. sınıfların test puanlarının ortalamaları; MTK için 4-5 ve 5-7. sınıfların yetenek parametre ortalamaları arasındaki fark manidardır. Araştırma sonunda MTK ile yapılan çalışmalarda iki kategorili puanlama, KTK ile yapılan çalışmalarda çok kategorili puanlama önerilmektedir.

Nartgün (2002)'ün araştırmasında, aynı tutumu ölçmek için hem Likert tipi hem de metrik ölçek olarak düzenlenen Matematikle ilgili düşünceler ölçeğinin madde ve ölçek istatistiklerini KTK ve MTK'ye göre incelenmiştir. Elde edilen bulgulara göre ölçeklerin iki kurama göre de kestirilen madde ayırt edicilik güçlerinin benzer olduğu, ölçeklerdeki maddelerin ölçtükleri tutum düzeylerinin, tutum ölçümlerinin belirtilen kuramlara göre kestirilen değerlerinin yüksek düzeyde benzer olduğu, ölçeklerin ölçüt grup geçerliklerine ve güvenilir ölçme sonuçlarına sahip oldukları sonucuna varılmıştır.

Doğan & Tezbaşaran (2003)'in araştırmasında örtük özellikler kuramının KTK'ye bir üstünlüğü olarak bilinen madde parametrelerinin değişmezliğinin incelenmesi amaçlanmıştır. Veriler 2001 yılında uygulanan her biri 25 sorudan oluşan dört alt test (Türkçe-matematik-fen ve sosyal bilgiler testlerini) içeren MEB Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavını alan 553.281 öğrenciden elde edilmiştir.

Araştırmada sağa, sola çarpık, random, basık, normal dağılım özelliği gösteren beş farklı örneklem kullanılmıştır. Araştırma sonunda; KTK'ye göre elde edilen p ve ayırt edicilik değerlerinin özellikle çarpık örneklem başta olmak üzere örneklemden etkilendiği; MTK'nin 1, 2 ve 3 parametrelili modellerine göre kestirilen b parametresinin de p parametresi ile paralel sonuçlar verdiği, a parametresinin ise model ve örneklem büyüklüğünden etkilendiği; madde değişmezliği açısından kuramların bir diğerine üstünlüğünün bulunamadığı tespit edilmiştir.

Kan (2006)'ın araştırmasında 2001 yılında uygulanan Ortaöğretim Kurumları Sınavı (OKS)'nin Türkçe alt testindeki maddelerin KTK ve MTK'ye göre incelenerek madde istatistikleri arasındaki farklılıkların ortaya konması amaçlanmıştır. Veriler bu merkezî sınava giren 553.108 öğrenciden elde edilmiştir. İki teoriye göre de kestirilen parametre değerleri birlikte değişim gösterdiği; ayırt ediciliği en düşük olan 16. madde ve güçlüğü en düşük (en kolay) olan 2. madde her iki teoride de aynı olarak tespit edilmiştir. Ayrıca iki teoriye göre de kestirilen madde güçlük indeksleri ayırt edicilik indekslerine göre daha tutarlıdır. Araştırma sonunda sadece madde seçerken iki kuramın da kullanılabileceği, test ve madde bilgi fonksiyonları gibi referans noktaları önemli olduğunda ise MTK'nin kullanılması önerilmektedir.

Morales (2009)'ın araştırmasında üniversite öğrencilerinin matematikteki başarısını ölçmek için oluşturulmuş bir ölçme aracının geçerlik ve güvenirliği için Klasik Test Kuramı ve Madde Tepki Kuramı kullanmıştır. Veriler Kuzey Samar'daki Doğu Filipinler Üniversitesi'ndeki yüksekokul ve kıdemli matematik öğretmenleri ile İngilizce öğretmen adaylarından oluşan 80 öğrenciden toplanmıştır. Araştırmacı tarafından geliştirilen üniversite öğrencileri için beş seçenekli 40 maddeden oluşan çoktan seçmeli Matematik Başarı Testi kullanmıştır. Veriler öncelikle psikometrik özellikleri karşılaştırmak için KTK ve MTK kullanılarak analiz edilmiş daha sonra madde yanlılığını incelemiştir. KTK'deki

analizler için SPSS ve Microsoft Excell, MTK'deki analizler için Winsteps'in Bigsteps yazılımı kullanmıştır. Testin iç tutarlılığı Cronbach'ın alfa değeri 0,77 ile yüksek bulunmuştur. Guttman katsayısı 0,72 ile sonuçlanmıştır. Bu değer, testteki cevapların iç tutarlılıklarını gösteren bir değerdir. Son olarak KR20, iç tutarlılığı 0,90 bulunmuştur. Analiz sonucunda başarı testinin iyi bir test olduğu KTK ve MTK birbirinden farklı olmasına rağmen KTK'de zayıf olan maddelerin Rasch modelinde de çalışmadığı görülmüştür.

Kim & Feldt (2010)'in araştırmasında büyük ölçekli gerçek verilerin manipüle edilmesiyle elde edilen çeşitli test koşulları altında MTK güvenilirlik istatistiklerinin ve iki KTK güvenilirlik istatistiklerinin (Cronbach alfa ve Feldt-Gilmer konjenerik katsayılarının) göreceli performanslarının incelenmesi amaçlanmıştır. Test verileri için İngilizce, matematik ve bilim testlerinin her biri için çoktan seçmeli ve açık uçlu sorulardan oluşan testler kullanılmış, 25.000 kişiye uygulanmıştır. Maddeler arasındaki heterojenlik derecesini değiştirmek için bu üç testin her biri üç alt teste (sadece çoktan seçmeli, sadece açık uçlu, ikisini birden içeren testlere) ayrılmıştır. Araştırmanın sonunda tüm test koşullarında MTK güvenilirlik katsayısının KTK güvenilirlik istatistiklerinden daha yüksek olduğu; MTK güvenilirlik katsayısının Cronbach alfa katsayısına göre Feldt-Gilmer katsayısına daha yakın olduğu; alfa katsayısının ise MTK güvenilirlik katsayısının alt limitine daha yakın olduğu görülmüştür. Bu nedenle KTK'ye dayalı yaklaşımların aşağıdaki sınırlılıkları olduğu tespit edilmiştir:

Birincisi, MTK çerçevesinde test geliştirme ve madde analizi yapıldığında, test puanı güvenilirliğini tahmin etmek için KTK yaklaşımlarını benimsemek makul değildir. Güvenilirliği hesaplamak için kullanılan MTK parametreleri var olduğu sürece, MTK tabanlı yaklaşımı kullanmak mantıklıdır. İkincisi, KTK güvenilirlik istatistikleri hem test hem de gruba bağlıdır. Buna karşılık, MTK güvenilirlik katsayısının yalnızca teste bağlı olması dışında MTK modellemesinin değişmezlik özelliği gereğince mevcut grubun

ötesinde güvenilirlik katsayısı popülasyonlar için hesaplanabilir. Üçüncüsü, KTK tabanlı yaklaşımlar sadece test seviyesinde güvenilirliği tahmin etmek için uygulanabilir, ancak MTK'ye dayalı yaklaşım, hem madde hem de test güvenilirliğini tahmin etmek için uygulanabilir. Bu nedenle, her bir maddenin KTK yaklaşımlarını kullanmaktan çok MTK yaklaşımını kullanarak test-puan güvenilirliğine katkısını anlamak mümkündür.

Sebille vd. (2010)'in araştırmasında hasta raporlarının sonuçlarının geçerlik ve güvenilirlikleri için gözlemlenen puanlara dayanan KTT ve modellere dayanan MTK kullanılmıştır. Ayrıca güç ve etki büyüklükleri için de bu iki kuram karşılaştırılmıştır. Enine kesit çalışması için simülasyonda iki grup için veri üretilmiştir. Bu verilerde IRT için, madde veya kişi parametrelerinin bilinip bilinmediği, madde parametrelerinin belirli bir dereceye kadar, iyi ya da zayıf olması dolayısıyla farklı tahminlere göre farklı senaryolar araştırılmıştır. İki kuram ile de elde edilen güçler karşılaştırılıp en güçlü etkiye sahip parametreler belirlenmiştir. Kişi parametrelerinin bilinmediği ve parametrelerinin bilinip bilinmediği varsayıldığında, KTK veya MTK kullanılarak elde edilen sonuçlar benzer bulunmuştur. Her iki yöntemde de maddelerin sayısı önemli ölçüde etkili olmuştur. Sonuç olarak kayıp veriler olmadan, KTK ve MTK karşılaştırılabilir durumdadır. KTK için klasik örneklem büyüklüğü formülleri yeterli olurken MTK için madde sayısı ve sınırlılıkların yerine getirilmesi şartları önemlidir.

Sünbül (2011)'ün araştırmasında boyutluluk özelliğine sahip psikolojik yapılarda madde parametrelerinin değişmezliğinin KTK, MTK ve ÇBMTK'ye göre incelenmesi amaçlanmıştır. Veriler ÇBMTK'ye göre, iki boyutlu; birinci boyuttaki madde sayısı 24, a_1 parametre değeri 1,00; ikinci boyutta 4, 8, 12, 16, 20, 24 madde sayılarında, a_2 parametre değerleri sırasıyla 0,25; 0,50; 0,75; 1,00 ve boyutlar arası korelasyon 0,00; 0,30; 0,60; 0,80 olacak şekilde üretilmiştir. Yine, değişimleme ölçütlerinin çaprazlanması sonucu 96 hücrenin her birinde 90.000 bireyin madde tepkisini içeren veriler üretilerek evren

belirlenmiştir. Parametre değişmezliğinin örneklem büyüklüğüne göre değişip değişmediğini görebilmek için 500, 1000 ve 3000’lik yüzer adet örneklem belirlenmiştir. KTK, MTK ve ÇBMTK’ye göre parametre değişmezliğini test etmek için standart hata ortalamaları incelenmiştir. Araştırma sonunda madde parametre değişmezliğinin KTK’ye göre bütün parametre ve koşullar için sağlandığı; MTK’ye göre a ve b parametrelerinde büyük ölçüde sağlandığı fakat ikinci boyutun birinci boyuta kuvvetçe eşdeğer ve boyutların birbirine dik konumlandığı deneysel hücrelerde sağlanmadığı (çok boyutluluğa karşı MTK’nin dayanıklı olduğu); ÇBMTK’ye göre ise büyük ölçüde sağlandığı fakat bu parametrelerin örneklem büyüklüğünden ve boyut oluşturamayan maddelerin boyut olarak tanımlanmasından etkilendiği tespit edilmiştir. Daha sonraki çalışmalarda; boyut sayısının arttırılması, madde parametreleri yerine yetenek parametrelerinin kullanılması, ÇBMTK’de doğrulayıcı yerine açıklayıcı faktör analizinin kullanılması, veri üretiminin farklılaştırılması, farklı analiz programları kullanılarak bu programların gücünün değerlendirilmesi, şans başarısının da araştırmaya dâhil edilmesi öneriler arasında yer almaktadır.

Özer Özkan (2012)’in araştırmasında KTK, tek ve çok boyutlu MTK modellerinin Öğrenci Başarılarının Belirlenmesi sınavının her biri 25 sorudan oluşan Türkçe ve Matematik alt testleri verilerine uygulanarak elde edilen başarı ölçülerinin karşılaştırılması amaçlanmıştır. Veriler 2008 yılında 36 ilde MEB’e bağlı 270 okulun 8. sınıflarında öğrenim gören 9.876 öğrenciden elde edilmiştir. Öncelikle testlerin boyutluluğu test edilmiş, iki testin de iki boyutlu olduğu tespit edilmiştir. Her iki test de çok boyutlu MTK ile elde edilen yetenek kestirimlerindeki hata miktarlarının KTK’ye ve tek boyutlu MTK’ye göre düşük düzeyde olduğu görülmüştür. Yine KTK’ye göre elde edilen puanlar ile tek boyutlu MTK’ye göre elde edilen yetenek kestirimlerinin benzerlik içinde olduğu tespit edilmiştir. Testlerin güvenilirliklerinde ise çok boyutlu MTK için en yüksek güvenilirlik katsayısı belirlenmiştir. Bu durumda çok boyutlu testlerin yetenek

kestirimlerinde çok boyutlu MTK kullanılmasının daha tanılayıcı bilgiler verdiği görülmektedir. Araştırma sonunda KTK yerine çok boyutlu MTK'ye göre geliştirilen bir testin kullanılması; boyutlar arası ilişkilere araştırmacı tarafından karar verilerek bir simülasyon çalışması yapılması; farklı istatistik programları kullanılması ve model veri uyumuna bakılması; aynı verilerle yetenek parametrelerinin tahmini yerine madde parametrelerinin kestirilmesine yönelik çalışmalar yapılması önerilmektedir.

Lee vd. (2012)'nin araştırmasında özellikle büyük ölçekli bir matematik değerlendirmesi olmak üzere bilişsel tanı modellemesini (CDM) kullanarak KTK, MTK ve CDM çerçevelerinde kullanılan madde istatistikleri arasındaki ilişkiyi araştırmak amaçlanmıştır. Bir CDM yaklaşımından madde ayırıcılığının bir ölçüsünü kullanarak değişen tanısal seviyelerine sahip testler arasındaki farklılıkları vurgulayarak bazı önemli KTK ve MTK madde özelliklerini ampirik olarak açığa çıkarmak da amaçlar arasındadır. Bu amaç doğrultusunda 3PL MTK modeline göre geliştirilen 2003 yılındaki Florida Kapsamlı Değerlendirme Testinin 9. Sınıf Matematik alt testinin uygulandığı gruptan rastgele 1500 kişi seçilmiştir. Araştırmanın sonucu MTK veya KTK tabanlı testin tanısal olarak bilgilendirici olmasına rağmen, testteki maddelerin farklı seviyelerde teşhis bilgileri sağlayabildiğini göstermektedir. Daha ayırt edici testlerin ise daha az değişken KTK ve MTK istatistiğine sahip ve böylece yetenek seviyesinin orta noktasındaki incelemeler için daha doğru ölçümlere yol açtığı saptanmıştır. Bu çalışmanın bir diğer önemli pratik bulgusu ise daha kısa testlerin KTK güvenilirliği ile ilgilidir - DINA ayırım indeksini kullanarak, maddelerin %40'ının test güvenilirliğine olumsuz bir şekilde etki etmeden silinebileceğini göstermektedir. Aynı uzunlukta oluşturulabilecek farklı bir testin ilk 30 maddesi testin KTK gereksinimini karşılamak için muhafaza edilebilir, son 20 madde, testin teşhis bilgisini en üst düzeye çıkarmak ya da kabiliyetin belirli bölümlerini daha doğru ölçmek için stratejik olarak seçilebilir. Bu çalışma, MTK veya KTK tarafından yapılan değerlendirmelerin tanısal değere sahip olabileceğini göstermiştir.

Değerlendirmenin tanısal olarak en faydalı olması için mevcut öğelerin özelliklerinin tanımlarını türetmek yerine, belirli bir alan için benimsenen özniteliklerin spesifik tanımlarına dayalı olarak öğeler oluşturulmalıdır. Bu tanımlar, öğrencilerin nasıl düşündükleri ve öğrendikleri ile ilgili en güncel anlayışı yansıtmalıdır. Araştırma sonuçlarının genellenebilirliği açısından sınırlıkları bulunmaktadır: Sonuçlar farklı bir alana veya inceleme grubuna uygulandığında bütünlüklerini koruyamayabilir, çıkarımların sayısı ve tanımları değiştikçe, çıkarımlar değişebilir. Aynı değerlendirme, öznitelikler seti ve inceleme grubu ile bile, farklı bir CDM kullanılırsa bulgular mevcut bulgulardan farklı olabilir. Bu sınırlıkları gidermeye yönelik araştırmaların yapılabileceği önerilmiştir.

Çelen & Aybek (2013)'in araştırmasında sınıf içi ders başarısını ölçmek için kullanılan bir testin KTK ve MTK'ye göre madde sayılarındaki değişim ve öğrenci başarıları arasındaki ilişkinin incelenmesi amaçlanmıştır. Bu amaç doğrultusunda veriler 2011-2012 bahar döneminde Eskişehir Osmangazi Üniversitesinde üç farklı programda öğrenim gören, ölçme ve değerlendirme dersi alan 231 öğrenciden toplanmıştır. Veriler KTK ve MTK'ye göre puanlanmış; bu puanlamalar arasındaki ilişki pozitif yönde, yüksek ve manidar bulunmuştur. Yapılan madde analizleri sonucu KTK'ye göre beş, MTK'ye göre (tek boyutluluğu ve model veri uyumunu sağlamayan) 14 madde testten çıkarılmış olmasına rağmen KTK ve MTK'ye göre yapılan puanlamalar maddeler çıkarılmadan önceki puanlamalar ile paralellik göstermiştir. Başka bir deyişle farklı uzunluktaki bu testlerin KTK'ye ya da MTK'ye göre geliştirilmiş olması öğrenci başarı sıralamalarında manidar bir farklılık yaratmamaktadır. Öğrenciler hakkında geçti-kaldı kararında ise MTK'ye göre puanlama yapıldığında KTK'ye göre geçmesi gereken 24 öğrenci, KTK'ye göre puanlama yapıldığında MTK'ye göre geçmesi gereken 5 öğrencinin kaldığı tespit edilmiş bu farklılığın değişen test uzunluklarının bir sonucu olarak test kapsamının da değişmesinden kaynaklanabileceği öne sürülmüştür. Araştırma sonunda geçti-kaldı kararları verileceğinde

KTK'nin, bireye uyarlanmış test kullanımı ve test eşitleme gibi amaçlar olduğunda MTK'nin kullanılması önerilmiştir.

Köse (2015)'nin araştırmasında Yapılandırmacı Öğrenme Ortamları Ölçeği kullanılarak KTK ve MTK'nin çoklu puanlamalar için uzantısı olan Aşamalı Tepki Modeli (DTM)'ne göre kestirilen madde ve test parametrelerinin karşılaştırılması, hangi kuramın daha keskin ve güvenilir sonuçlar verdiğinin belirlenmesi amaçlanmıştır. Veriler altı alt boyut, 28 maddeden oluşan ve toplamda 140 puan alınabilen Yapılandırmacı Öğrenme Ortamları Ölçeği ile MEB'e bağlı ortaokullarda öğrenim gören 306 öğrencinin analize katılımından elde edilmiştir. MTK'de test ve madde parametrelerinin kestirilebilmesi için tek boyutluluk şartı sağlanmıştır. Madde ayırt ediciliği incelendiğinde ayırt ediciliği düşük olan KTK'de 10 maddenin ölçekten çıkarılması gerekirken DTM'de 3 maddenin ölçekten çıkarılması gerekmektedir. Ayrıca madde ayırtıcılık parametreleri yüksek olduğundan maddeler farklı yetenek düzeylerindeki bireyler arasındaki farkları ortaya koyabiliyor demektir. İki kuramda kestirilen parametreler arasındaki korelasyon yüksek olduğundan kestirilen parametreler birbiri ile uyumludur. Araştırma sonucunda madde performansları ve test parametreleri hakkında DTM'nin daha fazla bilgi verdiği, KTK ile ulaşılan sonuçlar arasındaki uyumun da yüksek olduğu görülmüştür. Farklı örneklem ve farklı madde sayısı, beşten farklı noktada puanlanan bir ölçme aracının kullanılması önerilmiştir.

İlhan (2016)'ın araştırmasında matematik branşına ait açık uçlu sorular kullanılarak KTK ve çok yüzeyli Rasch modelinin yetenek kestirimlerindeki benzerlik ve farkların ortaya konması amaçlanmıştır. Veriler 8. sınıfta öğrenim gören 100 öğrenciden toplanmış, sorular dört matematik öğretmeni tarafından bütüncül rubrik kullanılarak puanlanmıştır. Araştırmada yetenek kestirimlerinde iki kuram arasındaki uyumun yüksek olduğu başka bir deyişle iki kuramda da öğrencilerin sıralamalarının çok benzer olduğu fakat yetenek kestirimine ait ortalamalar arasında (mutlak değerlendirme yapılması durumunda)

farklılıklar görüldüğü tespit edilmiştir. Ayrıca KTK'ye göre yetenek kestirimlerinin ölçüt geçerliğinin çok yüzeyle Rasch modelinde daha yüksek olduğu görülmüştür. Araştırma sonunda daha büyük örneklem, daha fazla madde ve analitik rubrik kullanılarak çalışmalar yapılması önerilmektedir.

2.1.4. Genellenebilirlik ve Madde Tepki Kuramlarını Karşılaştıran Araştırmalar

Kim ve Wilson (2009) araştırmalarında, öğrencilerin yazma becerilerinin değerlendirilmesi ile elde edilen verileri GK ve ÇDKRM ile karşılaştırmıştır. Her iki kuramın da araştırmacının amacına bağlı olarak önemli katkılarda bulunduğunu; yine her iki kuramın da bireyler, maddeler ve puanlayıcılar hakkında bilgi verdiğini, araştırmacıların çalışmak istediklerini bölümü net bir şekilde ortaya koymaları gerektiğini belirtilmiştir. Eğer araştırmacı, grupların nasıl hareket ettiği ile ilgileniyorsa GK'nin; bireysel performans ile ilgileniyorsa MTK yaklaşımlarının kullanılmasının daha iyi olacağını savunmaktadırlar.

Ure (2011)'nin araştırmasında misyoner öğretime ilişkin performansın değerlendirilmesinde, iki farklı puanlama koşulunun (kontrollü ve kontrollü olmayan) puanlayıcı davranışını ve değerlendirmenin güvenilirliğini nasıl etkilediğini araştırdığı çalışmada ÇDKRM ve GK'nin güvenilirlik belirleme performansları açısından karşılaştırmıştır. Bu çalışma sonucunda her iki modelden elde edilen güvenilirlik kestirimlerinin birbirinden istatistiksel olarak farklı olmadığını ancak bilgi edinmek amacı ile öncelikle GK'nin daha sonra ÇDKRM kullanılarak ayrıntıların incelenmesini önermektedir. Her iki modelin de farklı avantajlarından yararlanmanın önemi üzerinde durmuştur.

Arsan (2012)'ın araştırmasında buz pateninde hakem değerlendirmelerinin GK ve çok değişkenlik kaynaklı Rasch (ÇDKRM) modeli ile kestirilen güvenilirlik parametrelerinin incelenerek karşılaştırılması amaçlanmıştır. Veriler, Dünya buz pateni şampiyonalarında 2006-2011 yılları arasında tekler ve çiftler kategorilerinde yarışan toplam 397 sporcudan beş program bileşenindeki hakem değerlendirmeleri kullanılarak elde edilmiştir. 2006-2008 yıllarını puanlamak için 12, 2009-2011 yıllarını puanlamak için 9 hakem puanlamaya dâhil edilmiştir. Genellenebilirlik çalışması ve çok değişkenlik kaynaklı Rasch çalışması sonunda buz pateni performansının ölçülmesinde, hakemlerin tutarlı puanlamalar yaptığı ve sporcular arasındaki farklılıkların yansıtılabildiği görülmüştür. Ayrıca hakem sayısının 2009'dan sonra 9'a düşürülmesinin yapılan karar çalışması sonunda güvenilirlik katsayısını olumsuz yönde etkilemediği; GK ve çok değişkenlik kaynaklı Rasch modelinin birbirini tamamlar nitelikte sonuçlar verdiği tespit edilmiştir. Daha geniş kapsamlı değerlendirmeler yapabilmek için iki tekniğin de birlikte kullanılması önerilmektedir.

Alanyazın incelendiğinde üç kuramı karşılaştıran çalışmalara göre kuramları ikili karşılaştıran çalışmaların çoğunlukta olduğu görülmüştür. KTK ile GK'yi karşılaştıran araştırmalarda genellikle kullanılan ölçeklerden elde edilen puanların iç tutarlılık anlamında güvenirliliği, ikiden fazla puanlama yapıldığı durumlarda parametrik olmayan testler için daha uygun olan Kendall'ın uyum katsayısı ve GK'de daha çok çaprazlanmış desen kullanılarak elde edilen G ve Phi katsayıları hesaplanmıştır. Genel olarak GK'nin KTK'den daha ayrıntılı sonuçlar verdiği madde ve puanlayıcı sayıları arttığında genellenebilirlik ve güvenilirlik katsayılarının arttığı sonuçlarına ulaşılmıştır. Önerilerde ise madde, puanlayıcı ve desenlerin değiştirilerek yeni araştırmalar yapılması, MTK'de de analizlerin yapılarak karşılaştırılmasından bahsedilmiştir. KTK ile MTK'yi karşılaştıran araştırmalarda ise genellikle madde parametrelerinin karşılaştırılması amaçlanmış, iki kategorili maddelerin çoğunlukta olduğu testlerle büyük çalışma grupları ile çalışıldığı görülmüştür. Genel olarak madde parametrelerinde benzerlik olsa da MTK'nin KTK'ye

göre daha ayrıntılı sonuçlar verdiği, geçti-kaldı kararlarında KTK'nin kullanışlı olduğu, madde değişmezliği veya bireye uyarlanmış test gibi konularda MTK'nin üstün olduğu sonuçlarına ulaşılmıştır. Güvenirlik karşılaştırmayı temel alan çok fazla araştırma olmamasına rağmen madde bazında a ve b parametreleri incelenmiş, yalnız madde ve test fonksiyonları üzerinden güvenirlik yorumları yapıldığı görülmüştür. GK ve MTK'yi karşılaştıran araştırmalarda ise genellikle ÇDKRM'nin kullanıldığı, grup ve genel bilgi elde edilmek istendiğinde GK'nin, madde bazında ve değişkenlik kaynaklarının her biri için bilgi edinmek istendiğinde ise ÇDKRM'nin kullanılmasının daha uygun olduğu, birbirini tamamlar nitelikte sonuçlara ulaşıldığı gözlenmiştir. Değişkenlik kaynaklarının incelenmesi dışında, MTK için güvenirlik katsayılarının kestirime değinilmediği görülmüştür.

2.2. Kuramsal Çerçeve

Bu bölümde Klasik Test Kuramı, Genellenebilirlik Kuramı ve Madde Tepki Kuramının güvenirlğe bakış açılarından bahsedilmiştir.

2.2.1. Klasik Test Kuramı

KTK, gerçek dünyada var olan psikolojik özellikleri ölçme ile kuramsal yapılara bağlayarak bu özelliklerin açıklanmasını sağlar. Ölçülmek istenen, bu özelliğin bireyde ne derece var olduğudur. KTK'de bu özelliğin gerçek derecesi “gerçek puan” olarak tanımlanır ve bu puan teorik olarak bir bireyin aynı testi sonsuz kere alması sonucu elde edilen puanların aritmetik ortalamasını ifade eder. Uygulamada ise aynı testi bireye uygulama sayısı sınırlıdır. Bu nedenle, ideal olan bu kestirime kesin olarak ulaşmak söz konusu olamamaktadır. Bir sınav sonucunda elde edilen puan (her bir maddeden alınan puanların toplamı) bu sebeple kişinin gözlenen puanıdır. Gözlenen puan ise ölçmeye karışan (tesadüfi) hatalar sebebi ile hiçbir zaman pür hatasız bir sonuç belirtmez. KTK'ye

göre bazı sayıtlılar altında bu gözlenen puan (X), bireyin gerçek puanı (T) ile hata puanlarının (E) toplamına eşittir (Crocker & Algina, 2006, s. 105-111; Baykul, 2010, s. 107-123.).

$$X = T + E \quad \rightarrow \text{Klasik gerçek puan modeli}$$

Örneğin, Safa 10 soruluk bir testteki soruların 8 tanesini gerçekten biliyor olsun. Safa'nın gerçek puanını hesaplayabilmek için aynı testi üst üste uyguladığımızda 6, 9, 8, 7, 10 ... gibi puanlar alacaktır. Bu puanların Safa'nın gerçek puanı olan 8'den sapmasının sebepleri ölçme işlemine karışan tesadüfi hatalar (-2, +1, 0, -1, +2...) olacaktır. Bu durumda Safa'nın gerçek puanın yerine gözlenen puanlarından bahsetmek daha olası olacaktır. Safa'nın yüksek puan almış olması ölçülen özelliğe yüksek derecede sahip olduğunu, düşük puan alması ise aynı özelliğe düşük derecede sahip olduğu göstermektedir.

Bu ölçme işlemleri fiziksel bir ölçüm olduğunda gerçek puandan bahsetmek mümkün iken eğitimde ve psikolojide bu gerçeklik sapmakta, ölçme işlemine karışan hatalar sebebi ile ölçme sonucu özelliğin gözlenen değerinden bahsetmek daha doğru olmaktadır.

Tüm kuramlarda olduğu gibi KTK'de de gerçek puan ve hata puanının özelliklerinden bahsedilen ve test puanlarının güvenilirliğiyle ilgili çalışmalara fırsat sunan varsayımlar bulunmaktadır (Crocker & Algina, 2006, s. 111):

1. Bir evrendeki ölçmeler için hata puanlarının beklenen değeri sifıra eşittir.
2. Bir evrendeki ölçmeler için gerçek puanlarla hata puanları arasındaki korelasyon sifıra eşittir.
3. Bir evrendeki iki farklı ölçmeden birine ait hata puanları ile diğerine ait gerçek puanlar arasındaki korelasyon sifıra eşittir.
4. Bir evrendeki farklı ölçmelere ait hata puanları arasındaki korelasyon sifıra eşittir.

Görüldüğü gibi ölçme sonuçlarına karışan hatalar ve gerçek puan arasındaki ilişkiler (güvenirlik indeksi) KTK'nin temelini oluşturmaktadır. Ölçülen özelliğin gerçek değerine yaklaşabilmek ise hataların azlığını (hatalardan arınlığı) sağlamakla mümkün olmaktadır. Dolayısıyla ölçme aracından elde edilen puanın doğru yorumlanabilmesi için başka bir

deyişle güvenilirliğin yüksek olması için ölçümlerin tesadüfi hatalardan mümkün olduğu kadar arınık olması sağlanmalıdır. Bu arınlık KTK'ye göre teoride tekrarlı ölçümler yaparak pratikte ise paralel ölçmelerle sağlanabilmektedir (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, s. 109-111). Baykul (2010)'a göre iki ölçmenin paralellüğinden bahsedebilmek için (aynı özellik ya da benzer kapsam ölçülmek şartıyla) ölçmelerin hem aynı gerçek puanı vermesi hem de hata varyanslarının aynı olması gerekmektedir. Diğer bir deyişle, paralel ölçmeler “aynı değişkeni ölçmek şartıyla aritmetik ortalamaları ve varyansları eşit ölçmeler”dir (s. 132-133.). İşte bu paralel ölçmeler arasındaki korelasyon güvenilirlik katsayısını verecektir. Bu durumda da karşımıza paralel ölçümlerin gerçekte ne kadar paralel olabileceği sıkıntısı çıkmaktadır. Paralel ölçümler elde etmeye yaklaşabilmek için Crocker & Algina (2006), aşağıdaki önerilerde bulunmuşlardır.

- Test-tekrar test yöntemi: Aynı gruba iki ayrı durumda bir testin aynı formu uygulandığında test puanları arasındaki korelasyon kararlılık katsayısıdır. İki uygulama arasındaki zaman farkı grubun yeni öğrenmeler (olgunlaşma) gerçekleştiremeyeceği kadar kısa, test maddelerine verilen yanıtları hatırlayamayacakları (uygulama etkilerini yok etmeye izin verecek) kadar uzun olmalıdır (s. 133-134.).
- Paralel formlar yöntemi: Aynı gruba bir durumda, aynı içeriğe dayanan iki farklı test formu uygulandığında test puanları arasındaki korelasyon eşdeğerlik katsayısıdır. Formlar çok kısa sürede, sınava girenlere yorgunluk hissettirmeyecek kadar yeterli zamanda uygulanmalıdır (s. 132.).
- Karma yöntem: Bu iki yöntemin bir arada kullanıldığı durumdur. Grubun yarısına sırasıyla birinci ve ikinci formlar diğer yarısına ise ters sırayla formlar uygulandığında test puanları arasındaki korelasyon eşdeğerlik ve kararlılık katsayısıdır. Bu katsayı önceki iki katsayıdan daha düşük düzeyde değer verir (s. 134.).

- Gerçek paralel ölçmelerden elde edilebilecek korelasyon katsayısı ise kesinlik katsayısıdır ve diğer katsayılardan daha yüksek bir düzeyde değer verir (s. 117.).

Bu yöntemler birden fazla uygulamayı içerdiğinden kullanışlılığı düşürmektedir. Aşağıda tek uygulamaya dayalı yöntemlerden bahsedilmiştir. Bu yöntemde testin kendisiyle tutarlılığına bakılır ve bu katsayı iç tutarlık katsayısıdır (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 25).

- Yarıya bölme yöntemleri: Bu yöntemle testin kendi içinde alt testlerle paralel formlar oluşturulduğu varsayılır. Maddeler seçkisiz olarak ya da tek-çift diye gruplandırılarak ya da içerik olarak benzeşecek şekilde ikiye ayrılabilir. İki ayrı test oluşturulduğu varsayılarak testlerden birinin güvenilirliği Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanır. Tüm testin güvenilirliği ise Spearman-Brown formülüyle kestirilir.

Bu düzeltme formülünün kullanılmayacağı durumlarda testin güvenilirliği için fark puanlarını kullanan Rulon metodu ya da aynı formülden türetilen Guttman yöntemi ile tüm testin güvenilirliği hesaplanabilir (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, s. 26-28; Crocker & Algina, 2006, s. 136-138).

- Madde kovaryansına dayanan yöntemler: Bu yöntemde de testteki her bir maddenin bir alt test olduğu düşünülerek testin kendi içinde alt testlerle paralel formlar oluşturulduğu varsayılır. Her bir maddenin aynı ortalama ve varyansa sahip olduğu varsayımında ikili puanlanan maddeler için KR-20, buna ek olarak her maddenin gücü aynı ise KR-21 yöntemleri kullanılır. Çoklu puanlanan maddelerin bulunduğu testlerde ise güvenilirlik, Cronbach Alfa yöntemi ile kestirilir. Bireyler ve maddeler arasındaki değişimleri gözlemleme olanağı sunan Hoyt'un varyans analizi yönteminin ise paket programlarla güvenilirlik hesaplamada kullanışlı olduğu belirtilmiştir (Crocker & Algina, 2006, s. 135-141). Yazılı anlatım becerisi puanlama anahtarında maddeler 0-3 arasında puanlandığından KTK'ye göre iç

tutarlılık anlamında güvenilirlik kestiriminde Cronbach Alfa katsayısı hesaplanacaktır.

Puanlama anahtarları kullanılarak elde edilen ölçümlerin güvenirliliğinin yanında ölçmelerdeki hata kaynaklarından biri olan puanlayıcılar için de güvenilirlik incelemek gerekmektedir. Puanlayıcılardan kaynaklanan hatalarını azaltmak için ölçme araçları kullanılsa da puanlamaya hata karışması kaçınılmazdır. Puanlayıcılardan kaynaklanan hatalar, hale etkisi, öznellik hatası, cömertlik-cimrilik etkisi, sabit tepki hatası gibi farklılaşmaktadır (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 38).

Yazılı anlatım gibi bireylerin vereceği tepkilerin farklılaştığı ölçme durumlarında puanlayıcıların bireylere verdikleri puanlar arasındaki ilişki de incelenmelidir. Her ne kadar puanlama anahtarları (analitik, holistik, kontrol listesi, dereceleme ölçeği) kullanılsa da puanlayıcıların öznel yargıları da puanlamaya dâhil olacaktır. Bu durumda bir puanlayıcının aynı ürünü birden fazla kez puanlaması ya da aynı ürünleri birden fazla puanlayıcının puanlaması gibi farklı durumlarda puanlayıcı güvenirliliğinden bahsetmek gerekir. Puanlama güvenirliliği, puanlayıcıların puanlama üzerinde ne derece hemfikir olduğunu yansıtır (Ebel, 1965, s. 312-313). Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz'a (2017) göre puanlama güvenirliliğinin farklı durumlarda nasıl incelendiği aşağıda verilmiştir (s. 39-45):

Tablo 1.

Puanlama Güvenirliliğindeki Farklı Durumlarda Kullanılabilecek İstatistikler

Puanlayıcı sayısı	Puanlama sayısı ve türü	Kullanılması önerilen istatistik
1	2 kez ve toplam puan	Puanlar arası korelasyon, bağımlı gruplar için t-testi
	2 kez ve sıralama düzeyi	Çoklu uyum katsayısı
	2 kez ve sınıflama düzeyi	P ₀ uyum katsayısı, κ , Spearman'ın ϕ
	2 ve daha fazla ve toplam puan	Grup içi korelasyon katsayısı
2 veya daha çok	Birer kez ve sıralama düzeyi	Kendall'in uyum (W) katsayısı
	2 ve daha fazla ve toplam puan	Grup içi korelasyon katsayısı

Shoukri (2017) ise puanlayıcı güvenilirliğini kategori sayısı, ölçek türü ve puanlayıcı sayısına göre incelemiştir. İki puanlayıcının ya da iki puanlamanın olduğu durumlarda Cohen'in κ , ağırlıklandırılmış κ istatistiklerinin kullanılmasını önermiş, kategori sayısının sonsuza yaklaştığı durumda Cohen'in κ istatistiğinin grup içi korelasyon katsayısına yaklaştığını belirtmiştir (s. 213-230).

Erdoğan, Orekici Temel, Selvi & Ersöz Kaya (2017) yaptıkları araştırmada puanlama güvenilirliğinde bağımlı gruplar olduğunda Pearson korelasyon katsayısının, regresyon analizinin, bağımlı gruplar için t-testinin, kategorik ölçmelerde ki-kare ve Cohen'in kappa istatistiğinin kullanıldığını belirtmişlerdir. Bunlara ek olarak ölçme sayısı, puanlama sayısı, ölçek türü gibi değişkenlere göre kullanım alanı farklılaşan puanlama güvenirliliği hesaplama yöntemlerinden bazılarını grup içi korelasyon katsayısı, Scott'un p, Cohen's kappa, Fleiss kappa istatistiği, Krippendorff Alfa katsayısı, Ağırlıklandırılmış kappa istatistiği, Kendall'ın W katsayısı olarak belirtmişlerdir.

Sheskin (2003) ise ilişki (correlation/association) inceleneceğinde;

A. aralıklı-oranlı veriler için

- 1) iki değişken olduğunda Pearson korelasyon katsayısının,
- 2) puanlama sayısı ikiden fazla olduğunda grup içi korelasyon katsayısının,
- 3) çok değişken olduğunda çoklu ya da kısmi korelasyon katsayısının;

B. sıramala-dereceleme verileri için

- 1) iki değişken ya da puanlama sayısı iki olduğunda Spearman'ın sıra farkları korelasyonunun, Kendall'ın t, Goodman and Kruskal'ın gamma istatistiklerinin,
- 2) derecelemenin ikiden fazla gruptan elde edilmesinde Kendall'ın uyum katsayısının;

C. kategorik/sınıflama verileri için

- 1) iki kategorili/ayrık (dichotomous) (1-0, hep-hiç, vb.) değişken olduğunda olasılık, phi, Yule'nin Q, üstünlük oranı (the odds ratio), Cohen'in kappa, binom etki büyüklüğü (binomial effect size display) istatistiklerinin,
- 2) iki kategorili ayrık olmayan (nondichotomous) değişken olduğunda olasılık, Cramer'in phi, üstünlük oranı (the odds ratio) ya da Cohen'in kappa istatistiklerinin

D. diğer aralıklı-oranlı veriler için

- 1) değişkenlerden biri aralık-oran diğeri iki kategorik/nominal veri olduğunda Cohen'in d indeksinin,
- 2) değişkenlerden biri aralık-oran diğeri iki veya daha fazla kategorik/nominal veri olduğunda Eta kare korelasyonu

kullanılmasını önermiştir (s. 144).

2.2.2. Genellenebilirlik Kuramı

KTK'de bahsedilen yöntemler güvenilirlik hesaplamak için güçlü teknikler olsa da yeterli değildir. Bir grup öğrenciye yazdırılan makaleleri üç kişinin puanladığı durumda bireylere ait puanların ikiden fazla olması KTK'deki istatistiklerin uygulanmasını zorlaştıracaktır. Bir uzmanın özel eğitime ihtiyaç duyan bir bireyi tanılamak amacıyla uyguladığı alternatif formlar tamamıyla paralel olmadığı durumda güvenilirlik katsayısı ve ölçmenin standart hatası yanıltıcı sonuçlar verecektir. Hasta tutumlarının dört farklı danışman tarafından iki farklı seansta puanlandığı durumda hata varyansına puanlayıcılar arası, seanslar arası ve bunların ortak varyansları da dâhil olacaktır. Tüm bu sınırlılıklar hata kaynaklarını tek bir puan olarak ele alan KTK'nin tüm ölçme durumlarında kolaylıkla uygulanamadığının göstergesidir (Crocker & Algina, 2006, s. 157). Brennan (2011) bu durumu "Tek bir saatiniz varsa saatin kaç olduğunu bilirsiniz; iki farklı saatiniz varsa saatin kaç olduğundan asla emin olamazsınız!" sözleriyle anlatmıştır.

