



**KLASİK TEST KURAMINA GÖRE SAYISAL VE SÖZEL ALANLAR İÇİN
PUANLAMA GÜVENİRLİĞİNİN KAYIP VERİ KAPSAMINDA
İNCELENMESİ**

SİBEL ADA

**YÜKSEK LİSANS TEZİ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

KASIM, 2015

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek şartıyla tezin teslim tarihinden itibaren tezden fotokopi çekilebilir.

YAZARIN:

Adı : Sibel
Soyadı : ADA
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi : 19/11/2015

TEZİN:

Türkçe Adı : Klasik test kuramına göre sayısal ve sözel alanlar için puanlama güvenirliğinin kayıp veri durumu kapsamında incelenmesi

İngilizce Adı : Examining interater reliability considering missing value for numerical and verbal courses according to classical test theory

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazım 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.

Yazarın Adı Soyadı: Sibel ADA

İmza:

Jüri Onay Sayfası

Sibel ADA tarafından hazırlanan “Klasik test kuramına göre sayısal ve sözel alanlar için puanlama güvenirliğinin kayıp veri durumu kapsamında incelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. İsmail KARAKAYA

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi

Başkan: Prof. Dr. Selahattin GELBAL

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi

Üye: Doç. Dr. Şeref TAN

Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi

Tez Savunma Tarihi: 19/11/2015

Bu tezin Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Eğitim Bilimleri Enstitüsü Müdürü

Prof. Dr. Tahir ATICI

KLASİK TEST KURAMINA GÖRE SAYISAL VE SÖZEL ALANLAR İÇİN PUANLAMA GÜVENİRLİĞİNİN KAYIP VERİ KAPSAMINDA İNCELENMESİ

(Yüksek Lisans Tezi)

SİBEL ADA

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

KASIM, 2015

ÖZ

Bu çalışma puanlama güvenirliğini belirlemek için kullanılan uzlaşma katsayılarını karşılaştırmayı ve veri setindeki kayıp veri oranına göre bu katsayıları incelemeyi amaçlamaktadır. Aynı zamanda sayısal ve sözel ders için hesaplanan uzlaşma katsayılarının bir kıyaslaması yapılmıştır. Çalışma grubunu 71'i sekizinci sınıf ve 159'u dokuzuncu sınıf öğrencisi olmak üzere toplam 230 öğrenci oluşturmaktadır. Veriler Uluslararası Öğrenci Değerlendirme Sınavında kullanılan soruların açıklananları arasından seçilen beş matematik ile altı okuduğunu anlama sorusundan oluşturulan matematik okuryazarlığı ve okuduğunu anlama uygulama formları ile toplanmıştır. Toplanan öğrenci yanıtları üçü matematik ve üçü okuduğunu anlama için olmak üzere altı farklı puanlayıcı tarafından puanlanmıştır. Puanlama işlemleri esnasında araştırmacı tarafından hazırlanan dereceli puanlama anahtarı kullanılmıştır. Çalışma kapsamında elde edilen veriler üzerinden uzlaşma katsayıları ve sınıf içi korelasyon değerleri iki puanlayıcı ve üç puanlayıcı için hesaplanmıştır. Uzlaşma katsayıları dereceli puanlama anahtarlarındaki her bir ölçüt için hesaplanmıştır ve analiz edilen veriler sıralama düzeyinde olduğu için ağırlıklandırılmış katsayılar kullanılarak yapılmıştır. İki puanlayıcı için hesaplanan uzlaşma katsayıları Cohen'in kappa, Gwet'in AC₂, Scott'nin pi, Krippendorff'un alfa, Brennan-Prediger, uzlaşma yüzdesi; üç puanlayıcı için hesaplanan uzlaşma katsayıları Fleiss'nin kappa, Gwet'in AC₂, Krippendorff'un alfa,

Conger'in kappa, Brennan-Prediger ve uzlaşma yüzdesi şeklindedir. Sınıf içi korelasyon katsayısı hesaplanırken toplam puan üzerinden analiz yapılmıştır. Sınıf içi korelasyon katsayısı için belirlenen uygun modele göre iki yönlü karma ANOVA kullanılmıştır. Ayrıca kayıp veri oranının yüzde beş, on ve on beş olduğu durumlar için tüm analizler tekrar yapılmıştır. Elde edilen güvenilirlik katsayılarına ait değerlere, uzlaşma katsayıları için standart hata değerleri ve sınıf içi korelasyon katsayısı için hata varyansları dikkate alınarak karşılaştırmalar yapılmıştır. Uzlaşma katsayıları arasındaki ilişkilerin iki ve üç puanlayıcı durum için kayıp veri olmadığı ve olduğu durumlarda benzer olduğu görülmüştür. Uzlaşma yüzdesinin her zaman en az hataya sahip olduğu ve en yüksek güvenilirlik değerini gösterdiği; Gwet'in AC₂'i uzlaşma katsayısının ise uzlaşma yüzdesinden sonra genel olarak en az hataya ve en yüksek değere sahip olduğu belirlenmiştir. Brennan-Prediger uzlaşma katsayısı değerlerinin genellikle Gwet'in AC₂'ye yakın değerler verdiği tespit edilmiştir. Cohen'in/ Conger'in kappa, Scott'nin pi/Fleiss'nin kappa ve Krippendorff'un alfa değerlerinin de genel olarak birbirine yakın olduğu, diğer katsayılara göre nispeten daha büyük standart hata değerine sahip olduğu ve daha düşük güvenilirlik değerleri gösterdiği saptanmıştır. Veri setindeki kayıp veri oranı açısından incelendiğinde ise hem uzlaşma katsayılarının hem de sınıf içi korelasyon değerlerinin iki yüz elli kişilik veri grubu için kayıp veri oranı yaklaşık olarak yüzde on beşe kadar olduğu durumda kayıp veri olmadığı durumla benzer sonuçlar vermektedir. Genel olarak üç puanlayıcının olduğu durumdaki puanlama güvenilirliği için hesaplanan güvenilirlik katsayılarının iki puanlayıcı için hesaplanana göre görece olarak daha yüksek olduğu görülmüştür. Sayısal ve sözel ders bakımından incelendiğinde ise matematik okuryazarlığı uygulamasından elde edilen güvenilirlik değerlerinin okuduğunu anlama uygulamasına göre daha yüksek olduğu belirlenmiştir. Çalışmanın sonuçlarına göre puanlama güvenirliliğinin belirlenmek istendiği çalışmalarda en az iki puanlama güvenirliliği katsayısının raporlanması, sözel derslerde çalışma yapanların sorular hakkında farklı yorumlamaları önlemek için iyi tasarlanmış dereceli puanlama anahtarları kullanmaları ve puanlayıcı eğitimine başvurmaları önerilmektedir. Araştırmacılara ise en uygun uzlaşma katsayısı değerleri için ne kadar puanlayıcının bulunması gerektiği ve çalışmada kullanılandan daha küçük ya da daha büyük örneklem için kayıp verinin nasıl çalışacağına yönelik çalışmalar yapması önerilmektedir.

Bilim Kodu : 10100125

Anahtar Kelimeler : Puanlama Güvenirliliği, Puanlayıcılar Arası Güvenirlilik Katsayısı, Sınıf İçi Korelasyon Katsayısı, Kayıp Değer

Sayfa Adedi : xvii + 124

Danışman : Doç. Dr. İsmail KARAKAYA

**EXAMINING INTERRATER RELIABILITY CONSIDERING
MISSING VALUE FOR NUMERICAL AND VERBAL COURSES
ACCORDING TO CLASSICAL TEST THEORY**

(M.S. Thesis)

SİBEL ADA

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

NOVEMBER, 2015

ABSTRACT

The purpose of this research was to compare agreement coefficient that was used to determine the interrater reliability and to examine this coefficient according to missing value ratio in data set. Besides, the agreement coefficients calculated for numerical and verbal courses were made comparison. Study group involved 230 students (71 in 8th grade and 159 in 9th grade). Data were collected through mathematics literacy and reading comprehension application forms which involved five mathematics and six reading comprehension questions in Programme for International Student Assessment. Students' responses were rated by three mathematics experts and three reading comprehension experts. Rubric prepared by the researcher was used during the scoring process. Agreement coefficients and intra-class correlation coefficient were calculated using data obtained from the research for two and three raters. Agreement coefficients were calculated for each criterion of the rubrics and weighted agreement coefficients were used for analysis because data were ordinal. The agreement coefficients calculated for two raters were as follows: Cohen's Kappa, Gwet's AC₂, Scott's pi, Krippendorff alpha, Brennan-Prediger and Agreement Percentage. The agreement coefficients calculated for three raters were as follows: Fleiss' Kappa, Gwet's AC₂, Conger's kappa, Krippendorff alpha, Brennan-Prediger and Agreement Percentage.

When intra-class correlation coefficient was calculated, total scores were used. Two-way mixed ANOVA model was used according to the appropriate model determined for intra-class correlation coefficient. Furthermore, for missing value ratio consisting of five, ten and fifteen percent, each analysis was conducted again. The obtained values of reliability coefficients were compared considering standard error values for agreement coefficient and error variance for intra-class correlation. Relationship between agreement coefficient for two and three raters with or without missing value was seen same. Agreement Percentage always had minimum standard error and maximum reliability values. It was determined that Gwet's AC₂ usually had minimum standard error and maximum reliability values following Agreement percentage coefficient. In general, Brennan-Prediger and Gwet's AC₂ agreement coefficient had almost similar values. It was also showed that values of Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa and Krippendorff alpha had almost same values and these agreement coefficients had higher standard error and less reliability value than others. When examined in terms of missing value, it was concluded that agreement coefficient and intra-class correlation reliability had same results with both no missing value and missing value ratio around fifteen percent. Reliability coefficient calculated for rater reliability for three raters relatively had higher reliability value than two raters'. When examined in terms of numerical and verbal courses, reliability value calculated for application of mathematics had higher reliability value than application of reading comprehension. According to the results of the research, it is recommended that at least two agreement coefficients could be reported for determined rater reliability; researchers who studied about verbal courses could use rubric designed good to avoid different interpretation and could educate for raters. For the researchers, it is recommended to study, for optimal agreement coefficient values, how many raters are needed, how missing value for large and small sample affects the results of the analysis.

Science Code : 10100125

Key Words : Interrater Reliability, Inter-rater Reliability Coefficient, Intra-class Reliability Coefficient, Missing Value

Page Number : xvii + 124

Supervisor : Doç. Dr. İsmail KARAKAYA

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
Jüri Onay Sayfası	iii
ÖZ.....	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ	xv
SİMGELER VE KISALTMALAR LİSTESİ.....	xvii
BÖLÜM 1	1
GİRİŞ	1
Problem Durumu.....	1
Araştırmanın Amacı	8
Araştırmanın Önemi	9
Sınırlılıklar	11
BÖLÜM 2	13
KAVRAMSAL ÇERÇEVE.....	13
Puanlayıcı Güvenirliği	13
Klasik Test Kuramına Göre Puanlayıcı Güvenirliği.....	14

Genellenebilirlik Kuramına Göre Puanlayıcı Güvenirliği.....	16
Rasch Ölçme Modeline Göre Puanlayıcı Güvenirliği	17
Çalışma Kapsamında Kullanılan Güvenirlik Katsayıları	18
Ağırlıklandırmalar	18
Tablolar	18
Cohen'in Kappa	21
Gwet'in AC_1/AC_2	22
Scott'nin P_i	25
Krippendorff'un Alfa	26
Fleiss'nin Kappa	29
Conger'in Kappa	31
Brennan-Prediger	32
Uzlaşma Yüzdesi	34
Sınıf İçi Korelasyon (ICC)	35
İlgili Araştırmalar	36
BÖLÜM 3	43
YÖNTEM	43
Araştırmanın Türü.....	43
Çalışma Grubu	43
Veri Toplama Aracı	46
Matematik Okuryazarlığı Uygulaması.....	46
Okuduğunu Anlama Uygulaması.....	47
Uzman Görüş Formu.....	47
Dereceli Puanlama Anahtarı (DPA)	48
Veri Toplama Süreci	52
Verilerin Analizi	52

BÖLÜM 4	55
BULGULAR VE YORUM.....	55
Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Karşılaştırılması.....	55
Matematik Okurvazarlığı ve Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Kayıp Veri Oranı Dikkate Alınarak Karşılaştırılması	60
Matematik Okurvazarlığı ve Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının İncelenmesi	65
Matematik Okurvazarlığı ve Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Kayıp Veri Oranı Dikkate Alınarak İncelenmesi.....	68
Matematik Okurvazarlığı ve Okuduğunu Anlama Uygulamalarında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerlerinin Kayıp Veri Oranına Göre İncelenmesi.....	72
Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamalarında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerlerinin Kayıp Veri Oranına Göre İncelenmesi.....	73
Puanlayıcı Sayısının Artmasının Puanlama Güvenirliğine Etkisi.....	74
KTİK'ya Göre Puanlama Güvenirliği İçin Hesaplanan Güvenirlik Katsayılarının Sayısal ve Sözel Ders Olma Durumuna Göre Kıyaslanması	75
BÖLÜM 5	77
SONUÇ VE ÖNERİLER.....	77
Sonuçlar.....	77
Öneriler	78
KAYNAKLAR	81
EK-1:	87
MATEMATİK OKURYAZARLIĞI UYGULAMA FORMU	87

EK-2:	92
OKUDUĞUNU ANLAMA UYGULAMA FORMU	92
EK-3:	99
DERECELİ PUANLAMA ANAHTARLARI	99
EK-4:	104
YÖNERGE ÖRNEKLERİ	104
EK-5:	107
ARAŞTIRMA BULGULARINA AİT TABLOLAR	107
EK-6:	123
SINIF İÇİ KORELASYON İÇİN VARSAYIM İNCELEMESİ	123

TABLÖLAR LİSTESİ

Tablo 1. Puanlayıcılar Arası Güvenrilik Katsayıların Sınıflandırılması.....	15
Tablo 2. Kategorilere ve Puanlayıcılara Göre Ölçme Objelerinin Dağılımı	19
Tablo 3. r Puanlayıcı Tarafından n Ölçme Objesinin Puanlanmasına Yönelik Dağılım	20
Tablo 4. Ölçme Objelerine ve Kategorilere Göre r Puanlayıcının Dağılımı	20
Tablo 5. Puanlayıcı ve Kategorilere Göre n Kategorinin Dağılımı	21
Tablo 6. Kappa İstatistiğinin Değerlendirilmesinde Kıyaslama Kriteri	21
Tablo 7. Fleiss (1981) ve Altman (1991) Kappa Kıyaslama Ölçeği.....	22
Tablo 8. Çalışma Grubunun Betimsel Özellikleri.....	44
Tablo 9. DPA'lara İlişkin Görüş Bildiren Uzmanların Özellikleri.....	45
Tablo 10. Puanlayıcıların Özellikleri.....	46
Tablo 11. $\alpha = 0,05$ Anlamlılık Düzeyinde KGO ve KGİ'ler için Uzman Sayısına Göre Minimum Değerler.....	50
Tablo 12. Kullanılan DPA'lara Yönelik Olarak Hesaplanan KGO (Davis-Lawhse) ve KGİ'leri	51
Tablo 13. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri	73
Tablo 14. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri	73
Tablo 15. Matematik Okuryazarlığı Uygulamasında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri.....	74

Tablo 16. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri	74
Tablo 17. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	107
Tablo 18. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	108
Tablo 19. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	109
Tablo 20. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	110
Tablo 21. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	111
Tablo 22. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	112
Tablo 23. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	113
Tablo 24. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	114
Tablo 25. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	115
Tablo 26. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	116
Tablo 27. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	117
Tablo 28. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	118
Tablo 29. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	119

Tablo 30. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	120
Tablo 31. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	121
Tablo 32. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları	122
Tablo 33. Puanlayıcılara göre matematik okuryazarlığı toplam puanına ait betimsel istatistik değerleri.....	123
Tablo 34. Puanlayıcılara göre okuduğunu anlama toplam puanına ait betimsel istatistik değerleri	123
Tablo 35. Varyansların homojenliği için Levene Testi	123

ŞEKİLLER LİSTESİ

Şekil 1. Çalışmada kullanılacak puanlayıcı güvenirliği katsayıları sınıflaması	8
Şekil 2. İki puanlayıcı için matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC2, SPI = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma)	56
Şekil 3. İki puanlayıcı için Okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC2, SPI = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)	57
Şekil 4. İki puanlayıcı için kayıp veri olduğu durumda matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC2, SPI = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma Yüzdesi)	61
Şekil 5. İki puanlayıcı için kayıp veri olduğu durumda okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)	63
Şekil 6. Üç puanlayıcı için matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)	65
Şekil 7. Üç puanlayıcı için okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK =	

Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)67

Şekil 8. Üç puanlayıcı için kayıp veri olduğu durumda matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan-Prediger, UY = Uzlaşma Yüzdesi).....69

Şekil 9. Üç puanlayıcı için kayıp veri olduğu durumda okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan-Prediger, UY = Uzlaşma yüzdesi)71

SİMGELER VE KISALTMALAR LİSTESİ

DPA	Dereceli Puanlama Anahtarı
KTK	Klasik Test Kuramı
ICC	Intra Class Corelation (Sınıf İçi Korelasyon Katsayısı)
PISA	Programme for International Student Assessment (Uluslar Arası Öğrenci Değerlendirme Sınavı)
MEB	Milli Eğitim Bakanlığı

BÖLÜM 1

GİRİŞ

Bu bölümde problem durumu ortaya konulmaya çalışılmış, araştırmanın amacı, önemi ve sınırlılıkları hakkında bilgi verilmiştir.