KTK’de sıklıkla kullanılan güvenilirlik belirleme yöntemleri ve hata kaynakları Tablo 2’de verilmiştir.

Tablo 2.

Sıklıkla Kullanılan Güvenirlik Belirleme Yöntemleri ve Hata Kaynakları

Yöntem	Güvenirlik Katsayısı	Hata Kaynağı
Test-tekrar test	Kararlılık	Zaman
Paralel formlar	Eşdeğerlik	Form
KR-20, KR-21, Cronbach Alpha	İç tutarlık	Görevler ya da Maddeler

(Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 214)

Tabloda görüldüğü gibi KTK’de her bir yöntem için hata kaynağı da sadece bir tane olmaktadır. Eğitimde ve psikolojide ise bu durum çok da mümkün görünmemektedir çünkü hata kaynakları ölçülen yapıdan, ölçmeyi yapan kişiye, ölçme aracından ölçme işleminin kendisine kadar farklılaşmaktadır. Bu nedenle, KTK’nin bir uzantısı olarak, zaman, form, maddeler gibi hata kaynaklarını hem ayrı ayrı hem de birlikte ele alarak tek bir güvenilirlik katsayısı hesaplamaya yarayan GK geliştirilmiştir (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 214; Cardinet, Johnson & Pini, 2010, s. 1-2).

KTK’de puanlar hakkında yorum yapabilmek için bireyin gruba göre karşılaştırılması (Göreceli değerlendirme: Okulun %5’inden daha başarılıdır.) kullanılırken GK, bu yoruma ek olarak bireyin gruptan bağımsız olarak düzeyinin belirlenmesine (mutlak değerlendirme) de imkân sağlar. Üstelik bu bilgiye ulaşmak için farklı istatistik yöntemleri kullanmak yerine aynı istatistikle karar vermek için yeterli bilgiyi sunar (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 5).

KTK’de güvenilirlik örnekleminden elde edilen toplam puanlara ait bir özellikken GK’de yapılan gözlemin güvenilirliği sonuç çıkarılmak istenen evrene bağlıdır. Aynı zamanda GK, güvenilirlik ve geçerlik arasındaki farkın güvenilir ölçme durumlarını belirleyerek nasıl ortadan kaldırılabilceğini de gösterir. Çünkü GK araştırmacıya gözlenen puanların evren

puanını temsil etme düzeyini de verir. Bu yönüyle yapı geçerliğine de kanıt oluşturur (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 215).

Shavelson ve Webb (1991), GK'nin KTK'nin bir uzantısı olduğunu aşağıdaki maddelerle belirtmiştir (s. 2-15):

1. GK'de tek bir analizle birden fazla hata kaynağı kestirilebilir.
2. Her bir değişkenlik kaynağının büyüklüğü belirlenebilir.
3. Bireylerin performansları hakkında hem bağıl hem de mutlak kararlar alınabilir ve iki farklı güvenilirlik katsayısı hesaplanabilir.
4. En uygun güvenilirlik katsayısının hesaplanabileceği karar çalışmaları yapılabilir.

Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz (2017) ise KTK ve GK arasındaki farklılıkları aşağıdaki gibi özetlemişlerdir (s. 216):

- KTK'de tek hata kaynağı vardır, farklı hata kaynaklarının birbirini etkilemesi söz konusu değildir; GK'de bir çok değişkenlik kaynağından gelen hataların ayrı ayrı ve birlikte değerlendirilmesi söz konusudur.
- KTK'de her hata kaynağı için farklı güvenilirlik katsayısı hesaplanır; GK'de değişkenlik kaynaklarının etkileşimleri de dâhil olmak üzere geniş kapsamlı bir güvenilirlik kestirimi yapılır.
- Yalnız GK'de güvenilirlik belirlemede mutlak ve bağıl değerlendirmeler için katsayılar üretilebilir.
- KTK'de geçerlik ve güvenilirlik ayrı ayrı hesaplanır; GK'de geçerlik ve güvenilirlik arasındaki fark ortadan kalkar ve güvenilirliğe yeni bir bakış açısı getirilir.
- KTK'de güvenilirliği arttırabilmek için değişkenlik kaynağı sayısının ne olabileceği hesaplanır (Spearman-Brown düzeltme formülü ile); GK'de yapılan karar çalışmaları ile istenen düzeyde güvenilirlik elde etmek için her bir değişkenlik kaynağının sayısı belirlenebilir.

Genellenebilirlik Kuramı KTK'nin yukarıdaki sınırlılıklarını gidermek için varyans analizi yöntemini temel alarak hata kaynaklarını çoklu bir şekilde incelemeye olanak sağlayan bir kuramdır. Bu kuram, ölçme konusu birey ya da nesnelerin gözlenen puanlarının evren puanlarına genellenebilmesini sağlayabilmek için ölçme sonuçlarını farklı varyans kaynaklarına ayırmayı amaçlamaktadır (Shavelson ve Webb, 1991, s. 1-5). KTK'de gözlenen puan (X), bireyin gerçek puanı (T) ile hata puanlarının (E) toplamı; " $X = T + E$ " olarak ifade edilmişti. GK'de ise gözlenen puan (X), KTK'deki gerçek puan (T) gibi yorumlanan evren puanı (μ_p) ile k farklı değişkenlik kaynağından gelen hata puanları (E_i ; $i = 1, 2, 3, \dots$) toplamı; " $X = \mu_p + E_1 + E_2 + E_3 + \dots E_k$ " olarak ifade edilir (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 3-5). Burada araştırmacı, ölçümün boyutlarını, yöntemini etkili bir şekilde tanımlayabilmek için hangi hata kaynaklarının öne çıktığına karar vermelidir (Brennan, 2011). Bu yöntemleri incelemeden önce GK'yi daha iyi anlayabilmek için bazı tanımları incelemek yararlı olacaktır.

Varyans kaynakları, *değişkenlik kaynağı (facet, yüzey)* olarak adlandırılır (Brennan, 2001, s. 5); KTK'deki varyans analizinde "faktör"e karşılık gelmektedir; madde, puanlayıcı, zaman gibi benzerlik gösteren ölçme durumları setidir. Bu değişkenlik kaynaklarının düzeyleri *koşul (condition)* olarak isimlendirilir (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 6). Örneğin, 10 öğrencinin matematik başarısını ölçmek için 10 maddelik bir test aldıklarını varsayalım. Öğrencilerin kâğıtları 3 öğretmen tarafından puanlansın. Bu ölçme durumunda hata kaynakları öğrenciler, puanlamayı yapan öğretmenler ve maddelerdir. Fakat değişkenlik kaynakları öğretmenler ve maddeler, her bir öğretmen de öğretmen değişkenlik kaynağının koşuludur. Burada öğrenciler değişkenlik kaynağı (yüzey) olarak düşünülmemektedir; çünkü bireylerarası farklılıklar istenen bir durumdur. Bireyler *ölçmenin objesi (object of measurement)* olarak isimlendirilir. Burada puanlanan, bireyler yerine maddeler olsaydı bu durumda ölçme objesi maddeler olacak değişkenlik kaynağı olarak ele alınmayacaktı (Atılğan, 2005; Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 8).

GK’de yüzeyler için *evren (universe)*, ölçme objesi için *popülasyon* terimleri kullanılır. *Evren puanı*, KTK’deki gerçek puan gibi yorumlanmakla birlikte, bir ölçme objesinin genelleme evrenindeki tüm ölçme koşullarından aldığı beklenen puanıdır (Brennan, 2001, s. 5-6). Elde edilecek tüm ölçmelerin popülasyonu *genelleme evreni (the universe of generalization)*, eldeki gözlemler yerine kullanılabilecek tüm gözlem popülasyonu *kabul edilebilir gözlemler evreni (the universe of admissible observation)* olarak adlandırılır (Atılgan, 2005; Crocker & Algina, 2006, s. 158-159). Tüm bu tanımların, GK’de kullanılan yöntemlerin incelenmesinde terminoloji sorununu en aza indirmesi beklenmektedir.

GK’de araştırmanın amacına göre farklı yöntem ve desenler bulunmaktadır:

- G (genelleme) ve D (karar) çalışmaları
- Yüzeylerin rastgele (random) ya da sabit (fixed) olması
- Verilerin yapısının çaprazlanmış (crossed) ya da yuvalanmış (nested) olması
- Genellenebilirlik (G) katsayısının ve mutlak güvenirlik (Phi- ϕ) katsayısının hesaplanması

G ve D Çalışmaları

Genellenebilirlik çalışmaları G (genelleme) ve D (karar) çalışmaları olmak üzere ikiye ayrılır.

G çalışmasında amaç, ölçme yapılan örnekleme evrene genellemektir. Verilen tepkilerin zaman içindeki istikrarı, en az bir araçla elde edilen sonuçların tutarlılığı, alt ölçekler ya da maddeler arasındaki ilişki ile ilgilenen çalışmalar birer G çalışmasıdır. D çalışmasında ise karar vermek için G çalışmasındaki bilgilerden yararlanılır. Seçme-yerleştirme için adayları belirlemek, deney gruplarını kıyaslamak, en az iki değişken arasındaki ilişkiyi araştırmak için D çalışmaları araştırmacıya veri sağlar (Güler, Kaya Uyanık & Taşdelen Teker, 2012; s. 5-6).

G çalışmasının amacı, yeterli düzeyde genellenebilirliğe sahip olacak şekilde D çalışmasının nasıl yapılacağına yardımcı olmaktır. Bu nedenle G çalışmalarının mümkün olan tüm olası desenleri öngörmesi, varyans kaynakları ile ilgili yeterince bilgi vermesi gerekir. Örneğin, bir araştırmacı iki farklı okul türüne aynı ölçeği uygulamış ve iki farklı puanlayıcıdan da ölçümler elde etmiş olsun. Eğer araştırmacının amacı, iki okul türünde de elde edilen puanların güvenilirliğini belirlemek ise G çalışması; iki okul türünü ölçülen özellik yönünden karşılaştırmak ise D çalışması yapması gerekir. Bu nedenle tek bir araştırma G ya da D çalışmasıdır demek uygun olmamakla birlikte, amaca göre farklılaşmaktadır (Crocker & Algina, 2006, s. 158-160; Shavelson & Webb, 1991, s. 83).

G çalışmaları, yüzeyleri belirterek o ölçmedeki zayıf ve güçlü yanları belirtirken D çalışmaları bu eksikliklerin iyileştirilmesi ya da tüm bu yanların yorumlanmasını; D çalışmasında yüzeylerin koşul sayısının değişiminde güvenilirliğin ne olacağı sorusunun cevabını vererek en uygun deseni sağlamaktadır (Cardinet, Johnson & Pini, 2010; s. 34-35). Genellenebilirlik katsayısını belirlemenin bir yolu, evren puan varyansını, beklenen gözlenen puan varyansına oranlamaktır. Ayrıca evren puanı ve varyansı genellenebilirlik evrenine bağlıdır. Bu durumun bir sonucu olarak hem farklı D çalışmaları için farklı genellenebilirlik katsayıları hem de evreni farklılaşan her D çalışması için yine farklı genellenebilirlik katsayıları hesaplanabilir (Güler, 2008; Crocker & Algina, 2006, s.158-160).

Yüzeylerin Rastgele (Random) ya da Sabit (Fixed) Olması

Araştırmacı, kendi amacı doğrultusunda bir yüzeyin tesadüfi ya da sabit olup olmadığına karar verebilmektedir. Evren, örneklem büyüklüğünden çok daha genişse, örneklem tesadüfi olarak seçilmiş ya da aynı evrendeki başka bir örneklem ile değiştirilebilecek durumdaysa yüzeyler tesadüfi olarak adlandırılır. Eğer araştırmacının amacı, örneklemde öte bir genelleme yapmak değilse yüzey sabit olarak ele alınır. Sabit yüzeylerde, koşul

sayıları sınırlandırılarak değiştirilmesi düşünülmemektedir. Örneğin; 10 kişinin yazılı kâğıdını iki puanlayıcının puanladığı ölçme durumunu göz önüne alalım. Burada amaç iki puanlayıcıyı evrendeki tüm puanlayıcılara genellemekse yüzey tesadüfi, sadece bu iki puanlayıcının verdikleri puanlarla ilgileniliyorsa yüzey sabit olarak tanımlanır (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 21).

Yüzeylerin tesadüfi ya da sabit olarak seçilmesi genellenebilirlik derecesi için önemlidir. En az bir yüzeyin sabit olarak alınması hata varyansını düşürerek güvenilirlik katsayılarının artmasını sağlasa da yapılacak yorumların daha sınırlı kalmasına sebep olur (Brennan, 2011).

Veri Yapısının Çaprazlanmış (Crossed) ya da Yuvalanmış (Nested) Olması

GK çalışmalarında iki farklı desenden yararlanılır. Bu desenler: Çaprazlanmış (crossed) ya da yuvalanmış (nested) desenlerdir. Verilerin yapısının çaprazlanmış olduğu durumda, değişkenlik kaynaklarının her bir koşulu gözlemlenmiş demektir. Verilerin yuvalanmış olduğu durumda, değişkenlik kaynaklarının her biri diğer değişkenlik kaynağının bazı koşullarını gözlemlemiş demektir. Değişkenlik kaynaklarından ve bunlar arasındaki etkileşimlerden gelen hataların kestiriminde en çok veri çaprazlanmış desen kullanıldığında elde edilebilir (Güler, Kaya Uyanık & Taşdelen Teker, 2012, s. 21).

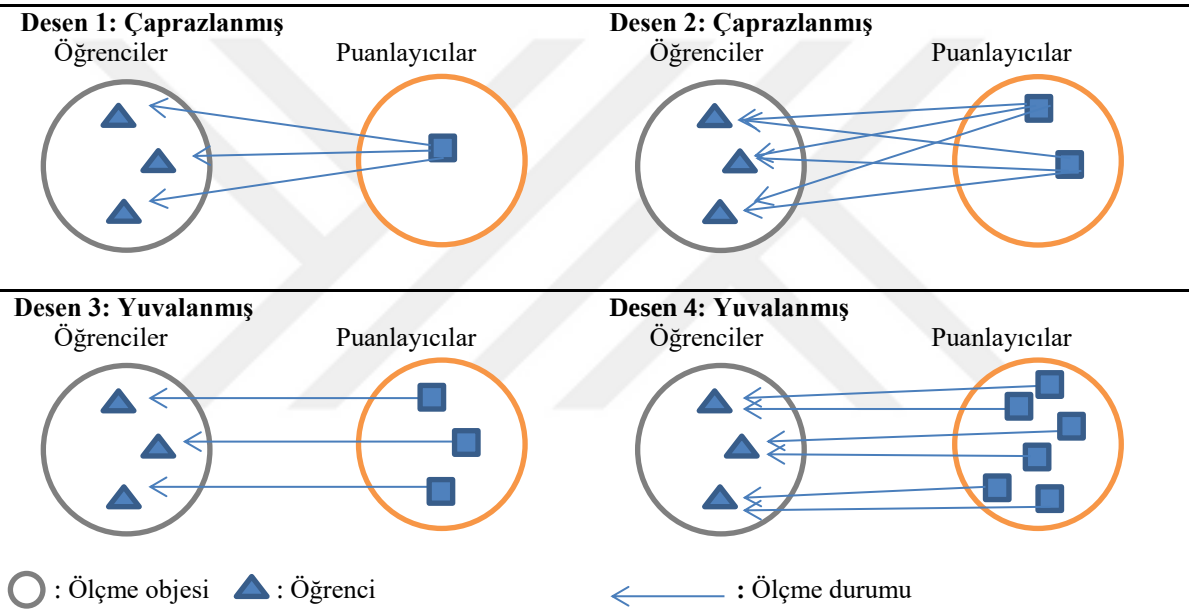
Bu durumu 3 öğrencinin notlarının puanlanacağını varsayalım. Crocker & Algina (2006) D çalışmaları için aşağıdaki desenleri belirtmişlerdir.

1. Her aday bir puanlayıcı tarafından puanlanır; bu puanlayıcı bütün adayları puanlar.
2. Her aday birçok puanlayıcı tarafından puanlanır; bütün puanlayıcılar her adayı puanlar.
3. Her aday farklı bir puanlayıcı tarafından puanlanır; her aday için sadece bir puanlayıcı vardır.

4. Her aday birçok puanlayıcı tarafından puanlanır; her aday için farklı puanlayıcılar vardır.

(s. 160-161)

Bu dört desende de tek yüzey (puanlayıcılar) ve bu yüzeyin üç koşulu vardır. İlk iki desende adaylar aynı ölçme durumuna, 3 ve 4. desende adaylar farklı durumlara tabidir. Burada birey (b), puanlayıcı (p) olmak üzere çaprazlanmış desen “ $b \times p$ ”; yuvalanmış desen “ $p : i$ ” şeklinde gösterilir. Daha net incelenebilmesi için bu durum Şekil 2’de şemalaştırılmıştır.



Şekil 2. Tek yüzeyli çaprazlanmış ve yuvalanmış desen gösterimi

Şekil 2’ye göre, öğrenciler aynı puanlayıcının ölçme durumlarına maruz kaldığında desen çaprazlanmış, farklı öğrenciler farklı puanlayıcıların ölçme durumlarına maruz kaldığında ise desen yuvalanmış olmaktadır.

Aşağıdaki tabloda iki yüzeyli çaprazlanmış ve yuvalanmış desenlere ait Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz (2017) tarafından hazırlanan puan tablosu verilmiştir (s. 219):

Tablo 3.

İki Yüzeyli Çaprazlanmış ve Yuvalanmış Desene Göre Puan Tablosu

Çaprazlanmış Desen								Yuvalanmış Desen									
$\mathbf{b \times m \times p}$								$\mathbf{b \times (m : p)}$									
		1. puanlayıcı				2. puanlayıcı						1. puanlayıcı		2. puanlayıcı			
Maddeler	1	2	3	4	1	2	3	4	Maddeler	1	2	3	4				
Birey 1	X	X	X	X	X	X	X	X	Birey 1	X	X	X	X				
Birey 2	X	X	X	X	X	X	X	X	Birey 2	X	X	X	X				
$2 \times 4 \times 2$								$2 \times (4 : 2)$									

Tablo 3'e göre, çaprazlanmış desende her iki puanlayıcı da iki bireyin dörder maddesini puanlamış ($2 \times 4 \times 2$) ve toplam 16 hücrelik veri elde edilmiştir. Yuvalanmış desende ise iki bireyin de ilk iki maddesini birinci puanlayıcı, 3 ve 4. maddesini ikinci puanlayıcı puanlamış ($2 \times (4 : 2)$) ve toplam 8 hücrelik veri elde edilmiştir.

Brennan (2011) iki yüzeyli çaprazlanmış desenlerde ($b \times m \times p$) değişkenlik kaynakları, serbestlik dereceleri, değişkenlik türleri, kestirilen varyans bileşenleri ve kestirilen kareler ortalamalarına ait tabloyu aşağıdaki gibi vermiştir (s. 7):

Tablo 4.

İki yüzeyli ($b \times m \times p$) Çaprazlanmış Desene İlişkin Değişkenlik Kaynağı, Serbestlik Dereceleri, Değişkenlik türü, Kestirilen Varyans Bileşenleri ve Kareler Ortalamaları

Değişkenlik Kaynağı/ Serbestlik Derecesi	Değişkenlik Türü	Kestirilen Varyans Bileşenleri	Kestirilen Kareler Ortalaması
Birey (b) $n_b - 1$	Evren puan varyansı, bireyler arasındaki farklılaşma	σ^2_b	$\sigma^2_{bnp.e} + n_m \sigma^2_{bp} + n_p \sigma^2_{bm} + n_m n_p \sigma^2_b$
Madde (m) $n_m - 1$	Bütün bireyler üzerindeki sabit etki, Maddelerin güçlük düzeylerindeki farklılaşma	σ^2_m	$\sigma^2_{bnp.e} + n_b \sigma^2_{mp} + n_p \sigma^2_{bm} + n_b n_p \sigma^2_m$
Puanlayıcı (p) $n_p - 1$	Bütün bireyler üzerindeki sabit etki, Puanlayıcıların puanları arasındaki farklılaşma	σ^2_p	$\sigma^2_{bnp.e} + n_b \sigma^2_{mp} + n_m \sigma^2_{bp} + n_b n_m \sigma^2_p$
$b \times m$ $(n_b - 1)(n_m - 1)$	Bir maddeden diğerine bireyin cevaplarındaki farklılaşma	σ^2_{bm}	$\sigma^2_{bnp.e} + n_p \sigma^2_{bm}$
$b \times p$ $(n_b - 1)(n_p - 1)$	Puanlama sırasında bireyler üzerinde puanlayıcılar arasındaki tutarsızlık	σ^2_{bp}	$\sigma^2_{bnp.e} + n_m \sigma^2_{bp}$
$m \times p$ $(n_m - 1)(n_p - 1)$	Bir maddeden diğerine puanlayıcı sıklığı arasındaki farkın neden olduğu sabit etki, Puanlamadaki katılık-cömertlik	σ^2_{mp}	$\sigma^2_{bnp.e} + n_b \sigma^2_{mp}$
$b \times m \times p.e$ $(n_b - 1)(n_m - 1)(n_p - 1)$	Artık/Kalan	$\sigma^2_{bnp.e}$	$\sigma^2_{bnp.e}$

(Brennan, 2011, s. 7)

Benzer şekilde yine Brennan (2011) iki yüzeyli yuvalanmış desenlerde ($b \times (m : p)$) değişkenlik kaynakları, serbestlik dereceleri, değişkenlik türleri, kestirilen varyans bileşenleri ve kestirilen kareler ortalamalarına ait tabloyu aşağıdaki gibi vermiştir (s. 436):

Tablo 5.

İki Yüzeyli ($b \times (m : p)$) Yuvalanmış Desene İlişkin Değişkenlik Kaynağı, Serbestlik Dereceleri, Değişkenlik türü, Kestirilen Varyans Bileşenleri ve Kareler Ortalamaları

Değişkenlik Kaynağı/ Serbestlik Derecesi	Değişkenlik Türü	Kestirilen Varyans Bileşenleri	Kestirilen Kareler Ortalaması
Birey (b) $n_b - 1$	Evren puan varyansı, bireyler arasındaki farklılaşma	σ^2_b	$\sigma^2_{b(m:p).e} + n_m \sigma^2_{bp} + n_m n_p \sigma^2_b$
Puanlayıcı (p) $n_p - 1$	Bütün bireyler üzerindeki sabit etki, Puanlayıcıların puanları arasındaki farklılaşma	σ^2_p	$\sigma^2_{b(m:p).e} + n_m \sigma^2_{bp} + n_b \sigma^2_{m:p} + n_b n_m$ σ^2_p
m : p $(n_m - 1) n_p$	Her bir puanlayıcının puanladığı maddelerden alınan puanlardaki farklılaşma	$\sigma^2_{m:p}$	$\sigma^2_{b(m:p).e} + n_b \sigma^2_{m:p}$
b × p $(n_b - 1) (n_p - 1)$	Puanlama sırasında bireyler üzerinde puanlayıcılar arasındaki tutarsızlık	σ^2_{bp}	$\sigma^2_{b(m:p).e} + n_m \sigma^2_{bp}$
b × (m : p).e $(n_b - 1) (n_m - 1) n_p$	Artık/Kalan	$\sigma^2_{bm:p.e}$	$\sigma^2_{b(m:p).e}$

(Brennan, 2011, s. 436)

Her genellenebilirlik katsayısı, evren puanın gözlenen puan varyansına oranı olmasına rağmen bu desenlerin her birinde gözlenen puan varyansının değişmesi sebebiyle genellenebilirlik katsayısı da değişmektedir (Crocker & Algina, 2006, s. 160).

Genellenebilirlik (G) Katsayısının ve Mutlak Güvenirlik (Phi-φ) Katsayısının Hesaplanması

GK’de görelî ya da mutlak değerlendirme yapıp yapılmayacağına göre hesaplanması gereken katsayılar da değişmektedir. G katsayısı, KTK’deki güvenirlik katsayısı gibi düşünülebilir; özellikle maddelerin tek yüzey olduğu, görelî değerlendirmenin yapılacağı tek değişkenlik kaynağının maddeler olduğu, karar çalışmasındaki düzeylerle madde

sayısının eşit olduğu durumlarda KTK'deki Cronbach alfa katsayısı ile aynı değeri vermektedir (Sudweeks, Reeve & Bradshaw, 2005). G ve alfa katsayıları arasındaki temel fark ise G katsayısında hata kaynaklarının ayrıştırılarak, hem ayrı hem de birlikte açıkladıkları varyans oranını vermesidir. (Cardinet, Johnson & Pini, 2010, s. 5-7).

G katsayısı 0-1 aralığında değer alır; 0,80 ve üzeri kabul edilebilir değerlerdir. G katsayısı görelî (bağıl) güvenilirliği kestirmek için kullanılır, aşağıdaki formülle hesaplanır:

$$g = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{Rel}^2}$$

Paydadaki görelî (relative) hata varyansı (σ_{Rel}^2) sadece ölçme objesiyle olan etkileşimleri temsil eder. Bu varyans bileşenleri göreceli kararlarla ilgili ölçme hatasına katkıda bulunur. Dolayısıyla görelî hata varyansı, modelin ölçme objesi ile yüzeylerin arasındaki etkileşimi temsil eden tüm varyans bileşenlerinin toplamıdır. Bu istatistiğin karekökü, KTK'deki standart ölçme hatasına benzemektedir (Brennan, 2000). Görelî hata varyansı $b \times m \times p$ deseni için aşağıdaki formülle hesaplanır:

$$\sigma^2(\delta) = \sigma^2(bm) + \sigma^2(bp) + \sigma^2(bmp)$$

Phi (φ) katsayısı da 0-1 aralığında değer alır. φ katsayısı mutlak güvenilirliği kestirmek için kullanılır, aşağıdaki formülle hesaplanır:

$$\varphi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{\Delta}^2}$$

Paydadaki mutlak hata varyansı (σ_{Δ}^2) öğrenciler için varyans bileşeni hariç tüm varyans bileşenlerinin toplamıdır ve hata varyansı olarak kabul edilmediğinden öğrencilere ait varyans bileşeni hariç tutulur. Çünkü; GK'de ölçme objesinden kaynaklanan varyans dışındaki tüm varyans bileşenleri, derecelendirmeler mutlak kararlar almak için temel olarak kullanıldığında ölçme hatasına katkıda bulunur. Mutlak hata varyansı, farklı öğrencilerin ortalama derecelendirmelerindeki sistematik değişimi temsil eder ve

KTK'deki gerçek puan varyansına benzer (Sudweeks, Reeve & Bradshaw, 2005). Mutlak hata varyansı, ölçme objesinin gözlenen ve evren puanları arasındaki farkın varyansıdır. $b \times m \times p$ deseni için aşağıdaki formülle hesaplanır:

$$\sigma^2(\Delta) = \sigma^2(m) + \sigma^2(p) + \sigma^2(bm) + \sigma^2(bp) + \sigma^2(mp) + \sigma^2(bmp)$$

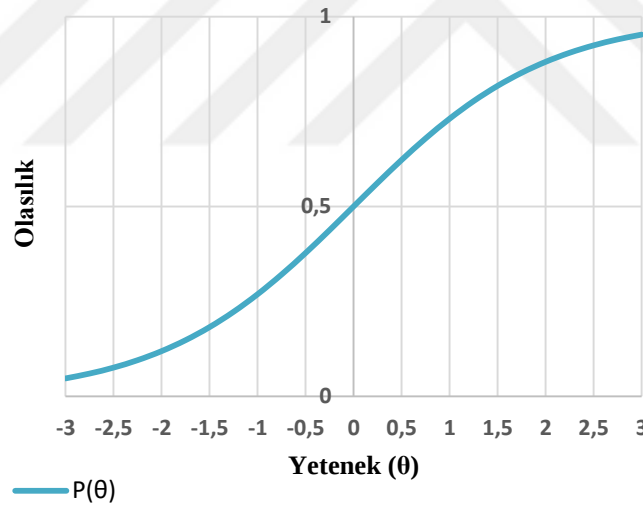
G kuramında bu iki katsayı dışında belli bir kriter puan (60 puan gibi, $\lambda=60$) için güvenilirliği kestiren $\phi(\lambda)$ katsayısı da mevcuttur. Yani bireylerin belli bir kriter puanının altında veya bu puana eşit ve daha iyi olma durumlarını belirleyebilmedeki güvenilirlik katsayısı da kestirilebilmektedir. $\phi(\lambda)$ katsayı yukarıda bahsedilenlerden farklı bir bakış açısı getiren kriter referanslı ölçmeler (criterion-referenced measurement) için hesaplanan güvenilirlik katsayısıdır. $\phi(\lambda)$ katsayısı, ölçme objelerinin görelî olarak aralarındaki mesafeyi ve ölçekteki yerini (λ 'dan ne kadar uzakta olduğunu) tahmin etmeyi sağlar. λ , ölçme kriteri veya kesme puanıdır (Cardinet, Johnson & Pini, 2010, 28-29).

2.2.3. Madde Tepki Kuramı

KTK, bireyde var olan özelliğın derecesini gerçek puan ile ilişkilendirerek gözlenen puanın, bu özelliğın varlığını ne kadar ortaya çıkardığı ile ilgilenir. Fakat gerçek puan ve gözlenen puanların hiç biri bireyin yeteneğî ile ilişkilendirilememektedir. Çünkü KTK'de zor bir testten düşük puan alanların gerçek puanı düşük olurken, kolay bir testten yüksek puan alanların gerçek puanı da yüksek olmaktadır. Oysa bireyin yeteneğî değişmemektedir (Hambleton & Jones, 1993). Bu nedenle bireyin yeteneğinin teste bağımlı olması sınırlılığını gidermek amacıyla MTK geliştirilmiştir. KTK'den farklı olarak, MTK'de bireylerin yetenek kestirimleri testten, madde parametreleri de testin uygulandığı gruptan bağımsızdır. MTK'de bireyin testten aldığı puan yerine her bir maddeye verdiği yanıtın doğru ya da yanlış olması ile ilgilenilir; her bir madde için o maddeyi doğru yanıtlama ihtimali hesaplanır ve bu hesaplamada madde parametreleri ile gizil yetenek de dâhil edilir. Ölçmenin standart hatası (SEM) KTK'de her zaman farklı yetenek seviyeleri (iyi-kötü

puan alan öğrenciler) için sadece hata puanları dağılımının birimi kestirilir, örneklem genelindeki hata kişinin hatasına eşittir; MTK’de ise her bir yetenek düzeyi için ölçmenin standart hatası da ayrı hesaplanır, madde güçlüğü yetenek seviyesine yaklaştıkça ve maddelerin ayırt ediciliği arttıkça ölçmenin standart hatası azalır (Hambleton, Swaminathan & Rogers, 1991, s.1-5).

MTK’de yeteneğe sahip olan kişilerin maddeyi doğru, sahip olmayan kişilerin ise yanlış yanıtlama olasılığının daha yüksek olması beklenir. θ -theta yetenek düzeyi olmak üzere $P(\theta)$, her bir yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığını göstermektedir. Dolayısıyla düşük yetenek düzeyindeki öğrenciler için $P(\theta)$ sıfıra yaklaşıırken, yüksek yetenek düzeyindeki öğrenciler için $P(\theta)$ bire yaklaşıır. Bu durum grafikte gösterildiğinde tek bir madde için oluşan “S” eğrisi, Madde Karakteristik Eğrisi (MKE) olarak adlandırılır (Baker, 2016, s. 1-5).



Şekil 3. Madde karakteristik eğrisi

Yukarıdaki şekilde yatay eksen yetenek düzeylerini, dikey eksen maddenin doğru yanıtlanma olasılığını göstermektedir, bireylerin yetenek düzeyleri arttıkça eğri de monoton artan özelliktedir.

Bu eğriye şeklini veren iki farklı özellik vardır: Madde güçlüğü ve madde ayırt ediciliği. Madde güçlüğü (b parametresi) maddenin hangi yetenek düzeyinde işlevsel olduğunu, eğrinin konumunu gösterirken; madde ayırt ediciliği (a parametresi) MKE’nin orta

noktasındaki eğimi ifade eder. Madde güçlüğü arttıkça eğri yetenek ekseninin sağ tarafına kayar; madde ayırt ediciliği arttıkça da eğrinin orta noktasındaki eğim artar (dikleşir) (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 131-132).

Maddeyi doğru yanıtlama olasılığı ile yetenek düzeyi arasındaki ilişkileri matematiksel olarak detaylı bir şekilde ifade edebilmek için normal ogive (sivri tepeli kemer şeklindeki eğri) ve lojistik fonksiyon üzerine temellenen MKE modelleri tanımlanmıştır. Normal ogive ve lojistik fonksiyon, birimlerin (1,7 lojit = 1 probit) farklı olması sebebiyle kullanılan formülü farklılaştırırsa da model değişimine gereklilik oluşturmamaktadır (Ayala, 2009, s. 21). Bu modeller maddelerin puanlanma kategorilerine göre farklılaşmaktadır. Normal ogive modelleri için Ayala (2009)'nın ilgili bölümleri incelenebilir (s. 360-372). Bu çalışmada lojistik fonksiyon üzerine geliştirilen formüller göz önünde bulundurulmuştur.

MTK modellerine geçmeden önce tüm modellerin kullanılabilmesi için gerekli olan normallik varsayımı dışında iki varsayımın da karşılanması gerekir: Yerel bağımsızlık (local independence) ve tek boyutluluk (unidimensionality). Yerel bağımsızlık, testte yer alan bir maddenin aynı yetenek düzeyindeki bireyler tarafından çözülebilmek ihtimalinin testteki diğer maddeye bağlı olmaması durumudur. Tüm modellerde maddelere verilen yanıtları etkileyen tek faktörün bireyin yeteneği olması gerekmektedir. Bu nedenle testteki bir maddenin diğer bir maddeye yanıt oluşturmayaacak şekilde düzenlenmesi beklenir. Tek boyutluluk ise testteki maddelerin birlikte tek bir yeteneği ölçmesi, tek bir örtük özellik için maddelerin istatistiksel olarak bağımlı olmasıdır (Hambleton, Swaminathan & Rogers, 1991, s. 9-10). Tek boyutluluk ile yerel bağımsızlık birbiri ile ilişkili kavramlardır. Genellikle testin boyut sayısı yerel bağımsız olabilen gizil özellik sayısına eşittir. Bir testin maddeleri tüm popülasyon için istatistiksel bağımlı ve tek bir örtük özelliğe göre homojen kişilerden oluşan bir alt grupta maddeler tek bir örtük özellik ortaya çıkarmışsa test tek boyutlu demektir. Tek boyutluluk varsayımı karşılandığında yerel bağımsızlık varsayımı da karşılanmış olur fakat tersi her zaman doğru değildir. Eğer madde çiftleri yerel bağımsız

olduğunda tek bir örtük özellik oluşuyorsa tek boyutluluk varsayımı da karşılanıyor denir (Crocker & Algina, 2006, s. 42-46).

İki Kategorili Maddeler İçin Modeller

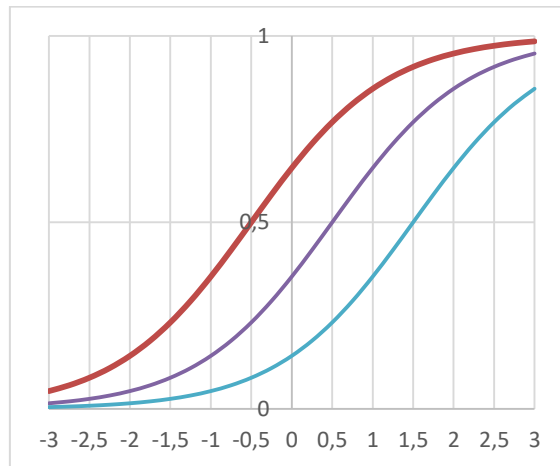
Maddeler doğru-yanlış, evet-hayır, katılıyorum-katılmıyorum gibi 1-0 şeklinde iki kategoride puanlandığında kullanılan tipik MKE modelleri verilmiştir.

Bir Parametrelili Lojistik Model (1PL) ve Rasch Modeli

Bir parametrelili modelde aşağıdaki lojistik fonksiyon formülü kullanılmaktadır.

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b_i)}}$$

1PL modelde tüm maddeler için a parametresinin (madde ayırt ediciliğinin) aynı (Rasch modelinde özel olarak $a = 1$) olduğu kabul edilir. Formüldeki b değeri madde güçlük derecesidir. 1PL modelde b parametresi maddeyi doğru yanıtlama olasılığının 0,5 olduğu yetenek düzeyini ifade eder; teorik olarak $(-\infty, +\infty)$, pratik olarak $[-3, +3]$ aralığında değerler alır.

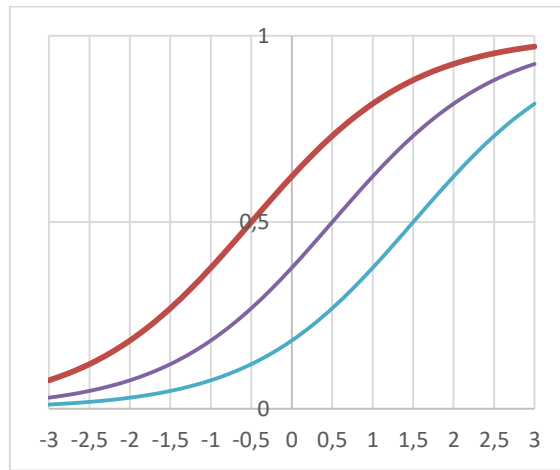


Şekil 4. b parametreleri farklı $a = 1,2$ olan üç maddeye ait MKE

Şekil 4’te madde güçlükleri farklı ($b = -0,5$; $b = 0,5$; $b = 1,5$) olan bir parametrelili model için MKE’ler gösterilmiştir. Burada en zor madde, doğru yanıtlanma olasılığının 0,5 olduğu durumda θ düzeyinin 1,5 olduğu başka bir deyişle $b = 1,5$ olan mavi eğridir. Benzer şekilde en kolay madde de kırmızı ile belirtilmiştir. Bu maddelerin her birinin kendi yetenek düzeyinin üzerinde yeteneğe sahip bireyler tarafından doğru yanıtlanma olasılığı 0,5’ten 1’e artarken, kendi yetenek düzeyinin altındaki bireyler tarafından doğru yanıtlanma olasılığı 0,5’ten 0,0’a azaltmaktadır. Bu durumda her bir eğri en fazla bilgiyi kendi yetenek düzeyi civarında vermektedir (Erkuş, Sünbül, Ömür Sünbül, Aşiret & Yormaz, 2017, s. 131-133). Zor maddelerin b değerleri daha yüksektir. KTK’de ise zor maddelerin madde güçlüğü (p değeri) daha düşüktür (DeMars, 2016, s. 12). Burada fonksiyonlar mavi eğriye sahip maddenin diğer iki maddeden daha zor olduğunu göstermektedir. Kırmızı eğriye sahip madde diğer iki maddeye göre düşük yetenek düzeyindeki bireyler için daha işlevseldir (Baker, 2016, s. 4-8).

Rasch modelinde ise tüm maddelerin ayırt edicilikleri aynı ve bire eşit olarak kabul edilir. $a=1$ için formül ise aşağıdaki gibi olur:

$$P(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$



Şekil 5. b parametreleri farklı $a = 1$ olan üç maddeye ait MKE

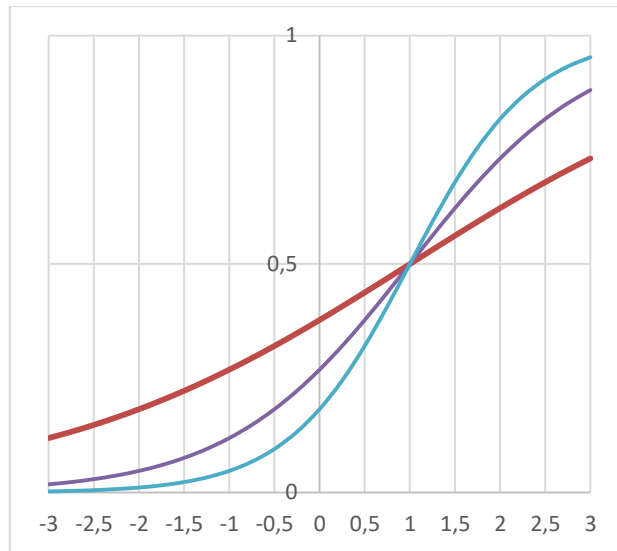
Şekil 4 ve 5 incelendiğinde 1PL ve Rasch modellerinde a parametrelerinin eşit olması sebebiyle maddelere ait eğriler kesişmezler. Bu durumda da farklı yetenek düzeyindeki bireyler için farklı iki maddenin güçlük düzeyi de aynıdır (objective measurement). Bir kişinin yeteneğini hesaplamak için toplam puanları kullanılır (sufficiency), toplam puanları aynı olan iki kişinin yetenek seviyeleri de aynıdır, dolayısı ile cevap örüntülerinin aynı olması beklenmemektedir (DeMars, 2010, s. 16-17).