Problem Durumu

Bir ölçme işleminde doğru kararlar verebilmek, yapılan ölçme işleminin karar vermek için uygun olup olmamasına bağlıdır. Aynı zamanda doğru bir şekilde karar alınmasında elde edilen ölçümlerin hatalardan arınık olması gerekir. Bu durum ise ölçme işlemine ve ölçümlere karışan hataların araştırılması gerekliliğini ortaya çıkarır. Diğer bir deyişle yapılan uygulamalarda doğru kararlar verebilmek için elde edilen ölçümlerin güvenilirliğinin incelenmesi gerekir (Yelboğa, 2008).

Bir test uygulandığında, testin uygulayıcıları sonuçların aynı bireyler aynı testi benzer koşullar altında aldığı anda tekrarlanabilir olmasını ister. Güvenirlik test puanlarının istenilen bu tutarlılığı ya da yeniden üretilebilirliğidir. Pratik bir terim olarak güvenirlik aynı ya da alternatif test formlarının tekrarlı uygulamaları üzerinde göreceli olarak tutarlı olan bireylerin sapma puanlarının ya da z-puanlarının derecesidir (Crocker ve Algina, 2006; s.105). Bir ölçme aracının güvenilirliğini belirleyebilmek için, bu ölçme aracı hangi özelliği ölçüyorsa onun bu özelliğinin gerçek değerlerine yakın ölçüler verdiği gösterilmelidir. Bunun için bu ölçme aracının “özelliği ölçülen varlık veya olayların bu özelliğinde bir değişme olmadıkça onları hep aynı sıraya koyan ölçüler vermesi” gerekir. Diğer yandan bir testin aynı kişilere benzer koşullar altında birden fazla uygulanmasıyla elde edilen ölçümler arasındaki korelasyonu da uygulamalardan elde edilen ölçümlerin sıralamasının ne ölçüde değişmez olduğunu gösteren kullanışlı bir ölçüdür (Özçelik, 2013a; s.123). Güvenirlik bir

kişinin testten aldığı puanda ne kadar hata bulunduğunu gösteren bir sayı değildir; ancak güvenilirlik bir testin puanlarına karışmış olabilecek ölçme hatasını gösterir. Kişinin bir testten aldığı puanın ölçme hatası nedeniyle belli olasılıklarla ne kadar değişmiş olabileceği güvenilirlik tahmini ve test puanlarının standart sapmasından yararlanılarak tahmin edilir. Diğer bir ifadeyle bir kişinin gerçek puanının belli olasılıklarla hangi puanlar arasında olabileceği hakkında yorum yapılabilir (Özçelik, 2013b; s.150).

Güvenirlik indeksi, gözlenen puanlarla gerçek puanlar arasındaki korelasyon veya gerçek puanlar varyansının gözlenen puanlar varyansına oranı olarak tanımlanır (Crocker ve Algina, 2006; s.115). Güvenirlik indeksi $[0,1]$ aralığında değer alır. Gözlenen puanlarda hiç hata yoksa yani gözlenen puan gerçek puana eşitse güvenilirlik indeksi 1 değerini alır. Eğer gözlenen puanların hatadan oluşuyorsa yani gözlenen puan hata puanına eşitse güvenilirlik indeksi 0 değerini alır. Güvenirlik indeksi bu tanımla teorik olarak kalır ve doğrudan hesaplanamaz. Güvenirliği sayısal olarak ifade etmeye yönelik yollar aranmış ve güvenilirlik katsayısı kavramına gidilmiştir. Bu amaçla test-tekrar test, paralel formlar, eşdeğer yarılar, maddelerin testin bütünüyle korelasyonu ve varyans analizine dayalı yöntemler geliştirilmiştir (Baykul, 2010; s. 161-162).

Güvenirlik tesadüfi hataların sebeplerine göre farklı anlamalarda kullanılır. Ölçmenin amacına uygun olarak yeterli duyarlılıkta olmayan bir araçla yapılan ölçmeler daha duyarlı bir araçla yapıldığına göre daha az güvenilir olur. Bu durumda güvenilirlik duyarlılık anlamındadır. Duyarlılık, ölçme aracının veya sonuçlarının biriminin büyüklüğü ile ilgili olup birimi küçük olan ölçme aracı veya sonucu, birimi büyük olandan daha duyarlıdır ve böylece daha güvenilirdir. Bir nitelik aynı ölçme aracıyla birden fazla ölçüldüğünde birbirine yakın ya da aynı sonuçları veriyorsa bu ölçme sonuçları kararlardır. Bu durumda güvenilirlik kararlılık anlamındadır. Eğitim ve psikolojide daha çok ölçme aracıyla ilgili olan, testi oluşturan maddelerin testin bütünüyle tutarlı olması şeklinde bir güvenilirlikten söz edilir. Bu durumda ise güvenilirlik tutarlılık anlamındadır (Baykul, 2010; s.159-160)

Test, cevaplayıcılar ve bu ikisinin dışındaki bazı faktörler ölçme aracını veya ondan elde edilen puanların güvenilirliğini etkiler. Testin uygulandığı grubun homojenliği, testin uzunluğu, verilen cevaplama süresi, öğrencilerin teste olan tutumları, sınav şartları vb. hususlar güvenilirliği etkileyen bazı faktörlerdir. Bu faktörlerden bazıları hesaplanabilir ve böylece kontrol edilebilir; ancak bazıları güvenilirliği etkilediği bilindiği halde miktarları belirlenemediğinden kontrol edilemez (Baykul, 2010; s.216).

Hata varyansının kaynakları (a) maddeler arası tutarsızlık, yani heterojen maddelere sahip olunması, (b) bir dizi paralel madde ve testler arasındaki performans farklılıkları, (c) aynı ya da paralel testlerden elde edilen performansın zaman aralıklarına göre farklılık göstermesi şeklinde sıralanabilir. Hata varyansı kaynaklarının bu şekilde sınıflandırılmasına dayalı olarak test-tekrar test, eşdeğer yarılar, Cronbach alfa gibi birçok güvenilirlik türü tanımlanmıştır. Ancak bunun dışında hata kaynakları iki ya da daha fazla puanlayıcının her bir test maddesine yönelik yaptıkları puanlamalardan kaynaklanabilir. Böyle durumlarda iki veya daha fazla puanlayıcının yaptıkları puanlamanın aynı olması durumunda puanlayıcı hatasının olmadığı söylenebilir. Eğitimde yazılı ve sözlü yoklamaların puanlamasına öznel etkiler karışabildiği için bu tür puanlara dayalı verilen kararların doğruluğunu değerlendirmek amacıyla farklı puanlayıcıların verdiği madde ve test puanlarının tutarlılığının değerlendirilmesi gerekir (Kan, 2014; s.48-49).

Sosyal bilimler, davranış ve tıp bilimlerinde veri kalitesinin belirlenmesinde kullanılan güvenilirlik işlemleri için ölçümlerin iki geniş kategorisi dikkate alınabilir. Birinci tip ölçümde herhangi bir birey ölçme birimlerini aynı benzerlikte anlamaktadır. Örneğin öğrencinin matematik dersindeki başarı puanı ya da bir hastanın kan basıncı düzeyi. İkinci tip ölçümde ise ölçme birimleri bireyler için aynı anlama gelmez. Örneğin bir hastanın memnuniyetinin beşli likert tipi bir ölçekle ölçüldüğü düşünölsün: (1) hiç memnun değöl, (2) memnun değöl, (3) ne memnun ne de memnun değöl, (4) memnun, (5) çok memnun. Böyle bir işlemde gerçök sınıflandırmayı tanımlamak çok zordur. Bu durumda genel yaklaşım şans yoluyla uzlaşma için hesaplanan bir yöntem kullanılarak birkaç puanlayıcının sınıflanmasının karşılaştırılmasıdır (Gwet, 2001; s.2). Ölçümlerin birinci tipinde ölçümdeki birimler herkes tarafından aynı şekilde anlaşıldığı, daha objektif olduđu için ölçme işlemine karışan hata puanlama yapan bireyler dışından gelen hata kaynaklarıdır. Örneğin çoktan seçmeli bir test puanlandırılırken puanlama yapan kişiden kaynaklanacak hata minimum düzeydedir. Ancak ölçümlerin ikinci tipinde ölçme birimlerinin kişiden kişiye farklı anlama gelebileceğı ya da ölçme birimlerinin farklı anlamlandırmaya açık olması nedeniyle puanlayıcıların değerlendirmesinde farklılıklar artacaktır. Bu durumda puanlayıcıların uzlaşma, ölçümlerin güvenilirliğine kanıt sunması açısından önemlidir.

Ölçmelerin tekrarı, hiç değölse ölçmenin tekrarı sayılabilecek bazı yöntemlerin olması eğitimdeki ölçmelerde hata miktarının veya güvenilirliğinin kestirilebilmesi için önemlidir. Ölçme aracının bireyler üzerinde birçok kez uygulanması çođu zaman mümkün

olmadığından ölçme aracının veya yönteminin bir kez uygulanması zorunlu hale gelir. Bu nedenle güvenilirliğe farklı anlamlar verilmiş ve buna bağlı olarak farklı hesaplama yöntemleri ortaya çıkmıştır (Turgut ve Baykul, 2012; s.123). Özellikle çoktan seçmeli, doğru yanlış madde türlerine göre açık uçlu, performans dayalı ölçümler gibi tek bir doğru cevabı olmayan üst düzey bilişsel özellikleri ölçen bir yapı bulunan durumların tekrarlanabilmesi çok daha zor olabilir. Böyle durumlarda elde edilen puanların güvenirliliği için bir uygulamaya dayalı toplanan verileri puanlandırırken birden fazla puanlayıcı kullanılarak bu puanlayıcıların verdikleri puanların yakın olduğunun gösterilmesi, ölçümlerde puanlayıcı yanlılığının az ya da ölçümlerin daha objektif olduğu anlamına gelmektedir. Bu durum elde edilen puanların güvenirliliği hakkında bir kanıt sunar.

Bireylerin göstermesi gereken alt düzey düşünceler ve üst düzey düşünceler arasında fark vardır ve bu farklılıklardan biri bu becerilerin kazandırılıp kazandırılmadığını belirleme yönündedir. Alt düzey düşünceler daha çok hatırlamaya veya ezberlemeye dayalı davranışları gerektirir ve yaygın olarak klasik ölçme araçlarıyla (çoktan seçmeli, kısa yanıtlı, eşleştirmeli, doğru-yanlış testleri vb.) ölçülür. Üst düzey zihinsel beceriler ise bireyin yeteneğini sergilerken kullandığı bilişsel, duyuşsal ve devinimsel özelliklerin bir bütünüdür ve bireyin birden fazla beceriyi kendi bireysel özellikleri ile ilişkilendirerek kullanmasına dayanır. Bireylerde gözlenmek istenen okuduğunu anlama, yazılı anlatım, sunu yapma, araştırma-inceleme vb. becerileri ve bunun daha üst şekli olan yeteneği belirlemede günümüzde bireyin davranışlarını belirlemek için kullanılan çoktan seçmeli, doğru-yanlış, eşleştirmeli, boşluk tamamlamalı gibi bazı klasik ölçme yöntemleri yetersiz kalmaktadır. Bireyin davranışları hakkında doğru bilgilere ulaşmak için bireyin hangi davranışları değerlendirilecekse o davranışlara uygun ölçme yolunun seçilmesi önemlidir (Kutlu, Doğan ve Karakaya, 2010; s: 7-9).

Ölçme araçlarının bazı türleri için maddelerin sadece bir seti kullanılır (örneğin bir performansın kontrolü üzerindeki davranışların bir listesi) fakat çoklu gözlemler her bir birey için aracı tamamlayan iki ya da daha fazla puanlayıcıyla toplanır. Bu durumda elde edilen gözlemlerin tutarlılığıyla ilgilenilir (Crocker ve Algina, 2006; s.143). Uzun yanıt gerektiren ve bir dereceye kadarda yanlış olan uzun cevapların doğruluk derecelendirmesine karar vermede önemli ölçüde öznel davranılabilmektedir. Bir değerlendiriciden diğerine hatta bir değerlendiricinin ayırdığı süreye göre de cevaba verilen puan değişebilmektedir (Özçelik, 2013b; s.42). Açık uçlu bir soruya verilen cevaplar için eğer iki uzman tarafından

belirlenen ölçütlere göre puanlandırma yapılmışsa değerlendiriciler arasındaki uyum puanların güvenilirliği hakkında bilgi verir (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz, Demirel, 2012; s.114).

Güvenirlik durum belirleme puanlarının tutarlılığı olarak adlandırılır. Örneğin, güvenilir bir testte bir öğrencinin durum belirlemesinin ne zaman tanımlandığına, yanıtların ne zaman ve kim tarafından puanlandığına aldırılmaksızın aynı puanı alması beklenir. Test-tekrar test, eşdeğer formlar, iki yarı gibi yöntemlerin her biriyle elde edilen güvenilirlik katsayıları belirli bir testte ya da birden fazla test için öğrenci performansının tutarlılığını belirlemesinde kullanılan istatistiksel yöntemlerdir. Bu tür güvenilirlik belirleme yöntemleri sınıf içindeki durum belirlemelerden daha çok standart ya da yüksek riskli testlerde söz konusu olur. Sınıf içi değerlendirmelerde ve dereceli puanlama anahtarı (DPA-rubric) kullanılan değerlendirmelerde genel olarak dikkate alınan güvenirlığın iki biçimi puanlayıcı güvenirligini içerir. Puanlayıcı güvenirligi genel olarak puanların tutarlılığı olarak adlandırılır ve en az iki bağımsız puanlayıcı ya da farklı zamanlarda aynı puanlayıcıların puanlandırılmasına dayanır (Moskal ve Leydens, 2000).

Ölçmede yer alan birden fazla puanlayıcının ölçülen aynı özelliğe verdikleri puanlar arasındaki tutarlılık “puanlayıcılar arası tutarlılık” olarak tanımlanır ve burada bulunan tek hata kaynağı puanlayıcıdır. Bir puanlayıcının verdiği puanlar arasındaki kararlılığın göstergesi puanlayıcı-içi (intrarater) güvenilirlik kapsamında belirlenir. Puanlayıcı-içi güvenilirlik bir puanlayıcının farklı zamanlarda aynı bireylerin kâğıtlarını iki defa ya da daha fazla puanlamasıyla elde edilen puanlar arasındaki ilişkinin hesaplanmasına dayanır. Burada puanlayıcının puanlamadaki güvenirliginin kestirilmesi söz konusudur. Diğer taraftan farklı puanlayıcılar tarafından birbirinden bağımsız olarak her bir öğrencinin kâğıdına aynı puanı verme tutarlılık derecesinin göstergesi olarak puanlayıcılar arası (interrater) güvenilirlik şeklinde ele alınır (Güler, 2008; s.17-18).

Puanlama güvenilirliği genel olarak puanlayıcılar tarafından verilen puanların tutarlılığına dayanır ve en az iki bağımsız puanlayıcı ya da farklı zamanlarda aynı puanlayıcıların puanlandırılmasıyla ilişkilidir. İki bağımsız puanlayıcının olduğu durum puanlayıcılar arası güvenilirlik (interrater reliability) olarak; farklı zamanlarda aynı puanlayıcının olduğu durum puanlayıcı içi güvenilirlik (inrarater reliability) olarak adlandırılır. Puanlayıcılar arası güvenilirlik öğrenci puanının puanlayıcıdan puanlayıcıya değişip değişmediğine odaklanır ve puanlarda değerlendiricinin subjektif yargıları olabileceğini düşünülür. Puanlayıcı içi

güvenirlikte ise öğrencilerin puanlarına dışsal bir faktörün etki ettiği söylenebilir (Moskal ve Leydens, 2000).

Neundorf (2002)'a göre puanlayıcıların (kodlayıcıların) bulunduğu bir çalışmada güvenilirlik puanlayıcılar arası güvenilirliğe diğer bir deyişle iki ya da daha fazla kodlayıcı arasındaki ilişki ya da uzlaşmanın miktarına dönüşür. Puanlayıcılar arası güvenirliliğin kabul edilebilir bir düzeyine ulaşması puanlama planının temel doğasını oluşturma ve çoklu puanlayıcı kullanmanın pratik avantajlarına sahip olma gibi nedenden dolayı önemlidir (s.141-142). Puanlayıcılar arası uyumsuzluğun bazı nedenleri olabilir. Bunlardan biri değerlendiricilerin (yani puanlayıcıların) ölçülmek istenen özelliğin tanımında ve/veya özel değerlendirme seviyeleri ya da kategorilerin tanımlamasındaki uyumsuzluk olabilir (Kılıç, 2009). Bu tür uyumsuzluğun önlenmesinde puanlayıcılara yönelik hazırlanacak ayrıntılı açıklamaların yer aldığı yönergelerin bulunması ve/veya puanlanacak özelliğe uygun olarak hazırlanmış DPA'ların kullanılması uyuşmayı artıracak önemli bir durum olarak karşımıza çıkabilir.