İki Parametrelili Lojistik Model (2PL)

a madde ayırt edicilik, b madde güçlük parametreleri olmak üzere lojistik fonksiyon aşağıdaki gibi ifade edilir:

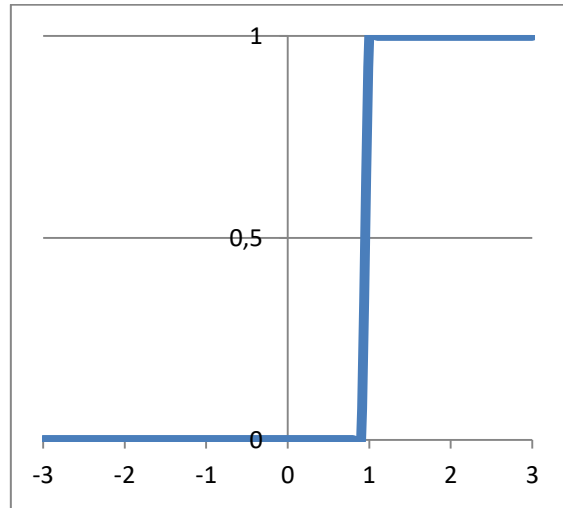
$$P(\theta) = \frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

Burada b parametresine ek olarak yetenek düzeyinde MKE'nin eğimi madde ayırt ediciliğini a parametresini verir; teorik olarak $(-\infty, +\infty)$, pratik olarak $(-2,80; +2,80)$ aralığında değerler alır. a parametresi θ için bir ağırlıklandırma yapmaktadır. Bu durumda ayırt ediciliği yüksek olan maddeler daha iyi bir yetenek kestirimi sağlar (Baker, 2016, s. 17-23).



Şekil 6. b parametreleri aynı a parametreleri farklı olan üç maddeye ait MKE

Şekil 6’da madde güçlükleri aynı ($b = 1$), madde ayırt edicilikleri farklı ($a = 0,5$; $a = 1$; $a = 1,5$) olan MKE’ler gösterilmiştir. Burada fonksiyonlar 3. maddenin (mavi eğri) diğer iki maddeden $\theta = 1$ yetenek düzeyinde bilen ile bilmeyeni daha iyi ayırt ettiğini göstermektedir. KTK’de de maddelerin örneklemede benzer doğru cevap verme oranlarına sahip oldukları düşünüldüğünde 3. maddenin nokta-çift serili korelasyon değeri diğer maddelerden daha büyük olacaktır (DeMars, 2016, s. 11). Fakat burada 1PL ve Rasch modellerinden farklı olarak maddelerin zorlukları farklı yetenek düzeyleri için farklılaşmaktadır. Örneğin; $b = 1$ yetenek düzeyinin altında yetenek düzeyinde olan bireyler için kırmızı renk ile belirtilen madde en kolay madde iken $b = 1$ yetenek düzeyinin üstünde yetenek düzeyinde olan bireyler için mavi renk ile belirtilen madde en kolay maddedir. Ayrıca aynı yetenek düzeyindeki bireylerin sadece toplam puanları değil tüm cevap örüntülerinin aynı olması gerekmektedir (DeMars, 2010, s. 9-12). Dolayısı ile 2PL modeli için “objective measurement ve sufficiency” kavramlarından bahsedilememektedir. Herhangi bir maddenin ayırt ediciliğinin mükemmel olması durumunda eğri, ilgili yetenek düzeyinde düz bir çizgi şeklinde olacaktır. Fakat bu durumda madde, o yetenek düzeyinin altında veya üstünde kalan yetenek seviyelerinde bireyleri ayırt edemiyor olacaktır (Baker, 2016, s. 4-8).



Şekil 7. $\theta = 1$ yetenek düzeyinde mükemmel ayırt ediciliğe sahip maddeye ait MKE

Şekil 7’de $\theta = 1$ yetenek düzeyinde mükemmel ayırt ediciliğe sahip bir maddenin MKE’si verilmiştir. Bu fonksiyona göre yetenek düzeyi birden küçük olan bireylerin maddeyi yanıtlama olasılığı her birey için sıfıra, birden büyük olan bireylerin maddeyi yanıtlama olasılığı her birey için bire eşittir. Dolayısı ile $\theta = 1$ yetenek düzeyi dışındaki bireyler arasındaki farklılaşmayı ortaya çıkaramamaktadır (Crocker & Algina, 2006, s.349-350).

Üç Parametrelili Model (3PL)

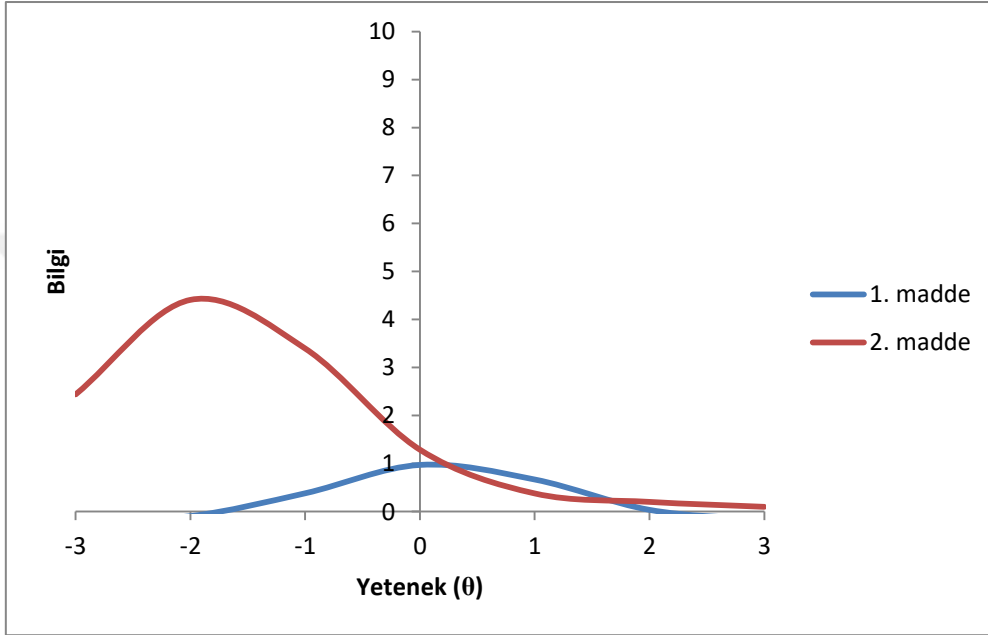
Diğer iki modelden farklı olarak bu modelde, olabilecek en düşük yetenek düzeyindeki bireylerin bile maddeyi cevaplayabilme olasılığı (pseudo guessing) olarak tanımlanan c parametresi (Kingston, 1985) de formüle dâhil edilerek iki parametrelili model genişletilmiştir.

$$P(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_i - b_i)}}$$

c parametresi ile yetenek düzeyi arasında teorik olarak bir ilişki olmaması sebebi ile yüksek ve düşük yetenek düzeylerindeki bireylerin maddeyi tesadüfi olarak doğru cevaplama olasılıkları aynıdır. c parametresi teorik olarak [0, 1], pratik olarak [0, 0,35] aralığında değerler alır, c = 0 olduğunda ise formül iki parametrelili lojistik modeldeki ile aynı olacaktır. Fakat b parametresi bir ve iki parametrelili modellerde eğrinin orta noktasını belirlemek için maddeyi doğru yanıtlama olasılığının 0,5 olduğu yetenek düzeyini belirtirken üç parametrelili modelde eğrinin orta noktası için c birimlik öteleme ile maddeyi doğru yanıtlama olasılığının $\frac{1+c_i}{2}$ olduğu yetenek düzeyini belirtecektir. Benzer bir yaklaşımla a parametresi için de $\theta = b$ noktasındaki eğim için $\frac{a_i(1-c_i)}{4}$ formülü kullanılmalıdır (Baker, 2016, s. 25-26).

Madde ve Test Bilgi Fonksiyonları

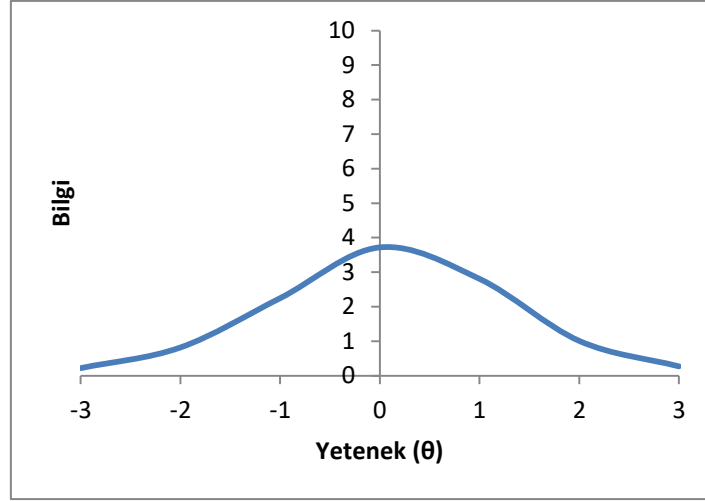
Her bir madde farklı θ değerleri için farklı miktarlarda bilgi verir; kolay maddeler daha düşük yetenek düzeyinde, zor maddeler ise daha yüksek yetenek düzeyindeki kişiler hakkında daha fazla bilgi verir. Madde bilgi fonksiyonu $I(\theta)$ en fazla bilgiyi ise o yetenek düzeyindeki bireyler için verir.



Şekil 8. Madde bilgi fonksiyonu

Şekil 8'deki madde fonksiyonları incelendiğinde 1. maddenin yaklaşık $\theta = 0,2$; 2. maddenin yaklaşık $\theta = -1,8$ yetenek düzeyinde en fazla bilgiyi verdiği görülür. Ayrıca, kendi yetenek düzeylerinde 2. madde 1. maddeden daha fazla bilgi vermektedir.

Test bilgi fonksiyonu ise madde setlerinden oluşan testin sağladığı bilgi miktarıdır ve o yetenek düzeyindeki madde bilgi fonksiyonlarının toplamıdır. Bu nedenle testteki madde sayısı arttıkça bilgi miktarı da artmaktadır (Baker, 2016, s. 101-104).



Şekil 9. Test bilgi fonksiyonu

Şekil 9’da verilen test bilgi fonksiyonu 10 maddelik bir testin yetenek düzeylerine göre sağladığı bilgi gösterilmektedir. Bu test, en fazla bilgiyi yaklaşık $\theta = 0,2$ düzeyinde vermektedir. Daha üst ve daha alt yetenek düzeylerinde uçlara doğru gidildikçe testin verdiği bilgi giderek azalmaktadır. İdeal bir test bilgi fonksiyonu düz bir çizgi gibi olsa da pratikte testin amacına göre fonksiyon farklılaşmaktadır: Burs vermek için yapılan bir testte kişileri ayırt eden kesme noktasında en yüksek değere sahip olan test bilgi fonksiyonu daha uygundur (Baker, 2016, s. 104-112).

Test bilgi fonksiyonu, katılımcılar için grafiğin zirve yaptığı yetenek düzeyi civarında oldukça duyarlıdır. Bu düzeylerin dışındaki katılımcılar için ise aynı hassasiyetle ölçüm yapamaz. Bu nedenle fonksiyonun sağladığı bilgi miktarı ile yetenek kestirimindeki değişkenlik arasındaki ilişki için ölçmenin standart hatası ($SE(\theta)$) hesaplanmaktadır.

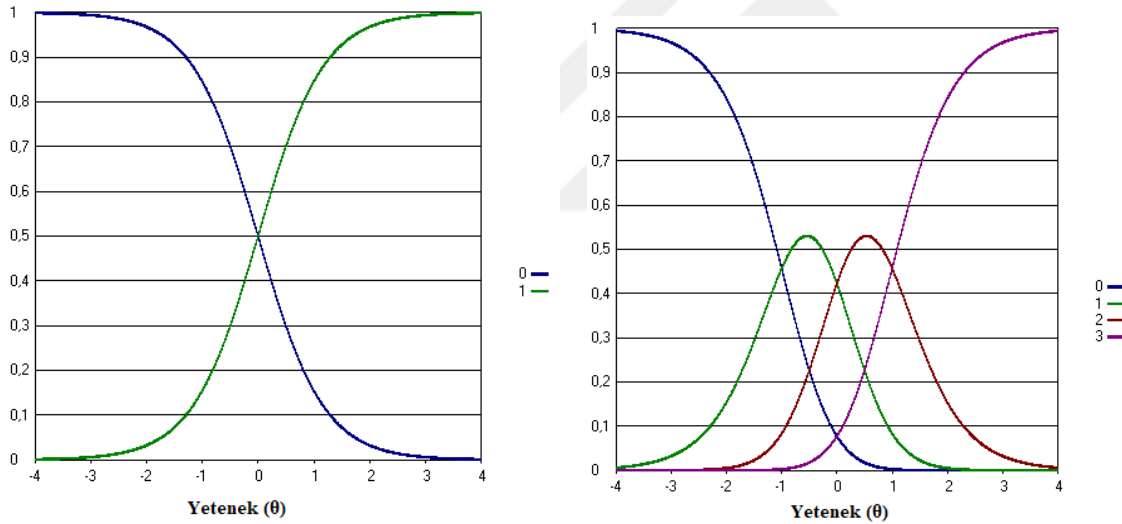
$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

2 PL’de a parametresi arttıkça maddenin verdiği bilgi miktarı artar, 3PL’de c parametresi arttıkça maddenin verdiği bilgi miktarı azalır. Madde bilgisi arttıkça test bilgisi artar, dolayısı ile ölçmenin standart hatası azalır.

Çok Kategorili Maddeler İçin Modeller

İkiden fazla puanlama yapılan maddelerde (likert tipi ölçekler, analitik/holistik rubrikler, yazılı yoklamalar, vb.) 1PL, 2PL ve 3PL modellerinin kullanılması uygun değildir çünkü bir kategorideki özellikleri bilmek üst kategorideki özelliklerin bilineceğini göstermemektedir. Bu sebeple her kategori fonksiyonu ayrı ayrı belirlenmelidir. Bu durumda çok kategorili MTK (polytomous) modelleri 1PL, 2PL ve 3PL modellerinin genişletilmiş hâli olarak karşımıza çıkmaktadır (Tang, 1996; Yürekli, 2010).

Çok kategorili MTK’de de iki kategorili maddeler için modellerde olduğu gibi madde ve test bilgi fonksiyonları kullanılarak madde hakkında bilgilere ulaşılmaktadır. Aşağıda iki kategorili puanlanan bir madde karakteristik eğrisi ile dört kategorili puanlanan bir madde karakteristik eğrisi gösterilmiştir:



Şekil 10. İki kategorili ve dört kategorili puanlanan maddelere ait madde bilgi fonksiyonları

Yukarıdaki şekiller incelendiğinde çok kategorili modeldeki en düşük ve en yüksek yetenek seviyelerinde grafiklerin iki kategorili modeldeki ile benzer olduğu farklılaşmanın ara kategorilerden kaynaklandığı görülmektedir. Çok kategorili puanlanan madde bilgi fonksiyonu incelendiğinde yetenek düzeyi yaklaşık -1 olduğunda (düşük yetenek düzeylerinde) bireylerin maddeyi 0; -1 ile 0 arasındaki yetenek düzeylerinde bireylerin maddeyi 1; 0 ile 1,5 arasındaki yetenek düzeylerinde bireylerin maddeyi 2; 1,5’tan yüksek

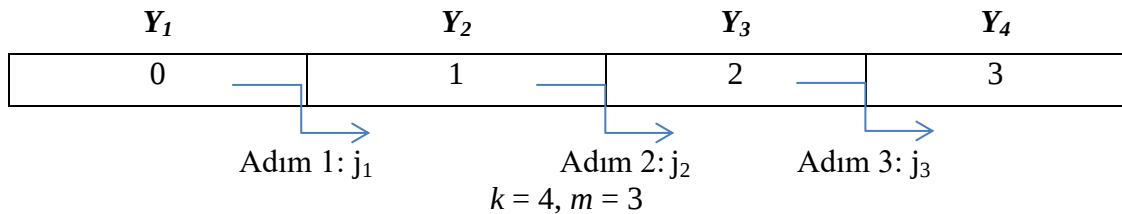
yetenek düzeylerindeki bireylerin ise maddeyi 3 olarak puanlama ihtimalinin daha yüksek olduğu görülmektedir (DeMars, 2010, s. 22-25).

İki kategorili MTK modellerinde maddenin doğru yanıtlanma olasılığı üzerinden yorumlar yapılırken çok kategorili MTK modellerinde maddenin üst yanıtları alabilme olasılığından bahsedilmektedir. Bu yorumları yapabilmek için temelde iki farklı yaklaşım vardır: Yığılmalı (cumulative) ve komşu kategori (adjacent categories) yaklaşımları. Yığılmalı yaklaşıma göre bir kişinin verdiği yanıtta daha büyük yanıtları verme olasılığı incelenirken; komşu kategori yaklaşımına göre kişinin verdiği yanıtta bir büyük yanıtı (komşu kategori) verme olasılığı incelenir (Ayala, 2009, s. 163; Ostini & Nering, 2006, s. 81-85).

Çok kategorili MTK modellerinden önce terminolojiden bahsetmek anlaşılabilirliği arttırmayı sağlayacaktır:

- Y = Çok kategorili puanlanan bir maddeye verilen yanıt
- k = Seçeneklerin (kategorilerin) sayısı
- j = Bir maddenin puanlanmasındaki bir adım
- m = Bir maddedeki adım sayısı

0, 1, 2 ve 3 şeklinde puanlanan bir maddedeki seçenek sayısı $k = 4$ 'tür. Herhangi bir birey bu maddeden 0 puan almışsa $Y = 0$ 'dır. 0 ile 1 ve 1 ile 2 ve 2 ile 3 arasındaki her aralık ise bir adım olarak tanımlanmaktadır. 0'dan 1'e bir adım, 1'den 2'ye bir adım, 2'den 3'e bir adım gibi. Dolayısı ile bir maddedeki adım sayısı $m = k - 1$ eşitliği ile bulunur (Hambleton, Swaminathan & Rogers, 1991, s. 26-28; (Ostini & Nering, 2006, s. 9-13).



Şekil 11. Dört kategorili bir maddedeki adımlar

Şekil 11 incelendiğinde dört kategorili bir maddenin adımları görülmektedir. Bu adımların anlamları kullanılan yaklaşıma göre farklılaşmaktadır.

Yığılmalı yaklaşıma göre

Adım 1= j_1 : 0 yanıtını veren bireyin 0'dan büyük ya da 1 ve üzerinde yanıt verme olasılığı,

Adım 2= j_2 : 1 yanıtını veren bireyin 1'den büyük ya da 2 ve üzerinde yanıt verme olasılığı,

Adım 3= j_3 : 2 yanıtını veren bireyin 2'den büyük yanıt ya da 3 yanıtını verme olasılığı (Ayala, 2009, s. 214-218);

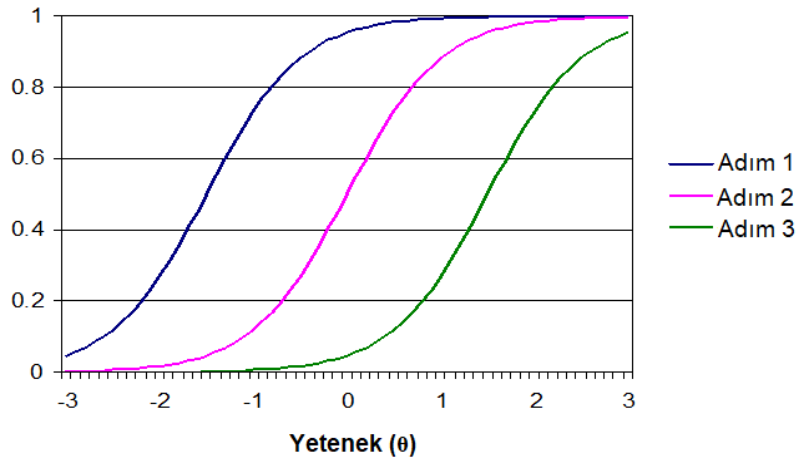
Komşu kategori yaklaşımına göre

Adım 1= j_1 : 0 yanıtını veren bireyin 1 yanıtını verme olasılığı,

Adım 2= j_2 : 1 yanıtını veren bireyin 2 yanıtını verme olasılığı,

Adım 3= j_3 : 2 yanıtını veren bireyin 3 yanıtını verme olasılığıdır (Stone & Zhu, 2015, s. 7-11).

Madde bilgi fonksiyonları ile bilgi sağlayan çok kategorili puanlanan maddeler için modellerin iki kategorililerden farklı olarak adım (step) fonksiyonları ile madde parametreleri kestirilir. Bir adım fonksiyonunda ilgili adıma ait fonksiyon, o adımın gerçekleşme ihtimali hakkında bilgi verir. Aşağıda 4 kategorili puanlan bir maddeye ait adım fonksiyonları gösterilmiştir:



Şekil 12. Dört kategorili puanlanan bir maddeye ait adım fonksiyonları

Şekil 12 incelendiğinde yapılan yorumlar kullanılacak yaklaşıma göre farklılaşmaktadır. Yığılmalı yaklaşıma göre Adım 1 için yaklaşık -1,5 yetenek düzeyinde maddeyi 0 puanlayan bireylerin maddeyi 1 ve üzerinde; Adım 2 için yaklaşık 0 yetenek düzeyinde maddeyi 1 puanlayan bireylerin maddeyi 2 ve üzerinde; Adım 3 için yaklaşık 1,8 yetenek

düzeyinde maddeyi 2 puanlayan bireylerin maddeyi 3 olarak puanlama ihtimali 0,5'tir (Reckase, 2009, s. 40-42).

Komşu kategori yaklaşımına göre ise Adım 1 için yaklaşık -1,5 yetenek düzeyinde maddeyi 0 puanlayan bireylerin maddeyi 1; Adım 2 için yaklaşık 0 yetenek düzeyinde maddeyi 1 puanlayan bireylerin maddeyi 2; Adım 3 için yaklaşık 1,8 yetenek düzeyinde maddeyi 2 puanlayan bireylerin maddeyi 3 olarak puanlama ihtimali 0,5'tir (Reckase, 2009, s. 33-36).

Çok kategorili MTK modelleri hem bu yaklaşımlara hem de tek-çok boyutlu ve sıralı-sıralı olmayan gibi durumlara göre farklılaşmaktadır: Samejima'nın derecelendirilmiş tepki modeli (graded response model), Muraki'nin genelleştirilmiş kısmî kredi modeli (generalized partial credit model) ve Samejima'nın ivme modeli (acceleration model) gibi. Ostini ve Nerig (2006)'in Samejima'nın bu modellerin hangi durumlar altında kullanılacağını verdikleri tablo aşağıdaki gibidir (s.62):

Tablo 6.

Çok Kategorili Puanlanan Maddeler İçin Samejima'nın Teorik Çerçevesi

Çok kategorili veri tipi	Modellerin genel sınıfı	Kullanılan spesifik model
Sürekli, nominal (gruplanmamış Rorschach veri gibi)	Heterojen (sürekli)	Yok
Sürekli, sıralı (grafik dereceleme ölçekleri gibi)	Heterojen (sürekli)	Sürekli Rasch model (Müller, 1987)
	Homojen (sürekli)	Sürekli tepki modeli (Samejima, 1973)
Ayrık nominal (çoktan seçmeli test maddeleri gibi)	Heterojen (kategorik)	Nominal yanıt modeli (Bock, 1972)
Ayrık sıralı (derecelendirme ölçeği maddeleri gibi)	Heterojen (kategorik)	İvme modeli (Samejima, 1995); Çok kategorili Rasch modelleri; Genelleştirilmiş kısmî kredi modeli (Muraki, 1992)
	Homojen (kategorik)	Derecelendirilmiş tepki modeli (Samejima, 1969)

(Ostini ve Nerig, 2006, s. 62)

Aşağıda çok kategorili, ayrık, sıralı ve homojen veri seti için kullanılan ve yığılmalı yaklaşımı temel alan "Samejima'nın derecelendirilmiş tepki modeli" kısaca açıklanmıştır.

Samejima'nın derecelendirilmiş tepki modeli

Samejima'nın derecelendirilmiş tepki modeli (DTM) yukarıda bahsedilen adımları tanımlamada yığılmalı yaklaşımı temel alır. Samejima'nın derecelendirilmiş tepki modeli yanıt kategorilerinin sıralanabileceğini (order) varsayar (Yürekli, 2010). Bu nedenle modelde bireyin yanıt verdiği kategorinin üzerindeki kategorilere tepki verme olasılığını belirlemeye çalışır. DTM fonksiyonu aşağıdaki gibi tanımlanır (Ostini & Nering, 2006, s. 64-65):

$$P(Y \geq j|\theta) = \frac{1}{1 + e^{-a_i(\theta_i - b_{ij})}}$$

Burada θ_i yetenek düzeyinde bireyin j ve üzerindeki adımları gerçekleştirme olasılığı hesaplanır. Bir madde için tek bir a parametresi ve adım sayısı kadar b parametresi kestirilir. a parametresi maddelerin dikliğini b parametreleri de eğrilerin yetenek eksenindeki konumunu belirler. a parametresi tek olduğundan adım fonksiyonlarının hepsi birbirine paraleldir. Tüm bu özellikleri ile DTM'nin 2PL'nin bir uzantısı olduğunu söylemek mümkündür (Hambleton, Swaminathan, & Rogers, 1991, s. 26-27; Park'tan aktaran Nartgün, 2002).

Maddelerden elde edilen bilgi fonksiyonlarının bir araya getirilmesi ile test bilgi fonksiyonu oluşturulur. Testten elde edilen bilgi arttıkça yetenek düzeyine ilişkin tahminin standart hatası azalır. Yetenek düzeylerine göre testin standart hatası kestirilebilir. θ yetenek seviyesinde yetenek kestiriminin standart hatası $SE^*(\theta)$ aşağıdaki formülle elde edilir:

$$SE^*(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Madde ve test bilgi fonksiyonu her bir yetenek düzeyinde ölçmenin standart hatasını örneklemden bağımsız olarak yansıtır. Kategori, madde ve test bilgi fonksiyonları ile standart hata θ 'ya bağlı olasılıklardır (Hambleton, Swaminathan, & Rogers, 1991, s. 94-95).

Tüm bu bilgiler ışığında KTK, bir grup bireyden elde edilen puanların toplamı ile elde edilen ölçümlerin güvenilirliğini ile ilgilenirken; GK, bir grup bireyden elde edilen ölçümlerin, kullanılan maddelerin ve puanlayıcıların da ötesine ölçüm sonuçlarının genellenebilmesi amacı taşır. Böylelikle tek bir analizle KTK’de tek bir güvenilirlik kestirimi yapılabilirken güvenilirlik ve genellenebilirlik başlıkları altında verilerin yorumlanabilmesini sağlar. MTK’de ise her bir maddeye verilen yanıt örüntüleri kullanılarak madde bazında hata kestirimleri yapılır. GK’de genellenebilirlik çalışması ile elde edilen sonuçlar karar çalışmalarının yapılmasına temel hazırlar, böylelikle puanlayıcı, madde vb. sayısının değişiminin güvenilirlik kestirimindeki değişiklikleri gözlenebilir. Tüm gözlem koşullarının ve değişkenlik kaynaklarının yer aldığı evrene ilişkin doğru kestirimlerin yapılması güvenilirlik ve geçerlik arasındaki farka da yeni bir bakış açısı kazandırmaktadır oysa KTK’de geçerlik ve güvenilirlik çalışmaları birbirinden farklı analizler yapılmasını gerektirir. MTK’de ise geçerlik ve güvenilirlik için a, b, c ve θ parametreleri ile test ve bigi fonksiyonları incelenmekte, marjinal güvenilirlik katsayıları kestirilmektedir. KTK’de tüm hata kaynaklarından gelen değişkenlik tek bir puan olarak karşımıza çıkarken; öğrenciler, maddeler, puanlayıcılar gibi hata kaynaklarını ayrı ayrı ve birlikte inceleme imkânı sağlayan GK’de öğrenciler arasındaki farklılaşma istenen bir durum olduğunda bireyler “ölçme objesi” olarak adlandırılmaktadır. Ölçme objesi araştırmanın amacına göre madde, puanlayıcı vb. da olabilmektedir. Değişkenlik kaynaklarının sabit ya da tesadüfi olarak seçilmesi o değişkenlik kaynağının olası tüm durumlara genellenebilmesini etkilemektedir. Sabit değişkenlik kaynağı, ölçme yapılan durum ile sınırlıdır. Bu nedenle değişkenlik kaynağı sabit olduğunda hata kaynağı azalır ölçme kesinliği artsa da ölçme sonuçlarının genellenmesi ile ilgili yorum yapmak güçleşecektir. Bu nedenle sabit değişkenlik kaynağını baz alan MTK’de genelleme yapma amacı bulunmamaktadır. Ayrıca bağıl ve mutlak kararlar alınacağına KTK’de tek bir güvenilirlik katsayısı kestirilebilirken GK’de bireylerin diğer bireylere göre ya da gruptan bağımsız durumunun ne olduğunu dikkate alınmasına göre iki farklı güvenilirlik katsayısı kestirilebilmektedir. MTK’de ise

örneklemeden bağımsız madde örnekleme ve maddeden bağımsız yetenek kestirimi ile bireyler hakkında kararlar verilebilmektedir. Aynı zamanda GK’de bir değişkenlik kaynağının diğer değişkenlik kaynağının tüm koşullarında gözlenip gözlenmemesi durumuna göre farklı desenler kullanılabilir. Bu desenlerden yalnız tamamıyla çaprazlanmış desen kullanıldığında tüm değişkenlik kaynakları için kestirimlerde bulunmak mümkün olmaktadır. Bunların yanında MTK’nin tek boyutluluk ve yerel bağımsızlık varsayımlarının sağlamlasının zorluğu bu kuramın kullanılmasını da güçleştirmektedir. Fakat bu varsayımlar karşılandığında test geliştirme, soru bankası oluşturma, bireye uyarlanmış test geliştirme, madde yanlılığının belirlenmesi, seçenekleri ağırlıklandırma ve test eşitleme konularında MTK’nin KTK’ye üstün olduğu savunulmaktadır.

Araştırmacının, araştırmanın amacına, kullanılan ölçme aracına, veri toplama yöntemine, elde edilen ölçümlere, ölçümlerin dağılımına, örnekleme, ölçümlerin hangi amaçla kullanılacağına, kuramların sınırlıklarına bağlı olarak kullanılması önerilen kuramlar da değişmektedir. Fakat ortak bir bakış açısı, en az iki kuramın birlikte kullanılmasının daha güvenilir sonuçlar ortaya koyduğu yönündedir. Bu çalışma da alanyazındaki çalışmaların önerileri doğrultusunda çok kategorili puanlanan bir ölçekten elde edilen puanların güvenilirlik kestiriminde KTK, GK ve MTK’yi karşılaştırmaktadır.



BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın deseni, çalışma grubu, veri toplama araçları ve verilerin analizi ile ilgili bilgilere yer verilmiştir.

3.1. Araştırma Deseni

Bu araştırmada yazılı anlatım becerileri puanlama anahtarı kullanılarak güvenirlik belirlemede KTK, GK ve MTK yöntemleri teker teker incelenecek; güvenirlik kestirimlerinde kullanılan teknikler karşılaştırılacaktır. Bu nedenle çalışma, betimsel bir araştırmadır.

Bir karşılaştırma grubu olmadan, ilişkiyi ya da farkı, nedenselliği sorgulamadan sadece sonuçların olduğu gibi saptanmasına yönelik yapılan çalışmalar, betimsel çalışmalardır (Erkuş, 2017, s. 123-124).

3.2. Çalışma Grubu

Çalışma grubunun belirlenmesinde seçkisiz olmayan örnekleme yöntemlerinden biri olan ve zaman, para ile işgücü kaybını önlemeyi hedefleyen uygun örnekleme yöntemi kullanılmıştır (Büyüköztürk vd., 2012). Araştırmanın çalışma grubunu 2017 yılı bahar döneminde Karabük ve Gaziantep'te bulunan birer okulda öğrenim gören toplam 523

ilkokul ve ortaokul öğrencisi oluşturmaktadır. Bu öğrencilerin 158 tanesi Karabük'te, 365 tanesi Gaziantep'te ikamet etmektedir. Yine bu öğrenciler 3, 4, 5, 6 ve 7. sınıfta öğrenim görmektedirler. Öğrenci sayılarının illere ve sınıf seviyelerine göre dağılımı aşağıdaki gibidir:

Tablo 7.

Örneklem: Öğrenci Sayıları

Öğrenci sayıları	3. Sınıf	4. Sınıf	5. Sınıf	6. Sınıf	7. Sınıf	Toplam
Karabük	50	28	18	36	26	158
Gaziantep	52	58	98	74	83	365
Toplam	102	86	116	110	109	523

Tablo incelendiğinde her sınıf seviyesinden en az 100 öğrenciye hikâye yazdırılması hedeflenmiş olsa da 4. sınıf için bu sayıya ulaşılamamış; 86 öğrenciye ulaşılmıştır.

Yine çalışma grubunda verileri puanlamak için 7 öğretmen gönüllü olmuştur. Öğretmenler ile ilgili bilgiler ise aşağıdaki tabloda verilmiştir.

Tablo 8.

Puanlayıcılar İle İlgili Bilgiler

Puanlayıcılar	İkamet ettikleri il	İş Tecrübesi	Görevi
1	Karabük	10 yıl üzeri	Türkçe öğretmeni
2	Bursa	10 yıl üzeri	Sınıf öğretmeni
3	Bursa	10 yıl üzeri	Sınıf öğretmeni
4	Karabük	3 yıl	Sınıf öğretmeni
5	Karabük	2 yıl	Sınıf öğretmeni
6	Gaziantep	3 yıl	Sınıf öğretmeni
7	Ankara	5 yıl	Ölçme ve Değerlendirme Uzmanı

Tablo 8 incelendiğinde öğretmenlerin tecrübelerinin en az iki yıl olduğu görülmektedir. Sınıf ve Türkçe öğretmenlerinin sayısının eşit olması hedeflense de puanlama işlemi için gönüllülük esas alındığından yalnız bir Türkçe öğretmeni ile çalışılabilmektedir.

3.3. Veri Toplama Araçları

Bu çalışmada öncelikle öğrencilerden belirlenen konularda verilen ölçütlere göre hikâye yazmaları istenmiştir. Bu uygulama ders saati içinde yapıldığından, öğrencilerin ve öğretmenlerin seçilmesinde gönüllülük esas alınmıştır. Yazılan hikâyeler, Ek-1’de verilen yazılı anlatım becerisi (analitik) puanlama anahtarına göre 7 puanlayıcı tarafından puanlanmıştır. Puanlayıcıların her birine puanlama anahtarının nasıl kullanılacağı ile ilgili gerekli eğitim verilmiştir.

3.3.1. 3-7. Sınıflar İçin Kullanılan (Hikâyeler) Formlar

Bu formlar 3 Türkçe öğretmeninden yazılı anlatım çalışmalarında kullandıkları konuları paylaşımları istenerek hazırlanmıştır. Gelen konu başlıkları incelenmiş, araştırmacı tarafından metinleştirilerek beş sınıf seviyesinde sıralanmıştır. Metinleştirme işlemi sınıf seviyelerindeki kelime sayıları ve konuların yoğunluğu göz önünde bulundurularak yapılmıştır. Daha sonra bu metinlerin Sınıf Öğretmenliği, Türkçe Eğitimi ve Eğitimde Program Geliştirme alanlarında çalışmalar yapan 3 akademisyen tarafından 3 ile 7.sınıf seviyeleri arasında sıralanması istenmiş ve ortak karar (oy birliği) ile formların son hâli oluşturulmuştur. Formların temaları 3. sınıf için orman (Ek-2a), 4. sınıf için renkler (Ek-2b), 5. sınıf için kitaplar (Ek-2c), 6. sınıf için öğretmen (Ek-2d), 7. sınıf için ayrımcılık (Ek-2e) olarak belirlenmiştir.

3.3.2. Yazılı Anlatım (Hikâye Yazma) Becerisi Değerlendirme Formu

Bu form özel okullarda görev yapan 7 öğretmen (1 tane 4. sınıf, 1 tane 3. sınıf, 3 tane 5. sınıf, 1 tane 6. sınıf, 1 tane 7. sınıf) ile haftada 4 saat çalışarak 4 haftada tamamlanmıştır. Araştırmacı ve öğretmenlerin ortak zamanı belirlenmiş ve her çalışmada en az 3 öğretmenin bulunmasına özellikle dikkat edilmiştir.

Öncelikle her bir öğretmenin "Öğrencilerinizin yazılı anlatım (hikâye yazma) becerilerini değerlendirirken hangi ölçütlere göre puanlama yapıyorsunuz?" sorusuna yanıt vermeleri istenmiştir. Yanıtlar birleştirildikten sonra benzer ve farklı yönleri yine öğretmenler tarafından incelenmiş ölçütler üç konu başlığında toplanmıştır: Düzen, yazım ve noktalama, içerik ve anlatım. Her bir başlığın altındaki ölçütler ise yine öğretmenlerin kriterleri baz alınarak araştırmacı tarafından önerilmiş ve yedi öğretmenimizle hemfikir olunmuştur. Daha sonra her bir ölçüt için öğretmenlerin öğrencilerinden bekledikleri en yüksek performanslar belirlenmiştir. Bu performanslar üç tam puan üzerinden puanlanmıştır. Öğrencilerin verebilecekleri diğer yanıtlar göz önünde bulundurularak kısmî puanlama için göstergeler de yine öğretmenlerle çalışarak karara bağlanmıştır.

0-3 arasında yapılan puanlamada 11 ölçüt için puanlama anahtarından alınabilecek en yüksek puan 33 en düşük puan 0 olarak belirlenmiştir. Alınan yüksek puan, yazılı anlatım (hikâye yazma) becerisinin bireyde yüksek olduğu anlamını taşımaktadır.

Hazırlanan analitik puanlama anahtarı öğretmenlerimiz tarafından bir eğitim-öğretim yılı içerisinde kullanılmış, puanlama anahtarının eksik olduğu düşünülen yerlerde ek açıklamalar eklenerek puanlama anahtarı daha anlaşılır hâle getirilmiştir.

3.4. Veri Analizi

Veriler araştırmanın alt problemlerine göre sırasıyla Klasik Test, Genellenebilirlik ve Madde Tepki Kuramları için kullanılan desenlere uygun olarak analiz edilmiştir. Güvenirlik belirlemede KTK'de Pearson korelasyon katsayısı, Cronbach Alfa (α) katsayısı, Eta korelasyon katsayısı için SPSS 22; GK'de G ve Phi (ϕ) katsayıları için Edu-G 6.1e ve MTK'de a, b, c ve θ parametreleri ile bilgi fonksiyonları için Multilog 7.03 programları kullanılmıştır. Elde edilen güvenirlik katsayılarının karşılaştırılması için ise Microsoft Excel programında Fisher Z' dönüştürmesi kullanılarak iki korelasyon katsayısı arasındaki farkın manidarlığı için t testi yapılmıştır. Normallik sayıltısı için SPSS 22'de grafikler Microsoft Excel programında çarpıklık ve basıklık katsayıları; tek boyutluluk ve yerel

bağımsızlık sayıtları için yine SPSS 22’de temel bileşenler analizi; model-veri uyumu için ise Multilog 7.03 programında gözlenen ve beklenen oranlar arasındaki farklar incelenmiştir.





BÖLÜM IV

BULGULAR VE YORUM

Bu bölümde araştırmaya ait bulgulara ve bulgulara dayalı yorumlara yer verilmiştir. Belirtilen bulgular ve yorumlar her bir alt problem için ayrıştırılmıştır.

4.1. Birinci Araştırma Sorusuna Yönelik Bulgular

1. alt problem: KTK'ye göre; yazılı anlatım becerisinin ölçülmesinde, yedi farklı puanlayıcının aynı öğrencilerin hikâyelerini (formlarını) puanlaması sonucu elde edilen puanlar arasındaki tutarlılık derecesi nedir?

Bu alt probleme daha ayrıntılı yanıt bulabilmek için iki farklı soru incelenecektir. Öncesinde ise yedi puanlayıcının verdikleri toplam puanların sınıflara göre ayrı ayrı betimsel istatistikleri aşağıda verilmiştir.