DPA bireylerin çalışmalarına ilişkin ölçütlerin, bu ölçütlere yönelik tanımların ve başarı düzeylerinin bulunduğu bir puanlama aracıdır. DPA puanlayıcıdan puanlayıcıya değişmeyen standart ve nesnel bir değerlendirme yapmaya yardımcı olur (Kutlu vd, 2010; s: 53). DPA'lar kavram haritalarını, literatür eleştirilerini, yansıtıcı yazmaları, bibliyografileri, sözlü sunumları, eleştirel düşünmeyi, alıntı analizlerini, portfolyoları, projeleri ve sözlü ya da yazılı iletişim becerilerini içeren öğrenci ürünlerini değerlendirmek için kullanılır (Reddy ve Andrade, 2010).

Sınıf ile ilgili değerlendirmelerde problem çözme, yaratıcılık, yazma süreci, kendine saygı, tutum ve diğer yapılar öğretmenler tarafından incelenmek istenir. Yapının ne olduğuna bakılmaksızın öğrencilerin öğrenmelerinin altında yatan süreçlerdeki ikna edici kanıtların oluşturulması ve gösterilmesi için yapının yüzeylerinin tanımlanması gerekir. Ölçme aracının geliştirilmesinde bu yüzeyler dikkatli bir şekilde belirlenmeli ve puanlama kriterleri olarak ele alınmalıdır. DPA'daki önemli nokta, yapının altında yatan uygun göstergelerin değerlendirildiği yanıtların faktörlerinin DPA'da olup olmadığı üzerinedir. Puanlanan DPA geliştirilirken belirlenen kriterler sadece ölçülmek istenen performansla yönelik olması DPA'nın içerik ilişkili kanıtı olur. Yapı-ilişkili geçerlilik kanıtı ise ölçme aracının sadece ve tam olarak istenilen yapıyı ölçtüğünü destekleyen kanıtlara dayanır. DPA'nın geçerliğini sağlayabilmek değerlendirmenin amacına bağlı olduğundan, öğretmenler öğrencilerin yanıtlarından ne öğrenmek istediklerini (yani amaç) ve bu yeterlilikleri öğrencilerin nasıl

göstereceğini (yani hedefler) açık olarak ortaya koymalıdır. Eğer puanlanan DPA verilen görev için öğrencinin cevaplarının değerlendirilmesinde rehber olarak kullanılırsa, bu DPA'nın hem ürünü hem de süreci ele alan kriterleri içerdiği söylenebilir (Moskal ve Leydens, 2000).

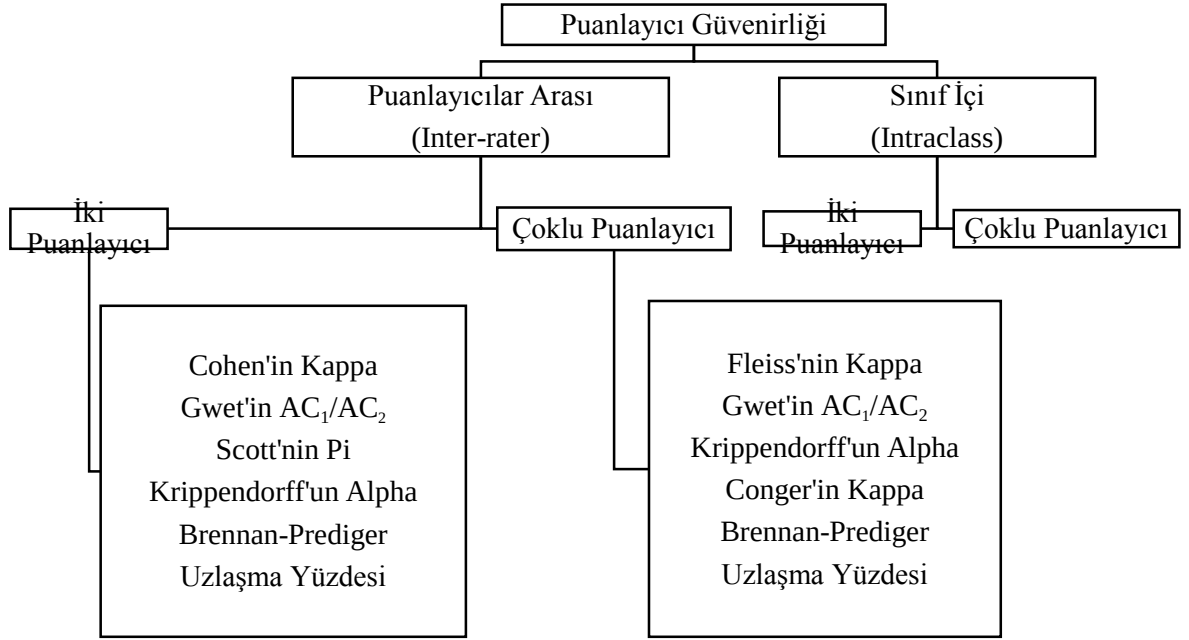
Literatürde DPA'ların araştırma projelerini değerlendirmek (Bresciani vd., 2009) ve öğrenci performansları ile programın eğitsel hedeflerini değerlendirmek (Newell, Dahm ve Neewell, 2002) için yapılan çalışmalar bulunmaktadır ve bu çalışmalarda puanlayıcılar arası güvenilirlik katsayıları DPA'lardan elde edilen puanların güvenilirliğini belirlemede kullanılmıştır. Puanlayıcı güvenirliliği çalışmaları DPA'ların öğrenci performansının yorumlanmasında nispeten ortak yoruma yol açabildiğini gösterme eğilimindedir (Reddy ve Andrade, 2010). Aynı şekilde puanlama anahtarlarının kullanılmasını açık uçlu soruların puanlanmasında, puanlayıcılar arası güvenirliliği arttırmaktadır (Tekindal, 1998)

Puanlama güvenirliliğine konu olan durumlar genel olarak puanlama işlemi yapılırken her bir puanlayıcı için aynı anlama gelmeyen ölçüm birimlerinin olabildiği zamanlarda ortaya çıkmaktadır. Bu tür durumlarda puanlayıcıların puanlamaya konu olan durum üzerinde çelişkiye düşmesi, kararsız kalması gibi nedenlerden dolayı bazı birimleri puanlandırmayabilir, yani boş bırakabilir. Diğer bir açıdan puanlamaya konu olan performansın tüm parçaları aday tarafından gösterilmediği için bazı puanlama birimleri boş kalabilir.

Eğitim ile ilgili çalışmalarda ve psikolojik testlerde madde düzeyinde kayıp veri sorunu yaygındır (Izquierdo ve Pedrero, 2014). “Kayıp veri (missing data)” kavramını amaçlanan gözlemlerin yapılamadığının ya da bulunamadığının göstergesi olan veri seti içindeki boşluklar olarak tanımlamaktadır (Little ve Rubin, 1987'den aktaran Demir ve Özbaşı, 2013). Kayıp değerlerin üstesinden gelmeyi amaçlayan geleneksel yöntemler ve modern yöntemler olarak adlandırılan çeşitli işlemler bulunmaktadır. Geleneksel yöntemler içinde silip çıkarma (listwise ya da pairwise), yerine basit atama yapma (ortalama atama, regresyon gibi) yöntemler yer almaktadır. Modern yöntemlerde ise maximum-likelihood ya da çoklu yerine atama işlemlerinin olduğu söylenebilir (Izquierdo ve Pedrero, 2014).

Yukarıda yapılan açıklamalardan da görüleceği üzere, puanlama güvenirliliği objektif olarak puanlandırılmayan testler, sözlü sınavlar, mülakatlar vb. durumlar için önemli ve gerekli bir işlem olarak karşımıza çıkmaktadır. Ayrıca bir çalışmada kayıp veri bulunması çalışma sonuçlarının incelenmesi noktasında önemlidir. Yapılan alan yazın incelemesi sonucunda

puanlama güvenilirliğinde kayıp veri sorununu dikkate alan çok fazla çalışmaya rastlanmamıştır. Bu çalışma kapsamında klasik test kuramına göre puanlama güvenilirliği katsayılarının kayıp veri oranının (%5, %10 ve %15) dikkate alınarak incelenmesi amaçlanmıştır. Diğer yandan puanlama güvenilirliği katsayıları sayısal ve sözel ders olma durumuna göre de karşılaştırılacaktır. Araştırma kapsamında Şekil 1 de yer verilen puanlayıcı güvenilirliği katsayıları dikkate alınmıştır.



Şekil 1. Çalışmada kullanılacak puanlayıcı güvenilirliği katsayıları sınıflaması

Araştırmanın Amacı

Çoktan seçmeli, doğru yanlış gibi objektif olarak puanlanabilen maddeler ve bu tür maddelerden oluşan testlerden elde edilen puanların güvenilirliği için yaygın olarak kullanılan bazı güvenilirlik katsayıları bulunmaktadır. Açık uçlu yazılı yoklama soruları ise kişinin cevabı kendisi yazmasına dayanan ve tek bir doğru cevabı olmayan madde türleridir ve bu tür maddelere verilen cevapların objektif olarak puanlanması düşüktür. Bu tür durumlarda daha objektif sonuçlar elde etmek için birden çok puanlayıcı kullanımı yoluna gidilebilir. Bu durumda ise bu tür maddelerden elde edilen puanların güvenilirliğini belirleme konusunda farklı puanlayıcıların verdiği puanlar arasındaki tutarlılığın belirlenmesi ön plana çıkmaktadır. Bu amaca yönelik olarak alanda birçok puanlayıcı güvenilirliği katsayısı ortaya çıkmıştır. Bu çalışmanın amacı puanlayıcı güvenilirliği katsayılarını kayıp veri durumunu dikkate alarak birbirleriyle karşılaştırmasını yapmaktır. Ayrıca araştırma kapsamında sayısal ve sözel derslerde hesaplanan uzlaşma katsayılarının bir kıyaslamasının yapılması

amaçlanmıştır. Böylece kayıp veri olmadığı durumda ve kayıp verinin değişen oranlarına göre hangi katsayıların birbirine yakın sonuçlar verdiği ortaya konulmuş olacaktır.

Bu kapsamda şu sorulara yanıt aranacaktır:

- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında kayıp veri olmadığında iki puanlayıcı için hesaplanan puanlayıcılar arası uzlaşma katsayılarından hangisi daha az hatalı sonuç vermektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında kayıp veri oranı (%5, %10 ve %15) dikkate alındığında iki puanlayıcı için hesaplanan puanlayıcılar arası uzlaşma katsayılarından hangisi daha az hatalı sonuç vermektedir ve kayıp değer oranının artması puanlayıcılar arası uzlaşma katsayılarını nasıl etkilemektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında kayıp veri olmadığında üç puanlayıcı için hesaplanan puanlayıcılar arası uzlaşma katsayılarından hangisi daha az hatalı sonuç vermektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında kayıp veri oranı (%5, %10 ve %15) dikkate alındığında üç puanlayıcı için hesaplanan puanlayıcılar arası uzlaşma katsayılarından hangisi daha az hatalı sonuç vermektedir ve kayıp değer oranının artması puanlayıcılar arası uzlaşma katsayılarını nasıl etkilemektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında iki puanlayıcı için hesaplanan sınıf içi korelasyon değerleri kayıp veri oranına (%5, %10 ve %15) göre nasıl değişmektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında üç puanlayıcı için hesaplanan sınıf içi korelasyon katsayıları kayıp veri oranına (%5, %10 ve %15) göre nasıl değişmektedir?
- *Matematik okuryazarlığı ve okuduğunu anlama* uygulamalarında puanlayıcı sayısının artması puanlama güvenilirliğini nasıl etkilemektedir?
- Puanlama güvenilirliği için hesaplanan güvenilirlik katsayıları sayısal ve sözel ders olma durumuna göre nasıl değişmektedir?

Araştırmanın Önemi

Eğitimde bireylere bilgi, kavrama gibi alt düzey becerilerinin kazandırılmasının amaçlanmasının yanı sıra analiz, sentez, problem çözme gibi üst düzey bilişsel yetenekler

de kazandırılmak istenmektedir. Bu durumda bu tür üst düzey bilişsel yetenekleri ölçmeye yönelik olarak kullanılacak bireyin kendinin yazmasına dayanan, uzun cevaplı olan ve tek bir doğru yanıtı olmayan madde türlerinin kullanımı söz konusu olmaktadır. Bu tür maddelerin puanlanması diğer yöntemlerin (kısa yanıtı, çoktan seçmeli vb.) puanlanmasına göre daha az objektif olduğu için verilen puanların güvenilirliği konusu ön plana çıkmaktadır. Ayrıca kompozisyon türü görevler, okullarda ya da mülakatlarda yapılan sözlü sınavlar, performansa dayalı durum belirlemeler gibi becerilerin ölçülmesinde objektiflik bir sorun olarak karşımıza çıkmaktadır. Ayrıca tıp, sağlık, spor vb. alanlarda birden fazla değerlendiricinin bulunduğu kararların alınması ön plana çıkmaktadır ve bu kararların tutarlılığı söz konusu olmaktadır. Bu sorunun üstesinden gelmek için DPA'larının kullanımı tercih edilse de farklı puanlayıcıların yaptığı ya da farklı zamanlarda yapılan puanlamalarda DPA kullanılmasına rağmen her zaman bire bir uyuma gözlenemeyebilir. Bu durumda yaratıcılık, problem çözme, yazılı ve sözlü anlatım becerileri gibi durumların belirlenmesinde objektifliği sağlamak için farklı puanlayıcılar ya da aynı puanlayıcıların farklı zamanlarda puanlamasının kullanılmasıyla elde edilen puanların güvenilirliği söz konusudur. Bu çalışma puanlayıcı güvenilirliği katsayılarına genel bir bakış çizmesi, durumlara göre farklı puanlayıcı güvenilirliği katsayılarını karşılaştırılması yönüyle elde bulgular açısından önemlidir. Ayrıca puanlayıcı sayısının artması, puanlama sırasında kayıp verinin bulunması ve kayıp veri miktarına göre belirlenen durumlardan yola çıkarak elde edilen bulgular daha tutarlı kararların alınmasında kullanılacak puanlayıcı güvenilirliği katsayıları hakkında fikir vermektedir.

Çalışmadan elde edilen bulgular büyük ölçekli sınavlarda da açık uçlu soru kullanımına geçen ve geçmeyi düşünen Milli Eğitim Bakanlığı (MEB) ve Ölçme, Seçme ve Yerleştirme Merkezi gibi kurumlara elde edilen puanların güvenilirliğini belirleme açısından bir bakış açısı sunabilir. Ayrıca okul uygulamalarında öğretmenler daha adil ve güvenilir puanlar verme konusunda bu çalışma kapsamında elde edilen bilgiler doğrultusunda amaçlarına ve kullanımlarına uygun olarak hangi katsayıyı kullanabilecekleri konusunda bilgi sahibi olabilir. Aynı zamanda çalışmanın bulguları ülkemizde lisansüstü eğitime öğrenci seçimlerinde ve işe alımlarda yaygın olarak kullanılan mülakatlarda farklı puanlayıcılar tarafından yapılan derecelendirmenin tutarlılığını belirleyebilecek en uygun güvenilirlik katsayılarını belirlemesi yönünden de önemlidir.

Sınırlılıklar

Bu çalışmada kolay kullanıcı ara yüzü sayesinde öğretmenlere ve puanlayıcı güvenilirliğini kullanan araştırmacılara örnek olması düşüncesiyle AgreeStat 2015.2 kullanımı tercih edilmiştir ancak bundan dolayı çalışma hesaplanan puanlayıcı güvenilirliği katsayıları bu programda hesaplanabilen katsayılar ile sınırlı tutulmuştur. Ayrıca kıyaslama kolaylığı açısından tek bir yazılımın kullanılması çalışmanın diğer bir sınırlılığıdır. Kayıp veri durumunun incelenmesi açısından ise çalışmada sadece kayıp veri oranı dikkate alınmış, kayıp veri üstesinden gelmeye yönelik yaklaşımlara değinilmemiştir. Kayıp veri oranı olarak çalışma %5, %10 ve %15 ile sınırlıdır.

BÖLÜM 2

KAVRAMSAL ÇERÇEVE

Bu bölümde puanlama güvenirliği hakkında bilgi verilip, puanlama güvenirliği klasik test kuramı (KTK) , genellenebilirlik kuramı ve Rasch ölçme modeli kısaca anlatılmıştır. Ardından çalışma kapsamında kullanılacak olan klasik test kuramına ait puanlayıcılar arası güvenirlik katsayıları ve sınıf içi güvenirlik katsayısı hakkında yüzeysel bilgi verilmiştir.

Puanlayıcı Güvenirliği

Ölçme işleminde kullanılan testlerin hepsi aynı şekilde tutarlı puanlamaya elverişli değildir. Diğer bir deyişle belirli testler diğerlerine göre daha tutarlı bir şekilde puanlanmaya elverişli olabilir. Bundan dolayı bir ölçme işleminde farklı puanlayıcılar tarafından yapılan puanlamalar sonucunda elde edilen ölçümler arasındaki tutarlılık derecesinin veya güvenirliğin sorgulanması anlamlı bir işlemdir. Puanlayıcılar arası güvenirlik, iki veya daha fazla puanlayıcının belirli bir ölçüm ile ilgili uzlaşmasının ya da tutarlılığının derecesi olarak tanımlanır. Puanlayıcı güvenirliği, yargıcı güvenirliği, gözlemci güvenirliği, derecelendiriciler-arası güvenirlik gibi çeşitli isimlerle adlandırılır (Cohen ve Swerdlik, 2013; s.150-151). Puanlayıcılar arası güvenirlik çoğu bilimsel ve bilimsel olmayan alanlarda puanlayıcılar olarak adlandırılan farklı veri üreten kaynaklar arasında uzlaşmanın düzeyinin bir ölçümünü belirleyen bir kelime olarak yaygın bir şekilde kullanılır. Puanlayıcılar arası güvenirlik katsayısı araştırmacılar tarafından veri toplama yöntemlerinin kalitesini belirlemede düzenli olarak kullanılan önemli bir istatistiksel araçtır. Literatürdeki alternatif terimler olarak gözlemciler arası (interobserver/inter-observer), yargılayıcılar arası (interjudge/inter-judge), kodlayıcılar arası (intercoder/inter-coder) kullanılır ve hepsi aynı nitelikte ölçümleri ifade eder (Gwet, 2001; s.vii).

Tüm deneysel psikolojik arařtırmalar için tutarlı bir řekilde doęru ölçümler yapmak gerekli bir amaçtır. Ölçümlerin uygunluęunu deęerlendirme kullanılan üç farklı yaklařım Klasik test kuramı, Genellenebilirlik kuramı ve Madde tepki kuramıdır. Genellenebilirlik kuramı ve madde tepki kuramı uzmanlar tarafından geniř ölçüde önerilse de klasik test kuramı yöntemleri yaygın bir řekilde bugüne kadar popülerlięini korumaktadır (Mushquash ve O'Connor, 2006). Tarihsel olarak, ölçümlerin güvenilirlięine yönelik olarak klasik test kuramına ait gerçek puan modeli 1900'lerin bařından 1940'lara kadar tartıřmasız bir egemenlięe sahip olmuřtur. Alan (davranıř) örnekleme kuramı olarak adlandırılan genellenebilirlik kuramı ise 1950'lerin bařında bir alternatif olarak ortaya çıkmıřtır (Cohen ve Swerdlik, 2013; s.157).

Atılğan (2005) genellenebilirlik kuramının korelasyonel teknięe dayalı olarak puanlayıcılar arası güvenilirlik hesaplayan bir yöntem olduęunu ve KTK'ya göre hesaplanan Kappa istatistięine göre daha güçlü ve kapsamlı bir alternatif olabileceęini dile getirmiřtir. Ayrıca genellenebilirlik kuramının KTK dayalı hesaplama yöntemlerinden olan test-tekrar test ve iç-tutarlılık güvenilirlik yöntemlerini birleřtirdięinden bahsetmiřtir. Dięer yandan Klasik Test Kuramı ve Genellenebilirlik Kuramına göre güvenilirlik katsayılarını karřılařtıran bazı çalıřmalarda her iki kuram için görelilik olarak uyumlu sonuçlar verdięi söylenebilir (Özder, 2012; Yelboęa ve Tavřancıl, 2010). KTK'da yapılan hesaplamalarda anket, test ya da gözlem formlarında kullanılan maddelerin seęeneklerinin aralıkları arasındaki farkların eřit olduęu varsayılırken (non-parametrik modeller hariç) Rasch analizi, seęenekler arası gerçek mesafeleri hesaplayarak daha hassas ve gerçekçi bir aralık birimi oluřturur. KTK'ya göre yapılan hesaplamalardaki örneklemden bağımsız madde güçlük indekslerinin ve testten bağımsız kiři yetenek düzeylerinin oluřturulmaması sorunu Rasch ölçme teknięinde dikkate alınır. Ayrıca Rasch ölçme modelinde her bir madde için kendine ait madde istatistikleri hesaplanarak analizde kullanılır (Bařtürk ve Iřıkoęlu, 2008). Rasch test güvenilirlięi, hata varyansında madde ve deęerlendirici varyansını içerdii için, genellenebilirlik kuramına göre daha yüksektir (Linacre, 1993).

Klasik Test Kuramına Göre Puanlayıcı Güvenirlięi

Klasik test kuramına göre puanlayıcılar arası uzlařmaya yönelik birçok katsayı hesaplanmaktadır. Cohen ve Swerdlik (2013) puanlayıcılar arasındaki tutarlılıęın derecesini belirlemenin en kolay yollarından birinin korelasyon katsayısı hesaplamak olduęunu ve bu

korelasyon katsayısının ise puanlayıcılar arası güvenilirlik katsayısı olarak adlandırıldığını belirtmişlerdir (s.151). Neuendorf (2002) kodlayıcıların değerlendirmeleri arasındaki uzlaşmanın düzeyini belirlemeye yönelik raporlanan çeşitli katsayıların ulaşılabilirliğinden bahsetmiştir. Sosyal ve davranış bilimlerinde en popüler katsayıların ise ham uzlaşma yüzdesi, Scott'nin pi, Cohen'in kappa, Krippendorff'un alpha, Spearman rho ve Pearson r olarak görüldüğünü dile getirmiştir (s.148). Klasik test teorisinde kullanılan puanlayıcılar arası güvenilirlik katsayılarına yönelik olarak farklı sınıflandırmalar bulunmaktadır.

Neuendorf (2002) kodlayıcılar arası güvenirliliğin ölçümlerin bir çeşiti arasındaki içsel tutarlığı değerlendirmedini belirtmiştir. Onun yerine kodlayıcılar arası güvenirlilikte, uzlaşma, şansla uzlaşma ve değişkenlerin birlikte değişimi gibi kriterlerle ilgilendiğine dikkat çekmiştir. Uzlaşma kriterini kullanan yöntemlerin anlaşma yüzdesi ve Holisti'nin yöntemi (1969) olarak belirtmiştir. Şansla uzlaşma katsayılarını ise Scott'in pi (π), Cohen'nin kappa (k), Krippendorff'un alpha (α) şeklinde sıralamıştır. Değişkenlerin birlikte değişimine yönelik olarak hesaplanan katsayıları ise Spearman rho, Pearson korelasyon katsayısı (r), Lin'in uyum korelasyon katsayısı (rc) şeklinde belirtmiştir (s.148-153).

Kutlu vd. (2010) tarafından DPA'lardan elde edilen puanların güvenirliliğini belirlemek için puanlayıcılar arası güvenirliliğe bakmanın gerekliliği dile getirilmiştir. Bu amaçla kullanılan puanlayıcılar arası güvenilirlik katsayılarını toplam puan üzerinde ve ölçüt düzeyinde hesaplanma durumu ile iki puanlayıcı ve ikiden fazla puanlayıcı durumuna göre hesaplanma dikkate alınarak sınıflayan bir tablo oluşturmuşlardır (s.83-87). Bu sınıflama Tablo 2'de verilmiştir. Bu sınıflamada klasik test kuramına dayalı puanlayıcı güvenirliliği katsayılarının bir kısmına yönelik farklı bir sınıflandırma sunmaktadır.

Tablo 1. Puanlayıcılar Arası Güvenirlilik Katsayıların Sınıflandırılması

Özellikler	Toplam Puan üzerinden uyum düzeyine bakar	Uyum düzeyine her ölçüt için ayrı ayrı baka.	2 puanlayıcı arasındaki uyum düzeyine bakar	2' den fazla puanlayıcı arasındaki uyum düzeyine bakar.
İstatistiksel Yöntemler				
Cohen's Kappa	-	+	+	-
Fleiss' Kappa	-	+	-	+
r	+	-	+	-
rs	+	-	+	-
Kendall W	+	-	-	+
Sınıf İçi Korelasyon Katsayısı	+	-	+	+
r: Pearson momentler çarpımı korelasyon katsayısı				
rs: Spearman Brown sıra farkları korelasyon katsayısı				
Kutlu vd, 2010; s.87				

Araştırmacılar her zaman en yüksek güvenilirliğe ulaşmayı amaçlar. Mükemmel güvenilirliğe ulaşmak zor olabildiğinden, özellikle kodlama görevlerinin karmaşık ve bu nedenle detaylı bilişsel süreçleri gerektirdiği zaman, araştırmacıların mükemmel güvenilirlik idealinin ne kadar veriden saptanacağını ve bu saptama işleminin kabul edilebilir güvenilirlik standartlarının üzerinde ya da altında olup olmadığını bilmesi gerekir. Bu iki soru herhangi bir uzlaşma ölçümünde cevaplanması gereken iki temel konudur (Krippendorff, 2004; s.221). Klasik test kuramında güvenilirlik katsayıları genellikle 0 ile 1 arasında değer alır ve güvenilirlik katsayısı 1'e yaklaştıkça güvenilirliğin yüksek olduğu, 0'a yaklaştıkça güvenilirliğin düşük olduğu söylenir (Tan, 2013; s.135). Bir test puanının çok önemli doğurgularının olduğu durumlarda testin güvenilirlik katsayısını yüksek standartlara çekmek gerekirken, eğer bir test puanı farklı test puanları ile birlikte kullanılıyorsa bu test puanı için güvenilirliği en yüksek standartlara taşımanın gereği yoktur. Ancak genel anlamda pratik bir bakış açısıyla güvenilirlik katsayılarını harf notlarına benzetirsek, 0,90 ve üzeri yüksek düzeyde güvenilirlik; 0,80 ile 0,90 arasında orta düzeyde güvenilirlik; 0,65 ile 0,70 arasındaki değerler ise zayıf ya da düşük güvenilirlik olarak adlandırılabilir (Cohen ve Swerdlik, 2013; s.151). Her bir değişken için puanlayıcılar arası güvenilirliğin kabul edilebilir düzeyinin ne olması gerektiği tartışmaya açık bir konudur. Sosyal bilimlerdeki araştırma yöntemlerindeki çoğu temel kitap belirli bir kriter tercih etmez ve bu nedenle tavsiye edilen çeşitli türde kriterler raporlamalıdır. Güvenirlik üzerindeki çalışmalar genel olarak gözden geçirildiğinde çalışmaların tümünde 0,90 ve üzerindeki güvenilirlik katsayıları hepsinde kabul edilebilir düzeyde, çoğu durumda ise 0,80 ve üzerinde ise kabul edilebilir düzeyde olarak alınmıştır (Neuendorff, 2002; s. 143).

Genellenebilirlik Kuramına Göre Puanlayıcı Güvenirliği

Genellenebilirlik kuramı, psikoloji ve eğitim gibi sosyal bilimlere yönelik geniş kavramsal çerçeve ve sayısız ölçme durumu için kapsamlı bir yapı oluşturur. Güçlü bir istatistiksel teknik olan varyans analizini kullanır. KTK ve varyans analizinin bir uzantısıdır. KTK da yapılan uygulamalarda, çoklu hata kaynakları tek bir seferde açık bir şekilde ortaya konulamaz ancak genellenebilirlik kuramı ile çoklu hata kaynakları dikkate alınabilir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012; s.3). KTK'nın sınırlılıklarına tepki olarak ortaya çıkmış olmasına rağmen yine de KTK'nın içinden doğmuş bir yaklaşım olan genellenebilirlik kuramında testten edilen puanları, testten elde edilebilecek olası tüm

puanların bir örneğidir. Genellenebilirlik kuramı hesaplamaları varyans analizine dayanır (Kan, 2014; s.49)

Genellenebilirlik kuramı, potansiyel hata kaynaklarının birden fazla olması durumunda güvenilirliğin hesaplanmasının tek bir analizle yapılmasına fırsat verir. Bu sayede ortak tek bir güvenilirlik katsayısının hesaplanabilmesine olanak sağlar. Genellenebilirlik kuramı yoluyla tek bir analiz ile ayrı ayrı hem bağıl hem de mutlak değerlendirme için bir tek güvenilirlik katsayısı hesaplanabilir. Daha güvenilir ve geçerli ölçme araçlarının hazırlanabilmesi için ölçme aracında kaç madde bulunması gerektiği, kaç puanlayıcının puanlama yapmasının uygun olacağı hakkında bilgi verir. G -çalışması ile ölçmedeki hata kaynakları hakkında olabildiğince bilgi edinilmesine olanak vererek ölçme deseninin seçilmesine ve ölçme araçlarının geliştirilmesine yardımcı olur. K -çalışması ise ölçme hatası ve güvenilirlikte artma ya da azalma durumu için madde, puanlayıcı gibi her bir değişkenlik kaynağının koşullarının nasıl değişmesi gerektiği hakkında bilgi verir. G-çalışması, bir ölçme aracının geliştirilmesi ya da ileriki uygulamalarda kullanımı için değişkenlik kaynaklarından gelen hataların azaltılmasına yönelik bilgi verirken; K-çalışması, ölçme aracının geliştirilmesi ya da ileriki uygulamalarda kullanımda ölçmenin nasıl olması gerektiği hakkında bilgi verir (Atılgan, 2005).

Rasch Ölçme Modeline Göre Puanlayıcı Güvenirliği

Rasch modeli madde tepki kuramını dayalı bir modeldir. Rasch model bir parametrelili lojistik model olarak da bilinir; iki ve üç parametrelili modelin özel bir halidir. Rash modelini kullanıcılar için çekici yapan bazı özellikleri vardır. Bunlardan biri daha az parametresi içerdiğinden dolayı daha kolay çalışma imkanı vermesidir. Bir diğeri ise parametre kestirimi ile ilgili problemler daha genel modellere göre oldukça daha azdır (Hambleton ve Swaminathan, 1991; s.47). Rasch ölçme modeli, objektif ölçme işlemi gerçekleştirebilmek amacıyla analizlerde ham puanı kullanmak yerine ham puanın bir tür standart dönüştürmesi olan logit değerlerini kullanır (Linacre, 1993'ten aktaran Baştürk ve Işıkoğlu, 2008). Rasch modeli objektiflik üzerine kurulmuş bir modeldir. Rasch analizi ile test, ölçek ve gözlem formları arasındaki gerçek uzaklıklar hesaplanarak daha hassas ve gerçekçi bir aralık birimi oluşturulabilir. Ayrıca Rasch ölçme modeli ile yanlılık analizi yapılabilir, yanlılıkların kaynakları ve yanlılıkların hangi değerlendirmeci ya da değerlendirmecilere ait olduğu gösterilebilir (Semerci, Semerci ve Duman, 2013).

Çalışma Kapsamında Kullanılan Güvenirlik Katsayıları

Çalışma kapsamında AgreeStat programında hesaplanan puanlayıcılar arası uzlaşma katsayıları ve sınıf içi korelasyon katsayısı kullanılmıştır. Programın el kitapçığında (manuelinden) faydalananak arka planında kullanılan matematiksel formüllerden bahsedilmiştir. Öncelikle formüllerde hesaplanan ağırlıklandırma açıklanmış ve gerekli olan tablolar hakkında bilgi verilmiştir. Daha sonra kullanılan uzlaşma katsayıları hakkında genel bilgiler bu alt bölümde sunulmuştur.

Ağırlıklandırmalar

AgreeStat 2015.2 programında ağırlıklandırma yapılmadan önce kullanılan kategorilerin numerik mi yoksa kategorik mi olduğuna dikkat edilmektedir. Bir güvenirlik çalışmasında 10 numerik kategori ($x_1, x_2, \dots, x_{10}$) kullanıldığını varsaydığımız bir çalışmada, x_1 en küçük sayı ve x_{10} en büyük sayı olsun. x_k ve x_l gibi herhangi iki sayı için kuadratik ağırlıklandırmalar şu şekilde hesaplanır:

$$w_{kl} = \begin{cases} 1 - \frac{(x_k - x_l)^2}{(x_{10} - x_1)^2}, & k \neq l \\ 1, & k = l \end{cases}$$

w_{kl} = k ve l kategorilerine ilişkin ağırlıklandırmalar

Bir güvenirlik çalışmasında 10 alfabetik kategori kullanıldığını varsayalım ve bu kategorileri ilk 10 tam sayı ile numaralandıralım. k ve l gibi iki kategori için kuadratik ağırlıklar şu şekilde hesaplanır:

$$w_{kl} = \begin{cases} 1 - \frac{(k - l)^2}{(10 - 1)^2}, & k \neq l \\ 1, & k = l \end{cases}$$

Tablolar *

AgreeStat 2015.2’de farklı uzlaşma katsayılarının ve bu katsayıların varyanslarının hesaplanmasındaki işlemler çapraz (contingency) tabloyana dayanır. Çapraz tablo iki farklı değişkene ilişkin ölçme objelerinin dağılımını gösteren bir tablodur. İki puanlayıcının bulunduğu durum için hesaplanan çapraz tablo örneği Tablo 2’de verilmiştir.

*Bu kısımdaki bilgiler AgreeStat, 2015.1’in el kitapçığından alınmıştır.