Tablo 9.

*Yedi Puanlayıcının Sınıf Seviyelerine Göre Öğrencilere Verdikleri Toplam Puanlara Ait
Betimsel İstatistikler*

Sınıf	Puanlayıcılar	Minimum	Maximum	Ortalama	Standart Sapma	Çarpıklık	Basıklık
3 N=102	1	2,00	32,00	11,510	6,456	0,865	0,735
	2	2,00	32,00	11,510	6,565	0,871	0,446
	3	1,00	32,00	11,804	6,236	0,647	0,257
	4	2,00	32,00	11,833	6,456	0,724	0,303
	5	2,00	33,00	12,284	6,692	0,688	0,097
	6	5,00	31,00	16,451	5,538	0,040	-0,679
	7	4,00	32,00	14,108	6,038	0,493	-0,168
4 N=86	1	1,00	33,00	12,698	8,889	0,866	-0,357
	2	1,00	33,00	12,419	8,982	0,873	-0,329
	3	1,00	33,00	12,186	9,145	0,771	-0,482
	4	1,00	33,00	12,826	9,131	0,747	-0,444
	5	1,00	33,00	12,512	9,176	0,737	-0,472
	6	1,00	31,00	17,058	6,886	-0,144	-0,716
	7	3,00	33,00	15,035	6,364	0,393	-0,266
5 N = 116	1	1,00	32,00	14,259	8,012	0,469	-0,881
	2	1,00	33,00	14,241	8,285	0,388	-1,026
	3	1,00	33,00	13,897	8,092	0,490	-0,747
	4	1,00	33,00	14,052	8,325	0,452	-0,990
	5	1,00	33,00	13,922	8,305	0,395	-0,976
	6	6,00	31,00	19,802	6,264	-0,188	-1,048
	7	4,00	32,00	18,371	7,198	0,048	-1,082
6 N = 110	1	0,00	33,00	17,836	9,059	-0,020	-1,092
	2	0,00	33,00	17,600	9,201	0,034	-1,175
	3	0,00	33,00	17,700	9,095	0,015	-1,115
	4	0,00	33,00	17,946	8,931	0,021	-1,091
	5	0,00	33,00	18,155	9,064	-0,027	-1,152
	6	4,00	32,00	21,009	6,905	-0,612	-0,469
	7	4,00	33,00	19,891	8,275	-0,117	-1,171
7 N = 109	1	4,00	33,00	23,275	7,277	-0,496	-0,597
	2	4,00	33,00	22,890	7,288	-0,569	-0,471
	3	4,00	33,00	22,890	6,915	-0,429	-0,742
	4	4,00	33,00	23,046	7,099	-0,424	-0,765
	5	4,00	33,00	22,963	6,987	-0,413	-0,736
	6	3,00	33,00	23,220	7,186	-0,519	-0,698
	7	3,00	33,00	21,312	7,267	-0,148	-1,101
Toplam N=523	1	0,00	33,00	16,098	9,018	0,213	-1,485
	2	0,00	33,00	15,918	9,080	0,215	-1,469
	3	0,00	33,00	15,882	8,948	0,221	-1,473
	4	0,00	33,00	16,111	9,002	0,221	-1,490
	5	0,00	33,00	16,145	9,032	0,204	-1,491
	6	1,00	33,00	19,664	7,007	-0,288	-1,164
	7	3,00	33,00	17,924	7,601	0,036	-1,345

Tablo 9 incelendiğinde öğrencilere verilen puanların 33,00 üzerinden 3, 4, 5 ve 7. Sınıf seviyelerinde 1,00 ile 33,00 arasında, 6. sınıflarda ise 0,00 ile 33,00 arasında değiştiği görülmektedir. Tüm seviyelerde 6. puanlayıcının diğer puanlayıcılardan daha yüksek bir ortalama ile puanlama yaptığı, sınıf seviyeleri arttıkça her bir puanlayıcının verdikleri puanların ortalamaları da beklendiği şekilde artmaktadır. En homojen puanlama 3, 5 ve 6. sınıflarla birlikte tüm öğrencilerde 6. puanlayıcı, 4. sınıflarda 7. puanlayıcı, 7. sınıflarda 3. puanlayıcı tarafından yapılmıştır.

Çarpıklık ve basıklık katsayıları örneklem büyüklüğünden etkilenmemesi için Microsoft Excel 2010 ile hesaplanmıştır. Kullanılan formüller aşağıdaki gibidir:

$$\text{Çarpıklık katsayısı} = \frac{\sum (X - \bar{X})^3}{N \cdot S_X^3}; \quad \text{Basıklık katsayısı} = \frac{\sum (X - \bar{X})^4}{N \cdot S_X^4} - 3$$

Tüm sınıf seviyelerindeki çarpıklık katsayıları (–0,612 ile 0,873 arasında) incelendiğinde değerlerin –1 ile +1 arasında olduğu görülmektedir. Basıklık katsayıları incelendiğinde (–1,491 ile 0,735 arasında) ise değerlerin –1,5 ile +1 arasında olduğu görülmektedir. Bu durumda verilerin dağılımının çarpık olmadığı, basıklığının ise kabul edilebilir seviyede olduğu söylenebilir. Bu verilere ait histogram grafikleri EK 3’te verilmiştir. Bu sonuçlara göre verilerin normal dağılım gösterdiği söylenebilir.

Yukarıda belirtilen betimsel istatistikler sadece puanların dağılımları ile ilgili bazı bilgiler vermektedir. Puanlamalardaki uyumu karşılaştırmak için ilave olarak aşağıdaki betimsel istatistikler de hesaplanıp yorumlanmıştır:

1.1. Puanlayıcıların verdikleri toplam puanlar arasındaki

(a) Pearson korelasyon katsayısı ve

(b) Eta korelasyon katsayıları ne düzeydedir?

(a) Yazılı anlatım (hikâye yazma) becerisinin ölçülmesinde kullanılan puanlama anahtarına göre 523 öğrencinin yedi puanlayıcının her birinden aldıkları toplam puanlar arasında bir

ilişki olup olmadığını test etmek için Pearson korelasyon katsayısı hesaplanmış, değerler aşağıdaki tabloda verilmiştir.

Tablo 10.

Yedi Puanlayıcının Sınıf Seviyelerine Göre Öğrencilere Verdikleri Toplam Puanlar Arasındaki Pearson Korelasyon Katsayıları

		Puanlayıcılar						
		1	2	3	4	5	6	7
3. sınıf N = 102	Puanlayıcılar							
	1	-	0,952	0,919	0,921	0,940	0,678	0,690
	2		-	0,940	0,946	0,963	0,658	0,681
	3			-	0,926	0,941	0,591	0,633
	4				-	0,947	0,653	0,684
	5					-	0,620	0,669
	6						-	0,827
4. sınıf N = 86	1	-	0,983	0,978	0,970	0,975	0,720	0,732
	2		-	0,985	0,976	0,979	0,702	0,714
	3			-	0,974	0,979	0,714	0,711
	4				-	0,979	0,686	0,707
	5					-	0,705	0,719
	6						-	0,826
5. sınıf N = 116	1	-	0,978	0,970	0,970	0,976	0,678	0,609
	2		-	0,980	0,979	0,982	0,675	0,628
	3			-	0,971	0,975	0,655	0,602
	4				-	0,980	0,691	0,628
	5					-	0,683	0,621
	6						-	0,782
6. sınıf N = 110	1	-	0,977	0,984	0,971	0,982	0,760	0,696
	2		-	0,989	0,978	0,981	0,766	0,705
	3			-	0,977	0,983	0,765	0,706
	4				-	0,984	0,765	0,696
	5					-	0,763	0,705
	6						-	0,873
7. sınıf N = 109	1	-	0,981	0,960	0,963	0,960	0,781	0,700
	2		-	0,960	0,960	0,957	0,792	0,698
	3			-	0,987	0,985	0,806	0,729
	4				-	0,989	0,813	0,730
	5					-	0,831	0,738
	6						-	0,840
Toplam N = 523	1	-	0,981	0,974	0,970	0,975	0,757	0,713
	2		-	0,980	0,976	0,979	0,755	0,716
	3			-	0,976	0,980	0,745	0,707
	4				-	0,982	0,755	0,714
	5					-	0,753	0,715
	6						-	0,849

Puanlayıcıların verdikleri puanların ikili korelasyonları incelendiğinde 3. sınıftan 7. Sınıfa kadar olan puanlamalarda ilk 5 puanlayıcı için her puanlayıcının diğer 4 puanlayıcı ile

arasındaki ilişkinin çok yüksek düzeyde (0,92 ile 0,99 arasında), 6 ve 7. puanlayıcılar arasındaki ilişkinin yüksek düzeyde (0,78 ile 0,87 arasında) , 6. ve 7. puanlayıcıların diğer 6 puanlayıcı arasındaki ilişkinin ise 3 ve 5. sınıflarda orta düzeyde (0,70'in altında) diğer sınıf seviyelerinde yine yüksek düzeyde (0,70 ile 0,82 arasında) olduğu görülmüştür.

Sınıf ayrımı yapılmaksızın tüm öğrenciler üzerinden yapılan istatistikte ise korelasyon katsayılarının 0,70 ile 0,98 arasında değişmesi yedi puanlayıcı arasındaki korelasyonun, yani puanlama uyumunun yüksek düzeyde olduğunu göstermektedir. Fakat bu uyum toplam puanlar üzerinden hesaplanarak görüldüğü için her bir maddeye verilen puanlarda da aynı uyum olduğunu söylemeyi mümkün kılmamaktadır; puanlar arasında bir uyum vardır demek yerine birlikte değişim gösterdiği söylenebilir. Çünkü herhangi bir puanlayıcının 523 öğrencinin puanlarında bir madde için verdikleri puan örüntüleri farklılaşsa da toplam puanları aynı olabilmektedir. Bu nedenle puanlayıcılar arasındaki uyum için Eta korelasyon katsayısının da hesaplanması gerekmektedir.

(b) Yazılı anlatım (hikâye yazma) becerisinin ölçülmesinde kullanılan puanlama anahtarına göre yedi puanlayıcı arasındaki puanlamaların tutarlılığı için Eta korelasyon katsayısı hesaplanmıştır. Bu işlem için SPSS 21 programındaki rastgele blok ANOVA deseni sonuçları kullanılarak aşağıdaki tablo elde edilmiştir.

Tablo 11.

Yedi Puanlayıcının Test Puanlarının Tutarlıklarına Ait Eta Korelasyon Katsayıları

Sınıflar	3	4	5	6	7	Toplam
η	0,880	0,913	0,881	0,928	0,936	0,926

Tablo 11 incelendiğinde yedi farklı puanlayıcının her bir öğrencinin hikâye yazma becerisini puanlaması sonucu elde edilen puanlar arasındaki uyumun derecesi 11 madde için 0,88 ve üzerinde olduğu görülür; bu durum bize puanlayıcılar arasındaki uyumun yüksek olduğunu gösterir.

Fakat bu iki istatistik de toplam puanlar arasındaki korelasyonu farklı şekillerde ifade ederken her bir puanlayıcının ne kadar tutarlı puanlama yaptığı hakkında bilgi vermemektedir.

1.2. Her bir puanlayıcının verdiği puanların sınıf düzeyleri ve tüm öğrencilerin puanları için Cronbach α katsayıları ne düzeydedir?

Yazılı anlatım (hikâye yazma) becerisinin ölçülmesinde kullanılan puanlama anahtarına göre yedi puanlayıcının 523 öğrencinin formlarına verdikleri puanların güvenilirliklerini hesaplamak için Cronbach Alfa (α) iç tutarlılık katsayısına bakılmıştır.

Tablo 12.

Her Bir Puanlayıcının Sınıf Seviyelerine ve Toplam Öğrenci Sayısına Göre Cronbach α İç Tutarlılık Katsayıları

	Puanlayıcılar						
	1	2	3	4	5	6	7
3. sınıf	0,892	0,885	0,869	0,881	0,891	0,796	0,862
4. sınıf	0,942	0,938	0,941	0,942	0,942	0,874	0,870
5. sınıf	0,914	0,924	0,917	0,923	0,922	0,846	0,879
6. sınıf	0,931	0,933	0,931	0,931	0,935	0,881	0,923
7. sınıf	0,907	0,910	0,908	0,916	0,911	0,916	0,893
Toplam	0,936	0,936	0,934	0,937	0,938	0,880	0,901

Tablo 12 incelendiğinde her puanlayıcının verdikleri puanların güvenilirlikleri oldukça yüksek (0,80'in üzerinde) olduğu görülür. Özellikle ilk beş puanlayıcı ve yedinci puanlayıcının puanlarına ait güvenilirlik katsayıları oldukça yüksektir (0,86 ve üzeri). Bu değerler her bir puanlayıcının kendi puanları arasındaki uyumu göstermekte olup puanlayıcılar arasındaki uyumu belirtmemektedir.

Yukarıdaki istatistiklerde öğrencilerin her birinin aldıkları toplam puanlar üzerinden hesaplama yapıldığından toplam puanlar aynı olsa bile puan örüntüleri

farklılaşabileceğinden puanlamalar arasındaki uyum da farklılaşabilecektir. Bu nedenle yukarıdaki gibi KTK’de yapılan istatistiklerde hata kaynakları yalnız puanlayıcılar olarak görülmektedir. Oysa hata kaynakları bireyler, puanlayıcılar, maddeler gibi farklılaşmakta ve ayrıca kendi aralarındaki etkileşimlerin sonucu da güvenilirliği etkilemektedirler. KTK’de her birinin güvenilirliğe etkisini incelemek için birden fazla istatistik kullanılması gereği aşikârdır. Bu nedenle KTK, analitik rubrikle elde edilen puanların güvenilirliğinin belirlenmesinde farklı hata kaynaklarının oranını ortaya koymada yetersiz kalmaktadır.

4.2. İkinci Araştırma Sorusuna Yönelik Bulgular

2. alt problem: GK'ye göre yazılı anlatım becerisinin ölçülmesinde kestirilen parametre değerleri nelerdir?

Bu alt probleme daha ayrıntılı yanıt bulabilmek için iki farklı soru incelenecektir.

2.1. Öğrenci (b), ölçüt (o) ve puanlayıcı (p) değişkenlik kaynaklarının tümüyle çaprazlandığı $b \times o \times p$ deseninin G çalışması sonucunda kestirilen hata varyansları ve toplam varyansı açıklama yüzdeleri nasıldır?

Yazılı anlatım (hikâye yazma) becerilerinin ölçülmesinin G çalışması ile elde edilen varyans ve yüzdelerini hesaplamak için yedi puanlayıcının beş sınıf seviyesinde toplam 523 öğrencinin hikâyelerini 11 ölçüt kullanılarak puanlaması istenmiş, $b \times o \times p$ deseni uygulanmıştır. Bu desende b, o ve p ana etkileri ile bo, bp, op, bop ortak etkilerinin sınıf seviyelerine göre dağılımı aşağıdaki tabloda verilmiştir.

Tablo 13.

*Öğrencilerinin G Çalışması Sonucunda Kestirilen Varyansları ve Toplam Varyansı**Açıklama Yüzdeleri*

	Varyans kaynağı	Sd	Toplam kareler	Kareler ortalaması	Varyans	Yüzde
3. Sınıf	B	101	2117,838	20,969	0,267	29,2
	O	10	298,850	29,885	0,025	2,7
	P	6	27,389	4,565	-0,006	0,0
	Bo	1010	745,332	0,738	0,038	4,2
	Bp	606	107,883	0,178	-0,027	0,0
	Op	60	714,711	11,912	0,112	12,3
	Bop	6060	2850,016	0,470	0,470	51,6
	Toplam					100
4. Sınıf	B	85	3346,375	39,370	0,503	43,9
	O	10	205,898	20,590	0,020	1,7
	p	6	16,447	2,741	-0,006	0,0
	bo	850	782,102	0,920	0,065	5,7
	bp	510	96,437	0,189	-0,025	0,0
	op	60	501,772	8,363	0,092	8,0
	bop	5100	2372,773	0,465	0,465	40,6
	Toplam					100
5. Sınıf	b	115	3769,907	32,782	0,418	34,6
	o	10	464,578	46,458	0,031	2,6
	p	6	29,610	4,935	-0,012	0,0
	bo	1150	1135,320	0,988	0,068	5,6
	bp	690	109,300	0,158	-0,032	0,0
	op	60	1256,841	20,947	0,176	14,6
	bop	6900	3550,535	0,515	0,515	42,6
	Toplam					100
6. Sınıf	b	109	4594,506	42,151	0,540	45,9
	o	10	169,749	16,975	0,013	1,1
	p	6	17,972	2,995	-0,002	0,0
	bo	1090	1022,355	0,938	0,061	5,2
	bp	654	121,353	0,186	-0,030	0,0
	op	60	372,301	6,205	0,052	4,4
	bop	6540	3335,232	0,510	0,510	43,4
	Toplam					100
7. Sınıf	b	108	3098,704	28,692	0,367	39,9
	o	10	176,630	17,663	0,011	1,2
	p	6	3,287	0,548	-0,007	0,0
	bo	1080	742,902	0,688	0,037	4,0
	bp	648	107,857	0,166	-0,024	0,0
	op	60	530,354	8,839	0,077	8,4
	bop	6480	2771,932	0,428	0,428	46,5
	Toplam					100
Toplam	b	522	21401,956	41,000	0,525	44,5
	o	10	1028,432	102,843	0,015	1,3
	p	6	72,255	12,043	-0,006	0,0
	bo	5220	4715,283	0,903	0,059	5,0
	bp	3132	565,278	0,181	-0,028	0,0
	op	60	2808,594	46,810	0,089	7,5
	bop	31320	15447,874	0,493	0,493	41,8
	Toplam					100

b: Öğrenci, o: Ölçüt, p: Puanlayıcı

Öğrenci (b) ana etkisi için kestirilen varyans değeri 3. sınıf öğrencileri dışındaki sınıf seviyelerinde 0,350'nin üzerinde ve toplam varyansı açıklama oranları da %40'ın üzerinde bulunmuştur. 3. sınıflarda ise varyans değeri (0,267) toplam varyansın %29,2'sini açıklamaktadır. Evren puanı için olan bu varyans bileşeni, öğrencilerin birbirinden ne derece sistematik bir şekilde farklılık gösterdiğini ifade etmektedir. Fakat toplam varyansı açıklama yüzdelerine bakıldığında bu ana etkinin 4 ve 6. sınıf öğrencileri ile tüm öğrencilerde ilk sırada; 3, 5 ve 7. sınıflarda ikinci sırada olduğu görülmektedir. Bu varyans bileşeni değerinin olabildiğince büyük olması istenen bir durumdur.

Ölçüt (o) ana etkisi için kestirilen hata varyansı değerlerinin (0,01 ile 0,03 arasında) toplam varyansı açıklama yüzdeleri %1,1 ile %2,7 arasında değişmektedir. Bu durum madde güçlükleri (zor, orta, kolay) arasında çok fark olmadığının göstergesidir.

Puanlayıcı (p) ana etkisi için kestirilen hata varyansı değerleri (-0,012 ile -0,002 arasında, varyans eksi değer alamayacağından bu değerler sıfıra yuvarlanır) negatif ve sıfıra oldukça yakın bir varyans değeridir, toplam varyansı açıklama yüzdeleri her seviyede ve toplamda %0 bulunmuştur. Bu değer puanlayıcıların puanlamaları arasındaki değişkenliğin derecesini vermektedir. Bu değerın sıfır çıkması ise puanlayıcıların puanlamaları arasındaki tutarlılığın mükemmel olduğunun bir göstergesidir.

Öğrenci-ölçüt (bo) ortak etkisinden kaynaklanan hata varyansı bileşeni, bir ölçütten diğerine öğrencilerin cevaplarındaki farklılıktır. Bu ortak etki için kestirilen varyans değerleri 0,037 ile 0,068 arasında, toplam varyansı açıklama oranları ise %4,2 ile %5,7 arasında değişmektedir. Buna göre, ölçütlerin güçlük düzeylerinin bir öğrenciden diğerine çok büyük farklılıklar göstermediği söylenebilir.

Öğrenci-puanlayıcı (bp) ortak etkisinden kaynaklanan hata varyansı bileşeni, birey ve puanlayıcı etkileşiminin etkisidir. bp varyans bileşenleri (-0,032 ile -0,024 arasında) yine her seviyede ve tüm öğrencilerde, toplam varyansın %0'ını açıklamaktadır. Bu değer, puanlayıcıların verdikleri puanların bireyden bireye hiç farklılık göstermediğini ifade eder.

Diğer bir deyişle, bir puanlayıcının yüksek puan verdiği öğrenciler diğer puanlayıcılardan da yüksek puan almıştır.

Ölçüt-puanlayıcı (op) ortak etkisinden kaynaklanan hata varyansı değerleri (0,052 ile 0,176 arasında) toplam varyansı açıklama yüzdeleri 4, 6 ve 7. sınıflar ile toplamda %8,5'in altında 3 ve 5. sınıflarda ise %12'nin üstündedir. Bu değer puanlayıcıların bir ölçütü puanlarken çok katı iken diğer ölçütte cömert olabilme derecelerini göstermektedir.

Öğrenci-ölçüt-puanlayıcı (bop) değişkenlik kaynağı (artık değişkenlik kaynağı) ise öğrenci, ölçüt ve puanlayıcıların etkileşiminin ve tesadüfi hataların yol açtığı değişkenliktir. Her sınıf seviyesinde ve tüm öğrencilerde en yüksek ya da en yüksek ikinci hata varyansı değeri olan bu hata varyansı değerlerinin (0,428 ile 0,515 arasında) toplam varyansı açıklama oranları %40,6 ile %51,6 arasında değişmektedir. Bu değer öğrenciler, ölçütler ve puanlayıcılar arası etkileşim ile bu çalışmada ölçülemeyen sistematik ya da tesadüfi değişkenlik kaynaklarının bulunduğu bir göstergesidir.

2.2. $b \times o \times p$ deseninde puanlayıcı ve ölçüt sayılarının arttırılıp azaltılmasıyla yapılan karar çalışması (K) sonucunda elde edilen G ve Phi katsayıları nasıldır?

Aşağıdaki tabloda çalışma grubundan elde edilen G ve Phi katsayıları koyu renk ile elde edilmiştir. Puanlayıcı sayısının 2, 4, 5 azaltılıp ve 1 arttırılarak; ölçüt sayısının iki katına çıkarılıp ve 2, 6 azaltılarak yapılan karar çalışmaları da aşağıdaki tabloda verilmiştir.

Tablo 14.

Öğrencilerin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin K Çalışması Sonucu Elde Edilen G ve Phi Katsayıları

	Ölçüt sayısı	Puanlayıcı sayıları									
		2		3		5		7		8	
		G	Phi	G	Phi	G	Phi	G	Phi	G	Phi
3. Sınıf	5	0,830	0,790	0,872	0,838	0,910	0,881	0,927	0,901	0,932	0,908
	9	0,898	0,871	0,925	0,903	0,948	0,930	0,958	0,943	0,961	0,946
	11	0,915	0,892	0,938	0,919	0,957	0,942	0,965	0,953	0,968	0,956
	22	0,955	0,943	0,968	0,958	0,978	0,970	0,982	0,976	0,984	0,977
4. Sınıf	5	0,894	0,874	0,920	0,903	0,941	0,928	0,950	0,939	0,953	0,942
	9	0,938	0,926	0,954	0,944	0,966	0,959	0,972	0,965	0,974	0,967
	11	0,949	0,939	0,962	0,953	0,972	0,966	0,978	0,971	0,978	0,973
	22	0,974	0,968	0,981	0,976	0,986	0,983	0,988	0,985	0,989	0,986
5. Sınıf	5	0,865	0,825	0,897	0,864	0,925	0,898	0,937	0,914	0,940	0,919
	9	0,920	0,894	0,940	0,920	0,957	0,941	0,964	0,950	0,966	0,953
	11	0,934	0,912	0,951	0,933	0,964	0,951	0,970	0,959	0,972	0,961
	22	0,966	0,954	0,975	0,965	0,982	0,975	0,985	0,980	0,986	0,980
6. Sınıf	5	0,895	0,884	0,921	0,912	0,943	0,935	0,953	0,946	0,956	0,949
	9	0,939	0,932	0,955	0,949	0,967	0,963	0,973	0,969	0,975	0,971
	11	0,949	0,943	0,963	0,958	0,973	0,969	0,978	0,975	0,979	0,976
	22	0,974	0,971	0,981	0,978	0,986	0,984	0,989	0,987	0,990	0,988
7. Sınıf	5	0,880	0,859	0,911	0,894	0,937	0,925	0,949	0,938	0,953	0,943
	9	0,929	0,917	0,948	0,938	0,964	0,957	0,971	0,965	0,973	0,967
	11	0,941	0,931	0,957	0,949	0,971	0,964	0,976	0,971	0,978	0,973
	22	0,970	0,964	0,978	0,974	0,985	0,982	0,988	0,985	0,989	0,986
Toplam	5	0,896	0,878	0,922	0,907	0,943	0,932	0,953	0,944	0,956	0,947
	9	0,939	0,928	0,955	0,946	0,968	0,961	0,973	0,968	0,975	0,970
	11	0,950	0,941	0,963	0,956	0,973	0,968	0,978	0,974	0,980	0,975
	22	0,974	0,969	0,981	0,977	0,987	0,984	0,989	0,987	0,990	0,987

Tablo 14'te her sınıf seviyesinde ve toplamda G ve ϕ katsayıları sağa ve aşağıya doğru gidildikçe artmaktadır. En küçük G ve ϕ katsayılarının sırasıyla 0,830 ve 0,790 olarak 3. sınıflarda 5 ölçüt ve 2 puanlayıcının olduğu durumda; en büyük G ve ϕ katsayılarının ise sırasıyla 0,990 ve 0,988 olarak 6. sınıflarda 22 ölçüt ve 8 puanlayıcının olduğu durumda elde edildiği görülür.

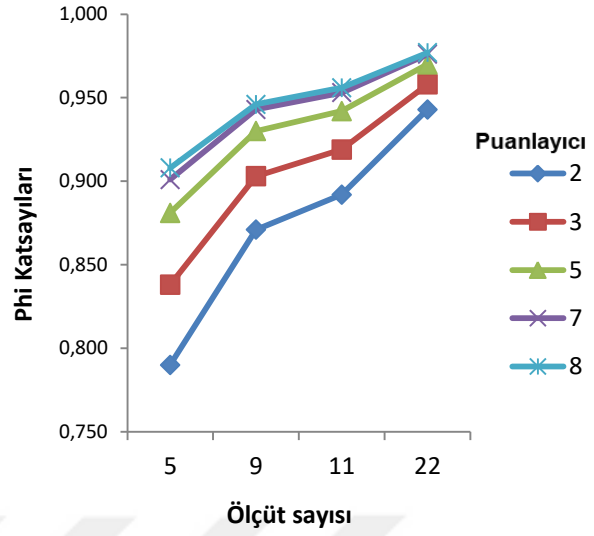
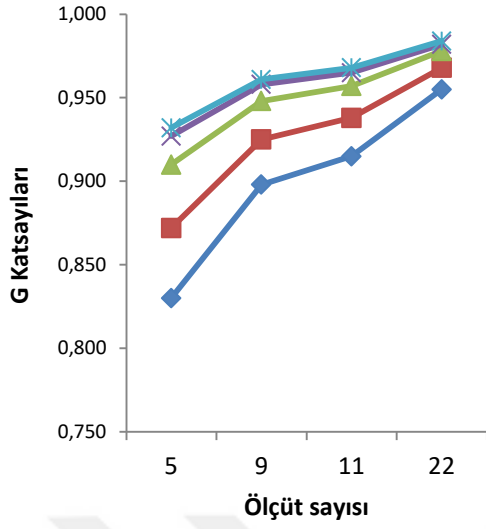
11 ölçütün 7 puanlayıcı tarafından puanlandığı asıl uygulama sonucunda 3. sınıflarda G katsayısının 0,965 ve ϕ katsayısının 0,953; 4. sınıflarda G katsayısının 0,978 ve ϕ katsayısının 0,971; 5. sınıflarda G katsayısının 0,970 ve ϕ katsayısının 0,959; 6. sınıflarda G katsayısının 0,978 ve ϕ katsayısının 0,975; 7. sınıflarda G katsayısının 0,976 ve ϕ katsayısının 0,971; toplamda ise G katsayısının 0,978 ve ϕ katsayısının 0,974 olarak kestirildiği görülmektedir. Aynı zamanda tablodan benzer durumlar altında her ϕ

katsayısının G katsayısından küçük olduğu da görülebilir. Uygulamada elde edilen sonuçların yüksek değerlerde olması sebebi ile K çalışmalarında ölçüt ve puanlayıcı sayılarının artışlarını incelemek yerine kullanılabilirlik (ekonomiklik) sağlanması adına daha az puanlayıcı ve ölçüt ile 0,80'e yakın değerler elde edilmeye çalışılmıştır.

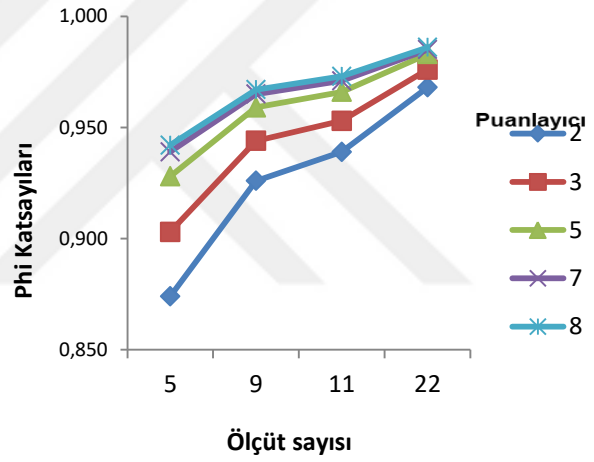
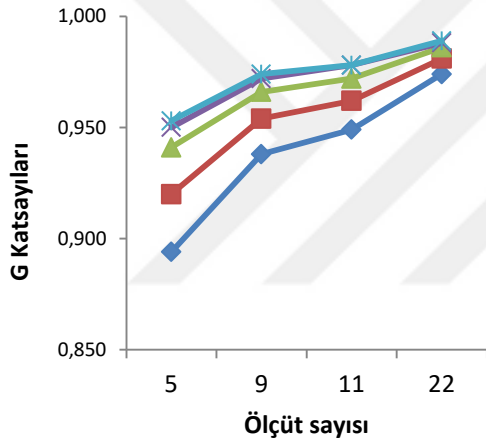
Her seviyede ve toplamda ölçüt sayısının sabit tutularak puanlayıcı sayısı azaltıldığında G ve ϕ katsayıları da azalırken; puanlayıcı sayısı artırıldığında G ve ϕ katsayıları da artmaktadır. Ölçüt sayısı sabit tutulup puanlayıcı sayısının arttığı karar çalışmalarında G katsayılarındaki artış, puanlayıcı sayısının sabit tutulup madde sayısının arttığı karar çalışmalarında G katsayılarındaki artıştan daha azdır. Fakat G katsayısı belli bir madde sayısı ve puanlayıcı sayısından sonra azalarak artmaktadır. Örneğin; 4. sınıflarda 2 puanlayıcının olduğu durumda madde sayısı arttıkça G katsayıları 0,894-0,938-0,949-0,974 şeklinde artmıştır. Yine aynı seviyede 5 maddenin olduğu durumda puanlayıcı sayısı arttıkça G katsayıları 0,894-0,920-0,941-0,950-0,953 şeklinde artmıştır. Aynı puanlayıcı ve ölçüt sayısında G katsayılarından daha düşük bir değer elde edilen ϕ katsayılarında da benzer durum görülmektedir. Yine benzer şekilde ϕ katsayılarının belli bir madde sayısı ve puanlayıcı sayısından sonra artış miktarı azalmaktadır. Örneğin; 4. sınıflarda 2 puanlayıcının olduğu durumda madde sayısı arttıkça ϕ katsayıları 0,874-0,926-0,939-0,968 şeklinde artmıştır. Yine aynı seviyede 5 maddenin olduğu durumda puanlayıcı sayısı arttıkça ϕ katsayıları 0,874-0,903-0,928-0,939-0,942 şeklinde artmıştır.

Artma miktarlarındaki değişimin daha net görülebilmesi için veriler çizgi grafiği ile de gösterilmiştir.

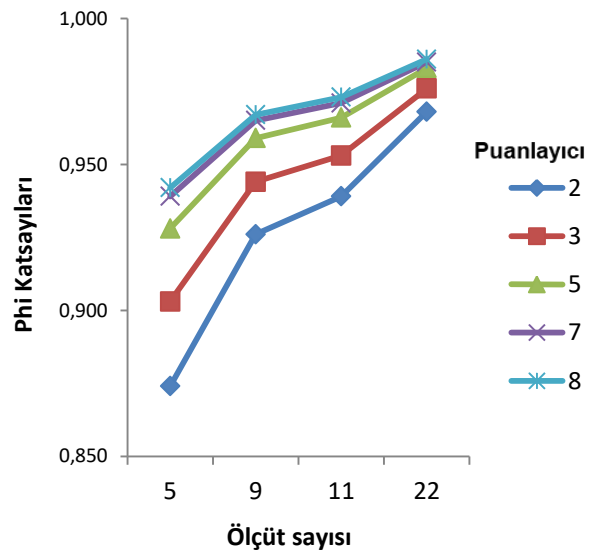
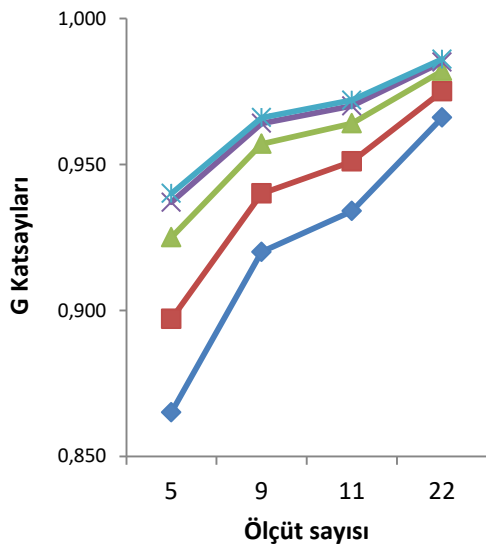
3. sınıf öğrencilerine ait G ve Phi katsayıları



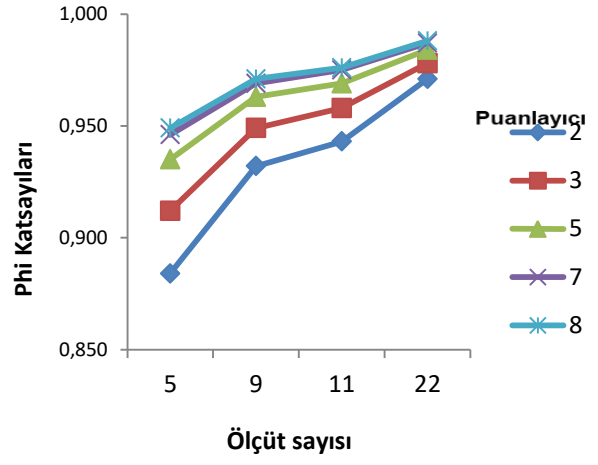
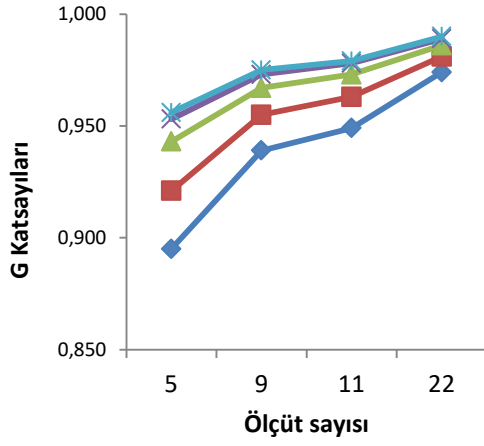
4. Sınıf öğrencilerine ait G ve Phi katsayıları



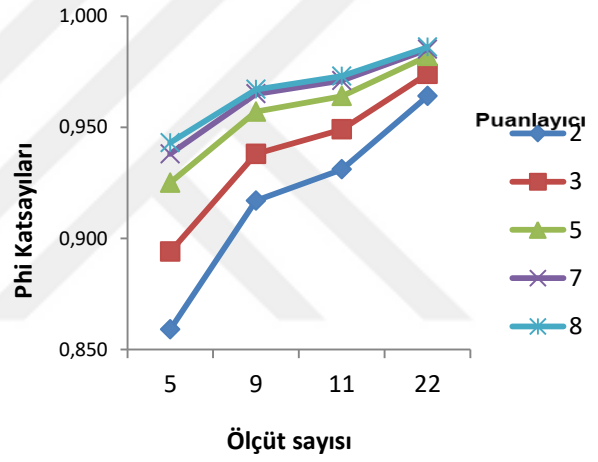
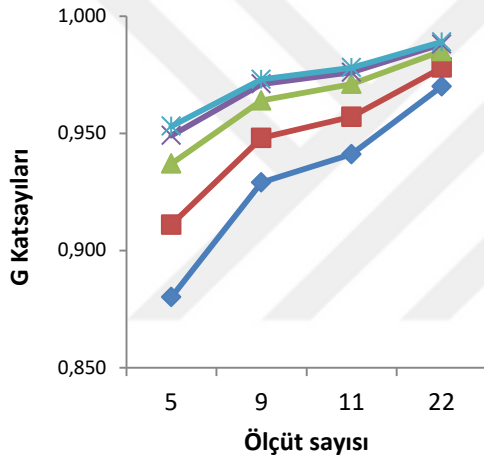
5. sınıf öğrencilerine ait G ve Phi katsayıları



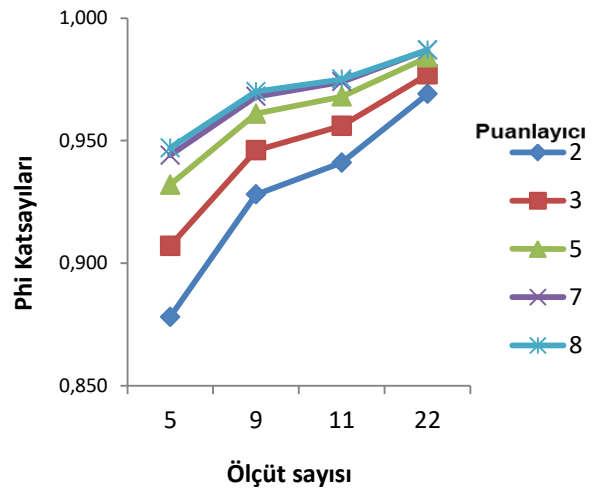
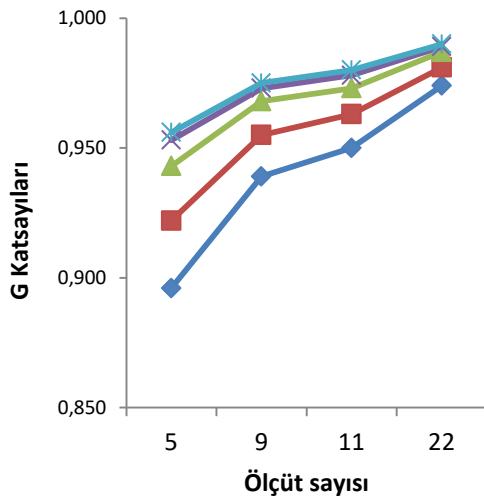
6. sınıf öğrencilerine ait G ve Phi katsayıları



7. sınıf öğrencilerine ait G ve Phi katsayıları



Tüm öğrencilere ait G ve Phi katsayıları



Şekil 13. G ve Phi katsayılarındaki değişimler

Şekil 13 incelendiğinde her seviyede ve toplamda G katsayılarının ϕ katsayılardan küçük olduğu, ölçüt sayısı ve puanlayıcıların sayısı arttıkça katsayılardaki artış miktarlarının azaldığı (çizgilerdeki eğimin azaldığı) görülmektedir.

Ayrıca KTK'ye göre diğer puanlayıcılarla arasındaki korelasyon düşük olan 6 ve 7. puanlayıcıları analizden çıkararak beş puanlayıcının sabit tutulup ölçüt sayısının 5, 9, 11 ve 22 olduğu durumda tüm öğrenciler için G ve Phi katsayıları tekrar incelenmiştir.

Tablo 15.