Tablo 2. Kategorilere ve Puanlayıcılara Göre Ölçme Objelerinin Dağılımı

Puanlayıcı A	Puanlayıcı B						Toplam
	1	2	...	l	...	q	
1	p_{11}	p_{12}	...	p_{1l}	...	p_{1q}	p_{1+}
2	p_{21}	p_{22}	...	p_{2l}	...	p_{2q}	p_{2+}
.	...	...	...	...	...	...	.
.							.
.							.
K	p_{k1}	p_{k2}	...	p_{kl}	...	p_{kq}	p_{k+}
.	...	...	...	...	...	...	.
.							.
.							.
q	p_{q1}	p_{q2}	...	p_{ql}	...	p_{qq}	p_{q+}
Toplam	p_{+1}	p_{+2}	...	p_{+l}	...	p_{+q}	1

p_{k+} = k. satırın marjinal oranı (proportion)

p_{+l} = l. sütunun marjinal oranı

Ayrıca bazı uzlaşma katsayılarının hesaplanmasında ağırlıklandırılmış oranlar ($\bar{p}_{+k}$ ve $\bar{p}_l$) tanımlanmaktadır:

$$\bar{p}_{+k} = \sum_{l=1}^q w_{kl} p_{+l}$$

$$\bar{p}_l = \sum_{k=1}^q w_{kl} p_{k+}$$

Üç ya da daha fazla puanlayıcının bulunduğu durumda ise “r puanlayıcı tarafından n ölçme objesinin puanlanmasına yönelik çapraz tablo” (Tablo 4), “ölçme objelerine ve kategorilere göre r puanlayıcının dağılımına yönelik çapraz tablo” (Tablo 5) ve “puanlayıcı ve kategoriye göre n ölçme objesinin dağılımına yönelik çapraz tablo” (Tablo 6) olmak üzere üç farklı tablo kullanılmaktadır.

Tablo 3 n ölçme objesini puanlayan r puanlayıcı tarafından ham puanların bir tablosunu temsil etmektedir. Örneğin, c_{ig} puanlayıcı g tarafından ölçme objesi i için belirli bir kategori (ya da puanı) temsil eder.

Tablo 3. r Puanlayıcı Tarafından n Ölçme Objесinin Puanlanmasına Yönelik Dağılım

Ölçme Objesi	Puanlayıcılar				
	1	...	g	...	q
1	c_{11}	...	c_{1g}	...	c_{1r}
.	...	...	...	...	...
.					
i	c_{i1}	...	c_{ig}	...	c_{iq}
.	...	...	...	...	...
.					
n	c_{n1}	...	c_{ng}	...	c_{nr}

Tablo 4 ölçme objelerine ve kategorilere (ya da puanlara) göre r puanlayıcının dağılımını temsil etmektedir. n ölçme objесinin q olası kategorilerinden biri içinde r puanlayıcı tarafından sınıflandırıldığı varsayılır. r_{ik} sembolü k kategorisi içinde ölçme objesi i'yi sınıflandıran puanlayıcıların sayısını temsil etmektedir ve r_i ölçme objesi i için a kategorisi atayan puanlayıcıların sayısıdır.

Tablo 4. Ölçme Objelerine ve Kategorilere Göre r Puanlayıcının Dağılımı

Ölçme Objesi	Kategori					Toplam
	1	...	K	...	q	
1	r_{11}	...	r_{1k}	...	r_{1q}	r_1
.	...	...	...	...	...	.
.						.
i	r_{i1}	...	r_{ik}	...	r_{iq}	r_i
.	...	...	...	...	...	.
.						.
n	r_{n1}	...	r_{nk}	...	r_{nq}	r_n
Ortalama	$\bar{r}_{+1}$	...	$\bar{r}_{+k}$		$\bar{r}_{+q}$	$\bar{r}$

Tablo 5 puanlayıcı ve kategorilere göre n ölçme objесinin dağılımını temsil etmektedir. n_{gk} sembolü puanlayıcı g tarafından kategori k içinde sınıflandırılan ölçme objelerinin sayısını ve n_g puanlayıcı g'nin puanlandığı ölçme objelerinin sayısını gösterir.

Tablo 5. Puanlayıcı ve Kategorilere Göre n Kategorinin Dağılımı

Puanlayıcı	Kategori					Toplam
	1	...	K	...	Q	
1	n_{11}	...	n_{1k}	...	n_{1q}	n_1
.	...	...	...	...	...	.
.						.
g	n_{g1}	...	n_{gk}	...	n_{gq}	n_g
.	...	...	...	...	...	.
.						.
r	n_{r1}	...	n_{rk}	...	n_{rq}	n_r
Ortalama	$\bar{n}_{+1}$	...	$\bar{n}_{+k}$		$\bar{n}_{+q}$	$\bar{n}$

Cohen'in Kappa

Kappa (κ) katsayısı basit bir şekilde meydana gelmeyen şans-beklentili (chance-expected) uzlaşmazlıkların bir oranıdır. Alternatif bir tanım olarak şans ile uzlaşmanın göz önüne alınan kısmı ortadan kaldırıldıktan sonra uzlaşmanın oranıdır. Aşağıdaki denklemle temel olarak ifade edilebilir (Cohen, 1960):

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

p_o = puanlayıcıların uzlaştığı birimlerin oranı

p_e = uzlaşmanın şans yoluyla beklendiği birimlerin oranı

Kappa istatistiğinin değerlendirilmesinde kappa istatistiğinin değerlerine göre uzlaşmanın gücünü tanımlayan kullanışlı bir kıyaslama kriteri Landis ve Koch (1977) tarafından aşağıdaki gibi verilmiştir:

Tablo 6. Kappa İstatistiğinin Değerlendirilmesinde Kıyaslama Kriteri

Kappa İstatistiği	Uzlaşmanın Gücü
<0,00	Zayıf
0,00-0,20	Önemsiz
0,21-0,40	Kayda değer
0,41-0,60	Orta
0,61-0,80	Önemli
0,81-1,00	Neredeyse mükemmel

Landis ve Koch, 1977

Ayrıca Fleiss (1981) tarafından ve Altman (1991) tarafından ileri sürülen kappa kıyaslama ölçeği Tablo 7’de verilmiştir (Fleiss, 1981 ve Altman 1991’den aktaran Gwet, 2012; s.125):

Tablo 7. Fleiss (1981) ve Altman (1991) Kappa Kıyaslama Ölçeği

Fleiss’in Kappa Kıyaslamam Ölçeği		Altman’nın Kappa Kıyaslama Ölçeği	
Kappa istatistiği	Uzlaşmanın gücü	Kappa istatistiği	Uzlaşmanın gücü
<0,40	Zayıf	<0,20	Zayıf
0,40-0,75	İyi için orta düzeyde	0,21-0,40	Kayda değer
0,75 üstü	Mükemmel	0,41-0,60	Orta
		0,61-0,80	İyi
		0,81-1,00	Çok iyi

AgreeStat 2015.2 programında ise iki puanlayıcı için Cohen’nin kappa katsayısı (Cohen, 1960; 1968 den aktaran AGREESTAT, 2015) aşağıdaki gibi tanımlanmaktadır:

$$\kappa = \frac{p_a - p_e}{1 - p_e}$$

$$p_a = \sum_{k,l} w_{kl} p_{kl} \text{ ve } p_e = \sum_{k,l} w_{k+} p_{k+} p_{+l}$$

Ağırlıklandırılmış kappa katsayısının varyansı ise şu şekilde hesaplanmaktadır (AGREESTAT, 2015):

$$v(\kappa) = \frac{1 - f}{n(1 - p_e)^2} \left\{ \sum_{k,l}^q p_{kl} [w_{kl} - (1 - \kappa)(\bar{p}_{+k} + \bar{p}_{l+})]^2 - [p_a - 2(1 - \kappa)p_e]^2 \right\}$$

f = ölçme objesi örnekleme kesri (the subject sampling fraction)

Yukarıdaki formüller incelendiğinde Cohen’nin kappa katsayısı hesaplanırken puanlayıcıların uzlaştığı birimlerin oranından şansla uzlaşılan birimlerin oranı çıkarılıp elde edilen değer şansla uzlaşma olmama olasılığına bölümünden elde edildiği görülmektedir. Uzlaşma katsayıları hesaplanırken iki puanlayıcının birleşik olasılıkları dikkate alınırken; şansla uzlaşma oranı bulunurken marjinal olasılıklar dikkate alınmaktadır.

Gwet’in AC₁/AC₂

AC₁ istatistiği araştırmacıların tahmin etmeye çalıştığı evren parametrelerinin olduğunu varsayan bir uzlaşma katsayısıdır. q yanıt kategorisinin bulunduğu bir kapsamda, AC₁ istatistiği şu şekildedir (Gwet, 2001; s. 79-80):

$$\kappa_{\gamma} = \frac{P_{\alpha} - P_{e,\gamma}}{1 - P_{e,\gamma}}$$

P_{α} toplam uzlaşma olasılığı, $P_{e,\gamma}$ ise şans yoluyla uzlaşma olasılığı olarak tanımlanır ve aşağıdaki şekilde hesaplanır.

$$P_{\alpha} = \sum_{q=1}^Q P_{qq} \quad \text{ve} \quad P_{e,\gamma} = \frac{1}{Q-1} \sum_{q=1}^Q \pi_q (1 - \pi_q)$$

P_{qq} = q kategorisinin her iki puanlayıcı tarafından aynı olarak sınıflandırılması beklenen konuların evren oranı

π_q = bir puanlayıcı tarafından q kategorisinde sınıflandırılan bir konunun olasılığı

AgreeStat 2015 programında Gwet'in AC_1 katsayısının ağırlıklandırılmış şekli olan Gwet'in AC_2 katsayısı (Gwet,2010'dan aktaran AGREESTAT, 2015) aşağıdaki gibi tanımlanır:

$$AC_2 = \frac{p_a - p_e}{1 - p_e}$$

$$p_a = \sum_{k,l} w_{kl} p_{kl} \quad \text{ve} \quad p_e = \frac{T_w}{q(q-1)} \sum_{k=1}^q \pi_k (1 - \pi_k)$$

$$\pi_k = \frac{p_{k+} + p_{+k}}{2}$$

T_w = Tüm kategoriler ile ilişkili tüm w_{kl} ağırlıklarının toplamı

AC_2 'nin varyansı ise şu şekilde hesaplanır (AGREESTAT, 2015):

$$v(AC_2) = \frac{1-f}{n(1-p_e)^2} \left\{ \sum_{k,l}^q p_{kl} \left[w_{kl} - 2 \frac{(1-AC_2)T_w}{q(q-1)} x \left(1 - \frac{\pi_k + \pi_l}{2} \right) \right]^2 - [p_a - 2(1-AC_2)p_e]^2 \right\}$$

f = ölçme objesi örnekleme kesri (the subject sampling fraction)

Yukarıdaki formüller incelendiğinde Gwet'in AC_2 katsayısı Cohen'nin kappa katsayısına benzer olarak puanlayıcıların uzlaştığı birimlerin oranından şansla uzlaşılan birimlerin oranının farkının şansla uzlaşılmayan birimlerin oranına bölümüdür. Gwet'in AC_2 katsayısında puanlayıcıların uzlaştığı birimler Cohen'nin kappa katsayısındakiyle aynı şekilde hesaplanırken şansla uzlaşma oranının hesaplanmasında farklılaşma bulunmaktadır. Cohen'nin kappa katsayısında şansla uzlaşmayı hesaplarken iki kategoriye ait marjinal olasılıkların çarpımının ağırlıklandırmasını kullanırken Gwet'in AC_2 katsayısı marjinal

olasılıkların aritmetik ortalamasından türettiği yeni bir oran üzerinden yapılan toplamın sonucunu tüm ağırlıklandırmaların toplamının kategori sayısının bir eksiğine bölerek elde ettiği sabit bir sayıyla çarparak hesaplamaktadır. Burada Cohen'nin kappa katsayısının toplamın içindeki her bir oran için farklı bir ağırlıklandırma yaptığı görülürken Gwet'in AC₂ katsayısı toplamın içindeki her bir oran için sabit bir ağırlıklandırma (her bir olasılık için hesaplanan ağırlıklandırmaların toplamından elde edilmiş bir toplam değerin kategori sayısına dayalı elde edilen bir değere bölümü ile elde edilmiş bir ağırlıklandırma) yapmaktadır.

Üç ya da daha fazla puanlayıcı için Gwet'in AC₁ katsayısı aşağıdaki gibi tanımlanır (Gwet, 2008a'dan aktaran AGREESTAT, 2015):

$$AC_1 = \frac{p_a - p_e}{1 - p_e}$$

Burada, p_a aşağıdaki gibi tanımlanır:

$$p_a = \frac{1}{n'} \sum_{i=1}^{n'} p_{a|i}, \quad p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(\bar{r}_{ik+}-1)}{r_i(r_i-1)}, \quad \bar{r}_{ik+} = \sum_{l=1}^q w_{kl} r_{il}$$

Burada n' iki ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısı olarak tanımlanmaktadır. Şansla uzlaşma olasılığı p_e şöyle tanımlanır:

$$p_e' = \frac{T_w}{q(q-1)} \sum_{k=1}^q \pi_k (1 - \pi_k), \text{ burada } T_w = \sum_{k,l} w_{kl}$$

Gwet'in AC₁ katsayısının standart hatası aşağıda tanımlanan varyansının kareköküdür:

$$v(AC_1) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n (AC_1^* - AC_1)^2$$

Burada, $AC_1^* = AC_{1|i} - 2(1 - AC_1) \frac{p_{e|i} - p_e}{1 - p_e}$ şeklindedir.

$$AC_{1|i} = \begin{cases} \frac{\binom{n}{n'} (p_{a|i} - p_e)}{1 - p_e}, & \text{eğer } r_i \geq 2 \\ 0, & \text{diğer durumlarda} \end{cases}$$

$$p_{e|i} = \frac{T_w}{q(q-1)} \sum_{k=1}^q \frac{r_{ik}}{r_i} (1 - \pi_k)$$

Gwet'in AC₁ katsayısı üç puanlayıcı için incelendiğinde genel formülde iki puanlayıcı için olana göre bir değişiklik görülmemektedir. Diğer bir deyişle üç puanlayıcı için de Gwet'in katsayısı hesaplanırken puanlayıcıların uzlaştığı birimlerin oranından şansla uzlaşılan

birimlerin oranı çıkarılıp elde edilen değerin şansla uzlaşma olmama olasılığına bölümünden elde edildiği görülmektedir. Ancak üç puanlayıcı için puanlayıcıların uzlaştığı birimlerin oranının ve şansla uzlaşılan birimlerin oranının hesaplanması iki puanlayıcıya göre farklılaşmaktadır. Gwet'in AC_1/AC_2 katsayılarının hesaplanışında iki puanlayıcı için kategorilere ve puanlayıcılara göre ölçme objelerinin dağılımına ait çapraz tablo kullanılırken; üç puanlayıcı için ölçme objelerine ve kategorilere göre puanlayıcıların dağılımını içeren çapraz tablo kullanılır.

Scott'nin Pi

Şans uzlaşmasının rolü için düzeltme yapan bu istatistik iki kodlayıcı için birleşik dağılım kullanır. Bu katsayı hesaplanırken kategorilerin sayısı değil bu kategorilerin kodlayıcılar tarafından nasıl kullanıldığına bakılır. Bu istatistiğin aralığı 0,00 ile 1,00 arasında değişmektedir. 0,00 şans düzeyinde uzlaşma; 1,00 ise mükemmel uzlaşmayı ifade etmektedir (Scott, 1955'den aktaran Neuendorf, 2002; s. 150). Scott's pi katsayısı ile Cohen's kappa katsayısı aynı kavramsal formülden türemektedir (Zwick, 1988'den aktaran Neuendorf, 2002; s.151).

$$pi ya da \kappa = \frac{p_o - p_e}{1 - p_e}$$

p_o = puanlayıcıların uzlaştığı birimlerin oranı

p_e = uzlaşmanın şans yoluyla beklendiği birimlerin oranı

AgreeStat 2015.2 programında iki puanlayıcı için Scott'nin pi katsayısı (Scott, 1955'ten aktaran AGREESTAT, 2015) aşağıdaki gibi tanımlanmaktadır:

$$P_i = \frac{p_a - p_e}{1 - p_e}$$

$$p_a = \sum_{k,l} w_{kl} p_{kl} \text{ ve } p_e = \sum_{k,l} w_{kl} \pi_k \pi_l$$

p_i 'nin varyansı ise aşağıdaki şekilde hesaplanır (AGREESTAT, 2015):

$$v(P_i) = \frac{1 - f}{n(1 - p_{e\pi})^2} \left\{ \sum_{k,l}^q p_{kl} [w_{kl} - (1 - P_i)(\bar{p}_k + \bar{p}_l)]^2 - [p_a^* - 2(1 - P_i)p_{e\pi}^*]^2 \right\}$$

Yukarıdaki formüller incelendiğinde Scott'nin pi katsayısı ile Cohen'nin kappa katsayısının hesaplamasının çok benzer olduğu görülmektedir. Sadece şansla uzlaşma oranını hesaplama

çok küçük bir fark olarak Cohen'nin kappa değeri marjinal olasılıkların çarpımını kullanırken Scott'nin pi katsayısı marjinal olasılıklar yardımıyla tanımlanan yeni bir oranı kullanmaktadır. Scott'in pi katsayısında da Cohen'nin kappa katsayısında olduğu gibi toplam içindeki her bir oran için farklı ağırlıklandırma kullanılmaktadır. Gwet'in AC₂ katsayısında ise çapraz tablodaki ağırlıklandırmaların toplamının kategori sayısına dayalı elde edilen bir değere bölünmesiyle elde edilen sabit bir ağırlıklandırma kullanılmaktadır.

Krippendorff'un Alfa

Krippendorff'un alfa katsayısı verinin büyük bir çeşidine uygulanabilen değişmeyen güvenilirlik standartlarına izin verir. Krippendorff'un alfa katsayısı ile aşağıdaki işlemler yapılabilir (Krippendorff, 2004; s.222):

- Her değişkende değerlerin herhangi bir sayısıyla kullanılabilir. Şans için düzeltmesini bu sayıdan bağımsız olarak oluşturur.
- Gözlemlerin herhangi bir sayısı ile kullanılabilir, sadece geleneksel olarak kullanılan iki gözlem ile değil.
- Benzer şekilde küçük ve büyük örneklemeler ile kullanılabilir. Güvenilir verinin çeşitlenen miktarı için kendi kendine düzeltme yapar.
- Çeşitli ölçek türleriyle (sınıflama, sıralama, aralıklı, oran) uygulanabilir.
- Bazı gözlemcilerin tüm kaydedilen birimler için katılmadığı kayıp değerlerin bulunduğu veriye uygulanabilir.

Krippendorff'un alfa katsayısı en genel şekliyle şöyle tanımlanır (Krippendorff, 2004; s.222):

$$\alpha = 1 - \frac{D_o}{D_e}$$

D_o gözlemlenen uzlaşmanın ölçümünü ve D_e şansın etkili olduğu durumda beklenebilen uzlaşmanın ölçümünü göstermektedir.

Yukarıda verilen formül Krippendorff'un alfa katsayısının hesaplanmasına ait en genel denklemi göstermektedir. Farklı değişken etkilerine göre D_o ve D_e yi hesaplama için gereken matematiksel işlemler farklılaşarak formüller daha karmaşık bir duruma gelmektedir.

Krippendorff'un alfa katsayısı mükemmel uzlaşma görüldüğünde 1 değerini alır. Alfa katsayısının örnekleme hatası ve sistematik uzlaşmazlık gibi nedenlerden dolayı negatif değer alması mümkündür. Bu nedenle alfa katsayısının limitleri $1 \geq \alpha \geq 0$ ($\pm$ örnekleme hatası; - sistematik uzlaşmama) şeklinde alınabilir (Krippendorff, 2004; s.222).

AgreeStat 2015.2 programında iki puanlayıcı için Krippendorff'un alfa katsayısı aşağıdaki gibi tanımlanır (AGREESTAT, 2015):

$$\alpha = \frac{p'_a - p_e}{1 - p_e}$$

$$p'_a = (1 - \varepsilon_n)p_a + \varepsilon_n \text{ ve } \varepsilon_n = \frac{1}{2n}$$

$$p_a = \sum_{k,l} w_{kl} p_{kl} \text{ ve } p_e = \sum_{k=1}^q \sum_{l=1}^q w_{kl} \pi_k \pi_l$$

n = ölçme objesi örneklem büyüklüğü (the subject sample size)

$$v(\alpha) = \frac{1 - f}{n(1 - p_e)^2} \left\{ \sum_{k,l} p_{kl} [(1 - \varepsilon_n)w_{kl} - (1 - \alpha)(\bar{p}_k + \bar{p}_l)]^2 - [(1 - \varepsilon_n)p_a - 2(1 - \alpha)p_e]^2 \right\}$$

Yukarıdaki formüller incelendiğinde Krippendorff'un alfa katsayısı puanlayıcılar arası uzlaşmayı Cohen'nin kappa, Gwet'in AC₂ ve Scott'nin pi katsayılarından farklı olarak ölçme objesi sayısını dikkate alarak hesaplanır. Krippendorff'un alfa katsayısı şansla düzeltme olmadan hesaplanan puanlayıcılar arası uzlaşmayı hesaplarken Cohen'nin kappa, Gwet'in AC₂ ve Scott'nin pi katsayılarının kullandığı hesaplama ölçme objesinin sayısından türettiği bir katsayı ile çarpılmaktadır. Ayrıca Scott'nin pi katsayısı ile Krippendorff'un alfa katsayısında şansla uzlaşma oranını ise aynı şekilde hesaplanmaktadır. Bu durum ve Krippendorff'un alfa katsayısındaki puanlayıcılar arasındaki uzlaşma oranındaki beklenen değer dikkate alındığında n sayısı (ölçme objesi sayısı) arttıkça Scott'nin pi katsayısı ile Krippendorff'un alfa katsayılarının benzer sonuçlar vereceği söylenebilir.

Üç ya da daha fazla puanlayıcı için Krippendorff'un alfa katsayısı şu şekilde hesaplanır (Krippendorff, 1980 ve Gwet, 2011'den aktaran AGREESTAT, 2015):

$$\alpha = \frac{p_a - p_e}{1 - p_e}$$

Burada, $p_a = (1 - \varepsilon_n)p'_a + \varepsilon_n$, $\varepsilon_n = \frac{1}{n' \bar{r}}$

n' , iki ya da daha fazla gözlemci tarafından puanlanan ölçme objelerinin sayısı (bir ya da daha az gözlemci tarafından puanlanan tüm ölçme objeleri dışarıda bırakılır). p'_a ise aşağıdaki gibi tanımlanır:

$$p'_a = \frac{1}{n'} \sum_{k=1}^{n'} p_{a|i}, \quad p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(\bar{r}_{ik+}-1)}{\bar{r}(r_i-1)}, \quad \bar{r}_{ik+} = \sum_{l=1}^q w_{kl} r_{il}$$

Şansla uzlaşma olasılığı p_e şöyle tanımlanır:

$$p_e = \sum_{k,l} w_{kl} \pi_k \pi_l, \text{ burada } \pi_k = \frac{1}{n'} \sum_{i=1}^{n'} \frac{r_{ik}}{\bar{r}} \text{ dir.}$$

Dikkat edelim ki $\bar{r}$ sadece $r_i \geq 2$ için r_i 'nin ortalamasıdır ve α ile ilişkili π_k Fleiss'nin kappa ve Gwet'in AC₁'den farklıdır; ancak çıktı kayıp değer olmadığı durumda aynıdır. Krippendorff'un alfanın varyansı aşağıdaki gibi tanımlanır (AGREESTAT, 2015):

$$v(\alpha) = \frac{1-f}{n} \frac{1}{n-1} \sum_{i=1}^n (\alpha_i^* - \alpha)^2$$

Burada, $\alpha_i^* = \alpha_i - (1 - \alpha)(p_{e|i} - p_e)/(1 - p_e)$ ve $\alpha_i = (p_{a\varepsilon_n|i} - p_e)/(1 - p_e)$ şeklindedir.

$$p_{a\varepsilon_n|i} = (1 - \varepsilon) \left[p_{a|i} - \frac{p_a(r_i - \bar{r})}{\bar{r}} \right] + \varepsilon_n, \text{ burada } p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(\bar{r}_{ik+}-1)}{\bar{r}(r_i-1)}, \quad p_a = \frac{1}{n} \sum_{i=1}^n p_{a|i},$$

$$p_{e|i} = \sum_{k=1}^q \bar{\pi}_k \frac{r_{ik}}{\bar{r}} - \frac{\pi_k(r_i - \bar{r})}{\bar{r}}, \quad \bar{\pi}_k = \frac{\bar{\pi}_{k+} + \bar{\pi}_{+k}}{2}, \quad \bar{\pi}_{k+} = \sum_{l=1}^q w_{kl} \pi_l \quad \text{ve}$$

$$\bar{\pi}_{+l} = \sum_{k=1}^q w_{kl} \pi_k$$

Krippendorff'un alfa katsayısının üç puanlayıcı için hesaplanış şekli Gwet'in AC₁/AC₂ katsayısının hesaplanışı gibi iki puanlayıcının hesaplandığı durumdaki genel ifadeye dayanır. Ancak, üç puanlayıcı için puanlayıcıların uzlaştığı birimlerin oranı ve şansla uzlaşılan birimlerin oranının hesaplanışında iki puanlayıcıya göre farklılaşmaktadır. Üç puanlayıcı için Krippendorff alfa katsayısı ile Gwet'in AC₁/AC₂ katsayısının hesaplanışları karşılaştırıldığında ise hem puanlayıcıların uzlaştığı birimlerin hesaplanışında hem de şansla uzlaşılan birimlerin oranında farklılaşma olduğu görülmektedir. Gwet'in AC₁/AC₂ katsayısında puanlayıcıların uzlaştığı birimler hesaplanırken iki ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısının (n') çarpma işlemi göre tersini kullanırken; Krippendorff'un alfa katsayısı bu değere belirli bir ölçme objesi için belirli bir kategoriye atayanların sayısının ikiden büyük olduğu değerlerin ortalamasının ($\bar{r}$) çarpma işlemine göre tersiyle çarpılarak elde edilen yeni bir katsayıyı (ε_n) kullanmaktadır. Aslında her iki katsayı için n' değeri büyüdükçe kullanılan değer sıfıra yaklaşacağı için

puanlayıcıların uzlaştığı birimlerin oranı birbirine yaklaşacaktır. Ancak bu durum Krippendorff'un alfa katsayısı ile Gwet'in AC_1/AC_2 katsayısının benzer sonuçlar vereceği söylenemez; çünkü uzlaşmanın şansla yoluyla beklendiği birimlerin oranının hesaplanışında da farklılaşma bulunmaktadır. Krippendorff'un alfa katsayısında uzlaşmanın şansla yoluyla beklendiği birimlerin oranı hesaplanırken her bir birim için farklı bir ağırlıklandırma belirlenerek çarpılırken; Gwet'in AC_1/AC_2 katsayısında her bir için elde edilen ağırlıklandırmaların toplamı ve kategori sayısına dayalı olarak elde edileb bir değer hesaplanarak birimlerin toplamı ile çarpılır. Bu durumun Krippendorff'un alfa katsayısı ile Gwet'in AC_1/AC_2 katsayılarının alacağı değerleri farklılaştıracağı söylenebilir.

Fleiss'nin Kappa

Fleiss (1971) kappa ve ağırlıklandırılmış kappa katsayılarının sadece iki puanlayıcıyla sınırlı olduğunu ve bundan dolayı kappa katsayısının ikiden fazla puanlayıcı için genelleştirilmesi gerekliliğinden bahsetmiştir.

Fleiss'nin kappa katsayısı aşağıdaki gibi hesaplanır:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$\bar{P} - \bar{P}_e$ = aslında şansın artmasıyla ulaşılan uzlaşmanın derecesi

$1 - \bar{P}_e$ = şansla tahmin edilenden ayrıca ulaşılan uzlaşmanın derecesi

AgreeStat 2015.2 programında üç ya da daha fazla puanlayıcı için Fleiss'nin ağırlıklandırılmış kappa katsayısı aşağıdaki gibi tanımlanır (Gwet, 2014'den aktaran AGREESTAT, 2015):

$$\hat{\kappa}_F = \frac{p_a - p_e}{1 - p_e}$$

Burada p_a aşağıdaki gibi tanımlanır:

$$p_a = \frac{1}{n'} \sum_{i=1}^{n'} p_{a|i}, \quad p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_i - 1)}, \quad r_{ik}^* = \sum_{l=1}^q w_{kl} r_{il}$$

Burada n' , 2 ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısıdır.

Şans yoluyla uzlaşmanın oranı p_e şöyle verilir:

$$p_e' = \sum_{k,l} w_{kl} \pi_k \pi_l \text{ burada } \pi_k = \frac{1}{n} \sum_{i=1}^n \frac{r_{ik}}{r_i}$$

Fleiss'nin kappa katsayısının standart hatası aşağıda tanımlanan varyansının kareköküdür (Gwet 2008a'dan aktaran AGREESTAT, 2015):

$$v(\hat{\kappa}_F) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\kappa_{F|i}^* - \kappa_F)^2$$

Burada

$$\kappa_{F|i}^* = \kappa_{F|i} - 2(1 - v_\pi) \frac{p_{e|i} - p_e}{1 - p_e}$$

$$\kappa_{F|i} = \begin{cases} \frac{\left(\frac{n}{n'}\right)(p_{a|i} - p_e)}{1 - p_e}, & \text{eğer } r_i \geq 2 \\ 0 & , \text{diğer durumlarda} \end{cases}$$

$$p_{e|i} = \sum_{k=1}^q \pi_k^* r_{ik} / r_i \quad \text{ile} \quad \pi_k^* = \sum_{l=1}^q w_{kl} \pi_l$$

Fleiss'nin kappa katsayısının üç puanlayıcı için hesaplanış şekli Gwet'in AC₁/AC₂ ve Krippendorff'un alfa katsayısının hesaplanışında olduğu gibi iki puanlayıcının hesaplandığı durumdaki genel ifadeye dayanır. Fleiss'nin kappa katsayısının hesaplanışını Gwet'in AC₁/AC₂ ve Krippendorff'un alfa katsayılarının hesaplanışı ile karşılaştırdığımızda puanlayıcıların uzlaştığı birimlerin oranı Gwet'in AC₁/AC₂ katsayısı ile Fleiss'nin kappa katsayısı için aynı şekilde hesaplanmaktadır. Fleiss'nin kappa katsayısı ile Krippendorff'un alfa katsayısının da ise şans yoluyla beklendiği birimlerin oranının hesaplanışının çok benzediği söylenebilir. Bu iki katsayı için uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanışındaki tek farkın, Fleiss'nin kappa katsayısında uzlaşmanın şans yoluyla beklendiği birimlerin oranı hesaplanırken ölçme objelerinin sayısı (n) dikkate alınırken; Krippendorff'un alfa katsayısında iki ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısı (n') kullanılması olduğu söylenebilir. Bu durumda n ve n' değerleri birbirine yaklaştıkça Fleiss'nin kappa değeri ile Krippendorff'un alfa değeri için uzlaşmanın şans yoluyla beklendiği birimlerin oranı neredeyse aynı değeri alacağı söylenebilir. Bir önceki alt başlıkta üç ve daha fazla puanlayıcı için Krippendorff'un alfa katsayı ile Gwet'in AC₁/AC₂ katsayıları için yapılan karşılaştırmada dikkate alındığında formüllerine dayalı olarak Fleiss'nin kappa katsayısı ile Krippendorff'un alfa katsayısının benzer sonuçlar vereceği ancak Fleiss'nin kappa ve Gwet'in AC₁/AC₂ katsayılarının farklılaşacağı söylenebilir.

Conger'in Kappa

Conger Fleiss'in açıklamasının içine yerleştirilmiş olduğuna inanılan, puanlayıcı etkisinin ortadan kaldırıldığı şanstan ötürü uzlaşmanın oranını oluşturdu. Conger's şans uzlaşma istatistiğinin formülü şu şekildedir (Dates ve King, 2008):

$$p_{eC} = p_{e\pi} - \sum_{q=1}^q s_{rq}^2 / (r-1)$$

p_{eC} = Şans uzlaşması için Conger'in tanımı

$p_{e\pi}$ = Şans uzlaşması için Fleiss'in tanımı

$$s_{rq}^2 = \left[r \sum_{r=1}^r (n_{rq})^2 - (\sum_{r=1}^1 n_{rq})^2 \right] / n^2 r^2$$

n_{rq} = puanlayıcı r tarafından kategori q da yer alan konuların sayısı

r = puanlayıcıların sayısı

AgreeStat 2015.2 programında üç ya da daha fazla puanlayıcı için Conger'in kappa katsayısı aşağıdaki gibi tanımlanır (Conger, 1980 ve Gwet, 2010'dan aktaran AGREESTAT, 2015):

$$\kappa_C = \frac{p_a - p_e}{1 - p_e}$$

Burada p_a aşağıdaki gibi tanımlanır:

$$p_a = \frac{1}{n'} \sum_{i=1}^{n'} p_{a|i}, \quad p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(\bar{r}_{ik+}-1)}{r_i(r_i-1)}, \quad \bar{r}_{ik+} = \sum_{l=1}^q w_{kl} r_{il}$$

Burada n' , 2 ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısıdır.

Şansla uzlaşmanın oranı p_e şöyle verilir:

$$p'_e = \sum_{k,l} w_{kl} (\bar{p}_{+k} \bar{p}_{+l} - s_{kl}/r) \text{ burada } \bar{p}_{+k} = \frac{1}{r} \sum_{g=1}^r p_{gk}$$

$$p_{gk} = \frac{n_{gk}}{n_g} \text{ ve } s_{kl} = \frac{1}{r-1} (\sum_{g=1}^r p_{gk} p_{gl} - r \bar{p}_{+k} \bar{p}_{+l})$$

Conger'in kappa katsayısının standart hatası aşağıda tanımlanan varyansının kareköküdür (AGREESTAT, 2015):

$$v(\kappa_C) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\kappa_C^{*(i)} - \kappa_C)^2$$

Burada

$$\kappa_C^{*(i)} = \kappa_C^{(i)} - 2(1 - \kappa_C) \frac{p_{e|i} - p_e}{1 - p_e}$$

$$\kappa_C^{(i)} = \begin{cases} \frac{\left(\frac{n}{n'}\right)(p_{a|i}-p_e)}{1-p_e}, & \text{eğer } r_i \geq 2 \\ 0, & \text{diğer durumlarda} \end{cases}$$

$p_{e|i} = \frac{1}{r(r-1)} \sum_{g=1}^r \sum_{k=1}^q \bar{w}_{gk}^{(i)} (r\bar{p}_{+k} - p_{gk})$ ve $\bar{w}_{gk}^{(i)} = \sum_{l=1}^q (w_{kl} + w_{lk}) \delta_{gl}^{(i)}$; $\delta_{gl}^{(i)} = 1$ olarak alınır eğer puanlayıcı g kategori l içinde ölçme objesi i'yi sınıflandırdıysa, aksi takdirde sıfır olarak alınır.