Tüm Öğrenciler İçin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin Beş Puanlayıcının Olduğu Karar Çalışmaları İle İlk Beş Puanlayıcının Sabit Tutulduğu Karar Çalışmalarının Sonuçları

Ölçüt sayısı	K çalışmasında 5 puanlayıcı için kestirilen sonuçlar		6 ve 7. puanlayıcılar dışındaki 5 puanlayıcı için kestirilen sonuçlar	
	G	Phi	G	Phi
5	0,943	0,932	0,963	0,956
9	0,968	0,961	0,979	0,975
11	0,973	0,968	0,983	0,979
22	0,987	0,984	0,991	0,990

Tablo incelendiğinde G ve Phi katsayılarının, madde sayısı arttıkça arttığı görülmektedir.

Ayrıca 6 ve 7. puanlayıcıların analizden çıkarılması sonucu ölçüt sayısı aynı olduğunda G ve Phi katsayılarının da arttığı fakat artışın 0,01-0,02 arasında değiştiği görülmektedir.

Benzer şekilde KTK'ye göre Cronbach Alfa değeri diğer puanlayıcılardan genel olarak düşük olan 6. puanlayıcının analizden çıkarılarak altı puanlayıcının olduğu ve ölçüt sayısının 5, 9, 11 ve 22 olduğu durumda tüm öğrenciler için G ve Phi katsayıları incelenmiştir.

Tablo 16.

Tüm Öğrenciler İçin Yazılı Anlatım Becerilerinin Ölçülmesine İlişkin 6 Puanlayıcının Sabit Tutulduğu Karar Çalışmaları İle 6. Puanlayıcı Dışındaki 6 Puanlayıcının Sabit Tutulduğu Karar Çalışmalarının Sonuçları

Ölçüt sayısı	K çalışmasında 6 puanlayıcı için kestirilen sonuçlar		6. puanlayıcı dışındaki 6 puanlayıcı için kestirilen sonuçlar	
	G	Phi	G	Phi
5	0,949	0,939	0,952	0,944
9	0,971	0,965	0,973	0,968
11	0,976	0,971	0,977	0,974
22	0,988	0,985	0,989	0,987

Tablo incelendiğinde G ve Phi katsayılarının madde sayısı arttıkça arttığı görülmektedir. Ayrıca 6. puanlayıcının analizden çıkarılması sonucu ölçüt sayısı aynı olduğunda G ve Phi katsayılarının da arttığı fakat artışın 0,001-0,005 arasında değiştiği görülmektedir.

4.3. Üçüncü Araştırma Sorusuna Yönelik Bulgular

3. alt problem: Çok kategorili MTK modellerinden biri olan Samejima'nın Derecelendirilmiş Tepki Modeli (DTM)'ne göre yazılı anlatım becerisinin ölçülmesinde kestirilen madde ve test istatistikleri nelerdir?

Bu araştırma problemine yanıt aramadan önce normal dağılım gösterdiği KTK'deki analizlerde hesaplanan verilerin MTK'nin tek boyutluluk ve yerel bağımsızlık varsayımlarını karşılayıp karşılamadığını ve model veri uyumunu incelemek gerekmektedir.

Çok kategorili MTK modellerinde örneklem büyüklüğü sayıtların sağlanması için önemlidir. De Ayala (2009) derecelendirilmiş tepki modeli sayıtlarının sağlanması için gereken cevaplayıcı sayısının alt sınırını 500 olarak önermiştir (s. 223). Bu nedenle KTK ve GK'den farklı olarak verilerin sınıf bazında analizleri yerine toplam öğrenci sayısı olan 523 üzerinden analizler yapılmıştır.

Tek boyutluluk varsayımını test etmek için SPSS 22 programında yedi puanlayıcının hepsi için ayrı ayrı Temel Bileşenler Analizi (TBA) yapılmıştır. Tüm yamaç birikinti grafikleri Ek-4'te; öz değerler, en düşük faktör yükleri ve açıklanan varyans oranları aşağıdaki tabloda verilmiştir.

Tablo 17.

Yedi Puanlayıcıya Ait Verilerin Tek Boyutluluk Sayıltısı İçin TBA Sonuçları

Puanlayıcı	1. faktöre ait öz değer	2. faktöre ait öz değer	Öz değerlerin oranı	Tek boyutluluk sayıltısı	En düşük faktör yükü	Tek boyutun açıkladığı varyans (%)
1	6,726	1,525	4,41	Sağlanır.	0,627	61,144
2	6,726	1,502	4,48	Sağlanır.	0,605	61,144
3	6,651	1,588	4,19	Sağlanır.	0,615	60,466
4	6,757	1,498	4,51	Sağlanır.	0,608	61,426
5	6,792	1,462	4,65	Sağlanır.	0,605	61,750
6	5,148	2,293	2,25	Sağlanmaz.		
7	5,627	2,028	2,77	Sağlanmaz.		

Ek-4'te verilen yamaç-birikinti grafikleri incelendiğinde yapının iki boyutlu olduğu fakat ilk 5 puanlayıcı için 1. faktöre ait eğimin 2. faktöre ait eğimden çok büyük olduğu görülmektedir. Benzer şekilde Tablo 17 incelendiğinde yapının iki boyutlu olduğu (öz değerleri 1'den büyük olan iki boyut vardır.) fakat birinci öz değerler ikinci öz değerlerin 4 katından fazla olduğundan ilk beş puanlayıcı için yapının tek boyutlu olduğunu söylemek mümkündür (Çokluk, Şekercioğlu & Büyüköztürk, 2014, s.226-227). 6 ve 7. puanlayıcılar için veriler tek boyutluluk sayıltısını sağlamadığı için DTM analizlerine dâhil edilememiştir.

Ölçek tek boyutluluk özelliği gösteriyorsa yerel bağımsızlık varsayımı da karşılanmış demektir (Crocker & Algina, 2006, s. 42-46). O hâlde ilk beş puanlayıcı için yerel bağımsızlık varsayımı da karşılanmaktadır.

Model veri uyumu için beş puanlayıcının puanladığı her bir maddeye ait gözlenen ve beklenen oranlar incelenebilir. Uyar, Öztürk Gübeş & Kelecioğlu (2013) gözlenen oranlar ile beklenen oranlar arasındaki farkların “artık” olarak isimlendirildiğini, artıkların sıfıra yaklaşması modelin veri ile uyum gösterdiğinin bir ölçütü olarak kullanılabileceğini

belirtmişlerdir. Ek-7'deki tabloda yazılı anlatım becerileri puanlama anahtarının madde kategorilerine ait beklenen ve gözlenen oranlar ve aralarındaki farklar verilmiştir. Ekteki tablo incelendiğinde en büyük artık değerin 0,0321 olduğu görülür.

Çok kategorili MTK modellerinden biri olan Samejima'nın Derecelendirilmiş Tepki Modeli (DTM)'ne göre yazılı anlatım (hikâye yazma) becerisinin ölçülmesinde Multilog 7.03 programı kullanılarak kestirilen madde parametreleri ve bu parametrelerin standart hataları aşağıdaki gibidir:



Tablo 18.

Yazılı Anlatım Becerileri Puanlama Anahtarının Maddelerine Ait Ayırt Edicilik İle Adım Fonksiyonları Parametreleri Ve Standart Hataları

		Puanlayıcılar																			
		1				2				3				4				5			
		a	b1	b2	b3	a	b1	b2	b3	a	b1	b2	b3	a	b1	b2	b3	a	b1	b2	b3
Maddeler ve Standart Hataları	1	1,45	-1,15	0,40	1,36	1,40	-0,97	0,29	1,44	1,30	-1,29	0,17	1,45	1,41	-1,15	0,25	1,48	1,52	-1,11	0,29	1,42
	SH	0,16	0,14	0,11	0,15	0,15	0,13	0,11	0,16	0,14	0,15	0,11	0,18	0,15	0,14	0,11	0,16	0,16	0,14	0,10	0,14
	2	1,54	-1,32	0,06	1,21	1,55	-1,28	-0,02	1,18	1,53	-1,25	-0,05	1,06	1,73	-1,31	0,02	1,05	1,67	-1,23	0,00	1,20
	SH	0,16	0,15	0,10	0,13	0,17	0,14	0,10	0,13	0,15	0,13	0,09	0,13	0,17	0,13	0,09	0,12	0,16	0,13	0,09	0,12
	3	1,71	-1,33	-0,28	0,68	1,60	-1,42	0,24	0,79	1,74	-1,50	-0,41	0,59	1,88	-1,39	-0,30	0,66	1,88	-1,34	-0,28	0,62
	SH	0,18	0,14	0,09	0,11	0,16	0,15	0,09	0,11	0,17	0,14	0,09	0,10	0,18	0,13	0,08	0,09	0,18	0,13	0,08	0,09
	4	1,57	-2,17	-0,61	0,44	1,36	-2,04	-0,65	0,52	1,37	-2,35	-0,80	0,41	1,41	-2,22	-0,79	0,45	1,53	-2,15	-0,72	0,46
	SH	0,18	0,25	0,10	0,10	0,16	0,24	0,12	0,12	0,16	0,25	0,12	0,11	0,16	0,25	0,12	0,11	0,16	0,23	0,11	0,10
	5	1,67	-1,47	-0,14	0,71	1,58	-1,47	-0,20	0,77	1,67	-1,44	-0,35	0,63	1,83	-1,43	-0,31	0,70	1,80	-1,41	-0,27	0,70
	SH	0,18	0,16	0,09	0,11	0,17	0,16	0,09	0,11	0,17	0,14	0,09	0,11	0,19	0,14	0,08	0,10	0,18	0,14	0,08	0,10
	6	3,05	-0,44	0,44	1,33	2,81	-0,46	0,56	1,51	3,05	-0,56	0,32	1,31	3,17	-0,48	0,35	1,24	2,96	-0,47	0,42	1,28
	SH	0,25	0,06	0,06	0,08	0,24	0,07	0,07	0,09	0,25	0,06	0,06	0,09	0,28	0,06	0,06	0,07	0,24	0,07	0,06	0,08
	7	4,14	-0,46	0,35	1,11	3,81	-0,47	0,31	1,17	3,74	-0,58	0,20	1,05	4,04	-0,50	0,28	1,04	3,85	-0,50	0,32	1,15
	SH	0,34	0,05	0,05	0,06	0,31	0,05	0,06	0,06	0,30	0,05	0,05	0,06	0,31	0,05	0,05	0,06	0,31	0,05	0,05	0,06
	8	5,98	-0,44	0,39	0,95	5,89	-0,41	0,34	0,98	5,48	-0,52	0,29	0,94	5,30	-0,43	0,38	0,93	5,52	-0,41	0,36	0,97
	SH	0,54	0,04	0,05	0,04	0,50	0,04	0,04	0,04	0,48	0,04	0,05	0,05	0,46	0,04	0,05	0,05	0,49	0,04	0,05	0,04
	9	5,73	-0,44	0,32	0,92	6,44	-0,39	0,33	0,97	4,94	-0,50	0,20	0,94	6,20	-0,45	0,22	0,90	6,14	-0,42	0,29	0,93
	SH	0,47	0,04	0,05	0,04	0,57	0,04	0,05	0,05	0,44	0,05	0,04	0,05	0,57	0,04	0,04	0,04	0,56	0,04	0,04	0,04
	10	5,05	-0,44	0,39	0,95	5,17	-0,46	0,37	0,98	4,81	-0,64	0,26	0,90	4,33	-0,59	0,29	0,95	4,91	-0,54	0,33	0,91
	SH	0,42	0,04	0,05	0,05	0,46	0,04	0,05	0,05	0,38	0,04	0,05	0,05	0,35	0,05	0,05	0,05	0,41	0,05	0,05	0,05
	11	1,38	-1,74	0,70	1,60	1,26	-1,90	0,72	1,82	1,28	-2,00	0,57	1,87	1,33	-1,88	0,64	1,90	1,24	-1,93	0,74	1,87
	SH	0,15	0,22	0,12	0,17	0,15	0,25	0,13	0,19	0,14	0,23	0,12	0,21	0,15	0,23	0,12	0,20	0,15	0,25	0,13	0,20

Tablo 18 incelendiğinde 11 maddenin a parametrelerinin tüm puanlayıcılarda 1,24 ile 6,44 arasında değiştiği görülmektedir. Baker (2001) ayırt edicilik parametrelerini çok düşük (0,01-0,34), düşük (0,35-0,64), orta (0,65-1,34), yüksek (1,35-1,69) ve çok yüksek (1,70 ve üzeri) olarak sınıflamıştır (s. 35). Beş puanlayıcıya göre tüm maddelerdeki a parametreleri yüksek ve çok yüksek düzeylerindedir. a parametresi iki kategorili MTK'de madde karakteristik eğrisinin eğimini gösterirken çok kategorili MTK'de buna ek olarak maddenin verdiği bilgi miktarını da göstermektedir (DeMars, 2010, s. 74). 1 ve 3. puanlayıcılara göre en fazla bilgiyi veren madde 8. ölçüt iken 2, 4 ve 5. puanlayıcılara göre en fazla bilgiyi 9. ölçüt vermektedir. En az bilgi veren madde ise tüm puanlayıcılara göre 11. ölçüt olarak bulunmuştur.

Yine Tablo 18 incelendiğinde adım fonksiyonlarına ait konum parametreleri (kategorilere ait eşik değerleri) de görülmektedir. b parametreleri ilgili kategoriyi 0,50 olasılıkla seçmiş olan bireylere ait yetenek düzeyini göstermektedir. Bireylerin düşük kategoriyi seçmeleri içi düşük yetenek seviyesi gerekirken daha üstteki kategoriler için daha büyük yetenek seviyesi gerekmektedir. Tüm puanlayıcılarda b_1 değerleri -2,35 ile -0,39; b_2 değerleri -0,80 ile 0,74; b_3 değerleri 0,41 ile 1,90 arasındaki yetenek düzeylerinde değişmektedir. Tüm puanlayıcılar için bireylerin %50 ihtimalle 0 kategorisinden daha yüksek puan almaları için en düşük -2,35 yetenek düzeyinde, 1 kategorisinden daha yüksek puan almaları için en düşük -0,80 yetenek düzeyinde, 2 kategorisinden daha yüksek puan almaları için en düşük 0,41 yetenek düzeyinde bulunmaları gerekmektedir.

Aşağıda her bir puanlayıcıya göre maddelerin en fazla bilgi verdikleri θ düzeyleri gösterilmiştir:

Tablo 19.

Yazılı Anlatım Becerileri Puanlama Anahtarındaki Maddelerin En Fazla Bilgi Verdiği Yetenek Düzeyleri

Puanlayıcılar	Maddeler										
	1	2	3	4	5	6	7	8	9	10	11
1	0,70	0,30	-0,20	-0,20	0,20	0,40	0,40	0,40	0,40	0,40	1,00
2	0,40	0,00	-0,10	-0,50	0,00	0,60	0,20	0,40	1,00	0,40	1,20
3	0,40	0,00	-0,30	-0,50	-0,20	0,20	0,20	1,00	0,20	0,40	1,20
4	0,40	0,30	-0,20	-0,50	-0,20	-0,40	0,20	0,40	0,20	0,40	1,20
5	0,60	0,00	-0,10	-0,40	0,00	0,40	0,40	0,40	0,20	0,40	1,20

Tablo 19 incelendiğinde 1. puanlayıcıya göre 3 ve 4. maddeleri; 2, 3, 4 ve 5. puanlayıcılara göre 4. maddeyi gruba göre en düşük yetenek düzeyindeki bireylerin, tüm puanlayıcılara göre 11. maddeyi gruba göre en yüksek yetenek düzeyine sahip bireylerin doğru yanıtlaması beklenmektedir. Dolayısı ile tüm puanlayıcılarda bireylere verilen puanlar 2, 3, 4 ve 5. maddelerde üst kategorilerde yer alırken 1, 6, 7, 8, 9 ve 11. maddelerde alt kategorilerde yer almaktadır. Beş puanlayıcıya göre öğrencilerin yetenek düzeyleri ise Ek-8’de verilmiştir.

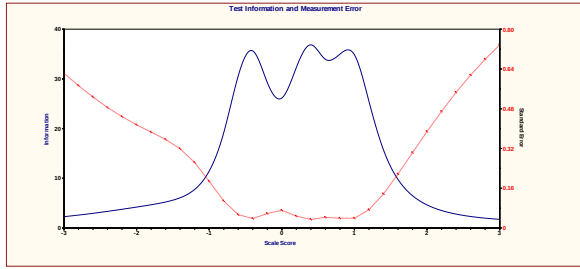
Maddelerin istatistiksel olarak analizlerine bakmak için karakteristik eğrileri ve bilgi fonksiyonları da kullanılmaktadır. Ek-9’da 1. puanlayıcıya ait madde karakteristik eğrileri ve bilgi fonksiyonları örnek olarak verilmiştir. Örnek teşkil etmesi için seçilen madde 3 ve madde 8’e ait grafikler incelendiğinde 8. maddeye ait eğrilerin daha dik olması sebebi ile ayırt ediciliğinin daha yüksek (5,98), 3. maddeye ait eğrilerin ise daha eğik olması sebebi ile ayırt ediciliğinin daha düşük (1,71) olduğu söylenebilir.

8. maddenin b parametreleri incelendiğinde 0,50 olasılıkla $-\infty$ ile yaklaşık -0,60 yetenek düzeyine kadar bireylerin 0 kategorisini, -0,60’tan yaklaşık 0,40 yetenek düzeyine kadar bireylerin 1 kategorisini, 0,40’tan yaklaşık 1,20 yetenek düzeyine kadar bireylerin 2 kategorisini, 1,20’den $+\infty$ yetenek düzeyine kadar olan bireylerin ise 3 kategorisini seçmeleri beklenmektedir. Aynı maddenin madde bilgi fonksiyonu incelendiğinde ise maddenin en fazla bilgi verdiği yetenek düzeylerinin yaklaşık -0,60 ile 1,20 arasında olduğu görülmektedir.

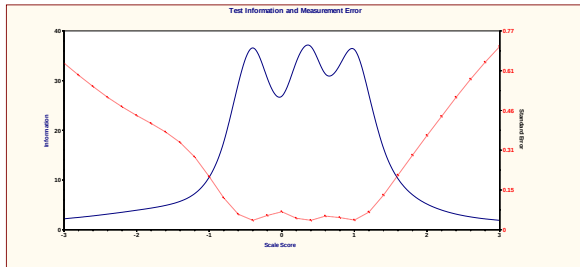
3. maddenin b parametreleri incelendiğinde 0,50 olasılıkla $-\infty$ ile yaklaşık -1,30 yetenek düzeyine kadar bireylerin 0 kategorisini, -1,30'dan yaklaşık -0,30 yetenek düzeyine kadar bireylerin 1 kategorisini, -0,30'dan yaklaşık 0,50 yetenek düzeyine kadar bireylerin 2 kategorisini, 0,50'den $+\infty$ yetenek düzeyine kadar olan bireylerin ise 3 kategorisini seçmeleri beklenmektedir. Aynı maddenin madde bilgi fonksiyonu incelendiğinde ise maddenin en fazla bilgi verdiği yetenek düzeylerinin yaklaşık -1,50 ile 1,00 arasında olduğu görülmektedir.

MTK'ye göre güvenirlik kestirimlerinde parametrelerin yanında her bir maddenin teste yaptığı katkıların toplamından oluşan test bilgi fonksiyonu ve marjinal güvenirlik katsayısı da hesaplanmaktadır. Beş puanlayıcıya ait test bilgi fonksiyonları aşağıdaki gibidir:

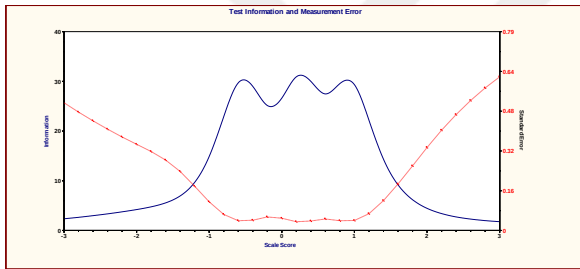
Birinci puanlayıcıya ait test bilgi fonksiyonu



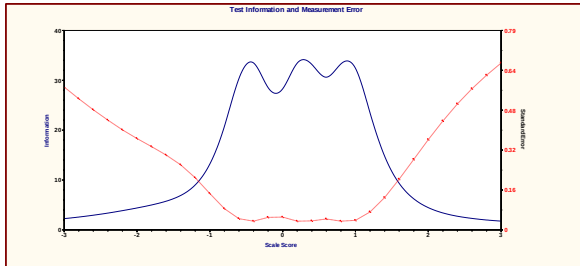
İkinci puanlayıcıya ait test bilgi fonksiyonu



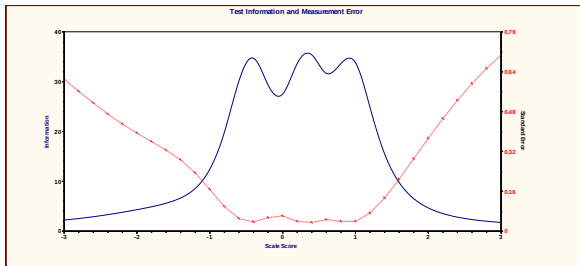
Üçüncü puanlayıcıya ait test bilgi fonksiyonu



Dördüncü puanlayıcıya ait test bilgi fonksiyonu



Beşinci puanlayıcıya ait test bilgi fonksiyonu



Şekil 14. Beş puanlayıcıya ait test bilgi fonksiyonları

Şekil 14'te görüldüğü gibi her puanlayıcı için test bilgi fonksiyonları, farklı yetenek düzeyleri için değişse de yaklaşık olarak -1,00 ile 1,50 arasındaki yetenek düzeylerinde bulunan bireyler için görece daha yüksektir. Test bilgi fonksiyonlarında bilgi miktarı arttıkça standart hata azalmaktadır. O hâlde -1,00 ile 1,50 arasında yeteneğe sahip olan bireyler için ölçme sonuçları daha az hata ile kestirilmiştir. Bu aralıktaki θ değerlerinden uzaklaştıkça ölçme sonuçlarının içerdiği hata miktarı da artmaktadır.

Marjinal güvenirlik katsayısı, ölçeğin bütünü için kestirilen bir güvenirlik katsayısıdır. 0-1 arasında değer alır; 1'e yaklaştıkça ölçekten elde edilen puanların güvenirliği artar. Aşağıda beş puanlayıcıya ait marjinal güvenirlik katsayıları verilmiştir:

Tablo 20.

Beş Puanlayıcıya Ait Marjinal Güvenirlik Katsayıları

	Puanlayıcılar				
	1	2	3	4	5
Marjinal Güvenirlik Katsayısı	0,9313	0,9304	0,9313	0,9330	0,9330

Tablo 20 incelendiğinde tüm katsayıların 0,93 civarında ve güvenirliğin oldukça yüksek olduğu görülmektedir.

4.4. Dördüncü Araştırma Sorusuna Yönelik Bulgular

4. alt problem: Bu üç teoride elde edilen güvenirlik kestirimleri arasında manidar bir farklılık var mıdır?

Önceki bölümlerde KTK'de her sınıf seviyesi ve her puanlayıcı için Cronbach Alfa katsayıları (35 tane), yedi puanlayıcı için Eta korelasyon katsayıları (7 tane) hesaplanmıştır. GK'de yedi puanlayıcı olduğunda sınıf seviyelerine göre ayrışan G (5 tane) ve Phi (5 tane) katsayıları hesaplanmıştır. MTK'de ise yedi puanlayıcı için marjinal güvenirlik katsayıları (5 tane) hesaplanmıştır.

Her bir puanlayıcıya ait KTK'deki Cronbach Alfa katsayıları ile MTK'deki marjinal güvenirlik katsayıları karşılaştırılmıştır. KTK'deki Cronbach Alfa katsayılarının GK'deki katsayılarla karşılaştırılabilmesi için her sınıf seviyesindeki yedi puanlayıcıya ait Cronbach Alpha (α) katsayılarının ortancası alınarak hesaplamalar yapılmıştır.

Tablo 21.

Sınıf Seviyeleri ve Toplamdaki Öğrenciler İçin α (Ortancaları), Eta, G ve Phi Katsayıları

	α (ortancaları)	Eta	G	Phi
3. sınıf	0,881	0,880	0,965	0,953
4. sınıf	0,941	0,913	0,978	0,971
5. sınıf	0,917	0,881	0,970	0,959
6. sınıf	0,931	0,928	0,978	0,975
7. sınıf	0,910	0,936	0,976	0,971
Toplam	0,936	0,926	0,978	0,974

Tablo 21'e göre dört katsayının ikişerli korelasyonları Fisher'in Z' testine göre karşılaştırıldığında %95 olasılıkla (0,05 anlamlılık düzeyinde) elde edilen Z' değerleri, Z testi ve anlamlılık (p) değerleri aşağıdaki tabloda verilmiştir:

Tablo 22.

Dört Katsayıya Ait Fisher Z' Testi Sonuçları

3. sınıf	Katsayılar		α (ortancaları)	Eta	G	p değerleri	α (ortancaları)		Eta	G
	Eta	G					Eta	G		
			1,38	1,375	2,014					

* $p < 0,05$.

Tablo 22 incelendiğinde her sınıf düzeyinde ve tüm öğrencilerde G ve Phi katsayıları arasında, α ve Eta korelasyon katsayıları arasında 0,05 düzeyinde anlamlı bir fark bulunamamıştır.

Her sınıf düzeyinde ve tüm öğrencilerde α ile G, α ile Phi, G ve Eta, Phi ve Eta korelasyon katsayıları arasında 0,05 düzeyinde anlamlı bir fark bulunmuştur.

Bu katsayılar ek olarak sınıf yerine puanlayıcı baz alındığında α ve Marjinal güvenirlik katsayıları aşağıdaki gibidir:

Tablo 23.

Puanlayıcılar İçin α ve Marjinal Güvenirlik Katsayıları

	Puanlayıcılar				
	1	2	3	4	5
Cronbach Alfa(α)	0,936	0,936	0,934	0,937	0,938
Marjinal güvenirlik	0,931	0,930	0,931	0,933	0,933

Tablo 23'e göre iki katsayı Fisher'in Z' testine ile karşılaştırıldığında 0,95 olasılıkla (0,05 anlamlılık düzeyinde) elde edilen Z' değerleri, Z testi ve anlamlılık (p) değerleri aşağıdaki tabloda verilmiştir:

Tablo 24.

İki Katsayıya Ait Fisher Z' Testi Sonuçları

Katsayılar	Puanlayıcılar				
	1	2	3	4	5
Fisher Z(α)	1,7047	1,7047	1,6888	1,713	1,721
Fisher Z (Marjinal güvenirlik)	1,6681	1,6614	1,6681	1,681	1,681
Z testi	0,5909	0,6996	0,3345	0,513	0,646
p değerleri	0,55	0,48	0,74	0,61	0,52

Tablo 24 incelendiğinde 0,05 anlamlılık düzeyinde katsayılar arasında anlamlı bir fark olmadığı görülür. Bu durumda KTK ve MTK'ye göre puanlayıcılar arası güvenirlik için benzer sonuçlar elde edilmiştir.



BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmada, tesadüfi hatayı en aza indirmek için çalışma alanlarına göre farklılaşan ölçme kuramlarından Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarının yazılı anlatım becerisi analitik puanlama anahtarı kullanılarak elde edilen ölçümlere ilişkin güvenirlik belirlemedeki yaklaşımları incelenmiş; araştırmanın bulgularına dayalı olarak sonuçlar, tartışma ve hem sonuçlara hem de bu çalışmanın kapsamına alınamayan durumlara yönelik önerilere yer verilmiştir.

5.1. Sonuç ve Tartışma

Araştırmanın alt problemlerine ait bulgulara ilişkin sonuçlar alt problemlerin veriliş sırasıyla açıklanmıştır.

5.1.1. Birinci Alt Problem

KTK'ye ait birinci alt problemi incelemeden önce normallik varsayımı test edilmiştir. Yedi puanlayıcının tüm öğrencileri puanlaması ile puanlama anahtarından elde edilen puanların normalliği için çarpıklık ve basıklık katsayıları sınıf seviyelerinde ve tüm öğrencilerde incelenmiş, dağılımların çarpık olmadığı basıklıkların ise kabul edilebilir düzeyde olduğu

görülmüştür. Dolayısı ile verilerin normal dağılımdan aşırı sapma göstermediği sonucuna ulaşılmıştır.

KTK'ye göre; yazılı anlatım becerisinin ölçülmesinde, yedi farklı puanlayıcının aynı öğrencilerin hikâyelerini (formlarını) puanlaması sonucu elde edilen puanlar arasındaki tutarlılık dereceleri için elde edilen iki farklı sonuç aşağıda açıklanmıştır.

Öncelikle puanlayıcıların verdikleri toplam puanlar arasındaki Pearson korelasyon katsayısı ile Eta korelasyon katsayısı hesaplanmıştır.

Yedi puanlayıcının puanları arasındaki basit korelasyonlar Pearson korelasyonu ile incelenmiş, ilk beş puanlayıcının basit doğrusal korelasyonlarının 0,92 ile 0,99 arasında değiştiği ve yüksek korelasyon gösterdiği; altı ve yedinci puanlayıcıların diğer beş puanlayıcı ile arasındaki basit korelasyonlarda ise 3 ve 5. sınıflarda diğer sınıf seviyelerinden daha düşük olduğu (0,59 ile 0,70 arasında) kendi aralarında ise tüm sınıf seviyelerinde ve toplamda diğer puanlayıcılardan daha yüksek (0,82'den yüksek) bir korelasyon gösterdiği görülmüştür. Bu bulgular, Güler (2008), Deliceoğlu (2009), Bağcı (2015) ve Şalgam'ın (2016) analitik rubrik, bütüncül rubrik, yazılı sınav, likert tipi ölçek kullandıkları çalışmalarında 72 ile 203 arasındaki örneklem büyüklüğü ve 3 ile 5 arasında değişen puanlayıcı sayıları ile yaptıkları araştırmalarda elde ettikleri Pearson korelasyon katsayıları ile benzer sonuçlar verirken, Öztürk'ün (2011) gözlem formu kullanarak 102 öğrencinin dört puanlayıcı tarafından gözlenmesi ile yaptığı çalışma ile farklı sonuçlar vermiştir: Öztürk (2011) Pearson korelasyon katsayılarının 0,177 ile 0,719 arasında olması sebebi ile puanlamalar arasında uyum olmadığı sonucuna ulaşmıştır. Öztürk (2011) bu durumun sebebinin kullanılan aracın her puanlayıcı için aynı şekilde anlaşılmadığı bu nedenle analitik rubriklerin kullanılması gerektiğini belirtmiştir. Bu çalışmada analitik rubrik kullanılmasının puanlayıcılar arasındaki tutarlılığın artmasını sağladığı söylenebilir.

Basit korelasyon katsayıları yalnızca puanların birlikte değişim gösterdiğini veya sıralamaların benzer olduğunu söylemekte, puanlayıcılar arasındaki uyum ile ilgili net bir bilgi vermemektedir. Çünkü korelasyon katsayısı için her bir öğrencinin ölçekten aldığı toplam puanlar kullanılmakta, yanıt örüntüleri farklı olan öğrenciler de aynı performansı göstermiş sayılmaktadır.

Buna ek olarak yedi puanlayıcı için puanlayıcılar arasındaki Eta korelasyon katsayısı hesaplanmıştır. Her sınıf seviyesinde ve tüm öğrencilerde η katsayısı 0,88 ve üzerinde bulunmuş ve puanlayıcıların öğrencileri puanlamadaki uyumlarının yüksek olduğu görülmüştür. Alanyazın incelendiğinde kuramları karşılaştıran çalışmalarda güvenilirlik belirlemek için Eta korelasyon katsayısının kullanıldığı çalışmaların oldukça az olduğuna rastlanmıştır. Sıralama ölçeğindeki veriler için puanlayıcıların öğrencileri sıralamadaki uyumunu gösteren Kendall'ın uyum katsayısının (Güler (2008), Yelboğa (2007), Deliceoğlu (2009) ve Öztürk'ün (2011) çalışmalarında) ve Eta korelasyonun katsayısı olan ANOVA tablolarından elde edilen sınıf içi korelasyon katsayılarının (Güler (2011), Büyükkıdık (2012) ve Bağcı'nın (2015) çalışmalarında) hesaplandığı görülmüştür.

Her bir puanlayıcının verdiği puanların sınıf düzeyleri ve tüm öğrencilerin puanlarının güvenilirliğini kestirmek için Cronbach α iç tutarlılık katsayıları hesaplanmıştır. İlk beş ve yedinci puanlayıcıların puanlamalarına ait Cronbach α iç tutarlılık katsayıları 0,86'nın üzerinde yüksek düzeyde; altıncı puanlayıcının puanlamalarına ait iç tutarlılık katsayıları ise 0,79'un üzerinde ve yüksek düzeyde bulunmuştur. Yedi puanlayıcının da test puanlarının iç tutarlılık olarak Cronbach α güvenilirlik katsayıları yüksek bulunmuştur. Bu bulgular Güler (2008), Yelboğa (2007), Deliceoğlu (2009), Öztürk (2011), Büyükkıdık (2012) ve Bağcı'nın (2015) çalışmalarında 0,77'nin üzerinde hesapladıkları Cronbach α iç tutarlılık katsayıları ile benzer sonuçlar vermiş, farklı olarak Güler'in (2011) rastgele veriler üreterek yaptığı çalışmada çok düşük düzeyde bulunmuştur. Güler (2011) bu durumun sebebini düşük geçerlik ve güvenilirliğe sahip rastgele verilerin de KTK ve GK'de benzer sonuçlar verip vermediğinin incelenmesinin amaçlanması olarak belirtmiştir.

5.1.2. İkinci Alt Problem

GK'ye göre yazılı anlatım becerisinin ölçülmesinde kestirilen parametre değerleri aşağıda açıklanmıştır.

Öğrenci (b), ölçüt (ö) ve puanlayıcı (p) değişkenlik kaynaklarının tümüyle çaprazlandığı $b \times \text{ö} \times p$ deseninin G çalışması sonucunda kestirilen hata varyansları ve toplam varyansı açıklama yüzdeleri incelenmiştir.

Öğrenci (b) ana etkisi için kestirilen varyans değeri toplam varyansı açıklamada her sınıf seviyesinde ve sınıf ayrımı yapılmadan tüm öğrencilerde en az ikinci sırada yer almaktadır. Olabildiğince büyük olması istenen bu varyans bileşeni öğrencilerin birbirinden ne derece sistematik bir şekilde farklılık gösterdiğini ifade etmektedir. Farklılaşma veya ölçme değişkenine ait hata varyansı fazla olduğundan yapılan puanlamaların öğrenciler arasındaki farklılaşmayı ortaya çıkardığını söylemek mümkündür.

Ölçüt (o) ana etkisi için kestirilen varyansın toplam varyansı açıklama yüzdeleri her sınıf seviyesinde ve tüm öğrencilerde %1,1 ile %2,7 arasında değişmektedir. Bu durum ölçütlerin kolay, orta ve zor gibi birbirinden güçlük bakımından çok farklılaşmadığını göstermektedir.

Puanlayıcı (p) ana etkisi için kestirilen varyans her sınıf seviyesinde ve tüm öğrencilerde %0 bulunmuştur. Bu durum, puanlayıcıların puanlamaları arasındaki tutarlılığın mükemmel olduğunun bir göstergesidir.

Öğrenci-ölçüt (bo) varyans bileşeni, bir ölçütten diğerine öğrencilerin cevaplarındaki farklılıktır. Bu ortak etkinin toplam varyansı açıklama oranları ise %4,2 ile %5,7 arasında değiştiğinden ölçütlerin güçlük düzeylerinin bir öğrenciden diğerine çok büyük farklılıklar göstermediği söylenebilir.

Öğrenci-puanlayıcı (bp) varyans bileşeni yine her seviyede ve tüm öğrencilerde, toplam varyansın %0'ını açıklamaktadır. Bu durumda, bir puanlayıcının yüksek puan verdiği öğrenciler diğer puanlayıcılardan da yüksek puan almıştır.

Ölçüt-puanlayıcı (op) etkileşiminin toplam varyansı açıklama yüzdeleri 4, 6 ve 7. sınıflar ile tüm öğrencilerde %8,5'in altında 3 ve 5. sınıflarda ise %12'nin üzerinde olduğundan puanlayıcıların bir ölçütü puanlarken çok katı, diğer ölçütte ise cömert olabildikleri görülmektedir.

Öğrenci-ölçüt-puanlayıcı (bop) değişkenlik kaynağının toplam varyansı açıklama oranları %40,6 ile %51,6 arasında değişmektedir. Ölçütler ve puanlayıcılar arası etkileşim ile bu çalışmada ölçülemeyen sistematik ya da tesadüfi değişkenlik kaynaklarının bulunduğu saptanmıştır.

Bu sonuçlar Güler (2008), Brennan (2011), Yelboğa (2007), Deliceoğlu (2009), Büyükkıdık (2012), Arsan (2012) ve Şalgam'ın (2016) tamamıyla çaprazlanmış desende 72 ile 397 arasında birey, 2 ile 9 arasında puanlayıcı, likert ölçekler, bütüncül ve analitik rubrik kullanarak elde ettikleri veriler ile örtüşmektedir.

$b \times \phi \times p$ deseninde puanlayıcı ve ölçüt sayılarının arttırılıp azaltılmasıyla yapılan karar çalışması (K) sonucunda elde edilen G ve Phi katsayıları incelenmiştir.

11 ölçütün yedi puanlayıcı tarafından puanlandığı asıl uygulama sonucunda G katsayılarının 0,96'nın üzerinde, ϕ katsayılarının 0,95'in üzerinde kestirildiği görülmektedir. Aynı zamanda teorik olarak olması gerektiği gibi benzer durumlar altında her ϕ katsayısı, G katsayısından küçük bulunmuştur. Elde edilen sonuçların yüksek değerlerde olması sebebi ile K çalışmalarında ölçüt ve puanlayıcı sayılarının artışlarını incelemek yerine kullanışlılık (ekonomiklik) sağlanması adına daha az puanlayıcı ve ölçüt ile 0,80'e yakın değerler elde edilmeye çalışılmıştır. Bu katsayılar sırasıyla en düşük 0,830 ve 0,790 olarak 3. sınıflarda 5 ölçüt ve 2 puanlayıcının olduğu durumda; en büyük G ve ϕ katsayılarının ise sırasıyla 0,990 ve 0,988 olarak 6. sınıflarda 22 ölçüt ve 8 puanlayıcının olduğu durumda elde edilmiştir.

Bu sonuçlar ise Güler (2011), Öztürk (2011)'ün çalışmaları ile farklı durumlar ortaya koymuş bu iki çalışmada G ve Phi katsayıları düşük düzeyde elde edilmiştir. Bu durumun sebebi ise Öztürk (2011)'ün çalışmasında gözlem formu, Güler (2011)'in çalışmasında geçerlik ve güvenilirliği düşük olması istenen rastgele veriler olarak belirtilmiştir.

Her seviyede ve tüm öğrencilerde ölçüt sayısının sabit tutularak puanlayıcı sayısı azaltıldığında G ve ϕ katsayıları da azalırken; puanlayıcı sayısı arttırıldığında G ve ϕ katsayıları da artmaktadır. Fakat katsayılar belli bir madde sayısı ve puanlayıcı sayısından sonra azalarak artmaktadır. Bu durumda ölçüt sayısı beş ve puanlayıcı sayısı iki olduğunda da genellenebilirlik ve güvenirliliğin yüksek düzeyde elde edileceği sonucuna ulaşılmıştır.

Bu sonuçlara ek olarak KTK'de görece daha küçük katsayıların elde edildiği altı ve yedinci puanlayıcılar analizden çıkarılarak elde edilen 5 puanlayıcıya ait G ve ϕ katsayıları ile K çalışmasında 5 puanlayıcının olduğu durumdaki G ve ϕ katsayıları karşılaştırılmış ve gerçek duruma ait katsayıların daha küçük olduğu görülmüştür. Benzer şekilde yalnız altıncı puanlayıcının analizden çıkarıldığı durumda da sonuç aynı olmuş katsayılar düşmüştür.

Bu durumda GK, değişkenlik kaynaklarını ayrıştırarak öğrenciler, ölçütler ve puanlayıcıları ayrı ayrı ve etkileşimlerini içeren sonuçlarla KTK'ye göre daha ayrıntılı sonuçlar vermiştir. Bunun yanında Pearson korelasyon katsayısı görece daha düşük olan puanlayıcılarla analiz yapıldığında G ve Phi katsayıları da bir miktar azalmıştır.

5.1.3. Üçüncü Alt Problem

De Ayala (2009) derecelendirilmiş tepki modeli sayıtlarının sağlanması için gereken cevaplayıcı sayısının alt sınırını 500 olarak önermiştir (s. 223). Buna ek olarak MTK'ye ait bu alt problem için tek boyutluluk ve yerel bağımsızlık varsayımları ile model veri uyumu incelenmiştir. Tek boyutluluk varsayımı için her puanlayıcının tüm öğrencileri puanlaması

sonucu elde edilen verilere Temel Bileşenler Analizi yapılmış ve ilk beş puanlayıcıya göre ölçeğin tek boyutlu, altı ve yedinci puanlayıcılara göre ise ölçeğin iki boyutlu olduğu görülmüştür. Çok boyutluluk bu çalışmanın konusunu oluşturmadığı için analizler ilk beş puanlayıcı üzerinden yapılmıştır. Bilindiği üzere ölçek tek boyutluluk özelliği gösterdiğinde yerel bağımsızlık varsayımı da karşılanmaktadır (Crocker & Algina, 2006, s. 42-46).