Yukarıdaki formüller incelendiğinde Conger'in kappa katsayısı için puanlayıcıların uzlaştığı birimlerin oranının hesaplanışının Gwet'in AC₁/AC₂ ve Fleiss'nin kappa katsayıları için aynı olduğu görülmektedir. Ancak Conger'in kappa katsayı için uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanışı Gwet'in AC₁/AC₂, Krippendorff'un alfa ve Fleiss'nin kappa katsayılarına göre farklılaşmaktadır. Conger'in kappa katsayısında uzlaşmanın şans yoluyla beklendiği birimlerin oranı hesaplanırken sadece marjinal olasılıklar dikkate alınmaz aynı zamanda birleşik olasılıklar da dikkate alınmaktadır. Diğer üç katsayıda ise uzlaşmanın şans yoluyla beklendiği birimlerin oranı hesaplanırken sadece marjinal olasılık dikkate alınır. Ancak Conger'in kappa katsayısında uzlaşmanın şans yoluyla beklendiği birimlerin oranı hesaplanırken ağırlıklandırmaları Krippendorff'un alfa ve Fleiss'nin kappadaki gibi her bir birim için Gwet'in AC₁/AC₂ katsayısından farklı olarak değişik ağırlıklandırmalar almaktadır. Formüllere dayalı inceleme yapıldığında Conger'in kappa katsayısı ile Krippendorff'un alfa ve Fleiss'nin kappa katsayılarının benzer sonuçlar vereceği söylenebilir.

Brennan-Prediger

Brennan-Prediger katsayısının ağırlıklandırılmamış versiyonu Brennan & Prediger (1981) tarafından; ağırlıklandırılmış versiyonu Gwet (2010) tarafından önerilmiştir (Brennan & Prediger, 1981'den ve Gwet, 2010'dan aktaran AGREESTAT, 2015):

AgreeStat 2015.2 programında iki puanlayıcı için Brennan-Prediger katsayısı aşağıdaki gibi tanımlanır (AGREESTAT, 2015):

$$BP = \frac{p_a - p_e}{1 - p_e}$$

$$p_a = \sum_{k,l} w_{kl} p_{kl}$$

$$p_e = \frac{1}{q^2} \sum_{k,l} w_{kl}$$

BP'nin varyansı ise şu şekilde hesaplanır:

$$v(BP) = \frac{1-f}{n(1-p_e)^2} \left(\sum_{k,l}^q w_{kl}^2 p_{kl} - p_a^2 \right)$$

Yukarıdaki formüller incelendiğinde Brennan-Prediger katsayısının hesaplanması ile Gwet'in AC₂'nin katsayısına çok benzemektedir. Puanlayıcıların uzlaştığı birimlerin oranı Brennan-Prediger ile Cohen'nin kappa, Gwet'in AC₂, Scott'nin pi için aynı şekilde hesaplanmaktadır. Ancak uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanışına bakıldığında Brennan-Prediger katsayısının sadece Gwet'in AC₂ katsayısı ile benzerlik göstermektedir. Her iki katsayıda uzlaşmanın şans yoluyla olduğu birimlerin oranını hesaplarken payda kısmında kategori sayısına yönelik işlem ve pay kısmında ise toplam kategoriler ile ilişkili tüm ağırlıkların toplamını kullanan bir işlemi dikkate almıştır. Bu nedenle Brennan-Prediger ile Gwet'in AC₂ katsayısının benzer sonuçlar verebileceği düşünülebilir.

AgreeStat 2015.2 programında üç ya da daha fazla puanlayıcı için Brennan & Prediger katsayısı aşağıdaki gibi tanımlanır (Brennan & Prediger, 1981'den aktaran AGREESTAT, 2015):

$$BP = \frac{p_a - p_e}{1 - p_e}$$

Burada p_a aşağıdaki gibi tanımlanır:

$$p_a = \frac{1}{n'} \sum_{i=1}^{n'} p_{a|i}, \quad p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(\bar{r}_{ik+}-1)}{r_i(r_i-1)}, \quad \bar{r}_{ik+} = \sum_{l=1}^q w_{kl} r_{il}$$

Burada n' , 2 ya da daha fazla puanlayıcı tarafından puanlanan ölçme objelerinin sayısıdır.

Şans yoluyla uzlaşmanın oranı p_e şöyle verilir:

$$p_e = \frac{1}{q^2} \sum_{k,l} w_{kl}$$

BP katsayısının standart hatası aşağıda tanımlanan varyansının kareköküdür (Gwet 2008a'dan aktaran AGREESTAT, 2015):

$$v(BP) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (BP_i - BP)^2$$

Burada

$$BP_i = \begin{cases} \frac{\left(\frac{n}{n'}\right)(p_{a|i}-p_e)}{1-p_e}, & \text{eğer } r_i \geq 2 \\ 0 & , \text{diğer durumlarda} \end{cases}$$

Üç ya da daha fazla puanlayıcı için Brennan-Prediger katsayısı Gwet'in AC_1/AC_2 , Krippendorff'un alfa, Fleiss'nin kappa ve Conger'in kappa katsayıları ile karşılaştırıldığında puanlayıcıların uzlaştığı birimlerin oranının hesaplanması Gwet'in AC_1/AC_2 , Fleiss'nin kappa ve Conger'in kappa katsayıları ile aynıdır. Ancak uzlaşmanın şans yoluyla beklendiği birimlerin oranı açısından Gwet'in AC_1/AC_2 katsayısındaki benzer şekilde kategori sayısı ve her birim için ağırlıklandırmaların toplamını (T_w) dikkate alınarak hesaplanmaktadır. Bu iki katsayı için uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanışındaki fark ise Brennan-Prediger katsayısının Gwet'in AC_1/AC_2 katsayısından farklı olarak marjinal olasılıkları dikkate almamasıdır.

Uzlaşma Yüzdesi

Uzlaşma sayısının, ölçümlerin toplamına bölümünü temsil eden basit bir yüzdendir. Bu istatistiğin değeri 0,00 ile 1,00 arasında değişmektedir. 0,00 olması uzlaşmanın olmadığını, 1,00 olması ise mükemmel uzlaşma olduğunu gösterir. Uzlaşma yüzdesinin kavramsal formülü şu şekildedir (Neuendorf, 2002; s. 149):

$$PA_o = \frac{A}{n}$$

PA_o = gözlemlenen uzlaşma yüzdesi

A = iki kodlayıcı arasında uzlaşmaların sayısı

N = iki kodlayıcı tarafından puanlanan birimlerin sayısı

AGREESTAT (2015)'de ağırlıklandırılmış uzlaşma yüzdesi şu şekilde tanımlanmıştır:

$$p_a = \sum_{k,l} w_{kl} p_{kl}$$

w_{kl} sembolü k ve l kategorileri ile ilişkili ağırlıkları belirtir. p_{kl} puanlayıcıya ve kategoriye göre konuların dağılımı tablosundaki l. sütun ve k. satıra ait birleşik olasılık olarak belirtilmiştir.

Uzlaşma yüzdesinin varyansı ise şu şekilde hesaplanır:

$$v(p_a) = \frac{1-f}{n} \left(\sum_{k,l}^q w_{kl}^2 p_{kl} - p_a^2 \right)$$

Uzlaşma yüzdesi hesaplanırken uzlaşmanın şans yoluyla olduğu birimler dikkate alınmamaktadır. Cohen'nin kappa, Gwet'nin AC₂, Scott'nin pi ve Brennan-Prediger için de puanlayıcıların uzlaştığı birimlerin oranı uzlaşma yüzdesi ile aynı şekilde hesaplanmaktadır. Ancak Cohen'nin kappa, Gwet'nin AC₂ ve Scott'nin pi katsayıları hesaplamalarda uzlaşmanın şans yoluyla belirlenmesine yönelik bir oran hesaplayarak düzeltme yapmaktadır. Bu nedenle uzlaşma katsayısının diğer katsayılardan daha yüksek değer verebileceği söylenebilir.

Sınıf İçi Korelasyon (ICC)

Sınıf içi korelasyon katsayısı homojenliği ölçmek için kullanılan alternatif bir istatistiksel tekniktir. Sadece ölçümlerin bir çifti için değil aynı zamanda ölçümlerin geniş bir seti içinde kullanılabilir (McGraw ve Wong, 1996).

Sınıf içi korelasyon katsayılarının farklı modelleri bulunmaktadır. Uygun sınıf içi korelasyon katsayısına belirlemede karar verilmesi gerek üç durum bulunmaktadır: (a) bir-yönlü varyans analizi (ANOVA) ya da iki yönlü varyans analizi güvenilirlik çalışmasının analizi için uygun mu? (b) yargılayıcıların ortalama puanları arasında fark ilgilenilen konunun güvenilirliğiyle ilişkili mi? (c) analiz birimleri bireysel mi puanlanmış ya da birkaç puanlamanın ortalaması mı? Birinci ve ikinci karar güvenilirlik çalışması için uygun modelin belirlenmesiyle; ikinci ve üçüncü karar sonuçların olası kullanımı ile ilgilidir (Shrout ve Fleiss, 1979).

Tipik bir puanlayıcı güvenilirliği çalışmasında n hedefin randum bir örnekleminin her biri k yargılayıcı tarafından bağımsız olarak puanlandırılır. Çalışmaların bu türü için üç farklı durum tanımlanabilir. Durum 1: her bir hedef, yargılayıcıların daha geniş bir evreninden randum olarak seçilen k yargılayıcının farklı bir kümesi tarafından derecelendirilir. Durum 2: k yargılayıcının bir randum örnekleme daha geniş bir evrenden seçilir ve her bir yargılayıcı her bir hedefi puanlandırır. Böylece her bir yargılayıcı n hedefi birlikte puanlandırır. Durum 3: her bir hedef sadece konunun yargılayıcılarından olan aynı k yargılayıcının her biri tarafından puanlandırılır. Çalışmanın bu durumlardan hangisine dayalı olduğuna bağlı olarak, çalışmanın modeli değişmektedir. Seçilen duruma göre tahmin edilebilir etkiler

hakkında farklı varsayımlar yapılır ve yapıyla ilişkili olarak ANOVA farklılaşacaktır. Durum 1 altında yargılayıcılardan dolayı yargılayıcı ve hedef arasındaki etkileşimin etkisi ve random hataların etkisi ayrılamaz. Teknik olarak bir-yönlü random etkiler modeli durum 1 altında görülür. Durum 2 için teknik olarak iki-yönlü random etkiler modelinin kullanımı ve durum 3 için teknik olarak iki yönlü karma etkiler modelinin kullanımı uygundur (Shrout ve Fleiss, 1979).

İlgili Araştırmalar

Güler (2008) performansın ölçülmesinde kullanılan maddelerin birden fazla puanlayıcı tarafından bütünsel DPA ile puanlanmasıyla elde edilen puanlar üzerinden klasik test kuramına, genellenebilirlik kuramına ve çok değişkenli Rasch ölçme modeline göre parametre kestirimleri yapıp, bu farklı yaklaşımlara göre karşılaştırmıştır. Ayrıca elde edilen puanların güvenilirliği farklı kuramlara göre incelenmiştir. Araştırmanın çalışma grubu 2007 yılı bahar döneminde öğrenim gören 203 sekizinci ve dokuzuncu sınıf öğrencisinden oluşmaktadır. Veriler 1999 yılında Uluslararası Matematik ve Bilim Çalışması Eğitimi (TIMSS-The Trends in International Mathematics and Science Study) uygulamasında yer alan matematik performansını belirlemek için kullanılan 24 madde aracılığı ile toplanmıştır. Toplanan veriler dört farklı puanlayıcı tarafından bağımsız olarak puanlamıştır. Güvenirlik için klasik test kuramına göre Cronbach alfa güvenirlilik katsayısı, puanlayıcılar arasında uyumun belirlenmesinde Kedall'ın konkordans katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcıların verdikleri puanların ortalamaları arası fark olup olmadığını belirlemek için F testi kullanılmıştır. Genellenebilirlik kuramında bxqxp tümüyle çaprazlanmış desen kullanılarak genellenebilirlik ve güvenirlilik katsayıları hesaplanmıştır. Çok değişkenli Rasch modeli ile birey, puanlayıcı ve madde boyutlarına ilişkin ayrı ayrı güvenirlilik hesaplamaları yapılmıştır. Araştırma sonucuna göre matematik başarısının ölçülmesinde güvenirliliğin belirlenmesinde en az iki kuramdan yararlanmanın daha uygun olacağı sonucuna varılmıştır.

Gwet (2008a) π (π) ve kappa (κ) istatistikleri nominal olarak ölçeklenmiş veri üzerinden puanlayıcılar arasındaki uzlaşmanın oranını hesaplamak için psikiyatri alanında ve psikolojik testlerde yaygın olarak kullanıldığını belirtmiştir. Ayrıca bu katsayıların kappanın çelişkisi olarak bilinen durumlarda bazen beklenmedik sonuçlar ürettiğinden bahsetmiştir. Bu nedenle çalışmada bu sınırlılıkların kökeni keşfetmeyi ve AC_1 katsayısı olarak

adlandırılan bir alternatif ve daha kararlı uzlaşma katsayıları tanımlamayı amaçlar. Aynı zamanda puanlayıcılar arasında bağımsızlık hipotezine dayanmayan geçerliğiyle çoklu-puanlayıcılar genelleştirilmiş π ve AC_1 istatistikleri için yeni varyans kestirimleri amaçlanmaktadır. Bu durum bağımsızlık varsayımına dayanan var olan alternatif varyanslar üzerinde bir gelişme olarak görünmektedir. Çalışma kapsamında bir monte-carlo simülasyon çalışmasıyla güven aralığı anlamı için bu varyans kestirimlerinin bir geçerliliği gösterilmiştir. Bu sayede araştırma sonucunda var olan puanlayıcılar arası güvenilirlik istatistikleri için bir geliştirilmiş alternatif olarak AC_1 in değerleri doğrulanmıştır.

Gwet (2008b) puanlayıcılar arası güvenilirlik katsayılarının varyansının hem konuların hem de puanlayıcıların örnekleme birleşik etkisinden kaynaklanabileceğinin görülebilmeye rağmen literatürde yapılan çalışmalarda sadece konu örnekleme değişkenliği için hesaplamaların yapıldığını fakat puanlayıcıların örnekleme birleşiminden kaynaklanacak değişkenliğin göz ardı edildiğine dikkat çekmiştir. Bu nedenle çalışmada hem konuların hem de puanlayıcıların evreninden çıkarım yapabilecek olası varyans kestirimlerinin tutarlılığı güven aralığı anlamı için geçerlik olarak kanıtlanır. Bu sonuçlar her bir puanlayıcının her bir konuyu puanladığı tam olarak çaprazlanmış desenler için sadece uygundur. Ayrıca küçük örneklem üzerindeki yaklaşımların büyük örneklem için doğruluğunu göstermek için monte-carlo simülasyon çalışması yapılmıştır ve büyük örneklem için de aynı yaklaşımların olduğu sergilenmiştir.

Kılıç (2009) çalışmasında değişik veri tipleri (kategorik ve sürekli) için geliştirilen uyum ölçütlerinin (uzlaşma katsayılarının) hangi durumlarda kullanıldığını göstermeyi amaçlamıştır. Ayrıca yaygın olarak kullanılmayan bazı ölçütlere dikkat çekmeye çalışmıştır. Çalışmada uyumu belirlemek için kullanılan ölçütler genel uyumluluk oranı, özel uyumluluk oranı, Kappa, Ağırlıklandırılmış Kappa, Kendall W, Goodman ve Kruskal γ , Kendal τ , sınıf içi korelasyon katsayısı ve Krippendorff alfa şeklindedir. Uyum ölçütlerine yönelik değerleri hesaplamak için literatürden alınan örnek veriler kullanılmıştır. Sonuç olarak, sıralı değerlendirme yapan çoklu değerlendirici için Kendall W'nin uyumu doğru olarak yansıtamadığı bu nedenle Krippendorff alfa değerinde hesaplanamaması ve Kendal W değeri ile karşılaştırılmasının doğru olacağı belirlenmiştir. Ölçek türlerinin hepsinde Krippendorff alfa eksik veri durumunda eksik değerlendirilmesi olan bireyleri hesaba kattığı için diğer ölçütlere göre daha avantajlı görülmüştür. Ek olarak, Goodman ve Kruskal γ , Kendall τ ,

Pearson ve Spearman korelasyon katsayılarının uyumluluğu doğru yansıtmama durumundan dolayı dikkatli kullanılmasının gerekliliği belirtilmiştir.

Yelboğa ve Tavşancıl (2010) farklı zamanlarda kullanılan bir iş performansı ölçeğinden elde edilen veriler üzerinden klasik test kuramı ve genellenebilirlik kuramına göre hesaplanan güvenilirlik katsayılarının düzeylerini ve birbiriyle olan ilişkilerini belirlemeyi amaçlamışlardır. Tarama türündeki çalışmalarında hizmet sektöründeki bir işletmenin birbirinden farklı on bir biriminde çalışan 176 personel ile çalışmışlardır. Çalışma kapsamında üç farklı değerlendirici 176'ı personeli tek tek değerlendirmiş, tek değerlendiricinin bulunduğu durumlar için sadece 2. değerlendiricinin değerlendirilmesi esas alınarak değerlendirme yapılmıştır. Araştırmanın sonuncunda aynı ölçme durumuna ait klasik test kuramı ve genellenebilirlik kuramı için elde edilen güvenilirlik katsayılarının birbiriyle uyumlu sonuçlar ürettiği tespit edilmiştir.