Bu nedenle sınıf ayrımı olmadan tüm öğrencilerin puanları ve ilk beş puanlayıcı için MTK’de analizler yapılmıştır.

Model veri uyumunda ise Multilog 7.03 programında elde edilen gözlenen ve beklenen oranların arasındaki fark incelenmiş en büyük artık değerin 0,0321 olduğu görülmüştür. Bu değerin sıfıra yakın ve küçük olması model veri uyumunun sağlandığını göstermiştir.

Alanyazın incelendiğinde Özdemir (2002), Nartgün (2002), Doğan & Tezbaşaran (2003), Kan (2006), Morales (2009), Gelbal (1994), Kelecioğlu (2001), Lee vd. (2012), Çelen & Aybek (2013)’in araştırmalarında başarı testleri ya da simülasyon ile üretilmiş veri kullanarak iki kategorili MTK modelleri; Sünbül (2011), Özer Özkan (2012)’in çok boyutlu MTK modelleri; İlhan (2016), Arsan (2012), Kim ve Wilson (2009), Ure (2011)’nin çok değişkenlik kaynaklı Rasch modeli kullanarak parametre kestirimleri yaptıkları görülmüştür.

Çok kategorili MTK modellerinden biri olan Samejima’nın Derecelendirilmiş Tepki Modeli (DTM)’ne göre yazılı anlatım becerisinin ölçülmesinde kestirilen madde ve test istatistikleri bulunmuştur.

11 maddenin a parametrelerinin tüm puanlayıcılarda 1,24 ile 6,44 arasında değiştiğinden her puanlayıcı için maddelerin ayırt ediciliklerinin ve verdikleri bilgilerin yüksek düzeyde olduğu görülmüştür. 1 ve 3. puanlayıcılara göre en fazla bilgiyi veren madde 8. ölçüt iken 2, 4 ve 5. puanlayıcılara göre en fazla bilgiyi 9. ölçüt vermektedir. En az bilgi veren madde ise tüm puanlayıcılara göre 11. ölçüt olarak bulunmuştur. Çok kategorili MTK modellerinin kullanıldığı Koch (1983), Nartgün (2002), Özdemir (2002) ve Köse

(2015)'nin çalışmalarında madde ayırt edicilik değerleri en yüksek Nartgün (2002)'ün çalışmasında erkek örnekleminde birinci madde için kestirilen 3,34 değeridir. Bu çalışmada ise ayırt edicilik değerinin 6,44 bulunması örneklemdaki okulların eğitim düzeyleri, öğrencilerin yazılı anlatım çalışmalarına aşinalığı, 3 ile 7. sınıflar arasındaki tüm öğrencilerin örnekleme dâhil edilmesinin olabileceği düşünülmektedir.

Tüm puanlayıcılarda b_1 değerleri -2,35 ile -0,39; b_2 değerleri -0,80 ile 0,74; b_3 değerleri 0,41 ile 1,90 arasındaki yetenek düzeylerinde değişmektedir. Tüm puanlayıcılar için bireylerin %50 ihtimalle 0 kategorisinden daha yüksek puan almaları için en düşük -2,35 yetenek düzeyinde, 1 kategorisinden daha yüksek puan almaları için en düşük -0,80 yetenek düzeyinde, 2 kategorisinden daha yüksek puan almaları için en düşük 0,41 yetenek düzeyinde bulunmaları gerekmektedir.

θ yetenek düzeylerinde ise 1. puanlayıcıya göre 3 ve 4. maddeleri; 2, 3, 4 ve 5. puanlayıcılara göre 4. maddeyi gruba göre en düşük yetenek düzeyindeki bireylerin, tüm puanlayıcılara göre 11. maddeyi gruba göre en yüksek yetenek düzeyine sahip bireylerin doğru yanıtlaması beklenmektedir.

Bu sonuçlara ek olarak test bilgi fonksiyonları ve hata fonksiyonları incelendiğinde de benzer sonuçların elde edildiği görülmüştür.

GK'ye göre maddelerin güçlük düzeyleri çok fazla farklılaşmıyor olarak bulunmasına rağmen MTK'ye göre b parametreleri -2,35 ile 1,90 ve θ düzeyleri -0,50 ile 1,20 arasında farklılaşmaktadır.

Marjinal güvenilirlik katsayıları incelendiğinde ise güvenirliliğin oldukça yüksek (0,93 civarında) olduğu görülmüştür. Bu bulgu Nartgün (2002) ve Köse (2015)'nin elde ettikleri marjinal güvenilirlik katsayıları ile çok yakınsa (0,97 ve 0,93) Özdemir (2002)'in elde ettiği katsayılardan (0,65 ile 0,73 arasında) farklılaşmaktadır. Marjinal güvenilirlik katsayısı dışında MTK'de güvenilirlik için tek bir katsayıya Morales (2009) -Person reliability (kişi

güvenirligi) 0,95- ve Çelen & Aybek (2013) -Empirical reliability (Görgül güvenirlilik) 0,80-'in çalışmalarında rastlanmıştır.

KTK ve GK ile benzer şekilde MTK'de de beş puanlayıcının öğrencilerin yazılı anlatım becerilerini puanlamaları sonucu elde edilen puanların güvenirligi yüksek düzeyde bulunmuştur. Sonuçta ölçümlerimiz için kullanılan üç güvenirlilik kuramlarından elde edilen güvenirlilik kestirimlerinin hepsi oldukça yüksek bulunmuştur.

5.1.4. Dördüncü Alt Problem

Bu çalışmada kullanılan üç ölçme teorisinden elde edilen güvenirlilik kestirimleri arasında manidar bir farklılık olup olmadığı iki şekilde incelenmiştir. İlk kısımda sınıf seviyelerine göre elde edilen KTK'deki Cronbach Alpha katsayıları, Eta korelasyon katsayıları ile GK'deki G ve Phi katsayıları karşılaştırılmıştır. Bu işlem için her sınıf seviyesindeki yedi puanlayıcıya ait Cronbach Alfa (α) katsayılarının ortancası alınmıştır. Her sınıf düzeyinde ve tüm öğrencilerde α ile G, α ile Phi, G ve Eta, Phi ve Eta katsayıları arasında 0,05 düzeyinde Genellenebilirlik Kuramı katsayıları lehine anlamlı bir fark bulunmuştur. Bu durumda güvenirlilik kestirimi için KTK ile GK katsayılarının farklılaştığı görülmüştür. Alanyazın incelendiğinde KTK ile GK'yi karşılaştıran çalışmalarda korelasyon katsayılarının arasındaki farkın manidarlığı için yapılan bir analizle karşılaşılmamıştır.

İkinci kısımda ise puanlayıcılara göre elde edilen KTK'deki Cronbach Alpha katsayıları ile MTK'deki marjinal güvenirlilik katsayıları karşılaştırılmıştır. 0,05 anlamlılık düzeyinde katsayılar arasında anlamlı bir fark olmadığı tespit edilmiştir. Bu durumda KTK ve MTK'ye göre puanlayıcılar arası güvenirlilik için benzer sonuçlar elde edilmiştir. Alanyazın incelendiğinde KTK ile MTK'yi karşılaştıran çalışmalarda Nartgün (2002) bu çalışma ile benzer olarak Cronbach Alfa iç tutarlılık katsayısı ile marjinal güvenirlilik katsayısı arasındaki farkı Fisher'in Z dönüşümü ile inceleyerek manidar bir fark olmadığı sonucuna ulaşmıştır. Doğan & Tezbaşaran (2003) ise bu çalışmadan farklı olarak KTK ve MTK'deki

madde ayırt edicilikleri ve güçlükleri arasındaki farkın manidarlığını Fisher'in Z dönüşümü ile incelemiş, manidar bir fark olmadığı sonucuna ulaşmıştır.

Sonuç olarak ölçümlerin güvenilirliğini kestirmeye yönelik olan bu çalışmaya göre, örneklem sayısı en az 500 olduğunda ve tek boyutluluk-yerel bağımsızlık varsayımları karşılandığında MTK'de Samejima'nın Derecelendirilmiş Tepki modeli ile madde düzeyinde hata kestirimleri yapmak madde ve test bilgi fonksiyonları aracılığıyla güvenilirlik kestirimleri yapmak KTK'ye göre daha ayrıntılı bilgiler sunmaktadır. Örneklem sayısı 500'den az, değişkenlik kaynakları ikiden fazla olduğunda, GK kullanılarak hata varyanslarının ayrı ayrı ve birlikte ele alınması ile mutlak ve bağıl kararlara göre farklılaşan genellenebilirlik ve güvenilirlik katsayıları hesaplamak KTK'den farklı olarak mümkün olmaktadır. Değişkenlik kaynağının tek olduğu çalışmalarda, geçti-kaldı kararları ya da araştırmacının sıralama yapma amacı olduğunda ise KTK'nin kullanılması daha kullanışlıdır.

5.2. Öneriler

Elde edilen sonuçlar doğrultusunda uygulayıcılar ve araştırmacılar için öneriler aşağıdaki gibidir:

1. Bu çalışmada yazılı anlatım becerilerinin ölçülmesi için öğrencilerin hikâye yazma becerileri üzerinde durulmuştur. Benzer çalışmalar problem çözme, akıl yürütme, bilimsel süreç becerileri gibi üst düzey becerileri ölçen araçlar kullanılarak yapılabilir.
2. Bu çalışmada yazılı anlatım becerilerinin puanlanmasında analitik puanlama anahtarı kullanılmıştır. Benzer çalışmalar hem analitik hem de bütüncül puanlama anahtarları kullanılarak ya da karşılaştırılarak yapılabilir.
3. GK'de öğrenci, ölçüt ve puanlayıcı değişkenlik kaynaklarının tamamıyla çaprazlandığı, ölçüt için 11, puanlayıcı için 7 koşulun bulunduğu desen

kullanılmıştır. Koşullar, değişkenlik kaynakları ve desen (yuvalanmış ya da karma desenler) değiştirilerek yeni çalışmalar yapılabilir.

4. MTK’de çok kategorili puanlanan maddeler için Samejima’nın DTM kullanılmıştır. Farklı puanlanan maddeler ve farklı modeller kullanılarak madde ve test istatistikleri incelenebilir.
5. MTK’de örneklem büyüklüğünün yeterli olmaması sebebiyle sınıf düzeyinde karşılaştırmalar yapılamamıştır. Her seviyede en az 500 öğrencinin bulunması sağlanarak Samejima’nın DTM’si çalışılabilir.
6. Yapılan çalışmaların sonucunda puanların evrene genellenmesi ya da mutlak veya bağıl karar verilmek istendiğinde GK; puanlayıcılar arasındaki güvenilirliği belirlemede ekonomik bir kuram olması sebebiyle GK; güvenilirliği madde bazında inceleyebilmek, bireylerin performanslarının maddelere göre değişmesini engellemek istendiğinde MTK kullanılması önerilmektedir.

KAYNAKLAR

- Arı, A. (2013). Bilişsel alan sınıflamasında yenilenmiş Bloom, SOLO, Fink, Dettmer Taksonomileri ve uluslararası alanda tanınma durumları. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 259-290.
- Arsan, N. (2012). *Buz pateninde hakem değerlendirmelerinin genellenebilirlik kuramı ve Rasch modeli ile incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Atılğan, H. (2005). Genellenebilirlik Kuramı ve Puanlayıcılar Arası Güvenirlik İçin Örnek Bir Uygulama. *Eğitim Bilimleri ve Uygulama*, 4(7), 95-108.
- Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. USA: The Guildford Press.
- Aydın, N., & Yılmaz, A. (2010). Yapılandırıcı yaklaşımın öğrencilerin üst düzey bilişsel becerilerine etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39, 57-68.
- Bağcı, V. (2015). *Matematiksel Muhakeme Becerisinin Ölçülmesinde Klasik Test Kuramı ile Genellenebilirlik Kuramındaki Farklı Desenlerin Karşılaştırılması (Yüksek Lisans Tezi)*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Baker, F. B. (2016). *Madde Tepki Kuramının Temelleri*. (N. Güler, Dü., & M. İlhan, Çev.) Ankara: Pegem Akademi.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.

- Brennan, R. L. (2011). Generalizability theory and classical test theory. *Applied Measurement And Education*, 24, 1-21.
- Büyükkıdık, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenirliğin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG*. USA: Routledge-Taylor & Francis Group.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. USA: Cengage Learning.
- Çelen, Ü., & Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2014). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları*. Ankara: Pegem Akademi.
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenirliklerinin karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- DeMars, C. (2010). *Item Response Theory*. USA: Oxford University Press.
- DeMars, C. (2016). Madde Tepki Kuramı. E. H. Özberk, & H. Kelecioğlu (Dü.) içinde, *Bölüm 1. Giriş* (s. 1-30). Ankara: Nobel Akademi.
- Doğan, N., & Tezbaşaran, A. A. (2003). Klasik test kuramının ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67.

- Ebel, R. L. (1965). *Measuring educational achievement*. . USA: Prentice-Hall, Inc.
- Erdoğan, S., Orekici Temel, G., Selvi, H., & Ersöz Kaya, I. (2017). An alternative method used in evaluating agreement among repeat measurements by two raters in education. *Educational Sciences: Theory & Practice*, 17, 217–236.
- Erkuş, A. (2017). *Bilimsel araştırma süreci*. Ankara: Seçkin.
- Erkuş, A., Sünbül, Ö., Ömür Sünbül, S., Aşiret, S., & Yormaz, S. (2017). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi.
- Gelbal, S. (1994). p Madde güçlük indeksi ile Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 10, 85-94.
- Göçer, A. (2011). Öğrencilerin yazılı anlatım çalışmalarının Türkçe öğretmenlerince değerlendirilmesi üzerine. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 30(2), 71-97.
- Güler, N. (2008). *Klasik test kuramı genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N. (2011). Rastgele veriler üzerinde genellenebilirlik kuramı ve klasik test kuramına göre güvenilirliğin karşılaştırılması. *Eğitim ve Bilim*, 36(162), 225-234.
- Güler, N., Kaya Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. Ankara: Pegem Akademi.
- Haladayna, T. M. (1997). *Writing test items to evaluate higher order thinking*. USA: Viacom Company.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement*(12), 38-47.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. USA: Sage Publications.

- 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.
- Kan, A. (2006). Klasik test teorisine ve örtük özellikler teorisine göre kestirilen madde parametrelerinin karşılaştırılması üzerine ampirik bir çalışma. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 2(2), 227-235.
- Karatay, H. (2015). Süreç temelli yazma modelleri: 4+1 Planlı yazma ve değerlendirme modeli. M. Özbay (Dü.) içinde, *Yazma eğitimi* (s. 21-48). Ankara: Pegem Akademi.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Kelecioğlu, H., & Göçer Şahin, S. (2014). Geçmişten günümüze geçerlik. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(2), 1-11.
- Kim, S., & Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review Journal*, 11, 179-188.
- Kim, S., & Wilson, M. (2009). A comparative analysis of the ratings in performance assessment using generalizability theory and many-facet rasch measurement. *Journal of Applied Measurement*, 10(4), 408-422.
- Kingston, N. M. (1985). *Assessing Guessing Behavior Using the Three-Parameter*. USA: Plus Postage.
- Koch, W. R. (1983). Likert scaling using the graded response latent trait model. *Applied Psychological Measurement*, 7(1), 15-32.
- Köse, A. (2014). Ölçmede Güvenirlilik. R. N. Demirtaşlı (Dü.) içinde, *Eğitimde ölçme ve değerlendirme* (s. 86-109). Ankara: Edge Akademi.

- Köse, A. (2015). Aşamalı tepki modeli ve klasik test kuramı altında elde edilen test ve madde parametrelerinin karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 184-197.
- Kutlu, Ö., Doğan, C., & Karakaya, İ. (2014). *Ölçme ve değerlendirme performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi.
- Lee, Y.-S., Torre, J. d., & Park, Y. S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices: an empirical investigation. *Asia Pacific Educ. Rev.*(13), 333–345.
- Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives*. USA: Corwin Press.
- Morales, R. A. (2009). Evaluation of Mathematics Achievement Test: A Comparison between CTT and IRT. *The International Journal of Educational and Psychological Assessment*, 1(1), 19-26.
- Morgan, C. (2013). *Psikolojiye giriş*. (S. Karakaş, & R. Eski, Çev.) Konya: Eğitim.
- Nartgün, Z. (2002). *Aynı tutumu ölçmeye yönelik likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. USA: Sage Publication.
- Özbek, Ö. Y. (2014). Ölçme araçlarında bulunması istenen nitelikler. S. Tekindal (Dü.) içinde, *Eğitimde ölçme ve değerlendirme* (s. 43-91). Ankara: Pegem Akademi.
- Özdemir, D. (2004). Çoktan seçmeli testlerin klasik test teorisi ve örtük özellikler teorisine göre hesaplanan psikometrik özelliklerinin iki kategorili ve ağırlıklandırılmış puanlanması yönünden karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 26, 117-123.

- Özer Özkan, Y. (2012). *Öğrenci başarılarının belirlenmesi sınavından (ÖBBS) klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Öztürk, M. E. (2011). *Voleybol becerileri gözlem formu ile elde edilen puanların genellenebilirlik ve klasik test kuramına göre karşılaştırılması*. . Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Popham, J. (2000). *Modern educational measurement*. USA: A Pearson Education Company.
- R.L.Brennan. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339–353.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. USA: Springer.
- Sebille, V., Hardouin, J.-B., Neel, T. L., Kubis, G., Boyer, F., Guillemin, F., & Falissard, B. (2010). Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulation study. *BMC Medical Research Methodology*(10), 24.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer*. USA: Sage Publications.
- Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. USA: CRC Press.
- Shoukri, M. M. (2010). *Measures of interobserver agreement and reliability*. USA: CRC Press.
- Stone, C. A., & Zhu, X. (2015). *Bayesian Analysis of Item Response Theory Models Using SAS*. USA: SAS Institute Inc.

- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9, 239–261.
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, tek boyutlu madde tepki kuramı ve çok boyutlu madde tepki kuramı çerçevesinde incelenmesi*. Doktora tezi, Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Mersin.
- Şalgam, A. (2016). *Kısa cevaplı matematik yazılı sınavının genellenebilirlik kuramı ve test tekrar test yöntemiyle güvenilirliğinin kıyaslanması*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Tan, Ş. (2014). *Öğretimde ölçme ve değerlendirme KPSS el kitabı*. Ankara: Pegem Akademi.
- Tang, K. (1996). *Polytomous item response theory (IRT) models and their applications in large-scale testing programs: review of literature*. New Jersey: Educational Testing Service.
- Tekin, H. (2015). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. USA: Wiley.
- Turgut, F., & Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM.
- Ure, A. C. (2011). *The effect of raters and rating conditions on the reliability of the missionary teaching assessment*. Master Thesis, University of Brigham Young, USA.
- Uyar, Ş., Öztürk Gübeş, N., & Kelecioğlu, H. (2013). PISA 2009 tutum anketi madde puanlarının aşamalı tepki modeli ile incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 2(4), 125-134.
- Yaşar, M. (2014). Ölçme ve değerlendirme ile ilgili temel kavramlar. S. Tekindal (Dü.) içinde, *Eğitimde ölçme ve değerlendirme* (s. 9-41). Ankara: Pegem Akademi.

- Yelboğa, A., & Tavşancıl, E. (2010). Klasik Test ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(3), 1825-1854.
- Yürekli, H. (2010). The Relationship Between Parameters from Some Polytomous Item Response Theory Models (Mater Thesis). *The Florida State University*. USA.
- Zoller, U. (1993). Are lecture and learning compatible? May be for locs; unlikely for HOCS. *Journal of Chemical Education*, 70(3), 195-197.



EKLER



EK 1.Yazılı Anlatım Becerisi Puanlama Anahtarı

YAZILI ANLATIM PUANLAMA ANAHTARI					
	Ölçütler	Üstün (3)	Kabul edilebilir (2)	Geliştirilmeli (1)	Notlar
DÜZEN	Kâğıt Kenarlarındaki Boşluk	Satır başında/sonunda ve sayfa üstünde/altında olmak üzere 4 kısımda uygun boşluk bırakılmıştır.	Bu 4 boşluktan 3 ya da 2 tanesinde uygun boşluk bırakılmıştır.	Bu 4 boşluktan sadece 1 tanesinde uygun boşluk bırakılmıştır.	
	Metin Düzeni	Paragraf başı, satır çizgisi takibi, harf büyüklüğündeki uyum ve kelimeler arasındaki boşluklar uygun şekilde yapılmıştır.	Bu 4 kriterden 3 ya da 2 tanesini uygun şekilde yapmıştır.	Bu 4 kriterden 1 tanesini uygun şekilde yapmıştır.	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı* -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	Metindeki bütün büyük harflerin yazımı doğru yapılmıştır.	Metindeki büyük harflerin yazımı %90 ve üzeri oranda doğru yapılmıştır.	Metindeki büyük harflerin yazımı %80 ve üzeri oranda doğru yapılmıştır.	* Kelime sayıları hesaplanırken bağlaçlar ve soru ekleri kelime sayısına dahil edilmeyecektir.
	Kelimelerin Yazımı* -Eksik/Fazla harfle yazım -Yanlış harfle yazım	Metindeki bütün kelimelerin yazımı doğru yapılmıştır.	Metindeki kelimelerin yazımı %95 ve üzeri oranda doğru yapılmıştır.	Metindeki kelimelerin yazımı %90 ve üzeri oranda doğru yapılmıştır.	
	Noktalama İşaretlerinin Kullanımı	Metindeki noktalama işaretleri %90 ve üzeri oranda doğru kullanılmıştır.	Metindeki noktalama işaretleri %80 ve üzeri oranda doğru kullanılmıştır.	Metindeki noktalama işaretleri %70 ve üzeri oranda doğru kullanılmıştır.	
İÇERİK VE ANLATIM	Ana Karakter -Fiziksel özellikler -Duyuşsal özellikler	Karakterin özellikleri ile ilgili en az 3 bilgi verilmiştir.	Karakterin özellikleri ile ilgili 2 bilgi verilmiştir.	Karakterin özellikleri ile ilgili 1 bilgi verilmiştir.	

	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemlerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	Hikâyenin gerçekleşme zamanı ve mekân belirtilmiş ve en az biri betimlenmiştir. Yüklemelerde kullanılan zaman ekleri anlatıma uygundur.	>Hikâyenin gerçekleşme zamanı ve mekân belirtilmiş fakat betimleme yapılmamıştır. Yüklemelerde kullanılan zaman ekleri anlatıma uygundur. >Hikâyenin gerçekleşme zamanı ve mekândan sadece biri belirtilip betimlenmiştir. Yüklemelerde kullanılan zaman ekleri anlatıma uygundur.	>Hikâyenin gerçekleşme zamanı ve mekândan en az bir tanesi belirtilmiş fakat betimleme yapılmamıştır. Yüklemelerde kullanılan zaman eklerinin anlatımına dikkat edilmemiştir. >Hikâyenin gerçekleşme zamanı ve mekândan bahsedilmemiştir. Yüklemelerde kullanılan zaman ekleri anlatıma uygundur.	
	Başlatıcı Olay ve Amaç	Hikâyenin amacına göre bir başlatıcı olay ayrıntılı olarak yazılmıştır.	Hikâyenin amacına göre bir başlatıcı olay yazılmıştır, ayrıntı verilmemiştir.	Başlatıcı olay amaca uygun olacak şekilde yazılmamıştır.	
	Kurgu ve Gelişme	Olay örgüleri özgün* bir biçimde anlatılmış, birbiri ile ilişkili olacak şekilde ve oluş sırasına uygun yazılmıştır.	Olay örgüleri birbiri ile ilişkili olacak şekilde ve oluş sırasına uygun yazılmıştır.	Olay örgüleri oluş sırasına göre yazılmamış ancak, özgün* bir biçimde anlatılmıştır.	*Olayların içeriğine dair farklı, detaylı, sürükleyici ve merak uyandırıcı anlatım özelliklerinden biri kullanılmıştır.
	Sonuç	Kurgu ve gelişmeyi amaca uygun olacak şekilde, mantık çerçevesinde ve farklı cümlelerle* sonlandırmıştır.	Kurgu ve gelişmeyi amaca uygun olacak şekilde, mantık çerçevesinde sonlandırmıştır.	Kurgu ve gelişmeyi amaca uygun olmayacak şekilde veya mantık çerçevesi dışında sonlandırmıştır.	*Atasözleri, ders çıkarma, benzetme veya kıyaslama yapma, sıradan cümlelerden kaçınma (sonsuz dek mutlu yaşadılar, vb.) örneklerinden biri kullanılmıştır.
	Başlık	Metnin konusunu ya da anafikrini <u>yurgulayacak</u> , en az kelime* kullanılarak <u>ilgi çekici</u> bir başlık yazılmıştır. Örn: Kaybolan Kardan Adam	Metnin konusuna ya da anafikrine uygun en az kelime kullanılarak bir başlık yazılmıştır. Örn: Eriyen Kardan Adam Örn: Kardan Adam	Bir başlık yazılmıştır, fakat konuyla tam olarak örtüşmemektedir. Örn: Kardan Adam ile Ece'nin Arkadaşlığı	*Cümle şeklinde yazılan başlıklar kabul edilmeyecektir.

EK 2a. Form 1: 3. sınıf hikâye yazma

Ön sayfa

ADI SOYADI :
OKULU-SINIFI :

Sevgili öğrenciler,
Aşağıda, bir yazılı anlatım konusu verilmiştir. Sizden bu konuya uygun ve arka sayfadaki değerlendirme ölçütlerini dikkate alarak bir hikâye yazmanız beklenmektedir.

Teşekkür eder başarılar dileriz.



Ormanlar bir yurdun zenginlik kaynağıdır. Ormanları olmayan bir ülke hayal ederek, bu ülkede canlıların yaşamının nasıl olacağını anlatan bir hikâye yazınız.

Arka sayfa

YAZILI ANLATIM PUANLAMA ANAHTARI		
	ÖLÇÜTLER	PUAN (0-3)
DÜZEN	Kâğıt Kenarlarındaki Boşluk	
	Metin Düzeni	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	
	Kelimelerin Yazımı -Eksik/Fazla harfle yazım -Yanlış harfle yazım	
	Noktalama işaretlerinin Kullanımı	
	Ana Karakter -Fiziksel özellikler -Duyusal özellikler	
İÇERİK VE ANLATIM	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemlerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	
	Bağlatıcı Olay ve Amaç	
	Kurgu ve Gelişme	
	Sonuç	
	Başlık	

EK 2b. Form 2: 4. sınıf hikâye yazma

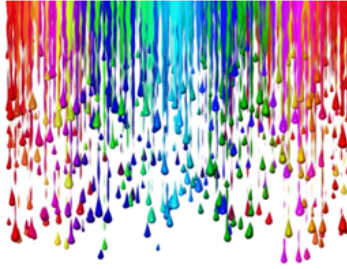
Ön sayfa

ADI SOYADI :

OKULU-SINIFI :

Sevgili öğrenciler,
Aşağıda, bir yazılı anlatım konusu verilmiştir. Sizden bu konuya uygun ve arka sayfadaki değerlendirme ölçütlerini dikkate alarak bir hikâye yazmanız beklenmektedir.

Teşekkür eder başarılar dileriz.



Renklerin insanlar üzerindeki etkileri çok çeşitlidir. Renkler bizi duygusal, sinirli, sakin, neşeli yapabilir. Yalnız bir rengin var olduğu bir ülke hayal ederek, bu ülkede insanların yaşamının nasıl olacağını anlatan bir hikâye yazınız.

Arka sayfa

YAZILI ANLATIM PUANLAMA ANAHTARI		
	ÖLÇÜTLER	PUAN (0-3)
DÜZEN	Kâğıt Kenarlarındaki Boşluk	
	Metin Düzeni	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	
	Kelimelerin Yazımı -Eksik/Fazla harfle yazım -Yanlış harfle yazım	
	Noktalama İşaretlerinin Kullanımı	
	Ana Karakter -Fiziksel özellikler -Duyusal özellikler	
İÇERİK VE ANLATIM	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemelerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	
	Başlatıcı Olay ve Amaç	
	Kurgu ve Gelişme	
	Sonuç	
	Başlık	

EK 2c. Form 3: 5. sınıf hikâye yazma

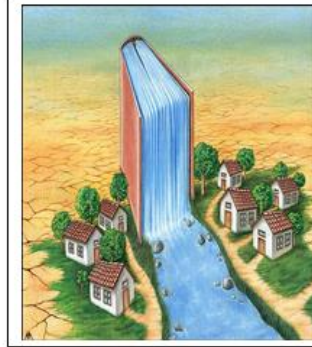
Ön sayfa

ADI SOYADI :
OKULU-SINIFI :

Sevgili öğrenciler,
Aşağıda, bir yazılı anlatım konusu verilmiştir. Sizden bu konuya uygun ve arka sayfadaki değerlendirme ölçütlerini dikkate alarak bir hikâye yazmanız beklenmektedir.

Teşekkür eder başarılar dileriz.

Kitaplar, bizim için en büyük hazinedir.
Kitapları olmayan bir ülke hayal ederek, bu ülkede canlıların yaşamının nasıl olacağını anlatan bir hikâye yazınız.



Arka sayfa

YAZILI ANLATIM PUANLAMA ANAHTARI		
	ÖLÇÜTLER	PUAN (0-3)
DÜZEN	Kâğıt Kenarlarındaki Boşluk	
	Metin Düzeni	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	
	Kelimelerin Yazımı -Eksik/Fazla harfle yazım -Yanlış harfle yazım	
	Noktalama işaretlerinin Kullanımı	
	Ana Karakter -Fiziksel özellikler -Duyusal özellikler	
İÇERİK VE ANLATIM	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemelerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	
	Başlatıcı Olay ve Amaç	
	Kurgu ve Gelişme	
	Sonuç	
	Başlık	

EK 2d. Form 4: 6. sınıf hikâye yazma

Ön sayfa

ADI SOYADI :
OKULU-SINIFI :

Sevgili öğrenciler,
Aşağıda, bir yazılı anlatım konusu verilmiştir. Sizden bu konuya uygun ve arka sayfadaki değerlendirme ölçütlerini dikkate alarak bir hikâye yazmanız beklenmektedir.

Teşekkür eder başarılar dileriz.



Sadece bir öğretmen verdiği emeklerle geleceği aydınlatabilir. Öğretmenlerimizin bu emeklerinin boşa gitmeyeceğini anlatan bir hikâye yazınız.

Arka sayfa

YAZILI ANLATIM PUANLAMA ANAHTARI		
ÖLÇÜTLER		PUAN (0-3)
DÜZEN	Kâğıt Kenarlarındaki Boşluk	
	Metin Düzeni	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	
	Kelimelerin Yazımı -Eksik/Fazla harfle yazım -Yanlış harfle yazım	
	Noktalama işaretlerinin Kullanımı	
	Ana Karakter -Fiziksel özellikler -Duyusal özellikler	
İÇERİK VE ANLATIM	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemelerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	
	Başlatıcı Olay ve Amaç	
	Kurgu ve Gelişme	
	Sonuç	
	Başlık	

EK 2e. Form 5: 7. sınıf hikâye yazma

Ön sayfa

ADI SOYADI :
OKULU-SINIFI :

Sevgili öğrenciler,
Aşağıda, bir yazılı anlatım konusu verilmiştir. Sizden bu konuya uygun ve arka sayfadaki değerlendirme ölçütlerini dikkate alarak bir hikâye yazmanız beklenmektedir.

Teşekkür eder başarılar dileriz.

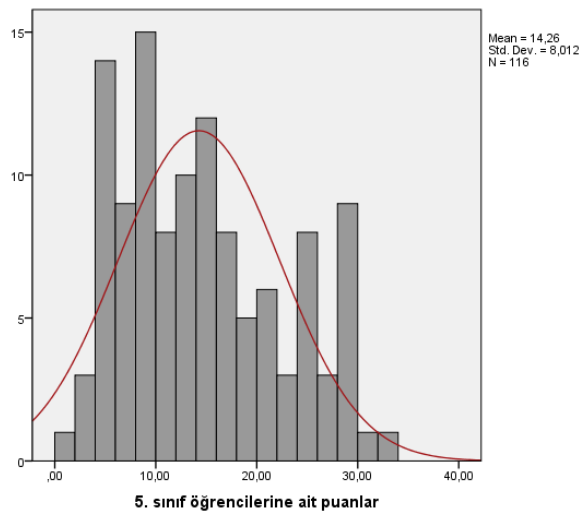
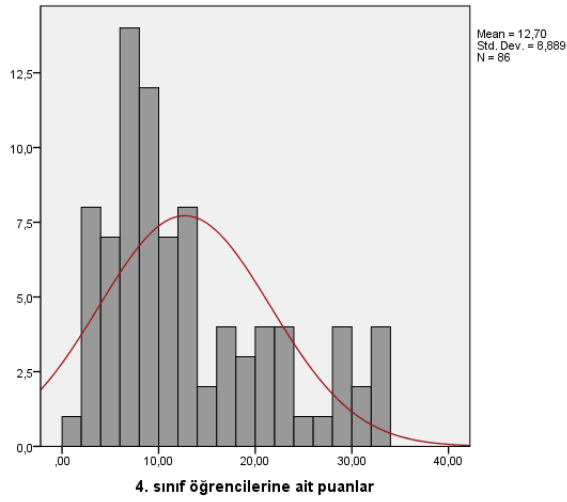
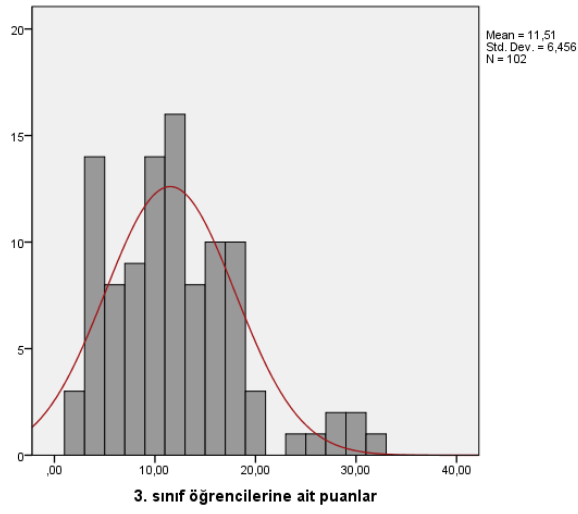


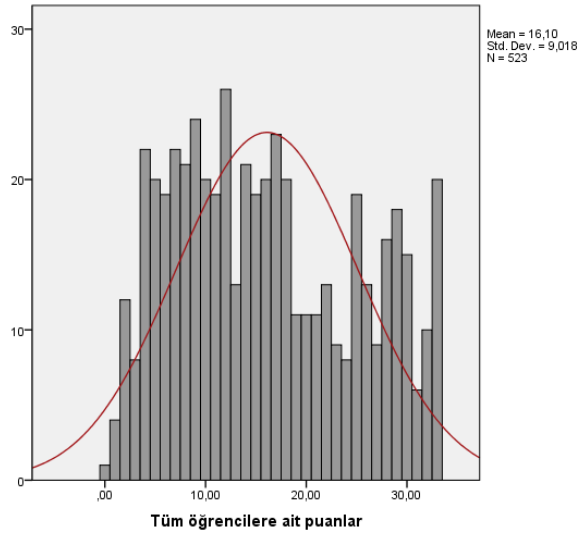
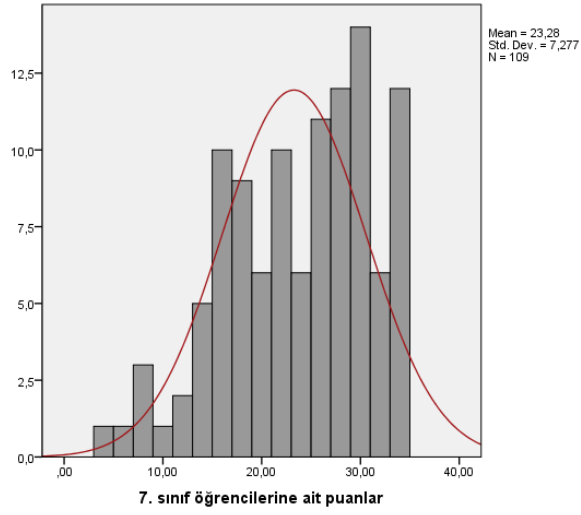
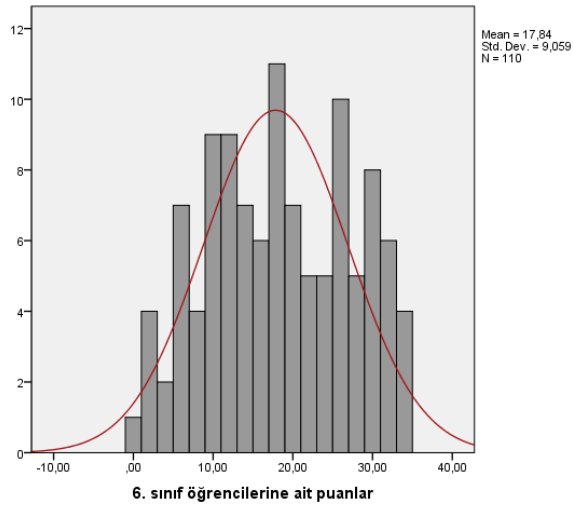
Ayrımcılık, insanların *din, dil, cinsiyet, ırk, siyasi düşünce, felsefi inanç, bedensel farklılık, sosyal sınıf vb.* nedenlerle diğer insanlar tarafından olumsuz muamele görmesi olarak tanımlanmaktadır. Ayrımcılığa uğrayan bir çocuğun başından geçenleri anlatan bir hikâye yazınız.

Arka sayfa

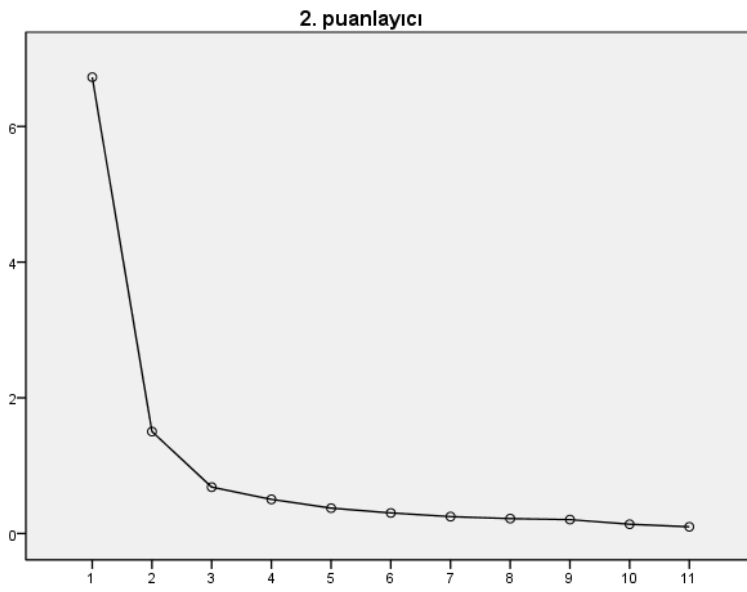
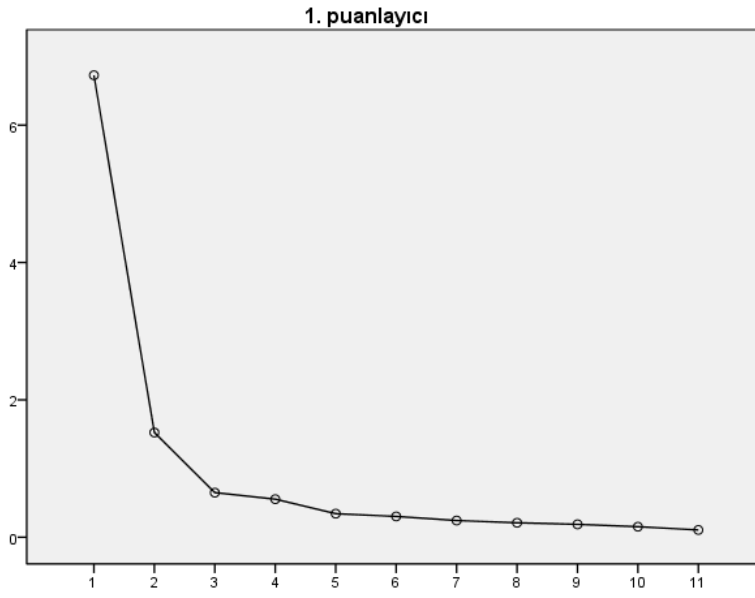
YAZILI ANLATIM PUANLAMA ANAHTARI		
	ÖLÇÜTLER	PUAN (0-3)
DÜZEN	Kâğıt Kenarlarındaki Boşluk	
	Metin Düzeni	
YAZIM VE NOKTALAMA	Büyük Harflerin Kullanımı -Özel isimlerde -Noktalama işaretlerinden sonra -Başlıkta	
	Kelimelerin Yazımı -Eksik/Fazla harfle yazım -Yanlış harfle yazım	
	Noktalama İşaretlerinin Kullanımı	
	Ana Karakter -Fiziksel özellikler -Duyusal özellikler	
İÇERİK VE ANLATIM	Zaman ve Mekân -Hikâyenin Gerçekleşme Zamanı -Yüklemlerde Kullanılan Zaman Ekleri -Hikâyenin Gerçekleştiği Yer	
	Bağlatıcı Olay ve Amaç	
	Kurgu ve Gelişme	
	Sonuç	
	Başlık	

EK 3. Verilerin dağılımına ilişkin histogram grafikleri

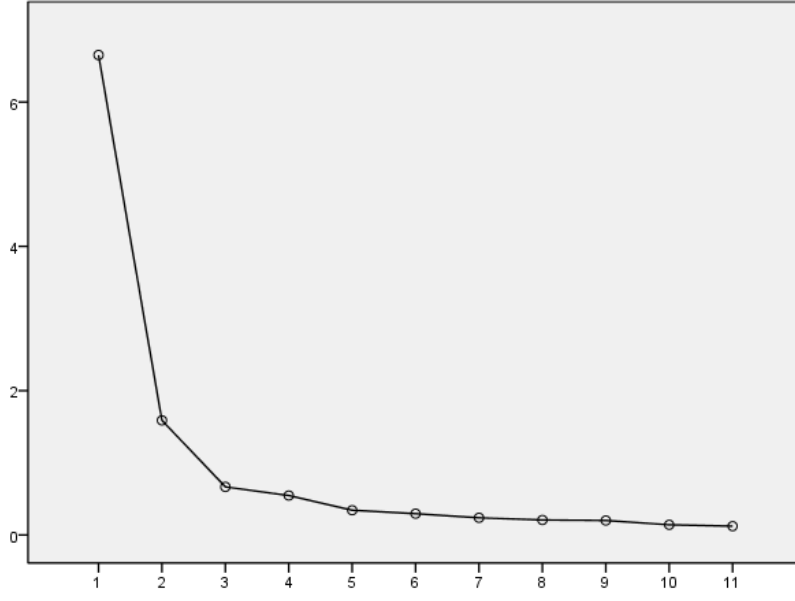




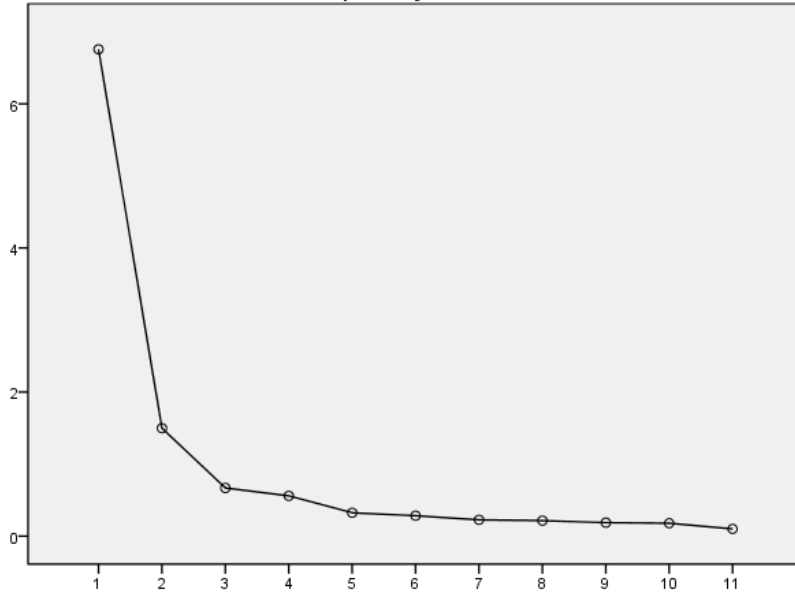
EK 4. Her bir puanlayıcıya ait tüm veriler için yamaç-birikinti grafikleri



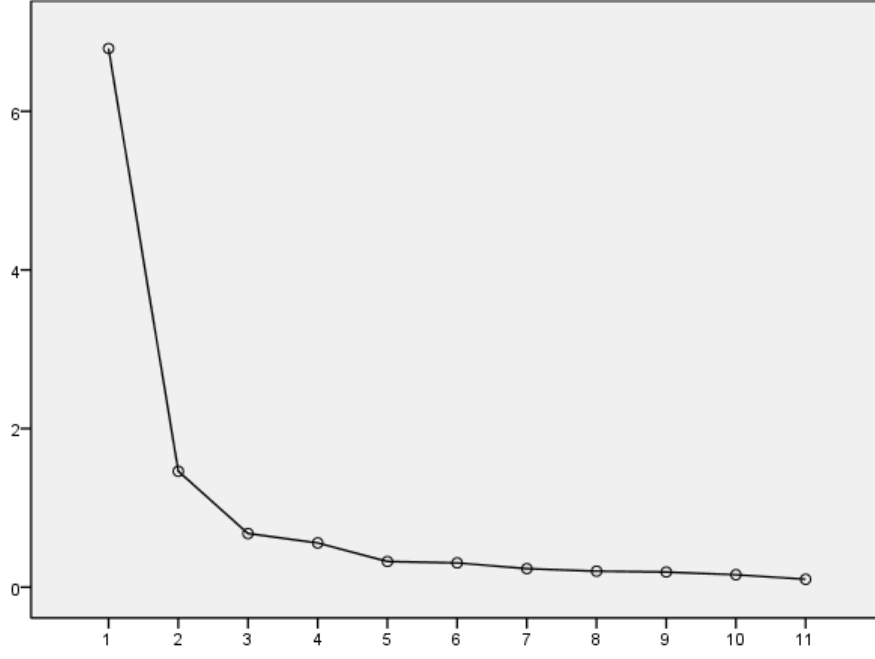
3. puanlayıcı



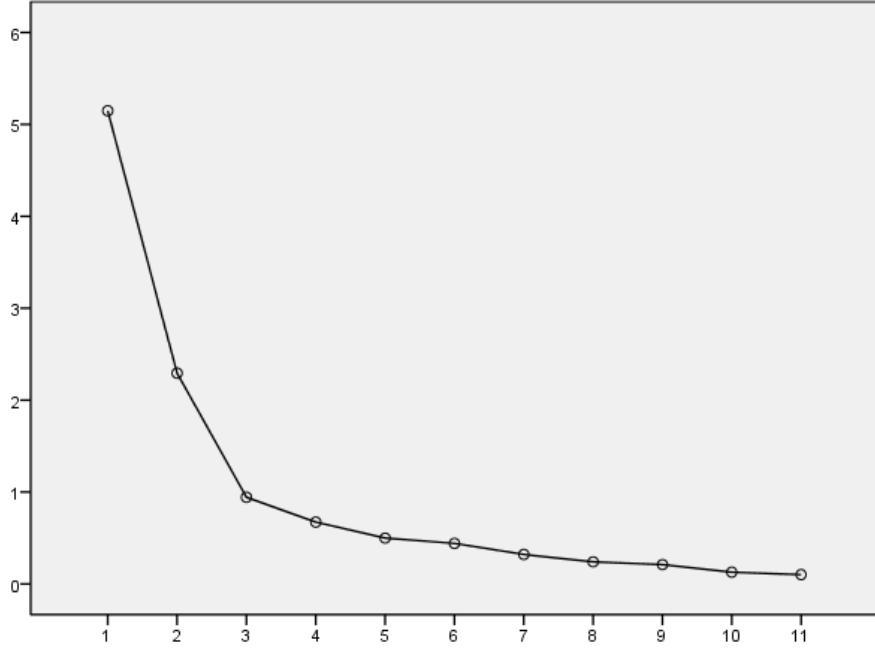
4. puanlayıcı

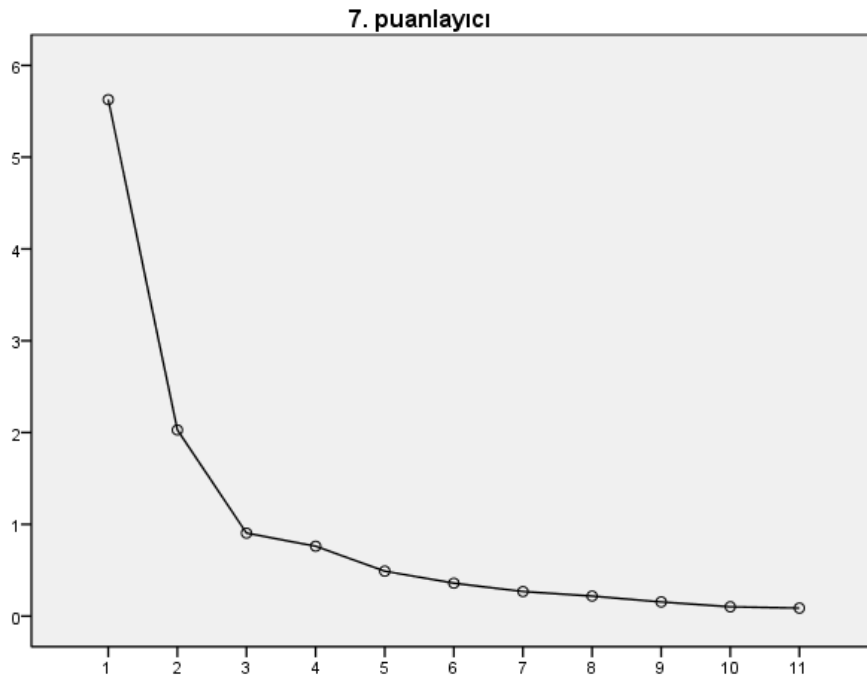


5. puanlayıcı



6. puanlayıcı





EK 5. GRM için MULTILOG komut dosyası örneği

MULTILOG for Windows 7.00.2327.2

Created on: 13 April 2018, 23:19:11

>PROBLEM RANDOM,

INDIVIDUAL,

DATA = 'C:\Users\merve.seheryeli\Desktop\MTK\Toplam\p1\p1.DAT',

NITEMS = 11,

NGROUPS = 1,

NEXAMINEES = 523,

NCHARS = 3;

>TEST ALL,

GRADED,

NC = (4(0)11);

>ESTIMATE,

NCYCLES = 99999,

ITERATIONS = 999;

>END;

4

1234

11111111111

22222222222

33333333333

44444444444

(3A1,T4,11A1)

EK 6. Yazılı anlatım becerileri puanlama anahtarının madde parametreleri kestirimi

MULTILOG çıktısı örneği – 1. Puanlayıcı

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION
7.0.3

MULTILOG for Windows 7.00.2327.2

Created on: 13 April 2018, 22:40:36

>PROBLEM RANDOM,

INDIVIDUAL,

DATA = 'C:\Users\merve.seheryeli\Desktop\MTK\Toplam\p1\p1.DAT',

NITEMS = 11,

NGROUPS = 1,

NEXAMINEES = 523,

NCHARS = 3;

DATA FILE NAME IS

C:\USERS\MERVE.SEHERYELI\DESKTOP\MTK\TOPLAM\P1\P1.DAT

TYPE OF INPUT:

INDIVIDUAL RESPONSE VECTORS

>TEST ALL,

GRADED,

NC = (4(0)11);

>ESTIMATE,

NCYCLES = 99999,

ITERATIONS = 999;

NUMBER OF CODES 4

1234

VECTOR OF CATEGORIES FOR CODE=1

1111111111

VECTOR OF CATEGORIES FOR CODE=2

2222222222

VECTOR OF CATEGORIES FOR CODE=3

3333333333

VECTOR OF CATEGORIES FOR CODE=4

4444444444

(3A1,T4,11A1)

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION
7.0.3

MULTILOG for Windows 7.00.2327.2

Created on: 13 April 2018, 22:40:36

DATA PARAMETERS:

NUMBER OF LINES IN THE DATA FILE: 523

NUMBER OF CATEGORICAL-RESPONSE ITEMS: 11

NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS: 1

TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS): 12

NUMBER OF CHARACTERS IN ID FIELDS: 3

MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM: 4

THE MISSING VALUE CODE FOR CONTINUOUS DATA: 9.0000

THE DATA WILL BE STORED IN MEMORY

ESTIMATION PARAMETERS:

THE ITEMS WILL BE CALIBRATED--

BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

MAXIMUM NUMBER OF EM CYCLES PERMITTED:*****

NUMBER OF PARAMETER-SEGMENTS USED IS: 11

NUMBER OF FREE PARAMETERS IS: 44
MAXIMUM NUMBER OF M-STEP ITERATIONS IS 999 TIMES
THE NUMBER OF PARAMETERS IN THE SEGMENT
THE M-STEP CONVERGENCE CRITERION IS: 0.000100
THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS: 0.9000
THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS: 1.0000
THE MAXIMUM ACCELERATION PERMITTED IS: 0.0000
THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED

QUADRATURE POINTS FOR MML,
AT THETA:

-4.500
-4.000
-3.500
-3.000
-2.500
-2.000
-1.500
-1.000
-0.500
0.000
0.500
1.000
1.500
2.000
2.500
3.000
3.500
4.000
4.500

MULTILOG for Windows 7.00.2327.2

READING DATA...

KEY-

CODE CATEGORY

```

1  1111111111
2  2222222222
3  3333333333
4  4444444444

```

FORMAT FOR DATA-

(3A1,T4,11A1)

FIRST OBSERVATION AS READ-

```

ID   001
ITEMS 11322132222
NORML   0.000

```

```

FINISHED CYCLE 109
MAXIMUM INTERCYCLE PARAMETER CHANGE=  0.00089 P( 33)

```

ITEM SUMMARY

MULTILOG for Windows 7.00.2327.2

ITEM 1: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

```

A      1  1.45 (0.16)
B( 1)   2 -1.15 (0.14)
B( 2)   3  0.40 (0.11)
B( 3)   4  1.36 (0.15)

```

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

```

-3.0 - -1.6 0.126 0.162 0.204 0.254 0.310 0.370 0.429 0.483
-1.4 -  0.0 0.528 0.560 0.578 0.586 0.590 0.592 0.598 0.608
 0.2 -  1.6 0.618 0.628 0.633 0.633 0.626 0.610 0.583 0.544
 1.8 -  3.0 0.495 0.438 0.376 0.315 0.258 0.208 0.164

```

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 107 196 112 108

OBS. PROP. 0.2046 0.3748 0.2141 0.2065

EXP. PROP. 0.2282 0.3740 0.2085 0.1893

ITEM 2: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 5 1.54 (0.16)
B(1) 6 -1.32 (0.15)
B(2) 7 0.06 (0.10)
B(3) 8 1.21 (0.13)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.154 0.200 0.255 0.319 0.389 0.462 0.530 0.589
-1.4 - 0.0 0.631 0.658 0.671 0.677 0.681 0.688 0.696 0.703
0.2 - 1.6 0.707 0.707 0.704 0.696 0.682 0.657 0.617 0.562
1.8 - 3.0 0.496 0.423 0.350 0.283 0.224 0.174 0.133

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 90 174 142 117

OBS. PROP. 0.1721 0.3327 0.2715 0.2237

EXP. PROP. 0.1890 0.3263 0.2754 0.2093

ITEM 3: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 9 1.71 (0.18)
B(1) 10 -1.33 (0.14)
B(2) 11 -0.28 (0.09)
B(3) 12 0.68 (0.11)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.150 0.202 0.269 0.350 0.443 0.543 0.642 0.728
-1.4 - 0.0 0.794 0.836 0.860 0.873 0.882 0.888 0.891 0.888
0.2 - 1.6 0.880 0.864 0.834 0.784 0.710 0.619 0.518 0.419
1.8 - 3.0 0.328 0.251 0.188 0.139 0.101 0.073 0.053

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 80 131 140 172

OBS. PROP. 0.1530 0.2505 0.2677 0.3289

EXP. PROP. 0.1747 0.2455 0.2643 0.3155

ITEM 4: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 13 1.54 (0.18)
B(1) 14 -2.17 (0.25)
B(2) 15 -0.61 (0.10)
B(3) 16 0.44 (0.10)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.406 0.477 0.542 0.594 0.629 0.647 0.652 0.651
-1.4 - 0.0 0.652 0.659 0.672 0.686 0.699 0.706 0.708 0.703
0.2 - 1.6 0.690 0.665 0.627 0.573 0.507 0.435 0.362 0.293
1.8 - 3.0 0.233 0.181 0.139 0.106 0.080 0.060 0.044

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 33 135 146 209

OBS. PROP. 0.0631 0.2581 0.2792 0.3996

EXP. PROP. 0.0771 0.2646 0.2752 0.3831

ITEM 5: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 17 1.67 (0.18)
B(1) 18 -1.47 (0.16)
B(2) 19 -0.14 (0.09)
B(3) 20 0.71 (0.11)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.186 0.246 0.319 0.403 0.494 0.585 0.665 0.725
-1.4 - 0.0 0.763 0.780 0.786 0.792 0.802 0.816 0.831 0.841
0.2 - 1.6 0.842 0.831 0.804 0.758 0.691 0.608 0.515 0.421
1.8 - 3.0 0.334 0.259 0.197 0.147 0.109 0.080 0.058

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 70 163 120 170

OBS. PROP. 0.1338 0.3117 0.2294 0.3250

EXP. PROP. 0.1541 0.3059 0.2294 0.3107

ITEM 6: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 21 3.05 (0.25)
 B(1) 22 -0.44 (0.06)
 B(2) 23 0.44 (0.06)
 B(3) 24 1.33 (0.08)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.004 0.007 0.013 0.023 0.043 0.078 0.142 0.255
 -1.4 - 0.0 0.448 0.759 1.210 1.757 2.242 2.481 2.472 2.428
 0.2 - 1.6 2.497 2.582 2.540 2.441 2.440 2.496 2.382 1.988
 1.8 - 3.0 1.444 0.939 0.567 0.327 0.184 0.102 0.056

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 173 157 120 73

OBS. PROP. 0.3308 0.3002 0.2294 0.1396

EXP. PROP. 0.3516 0.2976 0.2250 0.1258

ITEM 7: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 25 4.14 (0.34)
 B(1) 26 -0.46 (0.05)
 B(2) 27 0.35 (0.05)
 B(3) 28 1.11 (0.06)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.001 0.002 0.006 0.013 0.029 0.066 0.150
 -1.4 - 0.0 0.336 0.729 1.492 2.704 3.972 4.423 4.014 3.847
 0.2 - 1.6 4.319 4.560 4.190 4.072 4.421 4.223 3.085 1.786
 1.8 - 3.0 0.896 0.417 0.188 0.083 0.037 0.016 0.007

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4

OBS. FREQ. 164 156 114 89

OBS. PROP. 0.3136 0.2983 0.2180 0.1702

EXP. PROP. 0.3366 0.2909 0.2194 0.1531

ITEM 8: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 29 5.98 (0.54)
 B(1) 30 -0.44 (0.04)
 B(2) 31 0.39 (0.05)
 B(3) 32 0.95 (0.04)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.001 0.003 0.010 0.034
 -1.4 - 0.0 0.113 0.369 1.163 3.311 7.144 8.911 6.213 4.715
 0.2 - 1.6 6.974 9.298 8.261 8.654 8.886 5.339 2.117 0.701
 1.8 - 3.0 0.218 0.067 0.020 0.006 0.002 0.001 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4
 OBS. FREQ. 167 163 86 107
 OBS. PROP. 0.3193 0.3117 0.1644 0.2046
 EXP. PROP. 0.3382 0.3071 0.1721 0.1825

ITEM 9: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 33 5.73 (0.47)
 B(1) 34 -0.44 (0.04)
 B(2) 35 0.32 (0.05)
 B(3) 36 0.92 (0.04)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.001 0.004 0.014 0.043
 -1.4 - 0.0 0.133 0.411 1.225 3.281 6.700 8.247 6.230 5.539
 0.2 - 1.6 7.683 8.385 7.219 8.110 7.867 4.564 1.847 0.639
 1.8 - 3.0 0.209 0.067 0.021 0.007 0.002 0.001 0.000

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4
 OBS. FREQ. 167 151 93 112
 OBS. PROP. 0.3193 0.2887 0.1778 0.2141
 EXP. PROP. 0.3382 0.2818 0.1891 0.1909

ITEM 10: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 37 5.05 (0.42)
 B(1) 38 -0.44 (0.04)
 B(2) 39 0.39 (0.05)
 B(3) 40 0.95 (0.05)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.001 0.004 0.010 0.027 0.073
 -1.4 - 0.0 0.198 0.530 1.354 3.083 5.471 6.442 5.167 4.409
 0.2 - 1.6 5.586 6.807 6.583 6.693 6.446 4.344 2.119 0.869
 1.8 - 3.0 0.331 0.123 0.045 0.016 0.006 0.002 0.001

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4
 OBS. FREQ. 167 161 85 110
 OBS. PROP. 0.3193 0.3078 0.1625 0.2103
 EXP. PROP. 0.3390 0.3053 0.1689 0.1868

ITEM 11: 4 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 41 1.38 (0.15)
 B(1) 42 -1.74 (0.22)
 B(2) 43 0.70 (0.12)
 B(3) 44 1.60 (0.17)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.242 0.290 0.341 0.389 0.432 0.465 0.485 0.490
 -1.4 - 0.0 0.482 0.466 0.448 0.432 0.424 0.426 0.439 0.460
 0.2 - 1.6 0.486 0.513 0.536 0.554 0.563 0.562 0.553 0.532
 1.8 - 3.0 0.501 0.461 0.413 0.361 0.308 0.257 0.210

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3 4
 OBS. FREQ. 66 286 84 87
 OBS. PROP. 0.1262 0.5468 0.1606 0.1663
 EXP. PROP. 0.1391 0.5328 0.1690 0.1590

ITEM 12: GRP1, N[MU: 0.00 SIGMA: 1.00]

P(#);(S.E.): 46; (0.00) 47; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 -	-1.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-1.4 -	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.2 -	1.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.8 -	3.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TOTAL TEST INFORMATION

@THETA: INFORMATION:

-3.0 -	-1.6	2.268	2.585	2.945	3.340	3.760	4.197	4.662	5.222
-1.4 -	0.0	6.079	7.758	11.458	19.181	30.607	35.620	29.258	26.142
0.2 -	1.6	32.282	36.840	33.931	34.967	34.839	25.456	15.698	9.755
1.8 -	3.0	6.485	4.625	3.504	2.790	2.309	1.972	1.727	

@THETA: POSTERIOR STANDARD DEVIATION:

-3.0 -	-1.6	0.664	0.622	0.583	0.547	0.516	0.488	0.463	0.438
-1.4 -	0.0	0.406	0.359	0.295	0.228	0.181	0.168	0.185	0.196
0.2 -	1.6	0.176	0.165	0.172	0.169	0.169	0.198	0.252	0.320
1.8 -	3.0	0.393	0.465	0.534	0.599	0.658	0.712	0.761	

MARGINAL RELIABILITY: 0.9313

NEGATIVE TWICE THE LOGLIKELIHOOD= 5146.5

(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

NORMAL PROGRAM TERMINATION

START DATE: 04-13-2018

START TIME: 22:42:24

END TIME: 22:42:25

EK 7. Yazılı anlatım becerileri puanlama anahtarının madde kategorilerine ait beklenen ve gözlenen oranlar ve aralarındaki farklar

	1. puanlayıcı	1. kategori		2. kategori		3. kategori		4. kategori	
1	Gözlenen oran	0,2046	-0,0236	0,3748	0,0008	0,2141	0,0056	0,2065	0,0172
	Beklenen oran	0,2282		0,3740		0,2085		0,1893	
2	Gözlenen oran	0,1721	-0,0169	0,3327	0,0064	0,2715	-0,0039	0,2237	0,0144
	Beklenen oran	0,1890		0,3263		0,2754		0,2093	
3	Gözlenen oran	0,1530	-0,0217	0,2505	0,0050	0,2677	0,0034	0,3289	0,0134
	Beklenen oran	0,1747		0,2455		0,2643		0,3155	
4	Gözlenen oran	0,0631	-0,0140	0,2581	-0,0065	0,2792	0,0040	0,3996	0,0165
	Beklenen oran	0,0771		0,2646		0,2752		0,3831	
5	Gözlenen oran	0,1338	-0,0203	0,3117	0,0058	0,2294	0,0000	0,3250	0,0143
	Beklenen oran	0,1541		0,3059		0,2294		0,3107	
6	Gözlenen oran	0,3308	-0,0208	0,3002	0,0026	0,2294	0,0044	0,1396	0,0138
	Beklenen oran	0,3516		0,2976		0,2250		0,1258	
7	Gözlenen oran	0,3136	-0,0230	0,2983	0,0074	0,2180	-0,0014	0,1702	0,0171
	Beklenen oran	0,3366		0,2909		0,2194		0,1531	
8	Gözlenen oran	0,3193	-0,0189	0,3117	0,0046	0,1644	-0,0077	0,2046	0,0221
	Beklenen oran	0,3382		0,3071		0,1721		0,1825	
9	Gözlenen oran	0,3193	-0,0189	0,2887	0,0069	0,1778	-0,0113	0,2141	0,0232
	Beklenen oran	0,3382		0,2818		0,1891		0,1909	
10	Gözlenen oran	0,3193	-0,0197	0,3078	0,0025	0,1625	-0,0064	0,2103	0,0235
	Beklenen oran	0,3390		0,3053		0,1689		0,1868	
11	Gözlenen oran	0,1262	-0,0129	0,5468	0,0140	0,1606	-0,0084	0,1663	0,0073
	Beklenen oran	0,1391		0,5328		0,1690		0,1590	
	2. puanlayıcı	1. kategori		2. kategori		3. kategori		4. kategori	
1	Gözlenen oran	0,2505	-0,0174	0,3040	-0,0023	0,2447	0,0008	0,2008	0,0189
	Beklenen oran	0,2679		0,3063		0,2439		0,1819	
2	Gözlenen oran	0,1855	-0,0106	0,3021	0,0032	0,2811	-0,0098	0,2314	0,0172
	Beklenen oran	0,1961		0,2989		0,2909		0,2142	
3	Gözlenen oran	0,1530	-0,0136	0,2658	-0,0010	0,2696	-0,0034	0,3117	0,0181
	Beklenen oran	0,1666		0,2668		0,2730		0,2936	
4	Gözlenen oran	0,0937	-0,0103	0,2333	-0,0036	0,2849	-0,0033	0,3881	0,0172
	Beklenen oran	0,1040		0,2369		0,2882		0,3709	
5	Gözlenen oran	0,1472	-0,0127	0,2868	-0,0003	0,2467	-0,0059	0,3193	0,0189
	Beklenen oran	0,1599		0,2871		0,2526		0,3004	
6	Gözlenen oran	0,3327	-0,0149	0,3193	-0,0148	0,2294	0,0123	0,1185	0,0173
	Beklenen oran	0,3476		0,3341		0,2171		0,1012	
7	Gözlenen oran	0,3231	-0,0110	0,2658	-0,0096	0,2447	-0,0021	0,1663	0,0227
	Beklenen oran	0,3341		0,2754		0,2468		0,1436	
8	Gözlenen oran	0,3423	-0,0070	0,2620	-0,0136	0,1931	-0,0068	0,2027	0,0275
	Beklenen oran	0,3493		0,2756		0,1999		0,1752	

9	Gözlenen oran	0,3499	-0,0045	0,2562	-0,0134	0,1893	-0,0104	0,2046	0,0283
	Beklenen oran	0,3544		0,2696		0,1997		0,1763	
10	Gözlenen oran	0,3231	-0,0081	0,2906	-0,0129	0,1836	-0,0048	0,2027	0,0257
	Beklenen oran	0,3312		0,3035		0,1884		0,1770	
11	Gözlenen oran	0,1243	-0,0060	0,5430	0,0050	0,1816	-0,0097	0,1511	0,0107
	Beklenen oran	0,1303		0,5380		0,1913		0,1404	
3. puanlayıcı		1. kategori		2. kategori		3. kategori		4. kategori	
1	Gözlenen oran	0,2218	0,0059	0,3327	0,0062	0,2600	-0,0082	0,1855	-0,004
	Beklenen oran	0,2159		0,3265		0,2682		0,1895	
2	Gözlenen oran	0,2141	0,0109	0,2887	0,0058	0,2639	-0,0110	0,2333	-0,006
	Beklenen oran	0,2032		0,2829		0,2749		0,2391	
3	Gözlenen oran	0,1491	0,0050	0,2524	0,0107	0,2734	-0,0031	0,3250	-0,013
	Beklenen oran	0,1441		0,2417		0,2765		0,3376	
4	Gözlenen oran	0,0765	0,0018	0,2409	0,0095	0,2983	0,0023	0,3843	-0,014
	Beklenen oran	0,0747		0,2314		0,2960		0,3979	
5	Gözlenen oran	0,1644	0,0060	0,2581	0,0124	0,2600	-0,0055	0,3174	-0,013
	Beklenen oran	0,1584		0,2457		0,2655		0,3304	
6	Gözlenen oran	0,3365	0,0228	0,2906	-0,0054	0,2505	-0,0118	0,1224	-0,006
	Beklenen oran	0,3137		0,2960		0,2623		0,1280	
7	Gözlenen oran	0,3250	0,0260	0,2677	-0,0066	0,2409	-0,0147	0,1663	-0,005
	Beklenen oran	0,2990		0,2743		0,2556		0,1710	
8	Gözlenen oran	0,3442	0,0321	0,2868	-0,0109	0,1855	-0,0177	0,1836	-0,003
	Beklenen oran	0,3121		0,2977		0,2032		0,1871	
9	Gözlenen oran	0,3518	0,0315	0,2467	-0,0061	0,2161	-0,0215	0,1855	-0,004
	Beklenen oran	0,3203		0,2528		0,2376		0,1893	
10	Gözlenen oran	0,3002	0,0265	0,3212	-0,0022	0,1836	-0,0187	0,1950	-0,006
	Beklenen oran	0,2737		0,3234		0,2023		0,2006	
11	Gözlenen oran	0,1262	0,0080	0,5392	0,0224	0,2065	-0,0251	0,1281	-0,005
	Beklenen oran	0,1182		0,5168		0,2316		0,1334	
4. puanlayıcı		1. kategori		2. kategori		3. kategori		4. kategori	
1	Gözlenen oran	0,2218	-0,0103	0,3365	0,0046	0,2600	-0,0020	0,1816	0,0076
	Beklenen oran	0,2321		0,3319		0,2620		0,1740	
2	Gözlenen oran	0,1702	-0,0078	0,3289	0,0025	0,2658	-0,0024	0,2352	0,0078
	Beklenen oran	0,1780		0,3264		0,2682		0,2274	
3	Gözlenen oran	0,1434	-0,0097	0,2658	0,0062	0,2734	-0,0013	0,3174	0,0048
	Beklenen oran	0,1531		0,2596		0,2747		0,3126	
4	Gözlenen oran	0,0765	-0,0064	0,2218	-0,0018	0,3098	0,0037	0,3920	0,0047
	Beklenen oran	0,0829		0,2236		0,3061		0,3873	
5	Gözlenen oran	0,1396	-0,0099	0,2677	0,0070	0,2849	-0,0009	0,3078	0,0037
	Beklenen oran	0,1495		0,2607		0,2858		0,3041	
6	Gözlenen oran	0,3346	-0,0020	0,2753	-0,0089	0,2428	0,0026	0,1472	0,0083
	Beklenen oran	0,3366		0,2842		0,2402		0,1389	
7	Gözlenen oran	0,3212	-0,0038	0,2772	0,0004	0,2218	-0,0052	0,1797	0,0085

	Beklenen oran	0,3250		0,2768		0,2270		0,1712	
8	Gözlenen oran	0,3461	0,0035	0,2906	-0,0070	0,1625	-0,0082	0,2008	0,0117
	Beklenen oran	0,3426		0,2976		0,1707		0,1891	
9	Gözlenen oran	0,3346	0,0023	0,2543	0,0008	0,2046	-0,0144	0,2065	0,0113
	Beklenen oran	0,3323		0,2535		0,2190		0,1952	
10	Gözlenen oran	0,2887	-0,0033	0,3193	0,0049	0,1912	-0,0115	0,2008	0,0099
	Beklenen oran	0,2920		0,3144		0,2027		0,1909	
11	Gözlenen oran	0,1243	-0,0019	0,5449	0,0170	0,2046	-0,0183	0,1262	0,0032
	Beklenen oran	0,1262		0,5279		0,2229		0,1230	
5. puanlayıcı		1. kategori		2. kategori		3. kategori		4. kategori	
1	Gözlenen oran	0,2141	-0,0165	0,3461	0,0004	0,2543	0,0045	0,1855	0,0116
	Beklenen oran	0,2306		0,3457		0,2498		0,1739	
2	Gözlenen oran	0,1836	-0,0124	0,3040	-0,0013	0,2983	0,0016	0,2141	0,0121
	Beklenen oran	0,1960		0,3053		0,2967		0,2020	
3	Gözlenen oran	0,1491	-0,0140	0,2505	-0,0031	0,2639	0,0035	0,3365	0,0135
	Beklenen oran	0,1631		0,2536		0,2604		0,3230	
4	Gözlenen oran	0,0707	-0,0082	0,2294	-0,0064	0,3098	0,0029	0,3901	0,0117
	Beklenen oran	0,0789		0,2358		0,3069		0,3784	
5	Gözlenen oran	0,1415	-0,0129	0,2677	0,0010	0,2715	-0,0006	0,3193	0,0125
	Beklenen oran	0,1544		0,2667		0,2721		0,3068	
6	Gözlenen oran	0,3289	-0,0152	0,2906	-0,0071	0,2333	0,0117	0,1472	0,0106
	Beklenen oran	0,3441		0,2977		0,2216		0,1366	
7	Gözlenen oran	0,3117	-0,0126	0,2868	-0,0026	0,2409	0,0029	0,1606	0,0123
	Beklenen oran	0,3243		0,2894		0,2380		0,1483	
8	Gözlenen oran	0,3403	-0,0089	0,2753	-0,0083	0,1893	0,0012	0,1950	0,016
	Beklenen oran	0,3492		0,2836		0,1881		0,1790	
9	Gözlenen oran	0,3365	-0,0085	0,2562	-0,0079	0,2027	-0,0002	0,2046	0,0165
	Beklenen oran	0,3450		0,2641		0,2029		0,1881	
10	Gözlenen oran	0,2945	-0,0129	0,3117	-0,0023	0,1816	-0,0005	0,2122	0,0158
	Beklenen oran	0,3074		0,3140		0,1821		0,1964	
11	Gözlenen oran	0,1262	-0,0037	0,5507	0,0102	0,1797	-0,0120	0,1434	0,0056
	Beklenen oran	0,1299		0,5405		0,1917		0,1378	

EK 8. Öğrencilerin ilk beş puanlayıcıya göre kestirilen yazılı anlatım becerisi düzeyleri (Theta Yetenek Düzeyleri)

Öğrenci No	1. puanlayıcı	2. puanlayıcı	3. puanlayıcı	4. puanlayıcı	5. puanlayıcı
001	-0,628	-0,716	-0,933	-0,848	-0,716
002	-0,564	-0,800	-0,933	-0,677	-0,677
003	-1,417	-1,417	-1,240	-1,240	-1,417
004	-0,570	-0,667	-0,570	-0,667	-0,711
005	-1,614	-1,417	-1,614	-1,417	-1,614
006	-1,533	-1,533	-1,533	-1,533	-1,533
007	-1,417	-1,240	-1,240	-1,240	-1,079
008	-0,628	-0,628	-0,760	-0,905	-0,760
009	-1,417	-1,614	-1,417	-1,614	-1,417
010	-1,245	-1,245	-1,245	-1,245	-1,245
011	-0,787	-1,155	-1,155	-0,787	-0,787
012	-0,994	-0,994	-1,155	-0,848	-0,848
013	-0,564	-0,594	-0,508	-0,380	-0,564
014	-0,380	-0,399	-0,399	-0,380	-0,483
015	-0,848	-1,100	-1,100	-0,994	-0,848
016	-1,417	-1,417	-1,614	-1,417	-1,417
017	-1,240	-1,417	-1,614	-1,240	-1,417
018	-0,214	-0,483	-0,419	-0,312	-0,214
019	-1,240	-1,240	-1,417	-1,079	-1,417
020	-1,836	-1,836	-1,836	-1,836	-1,836
021	0,149	0,141	0,235	0,248	0,248
022	-0,129	-0,312	-0,399	-0,129	-0,312
023	-1,240	-1,240	-1,079	-1,240	-1,240
024	-0,513	-0,593	-0,423	-0,423	-0,513
025	-0,716	-0,716	-0,716	-0,716	-0,716
026	0,141	0,141	0,141	0,141	0,141
027	-0,491	-0,593	-0,593	-0,491	-0,491
028	-0,462	-0,593	-0,462	-0,342	-0,462
029	-0,399	-0,483	-0,677	-0,716	-0,399
030	0,052	-0,047	0,141	-0,052	0,049
031	-0,810	-0,711	-0,810	-0,810	-0,711
032	-0,628	-0,628	-0,628	-0,628	-0,628
033	-0,677	-0,933	-0,800	-0,564	-0,800
034	-0,380	-0,483	-0,483	-0,380	-0,380
035	-1,836	-1,614	-2,090	-1,836	-1,836
036	0,141	-0,055	-0,155	0,045	0,141
037	-1,087	-1,087	-1,087	-1,087	-1,087
038	-1,614	-1,836	-1,836	-1,836	-1,614
039	-1,240	-1,417	-1,240	-1,836	-1,240
040	-1,079	-1,155	-0,994	-1,240	-1,079

041	-0,460	-0,564	-0,800	-0,677	-0,800
042	-0,129	-0,043	0,045	-0,043	0,045
043	-0,141	-0,342	-0,129	-0,436	-0,155
044	-0,342	-0,760	-0,342	-0,379	-0,462
045	-0,460	-0,677	-0,800	-0,677	-0,564
046	-0,483	-0,564	-0,483	-0,564	-0,564
047	-0,800	-1,240	-1,240	-1,240	-0,800
048	-0,380	-0,677	-0,716	-0,594	-0,564
049	-0,594	-0,800	-0,399	-0,399	-0,677
050	-0,483	-0,848	-0,508	-0,380	-0,800
051	-0,380	-0,483	-0,399	-0,380	-0,380
052	-0,593	-0,558	-0,628	-0,593	-0,593
053	-0,141	-0,234	-0,047	-0,342	-0,141
054	0,246	-0,052	0,246	-0,148	0,328
055	0,667	0,667	-0,043	0,760	0,537
056	-0,141	-0,141	0,328	-0,235	0,047
057	0,173	-0,156	-0,164	0,173	0,052
058	-0,129	0,537	0,045	0,935	0,905
059	0,043	0,248	-0,043	0,346	0,443
060	0,045	0,234	0,045	0,045	0,234
061	1,079	0,760	0,508	1,417	0,848
062	2,090	2,090	2,090	2,090	2,385
063	0,049	0,312	0,049	0,312	0,328
064	0,164	0,155	-0,047	0,155	-0,049
065	-0,214	-0,129	-0,129	-0,328	-0,135
066	-1,065	-0,491	-0,628	-0,593	-0,628
067	-0,045	-0,045	0,045	0,312	0,224
068	-0,043	-0,047	0,129	-0,399	0,312
069	0,052	0,692	0,362	0,049	0,260
070	-0,129	0,043	0,129	-0,135	0,129
071	0,848	0,848	1,079	0,994	0,760
072	-0,290	-0,651	-0,047	-0,558	-0,342
073	1,533	1,614	1,716	0,935	1,759
074	0,224	0,045	0,346	0,135	0,224
075	-0,273	0,052	0,362	0,148	0,148
076	0,276	0,443	0,273	0,443	0,537
077	1,533	1,417	0,933	1,240	1,417
078	1,334	0,994	0,224	0,224	1,065
079	-0,312	-0,135	0,399	0,214	-0,141
080	-1,417	-1,486	-0,994	-1,240	-0,787
081	-0,760	-1,188	-0,994	-1,240	-0,905
082	-0,047	-0,055	-0,049	0,047	-0,156
083	-0,994	-1,240	-0,711	-1,240	-0,905
084	-0,462	-0,134	0,043	-0,043	0,135
085	-1,334	-0,994	-0,570	-0,933	-1,334

086	-1,417	-1,417	-1,245	-1,836	-1,245
087	-0,829	-0,403	-0,431	-0,403	-0,185
088	-1,048	-0,795	-0,795	-0,869	-0,443
089	-0,384	-0,149	0,058	-0,185	0,273
090	-1,533	-1,245	-1,065	-0,570	-0,224
091	-1,759	-1,065	-1,065	-1,065	-1,334
092	-1,240	-1,614	-0,905	-0,810	-0,810
093	-0,869	-0,795	-0,524	-0,066	-0,185
094	-0,994	-0,760	-0,760	-0,667	-0,848
095	-1,417	-1,240	-1,240	-1,155	-1,240
096	-0,933	-0,994	-0,994	-1,079	-0,933
097	-1,417	-0,848	-1,155	-1,155	-1,417
098	-1,836	-1,836	-1,836	-1,836	-1,836
099	-0,462	-0,491	-0,491	-0,593	-0,462
100	-0,760	-0,443	-0,419	-0,628	-0,419
101	-0,260	-0,049	-0,047	-0,049	-0,148
102	-1,417	-1,417	-0,905	-0,905	-1,417
103	-1,394	-1,337	-1,231	-1,231	-1,231
104	-2,090	-2,090	-2,090	-2,090	-2,090
105	1,836	2,385	2,090	2,385	2,090
106	-0,905	-0,905	-0,810	-0,740	-0,840
107	0,593	0,740	0,903	0,903	0,740
108	-1,417	-1,614	-1,417	-1,614	-1,614
109	-1,282	-1,282	-1,282	-1,282	-1,282
110	2,385	2,385	2,385	2,385	2,385
111	-0,052	-0,052	-0,052	0,052	-0,052
112	0,905	1,155	0,848	0,760	0,994
113	0,148	-0,049	0,155	0,052	0,052
114	-1,614	-1,614	-1,836	-1,836	-1,836
115	-0,935	-1,188	-1,294	-1,446	-1,446
116	-0,933	-1,533	-1,759	-1,533	-1,759
117	-1,079	-1,240	-1,240	-1,240	-1,240
118	-1,079	-1,079	-1,240	-1,240	-1,240
119	-1,417	-1,417	-1,836	-1,836	-1,836
120	-1,417	-1,417	-1,836	-2,090	-2,090
121	-1,417	-1,417	-2,090	-2,090	-2,090
122	-1,240	-1,417	-2,090	-2,090	-2,090
123	0,419	0,537	0,716	0,419	0,419
124	0,141	0,047	0,043	-0,043	0,043
125	-0,129	-0,135	-0,045	-0,135	-0,141
126	-0,994	-1,334	-1,533	-1,417	-1,614
127	-0,380	-0,380	0,043	-0,214	-0,214
128	-0,760	-0,905	-1,245	-1,245	-1,245
129	-0,419	-0,235	-0,235	-0,235	-0,328
130	0,399	0,224	0,141	0,141	0,141

131	0,129	0,312	0,312	0,399	0,399
132	0,399	0,312	0,419	0,399	0,399
133	-0,298	-0,214	-0,214	-0,214	-0,214
134	-0,933	-1,079	-1,240	-1,240	-1,240
135	-0,800	-0,933	-0,800	-0,800	-0,800
136	-1,079	-1,614	-1,836	-1,614	-1,614
137	-0,905	-0,760	-0,905	-0,760	-0,760
138	-0,810	-0,905	-0,905	-0,905	-0,905
139	-1,079	-1,079	-1,240	-1,240	-1,240
140	-1,614	-2,090	-2,090	-2,090	-2,090
141	-0,399	-0,399	-0,224	-0,129	-0,129
142	-0,380	-0,380	-0,214	-0,214	-0,129
143	-0,594	-0,994	-0,994	-0,994	-0,994
144	-0,594	-0,716	-0,716	-0,594	-0,848
145	0,994	0,848	0,905	1,065	1,065
146	-0,403	-0,740	-0,740	-0,634	-0,570
147	-0,076	-0,076	-0,076	-0,076	-0,076
148	-0,628	-0,628	-0,628	-0,760	-0,716
149	0,628	0,848	0,491	0,384	0,491
150	1,334	1,245	1,245	1,334	1,334
151	-0,045	-0,399	-0,224	0,045	-0,045
152	-0,994	-1,240	-1,155	-0,905	-0,994
153	-0,328	-0,328	-0,419	-0,419	-0,419
154	-0,994	-1,004	-1,004	-0,994	-1,065
155	-0,134	-0,508	-0,628	-0,234	-0,342
156	-0,994	-0,933	-0,848	-1,155	-1,079
157	-1,417	-1,417	-1,417	-1,417	-1,417
158	-0,760	-1,004	-1,100	-0,848	-0,760
159	-0,840	-1,334	-1,155	-0,840	-1,004
160	-0,045	0,149	-0,047	-0,045	0,049
161	-0,711	-0,662	-0,810	-0,156	-0,549
162	-1,836	-1,836	-1,836	-1,614	-1,836
163	-1,836	-1,417	-1,836	-1,155	-1,836
164	-1,836	-1,614	-1,417	-1,417	-1,417
165	-0,524	-0,310	-0,348	-0,491	-0,185
166	-1,155	-0,970	-1,004	-1,004	-1,155
167	-1,836	-1,836	-1,836	-2,090	-1,836
168	1,533	1,079	0,760	1,614	1,614
169	-0,628	-0,164	-0,848	-0,431	-0,045
170	-0,760	-0,933	-0,848	-0,848	-0,470
171	2,385	2,385	2,385	2,385	2,385
172	1,759	1,533	1,155	1,614	1,533
173	2,385	2,385	2,385	2,385	2,385
174	2,385	2,385	2,385	2,090	2,385
175	1,417	1,240	1,240	1,614	1,417

176	1,240	1,240	1,417	1,417	0,994
177	0,711	0,399	0,570	0,593	0,246
178	-1,836	-1,836	-1,836	-1,065	-1,836
179	-0,848	-1,240	-1,240	-1,079	-0,760
180	-1,245	-0,677	-0,628	-0,994	-0,508
181	-0,800	-0,848	-0,800	-0,135	-0,800
182	-1,240	-0,933	-1,417	-0,593	-1,417
183	-1,614	-1,759	-1,836	-1,155	-0,667
184	-0,711	-0,491	-0,379	-0,246	-0,570
185	-0,840	-0,903	-0,840	-0,810	-1,004
186	0,570	0,462	0,537	0,848	0,667
187	-0,508	-0,443	-0,594	-0,224	-0,594
188	0,443	0,141	0,419	0,141	0,443
189	-0,933	-0,933	-1,079	-1,079	-1,079
190	0,043	0,043	0,043	0,043	-0,045
191	0,443	0,570	0,570	0,443	0,570
192	-0,677	-0,933	-1,079	-0,800	-0,677
193	-0,905	-1,155	-0,994	-0,905	-1,155
194	-0,810	-0,711	-0,711	-0,810	-0,810
195	-1,334	-1,334	-1,245	-1,245	-1,245
196	-1,240	-1,417	-1,417	-1,240	-1,417
197	-0,740	-0,524	-0,634	-0,740	-0,740
198	-0,312	-0,399	-0,399	-0,399	-0,312
199	-1,240	-1,079	-1,079	-1,240	-1,240
200	-0,810	-0,810	-0,905	-0,760	-0,667
201	-1,240	-1,417	-1,417	-1,417	-1,240
202	-0,905	-0,905	-0,711	-0,607	-0,607
203	-0,436	-0,436	-0,436	-0,436	-0,436
204	-0,800	-0,800	-0,800	-0,800	-0,800
205	-0,290	-0,290	-0,185	-0,185	-0,185
206	-0,129	-0,043	-0,214	-0,129	-0,214
207	-0,628	-0,628	-0,667	-0,760	-0,760
208	1,155	1,334	1,155	1,155	1,155
209	0,848	1,155	0,905	0,905	0,905
210	-0,076	-0,210	-0,210	-0,328	-0,210
211	-0,173	-0,058	-0,058	-0,185	-0,306
212	1,533	1,533	1,334	1,334	1,334
213	1,533	1,533	1,533	1,533	1,533
214	1,334	1,533	1,533	1,334	1,334
215	0,905	1,065	0,905	0,905	0,905
216	-0,273	-0,403	-0,513	-0,513	-0,403
217	-0,905	-0,810	-0,994	-0,905	-0,810
218	-1,417	-1,065	-0,905	-1,155	-1,155
219	-1,240	-1,240	-1,836	-1,836	-1,836
220	-0,848	-1,155	-1,334	-1,614	-2,090

221	-0,246	-0,403	-0,403	-0,549	-0,155
222	-0,537	-1,065	-1,352	-1,245	-1,245
223	0,810	0,760	0,462	0,692	0,462
224	1,417	1,240	1,079	1,079	1,417
225	0,328	0,470	0,443	0,149	0,367
226	-0,933	-1,836	-1,240	-1,417	-1,240
227	1,334	0,760	1,065	1,334	1,155
228	0,052	0,155	0,379	0,052	0,049
229	0,346	0,419	0,235	0,224	0,045
230	-0,419	-0,537	-0,594	-0,628	-0,628
231	-0,810	-0,970	-1,065	-0,994	-1,079
232	-0,933	-1,155	-0,994	-0,970	-0,970
233	0,129	0,312	0,443	0,248	0,443
234	-0,634	-0,634	-0,740	-0,740	-0,740
235	-1,240	-1,417	-1,417	-1,240	-1,614
236	-1,836	-1,836	-1,836	-1,836	-1,417
237	-1,079	-0,994	-0,933	-0,933	-0,933
238	-0,848	-0,537	-0,692	-0,787	-0,905
239	-1,079	-1,417	-1,155	-1,155	-1,334
240	-1,240	-1,417	-1,417	-1,334	-1,334
241	-1,079	-0,848	-0,994	-0,594	-1,240
242	-0,800	-0,800	-0,933	-0,933	-0,933
243	-0,134	-0,312	-0,312	-0,214	-0,312
244	0,045	0,234	0,214	0,135	0,135
245	-0,214	-0,043	-0,224	-0,399	-0,312
246	-0,933	-1,155	-1,417	-1,417	-1,417
247	-1,417	-1,240	-1,417	-1,614	-1,836
248	-0,810	-0,994	-0,810	-0,905	-0,994
249	0,346	0,346	0,346	0,443	0,537
250	-0,196	-0,290	-0,455	-0,549	-0,348
251	0,148	0,414	0,047	0,148	0,483
252	0,328	0,312	0,235	0,235	0,399
253	0,740	0,651	0,328	0,740	0,692
254	1,417	1,065	0,933	1,417	1,334
255	-0,662	-0,431	-0,634	-0,549	-0,310
256	-0,623	-0,795	-0,491	-0,682	-0,592
257	0,076	-0,070	-0,070	-0,210	-0,197
258	-0,714	-0,592	-0,829	-0,829	-0,592
259	0,148	0,047	-0,148	-0,052	0,047
260	-0,714	-0,970	-0,714	-1,065	-1,155
261	-1,240	-1,614	-1,417	-1,614	-2,090
262	-1,065	-1,004	-0,970	-1,282	-0,970
263	0,043	-0,224	-0,135	-0,043	-0,045
264	-0,260	-0,436	-0,810	-0,491	-0,634
265	-0,848	-1,334	-0,994	-1,155	-0,716

266	-0,328	-0,537	-0,235	-0,470	-0,470
267	-0,549	-0,524	-0,662	-0,740	-0,740
268	0,328	0,628	0,570	0,491	0,570
269	-0,052	0,276	0,062	0,174	0,055
270	0,570	0,667	0,135	0,328	0,537
271	0,141	0,045	0,045	0,141	0,045
272	0,716	0,716	0,483	0,483	0,594
273	-2,090	-2,090	-2,090	-2,090	-2,090
274	-0,328	-0,328	-0,328	-0,328	-0,328
275	-0,298	-0,224	-0,135	-0,129	-0,129
276	0,905	0,994	0,628	1,155	0,716
277	-0,462	-0,462	-0,462	-0,537	-0,537
278	-1,240	-1,079	-1,079	-1,240	-1,240
279	-0,905	-0,905	-0,994	-0,905	-0,994
280	-1,417	-1,614	-1,614	-1,240	-1,417
281	-0,994	-0,994	-0,994	-0,994	-0,994
282	-1,240	-1,240	-1,417	-1,417	-1,417
283	-1,533	-1,533	-1,614	-1,533	-1,533
284	-1,836	-1,836	-1,614	-1,614	-1,836
285	-0,384	-0,384	-0,384	-0,384	-0,384
286	-1,614	-1,614	-1,614	-1,417	-1,417
287	1,836	1,836	2,385	2,090	1,836
288	2,090	2,385	2,385	2,385	2,385
289	1,486	1,486	1,282	1,282	1,486
290	0,055	0,047	-0,224	0,328	0,055
291	-0,760	-0,399	-0,564	0,141	-0,628
292	0,570	1,417	1,836	1,240	0,905
293	0,287	0,403	0,290	0,935	0,173
294	-0,399	-0,214	-0,043	-0,298	-0,214
295	0,905	0,716	0,716	0,994	0,716
296	0,994	0,994	0,628	0,905	1,155
297	1,533	1,155	1,417	1,334	1,533
298	0,848	0,716	0,628	0,760	0,508
299	0,800	0,800	0,312	0,800	0,716
300	-0,129	0,235	-0,047	0,628	0,537
301	-0,483	-0,564	-0,135	-0,677	-0,594
302	-0,234	-0,141	-0,141	-0,380	-0,234
303	1,417	0,994	1,240	1,240	1,079
304	0,994	1,079	1,334	0,994	0,994
305	0,994	0,848	1,240	1,079	1,240
306	1,417	1,417	1,240	1,417	1,240
307	0,328	0,537	0,810	0,328	0,328
308	0,933	1,240	0,994	0,848	1,079
309	-0,677	-0,594	-0,716	-0,677	-0,677
310	0,419	0,419	0,537	0,399	0,594

311	0,848	0,594	0,594	0,716	0,848
312	-0,043	0,045	0,045	0,045	0,045
313	0,047	0,141	0,141	0,047	0,047
314	1,417	1,417	1,240	1,417	1,614
315	0,346	0,470	0,346	0,141	0,346
316	0,043	0,043	-0,043	0,043	0,214
317	0,045	-0,045	-0,129	-0,043	0,045
318	0,848	0,848	0,848	0,848	0,848
319	-2,385	-2,385	-2,385	-2,385	-2,385
320	0,049	-0,047	0,049	-0,047	-0,047
321	-0,594	-0,594	-0,594	-0,594	-0,594
322	-1,240	-1,417	-1,240	-1,417	-1,240
323	1,334	1,614	1,675	1,836	1,533
324	0,328	0,328	0,197	0,431	0,310
325	2,090	1,836	1,836	1,614	1,614
326	1,065	0,760	0,905	1,065	1,150
327	0,667	0,593	0,593	0,491	0,570
328	0,047	-0,049	-0,155	-0,362	-0,273
329	0,667	0,667	0,570	0,328	0,810
330	-0,214	-0,298	-0,129	-0,214	-0,214
331	-0,483	-0,312	-0,325	-0,462	-0,716
332	-0,716	-0,848	-1,079	-0,933	-0,933
333	-1,240	-1,079	-0,848	-1,240	-1,240
334	-0,537	-0,628	-0,840	-0,787	-0,419
335	-0,129	-0,223	-0,309	-0,223	-0,043
336	-0,129	-0,246	-0,148	-0,045	-0,342
337	-0,298	-0,380	-0,129	-0,129	-0,214
338	1,614	1,836	1,836	1,614	1,614
339	1,245	1,533	1,533	1,533	1,533
340	1,240	1,240	1,079	1,079	1,079
341	2,090	1,836	2,090	2,090	2,090
342	0,342	0,047	0,045	0,047	0,047
343	1,979	1,979	1,979	1,979	1,979
344	1,079	0,994	1,079	0,994	0,994
345	-0,994	-0,905	-0,905	-1,079	-1,079
346	1,533	1,836	1,533	1,533	1,533
347	0,508	0,419	0,399	0,399	0,224
348	2,385	2,385	2,385	2,385	2,385
349	1,836	2,090	1,836	1,614	2,090
350	-0,594	-0,800	-0,677	-0,848	-1,079
351	2,385	2,385	2,385	2,385	2,385
352	2,385	2,385	2,385	2,385	2,385
353	-1,155	-1,065	-1,065	-1,065	-1,065
354	0,348	0,348	0,348	0,348	0,520
355	2,090	2,090	2,385	2,385	2,385

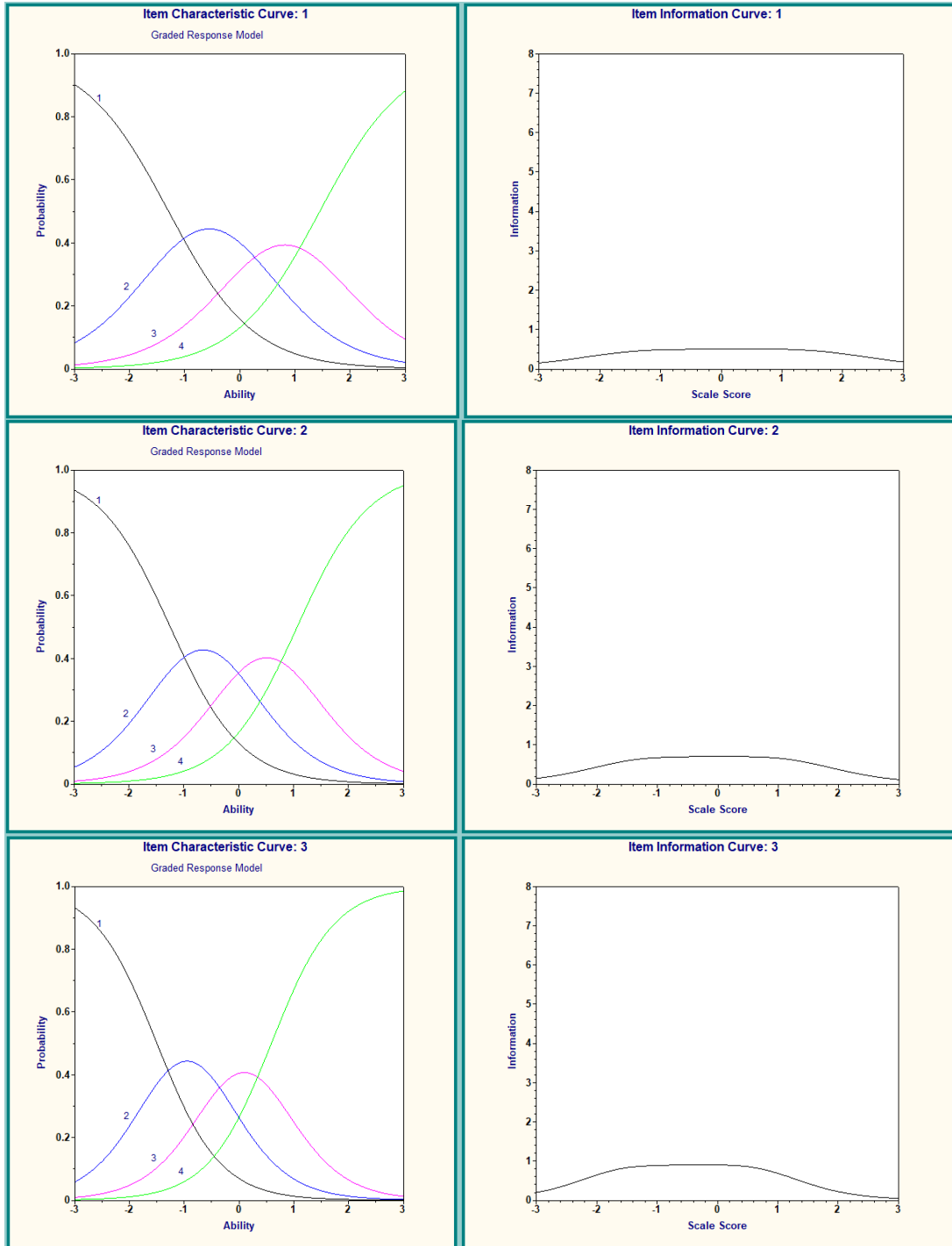
356	-0,070	-0,070	0,076	0,076	0,076
357	-0,410	-0,197	-0,197	-0,197	-0,197
358	0,905	0,994	0,994	1,079	0,994
359	-0,312	-0,312	-0,312	-0,312	-0,312
360	-0,399	-0,628	-0,399	-0,628	-0,399
361	0,527	0,651	0,527	0,787	0,787
362	-0,399	-0,483	-0,508	-0,594	-0,508
363	-1,614	-1,614	-1,614	-1,614	-1,614
364	-0,235	-0,235	-0,235	-0,235	-0,235
365	0,994	0,994	0,716	0,905	1,065
366	1,614	1,614	1,614	1,836	1,836
367	0,052	-0,156	-0,055	0,052	0,049
368	0,155	0,273	0,258	0,164	0,155
369	-0,933	-0,933	-1,079	-1,079	-1,079
370	0,258	0,155	0,129	0,537	0,537
371	0,740	0,662	0,634	0,667	0,810
372	-1,836	-1,417	-1,614	-1,836	-1,836
373	0,716	0,716	0,716	0,594	0,716
374	1,100	1,004	1,188	1,282	1,486
375	-0,651	-0,935	-0,651	-0,651	-0,651
376	-0,994	-0,994	-1,155	-0,994	-1,155
377	-0,380	-0,994	-0,994	-0,905	-0,905
378	0,043	-0,135	-0,141	-0,043	0,043
379	0,273	-0,298	-0,043	-0,164	0,570
380	-0,129	-0,460	-0,298	-0,460	-0,129
381	1,836	1,836	1,836	1,836	2,090
382	1,759	2,017	1,836	2,017	1,759
383	0,903	0,810	0,810	0,667	1,065
384	-0,576	-0,576	-0,455	-0,423	-0,287
385	-0,260	0,141	-0,462	-0,047	-0,148
386	0,508	0,537	0,628	0,537	0,628
387	-0,748	-0,714	-0,748	-0,623	-0,829
388	1,533	1,836	1,533	1,759	1,836
389	0,970	1,065	1,150	0,869	1,065
390	-0,055	0,384	0,362	0,156	0,276
391	-1,282	-1,486	-1,486	-1,282	-1,486
392	0,196	0,210	0,062	0,196	0,196
393	2,090	1,836	2,090	2,090	2,090
394	2,385	1,533	1,417	1,836	1,836
395	-0,770	-0,662	-0,549	-0,740	-0,549
396	0,244	-0,197	0,076	0,148	0,290
397	0,436	0,045	0,134	-0,045	0,223
398	1,240	1,417	1,836	1,240	1,240
399	0,933	0,716	0,716	1,240	0,716
400	-0,246	-0,399	-0,224	0,141	-0,135

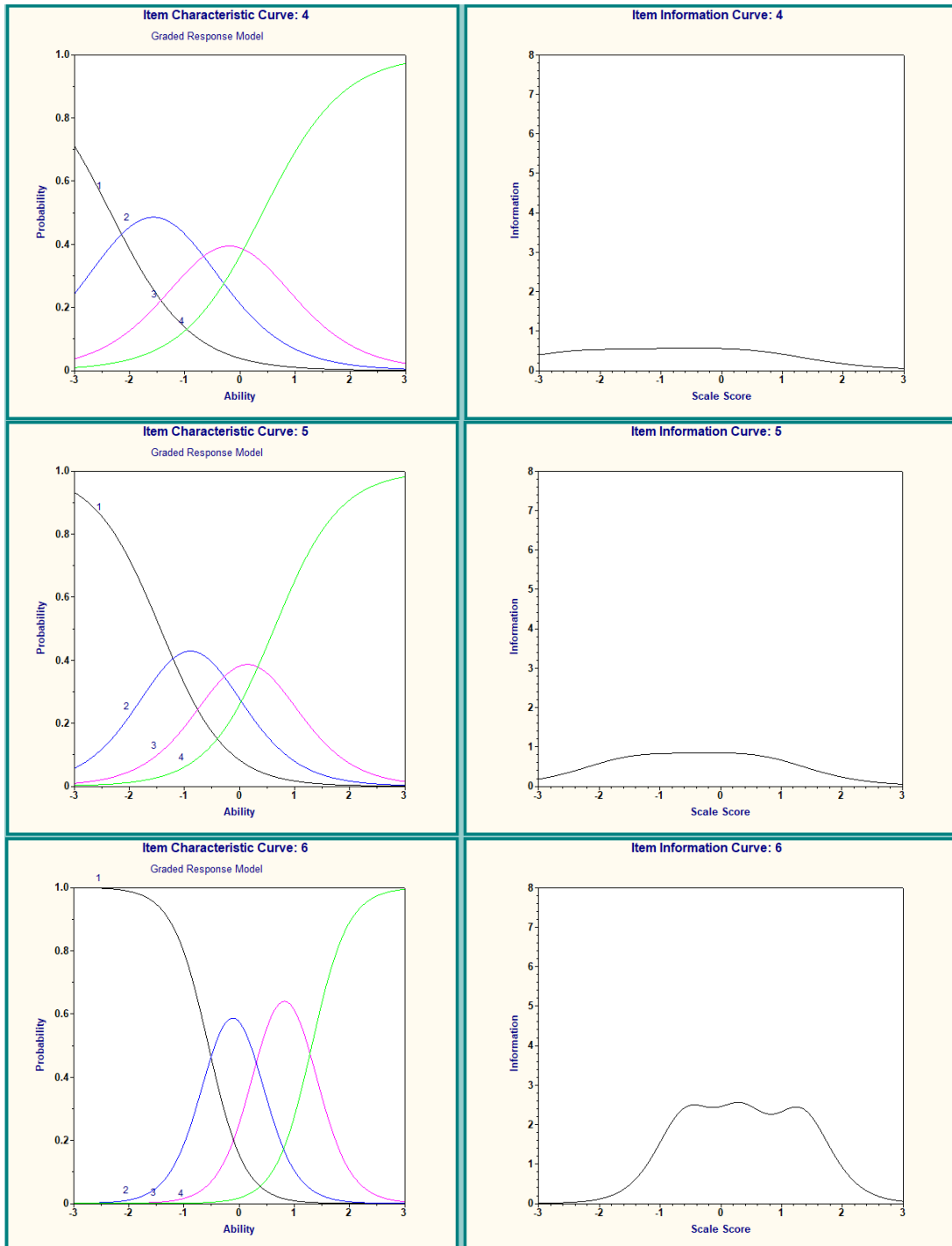
401	-0,537	-0,419	-0,508	-0,045	-0,419
402	-1,417	-1,065	-0,933	-0,933	-0,800
403	0,141	-0,045	0,047	0,234	0,045
404	-0,564	0,223	-0,045	-0,045	0,045
405	-0,933	-0,933	-0,933	-0,677	-0,564
406	-2,090	-2,090	-2,090	-1,417	-1,417
407	-0,677	-0,594	-0,594	-0,508	-0,508
408	-1,836	-1,417	-1,614	-0,628	-0,933
409	-0,558	-1,240	-0,628	-0,800	-0,677
410	-2,090	-2,090	-2,090	-2,090	-2,090
411	-1,079	-1,417	-1,240	-1,417	-1,240
412	-1,079	-1,240	-1,417	-0,933	-1,079
413	-0,716	-1,240	-0,994	-0,933	-0,933
414	-1,079	-1,079	-1,240	-0,047	-0,483
415	0,840	0,462	0,692	0,558	0,651
416	0,570	0,667	0,443	0,443	0,443
417	0,377	0,491	0,185	0,185	0,328
418	0,045	0,135	-0,214	-0,129	-0,129
419	0,436	0,246	0,436	0,558	0,558
420	-0,135	-0,045	-0,224	-0,224	-0,224
421	0,076	0,076	-0,276	-0,276	-0,276
422	-0,848	-0,848	-0,537	-0,537	-0,628
423	1,155	0,905	0,994	0,994	0,994
424	-0,066	-0,066	-0,328	-0,328	-0,328
425	-0,214	-0,214	-0,380	-0,380	-0,298
426	0,869	0,869	0,570	0,570	0,570
427	1,155	0,848	0,716	0,594	0,483
428	0,592	0,592	0,592	0,431	0,592
429	0,362	0,443	0,419	0,419	0,419
430	1,240	1,417	1,417	1,240	1,240
431	1,150	0,869	0,810	0,810	0,970
432	0,235	0,235	0,045	-0,043	0,043
433	0,848	0,760	0,628	0,399	0,628
434	1,417	1,614	1,079	0,933	0,933
435	0,628	0,628	0,711	0,711	0,711
436	0,527	0,527	0,935	0,994	0,848
437	0,905	0,905	0,848	0,800	0,628
438	-0,235	-0,328	0,760	0,848	0,848
439	-0,076	-0,076	-0,384	-0,491	-0,384
440	-0,043	-0,045	-0,129	0,043	-0,043
441	1,614	1,614	1,240	1,240	1,240
442	-0,905	-0,905	-0,667	-0,810	-0,760
443	0,460	0,460	0,564	0,564	0,564
444	1,614	1,614	1,240	1,417	1,614
445	1,334	0,994	0,905	1,065	0,994

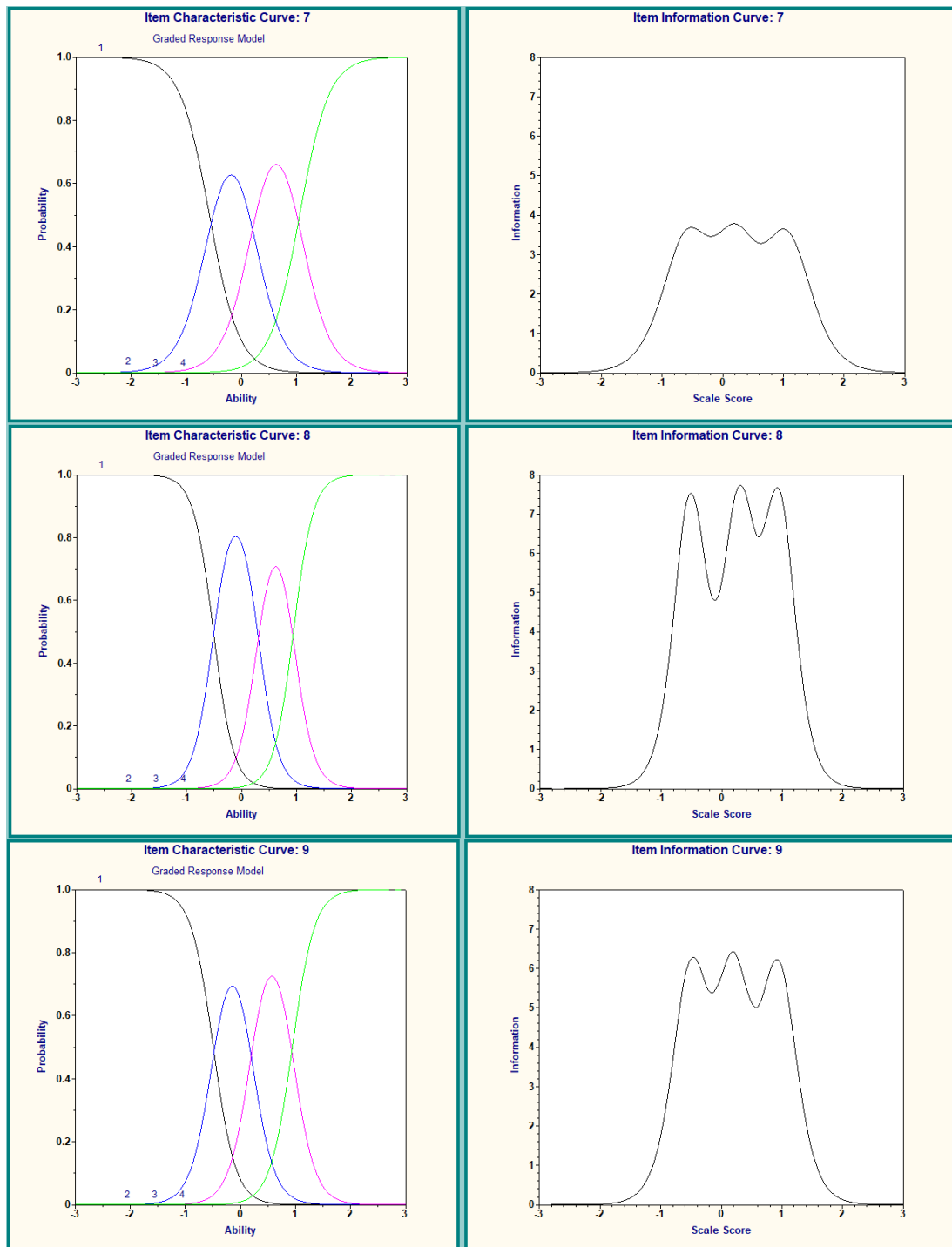
446	-1,079	-1,079	-0,760	-0,848	-0,760
447	-0,045	0,045	0,248	0,248	0,248
448	-1,065	-1,245	-0,760	-0,760	-0,760
449	-0,537	-0,760	-0,214	-0,399	-0,214
450	-0,310	-0,740	-0,384	-0,491	-0,570
451	1,614	1,417	1,614	1,836	1,836
452	1,334	1,155	0,810	1,065	1,245
453	1,240	1,240	0,933	1,240	1,079
454	1,614	1,417	1,417	1,836	1,836
455	1,155	0,994	1,155	1,334	1,334
456	0,129	0,214	0,043	0,043	0,129
457	0,970	0,869	0,869	0,970	0,970
458	-0,047	-0,134	-0,214	-0,224	0,045
459	0,848	0,810	0,667	0,848	0,848
460	1,150	0,970	1,065	1,065	0,970
461	0,135	0,537	0,328	0,328	0,235
462	-0,049	-0,129	-0,134	-0,141	-0,141
463	-0,047	-0,141	-0,045	-0,045	0,047
464	-0,224	-0,342	-0,328	-0,234	-0,312
465	1,334	0,994	1,155	1,334	0,848
466	0,994	0,994	1,065	1,065	0,905
467	0,055	-0,155	0,062	0,174	-0,052
468	0,049	0,049	-0,047	0,052	-0,047
469	2,385	2,385	2,385	2,385	2,385
470	2,385	2,385	2,385	2,385	2,385
471	2,385	2,385	2,385	2,385	2,385
472	1,486	1,486	1,486	1,486	1,486
473	1,079	1,240	1,240	1,240	1,240
474	1,533	1,533	1,533	1,533	1,533
475	1,836	1,836	2,090	2,090	1,836
476	0,483	0,594	0,677	0,677	0,677
477	0,443	0,443	0,537	0,537	0,537
478	1,417	1,417	1,417	1,240	1,240
479	2,385	2,385	2,385	2,385	2,385
480	0,716	0,594	0,677	0,677	0,677
481	1,836	1,836	1,836	2,090	2,090
482	0,328	0,508	0,508	0,419	0,328
483	0,419	0,848	0,716	0,848	0,994
484	2,385	2,385	2,385	2,385	2,385
485	2,090	1,240	1,079	1,417	1,079
486	0,045	0,224	0,043	0,043	-0,129
487	-0,760	-0,848	-0,419	-0,483	-0,312
488	1,614	1,079	1,614	1,240	1,417
489	0,312	0,224	-0,043	0,129	0,224
490	1,979	1,282	1,282	1,486	1,282

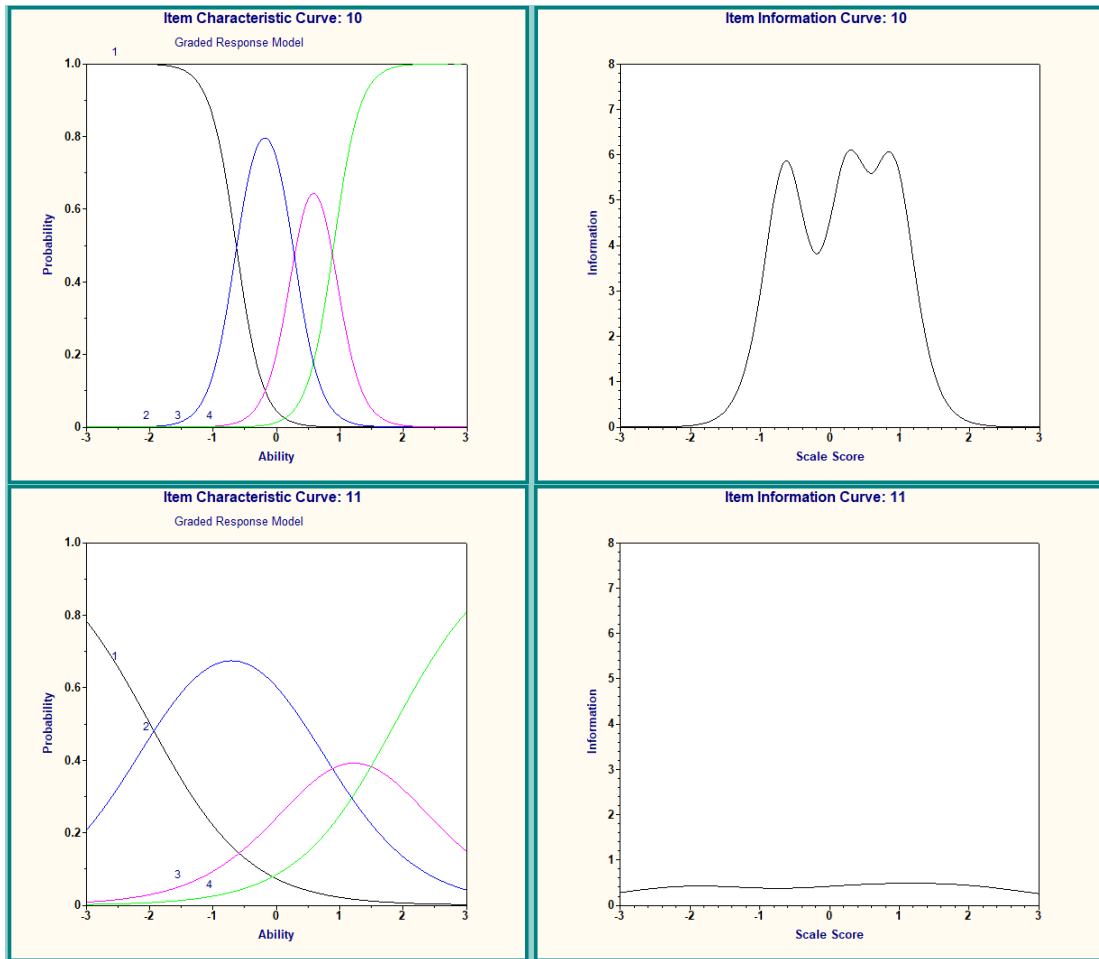
491	1,150	0,848	0,667	0,508	0,537
492	1,614	1,240	1,417	1,836	1,614
493	2,090	2,090	1,836	1,836	2,090
494	-0,129	-0,594	-0,214	-0,129	-0,129
495	2,385	1,836	1,836	1,836	1,836
496	2,385	1,836	2,090	2,385	2,090
497	1,240	1,417	1,240	1,079	1,417
498	-0,052	-0,260	-0,593	-0,164	-0,513
499	0,760	0,537	0,628	0,594	0,508
500	2,385	1,614	1,417	1,614	1,417
501	1,417	1,836	1,759	1,614	1,155
502	0,770	0,840	0,618	1,100	0,770
503	0,935	1,100	0,840	0,651	0,692
504	0,141	0,328	0,235	0,129	0,235
505	0,905	0,905	1,065	0,848	0,994
506	1,614	1,614	1,417	1,836	1,836
507	2,385	2,385	2,385	2,385	2,385
508	0,508	0,312	0,508	0,312	0,399
509	-1,417	-1,417	-1,417	-1,417	-1,417
510	0,848	0,848	0,994	0,994	0,994
511	-0,224	-0,224	-0,224	-0,224	-0,224
512	-0,513	-0,513	-0,513	-0,662	-0,662
513	2,090	2,385	1,836	2,385	1,836
514	1,759	1,759	2,017	1,759	1,759
515	2,090	2,090	2,090	2,090	2,090
516	0,197	0,310	0,431	0,431	0,431
517	1,245	1,245	1,150	1,150	1,245
518	1,614	1,614	1,614	1,614	1,417
519	2,385	1,836	1,240	1,614	2,090
520	0,570	0,443	0,443	0,443	0,328
521	2,385	2,385	2,385	2,385	2,385
522	2,385	2,385	2,385	2,385	2,385
523	0,234	0,234	0,527	0,134	0,134

EK 9. Maddelere ait karakteristik eğrileri ve bilgi fonksiyonları örneği-1. puanlayıcı











GAZİLİ OLMAK AYRICALIKTIR..