Kanık, Orekici Temel ve Ersöz Kaya (2010) Fleiss kappa ve Krippendorff alfa uyum katsayılarının örneklem genişliği, değerlendirici sayısı ve kullanılan ölçeğin kategori sayısından etkilenme durumlarını incelemiştir. Örneklem genişliği 30, 100 ve 1000 olduğu durumlar için; değerlendirici sayısının iki, beş ve yedi olduğu durumlar için ve kategori sayısının ikili, beşli, yedili ve onlu olduğu durumlar için incelemeler yapılmıştır. Örneklem büyüklüğünün 30, 100 ve 1000 olduğu durumların hepsinde sonuçların birbirine çok yakın olduğu gözlemlenmiştir. Bu durumda örneklem büyüklüğünün en az 30 olmasının Fleiss kappa ve Krippendorff alfa için uyum katsayılarında doğru parametre kestirimi için yeterli olduğu tespit edilmiştir. Ayrıca Krippendorff alfa katsayısının değerlendirici sayısı ve tanı testinin kategori sayısından etkilenerek beklenenden daha yüksek tahminlerde bulunduğu tespit edilmiştir.

Gwett (2011) yaptığı çalışmasında standart hatalar ve güven aralıklarıyla çeşitli uzlaşma katsayılarını hesaplamak için kullanılan AgreeStat 2011 programını kısaca tanıtmıştır. Ayrıca bu program puanlayıcı örnekleme üzerinden koşullu ya da koşulsuz olarak istatistiksel çıkarım yapmada kullanılabilir. Koşulsuz çıkarımda da hem konuların hem de puanlayıcıların evrenleri için analiz sonuçlarını yansıtmaya gibi bir avantaja sahip olunur. Çalışmada aynı zamanda literatürde herhangi diğer uzlaşma katsayılarının hesaplanmasına göre daha caydırıcı olmayan Krippendorff'un alfa katsayısının hesaplanmasında kullanılması beklenen temel bir yaklaşım amaçlanmıştır. Varyansının hesaplanması için amaçlanan yakın bir açıklamaya ek olarak, diğer bilinen katsayılarla daha karşılaştırılabilir

olacak şekilde bir alpha sunulmuştur. Cohen'in kappa ile ilgili çelişkili problemleri açılmaya ek olarak kayıp değerlerinde ele alındığı AC_1 ve AC_2 katsayıları ayrıca tanıtılmıştır. AC_2 için çeşitli veri türlerinde yer alan Krippendorff'un metrik farklılıklarına dayalı tüm ağırlıklandırmalar kullanılmıştır.

Bıkmaz (2011) puanlayıcı anahtarı türü ve puanlayıcı sayısı değişiminin puanlayıcı güvenilirliğini belirlemede kullanılan tekniklere göre elde edilen sonuçlardaki etkisini incelemeyi amaçlamıştır. Temel türden bir betimsel araştırma tekniği olan araştırmanın çalışma grubunu beşinci sınıfta okuyan 50 öğrenci ve puanlama yapan 10 öğretmen oluşturmaktadır. Fen ve Teknoloji dersine ait performans görevleri ile toplanan veriler 10 puanlayıcı tarafından ilk önce analitik puanlama anahtarı yardımıyla, aradan zaman geçtikten sonra ise bütünsel puanlama anahtarı ile puanlanmıştır. Analitik puanlama anahtarı kullanımında puanlayıcı sayısı değişiminin puanlayıcı güvenilirliğine etkisini belirlemek için kappa istatistiği, log lineer analiz tekniği ve Krippendorff alfa tekniği kullanılmış; bütünsel puanlama anahtarı için ise kappa istatistiği, Kendall uyum istatistiği ve Krippendorff alfa tekniği kullanılmış ve her iki durum içinde iki, beş ve on puanlayıcı arasındaki uyum incelenmiştir. Araştırmada kullanılan üç teknikten elde edilen analiz sonuçları her iki durum için güvenilir bulunmuştur. Diğer yandan hem analitik hem de bütünsel puanlama anahtarı kullanılarak toplanan puanların üç teknikle analizinden elde edilen sonuçlarda puanlayıcı sayısı arttıkça güvenirliliğin düştüğü, en yüksek güvenirlilik değerlerinin iki puanlayıcının kullanıldığında elde edildiği görülmüştür. Kullanılan tekniklerden kappa ve Krippendorff alfanın benzer sonuçlar verdiği görülmekle birlikte Krippendorff alfa tekniğinin puanlayıcı sayısı değişiminden kappaya göre daha az etkilendiği tespit edilmiştir.

Büyükkıdık (2012) sıradışı problem çözme becerilerinin değerlendirilmesi için hazırlanan performans görevlerinden elde edilen puanlar üzerinden Klasik Test Kuramı (KTK) ve Genellenebilirlik Kuramı ile puanlayıcılar arası güvenirlilik katsayılarını hesaplamış ve iki kuramın uygulanması sonucunda elde edilen katsayılar arası farklılaşma olup olmadığı ve hangi kuramın hangi yönlerden avantajlı olduğunu incelemiştir. Klasik test kuramı kapsamında sınıf içi ilişki katsayısı ve puanlayıcılar arası ilişki katsayısı hesaplanmıştır. Genellenebilirlik kuramında ise her iki puanlama anahtarı için $bxqxp$ ve anahtar çeşidinin değişkenlik kaynağı olarak alındığı $bxqxp_{xa}$ tümüyle çaprazlanmış desenleri kullanılarak analiz yapılmıştır. Araştırma sonucunda klasik test kuramından elde edilen katsayıların genellenebilirlik kuramına göre göreceli olarak daha düşük olduğu bulunmuştur.

Genellenebilirlik kuramı uygulamasıyla klasik test kuramına göre daha detaylı bilgiler sağlandığı görülmüştür. Güvenirlik açısından bakıldığında her iki kuramda elde edilen sonuçların güvenilirliğinin yüksek olduğunu desteklemiştir. Ancak genellenebilirlik kuramında kullanılan desenlerin güvenilirliği etkilediği belirlenmiştir.

Özder (2012) çalışmasında Kuzey Kıbrıs Türk Cumhuriyeti (KKTC)'nde öğretmen adaylarının seçiminde uygulanan mülakat sınavının güvenilirliğini incelemiştir. KKTC'de sınıf öğretmeni ve okulöncesi öğretmeni yetiştiren bir kurum olan Atatürk Öğretmen Akademisi öğrenci seçimini yazılı ve sözlü sınavla yapmaktadır. Yazılı sınavdan başarılı olan adayların alındığı sözlü sınavda, adayların öğretmenliğe yönelik tutum ve davranışları belirlenen beş kritere göre incelenmektedir. Sözlü mülakatın güvenilirliğini belirlemek için beş puanlayıcı tarafından yapılan puanlama sonucunda klasik test kuramına yönelik olarak Fleiss'in kappa değeri hesaplanmıştır. Genellenebilirlik kuramı için ise her bir hesaplamada bir puanlayıcının ve aynı zamanda toplam beş puanlayıcının güvenirlilik katsayıları incelenmiştir. Sonuç olarak Fleiss'in kappa değerine göre beş puanlayıcının yaptıkları puanlama güvenilirliğinin yüksek olduğu bulunmuştur. Genellenebilirlik kuramına göre de sonuç benzerdir. Ayrıca beş puanlayıcının bulunmasının puanlama güvenilirliğini arttırdığı sonucu ortaya konmuştur.

Wongpakaran, Wongpakaran, Wedding ve Gwet (2013) puanlayıcı uzlaşmasının önemli olduğu klinik araştırmalar için puanlayıcılar arası güvenirlilik çalışması yapmış ve Cohen'in Kappa ve Gwet'in AC_1 katsayılarını karşılaştırmıştır. Çalışma 67 hasta ve 9 puanlayıcı üzerinden yürütülmüştür. Cohen'in Kappa ve Gwet'in AC_1 katsayıları şans düzeltmeli uzlaşmanın düzeyini hesaplamak için farklı formüller kullanır. Araştırmanın sonucunda Gwet'in AC_1 katsayısının Cohen'in Kappa katsayısına göre daha kararlı bir güvenirlilik katsayısı olduğunu göstermiştir. Ayrıca Gwet'in AC_1 katsayısının Cohen'in kappaya göre yaygınlık (prevalence) ve marjinal olasılıklardan daha az etkilendiği bulunmuş ve böylece Gwet'in AC_1 katsayısının puanlayıcılar arası güvenirlilik analizleri ile kullanımının dikkate alınması gerekliliği görülmüştür.

Yıldıztekin (2014) çalışmasında uygulama sonucunda elde edilen puanlar üzerinden klasik test kuramı ve genellenebilirlik kuramına göre kestirimler yapmış ve puanlayıcılar arası tutarlılık derecesini incelemiştir. Klasik test kuramında kullanılan güvenirlilik analizleri Pearson momentler çarpımı korelasyon katsayısı, Spearman sıra farkları korelasyon katsayısı, cronbach alpha, kappa ve Krippendorff alpa katsayıları şeklindedir.

Genellenebilirlik kuramında ise bmxp deseninde deęişkenlik kaynakları ve yüzdeleri belirlenerek güvenilirlik analizi yapılmıştır. Sonuç olarak klasik test kuramı ve genellenebilirlik kuramına göre güvenilirlik katsayısı hesaplanırken birbirine paralel ve oldukça yüksek olduğu görülmüştür.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın türü, çalışma grubu, veri toplama araçları, verilerin toplanması ve analizi hakkında bilgi verilmiştir.

Araştırmanın Türü

Araştırma, klasik test kuramında puanlayıcılar arası ya da puanlayıcılar içi güvenilirlik katsayılarının kayıp veri olup olmama durumuna göre uygulanması, hangi durumlarda hangi katsayıların daha duyarlı sonuçları verdiğinin belirlenmesine dayanmaktadır. Bu çalışma esnasında hem iki puanlayıcı hem de çoklu puanlayıcı durumu dikkate alınmıştır. Klasik test kuramında yer alan puanlayıcılar arası ve puanlayıcılar içi güvenilirlik katsayılarının kıyaslanarak var olan bir durumun ortaya konulması amaçlandığından araştırma betimsel bir çalışma niteliğindedir.

Çalışma Grubu

Bu çalışmada elde edilen veriler üzerinden farklı puanlayıcılar arası güvenilirlik katsayılarının karşılaştırılması amaçlandığı için evreni temsil edecek şekilde bir örneklemin seçilmesine ihtiyaç duyulmamıştır. Çalışma grubu üzerinden veriler toplanmıştır. Araştırma kapsamında seçilen çalışma grubu araştırmada Uluslararası Öğrenci Değerlendirme Sınavında (PISA- Programme for International Student Assessment) kullanılan soruların kullanılması nedeniyle PISA örneklemindekine uygun şekilde yaklaşık olarak 15 yaş grubunda bulunan öğrencilerinden seçilmiştir. Bu amaçla 8. sınıf ve 9. sınıf öğrencilerinden veri toplanmıştır. Araştırmanın çalışma grubunu 2014-2015 eğitim-öğretim yılı bahar döneminde Ankara’da bir

ortaokulda öğrenim görmekte olan 8. sınıf öğrencilerinden ve yine Ankara'daki bir Anadolu lisesinde öğrenim gören 9. sınıf öğrencileri oluşturmaktadır. Çalışma grubunun betimsel özellikleri Tablo 8'de verilmiştir. Matematik okuryazarlığı uygulamasına 71'i 8.sınıf öğrencisi (36 kız, 35 erkek) ve 157'si 9.sınıf öğrencisi (101 kız, 56 erkek) olmak üzere toplam 228 öğrenci katılmıştır. 8. sınıfta bulunan öğrencilerin yaş ortalaması 14,16 iken 9. Sınıfta bulunan öğrencilerin yaş ortalaması 14,99 dur. Genel olarak matematik uygulamasına katılan tüm öğrencilerin yaş ortalaması 14,74'dür. Şekil 2'den de incelendiğinde matematik uygulamasına katılan öğrencilerin çoğunluğunun (154; %66) 15 yaş grubu öğrencisi olduğu görülmektedir. Okuduğunu anlama uygulamasına 71'i 8. sınıf öğrencisi (36 kız, 35 erkek) ve 159'u 9. sınıf (101 kız, 58 erkek) öğrencisi olmak üzere toplam 230 öğrenci katılmıştır. 8. sınıfta okuyan öğrencilerin yaş ortalaması 13,97 iken 9. sınıfta okuyan öğrencilerin yaş ortalaması 15'dir. Genel olarak okuduğunu anlama uygulamasına katılan öğrencilerin yaş ortalaması 14,76'dır.

Tablo 8. Çalışma Grubunun Betimsel Özellikleri

	Matematik okuryazarlığı uygulaması						Okuduğunu anlama uygulaması					
	8. Sınıf			9. Sınıf			8. Sınıf			9. Sınıf		
	K	E	Top	K	E	Top	K	E	Top	K	E	Top
N	36	35	71	101	56	157	36	35	71	101	58	159
%	51	49	100	64	36	100	51	59	100	64	36	100
Yaş Ortalaması	14,03	14,30	14,16	14,99	14,98	14,99	14,15	13,97	14,35	15	15	15

Çalışmada açık uçlu soruların yanıtlarını puanlandırmada kullanmak için beş farklı DPA geliştirilmiştir ve geliştirilen bu DPA'ların uygunluğuna yönelik olarak on farklı Eğitimde Ölçme ve Değerlendirme alan uzmanından görüş alınmıştır. Görüş alınan uzmanların özellikleri Tablo 9'da özetlenmiştir. Tablo 9 incelendiğinde uzmanların dokuzunun kadın, birinin erkek olduğu görülmektedir. Eğitim düzeyleri açısından incelendiğinde beş doktorasını bitirmiş, iki doktora tez dönemi aşamasında ve üç yüksek lisan tez aşamasında uzman bulunduğu görülmektedir. Uzmanlarının çoğunun ilgi alanlarının ise puanlama güvenilirliği yönünde olduğu söylenebilir.

Tablo 9. DPA ’lara İlişkin Görüş Bildiren Uzmanların Özellikleri

	Cinsiyet	Kıdem	Eğitim Düzeyi	Görevi	İlgi Alanları
U1	K	7 yıl	Doktora	Öğretim Görevlisi Dr.	Computer adaptive testing, Genellenebilirlik, differential item functioning
U2	K	2 yıl (öğretmen) 5,5 yıl (akademisyen)	Doktora	Öğretim Görevlisi Dr.	Genellenebilirlik kuramı, differential item functioning, Standart belirleme
U3	K	4 yıl	Doktora		PISA, Computer adaptive testing
U4	K	6 ay (öğretmen) 2 yıl (akademisyen)	Lisans (Yüksek lisans tez dönemi)	Araştırma Görevlisi	Çok yüzeyle rasch, Öz-akran değerlendirme, Puanlayıcı güvenilirliği
U5	K	2 yıl	Yüksek Lisans (Doktora tez dönemi)	Araştırma görevlisi	Puanlayıcılar arası tutarlık, Kuram karşılaştırmaları
U6	K	1 yıl (öğretmen) 2 yıl (akademisyen)	Lisans (Yüksek lisans tez dönemi)	Araştırma Görevlisi	Üst düzey becerilerin ölçümü, Puanlama güvenilirliği
U7	K	3 yıl	Lisans (Yüksek lisans tez dönemi)	Araştırma Görevlisi	
U8	K	2,5 yıl (akademisyen)	Yüksek lisans (Doktora tez dönemi)	Araştırma Görevlisi	Öğretmen yetiştirme, Kavram yanılgıları, Matematik öğrenme süreci (RBC), Yapısal eşitlik modelleme (YEM)
U9	K	15 yıl	Doktora	Öğretim Üyesi	Çok değişkenli istatistiksel analiz teknikleri, Yapısal eşitlik modelleme (YEM), Madde tepki kuramı (MTK)
U10	E	10 yıl (MEB) 8 yıl (akademisyen)	Doktora	Öğretim Üyesi	Performansa dayalı durum belirleme, e-portfolyo

Ayrıca öğrencilerden elde edilen cevapların puanlanması için altı farklı puanlayıcı bulunmaktadır. Puanlayıcıların üçü matematik okuryazarlığı uygulamasından elde edilen yanıtları puanlandırmış, diğer üçü okuduğunu anlama uygulamasından elde edilen yanıtları puanlandırmıştır. Puanlayıcıların özellikleri Tablo 10’da verilmiştir.

Tablo 10. Puanlayıcıların Özellikleri

Matematik Okuryazarlığı Uygulaması						
	Lisans		Yüksek Lisans		Doktora	
Puanlayıcı 1	İlköğretim öğretmenliği	matematik	Matematik (ilköğretim)	eğitimi	Matematik (ilköğretim dönemi)	eğitimi (tez)
Puanlayıcı 2	İlköğretim öğretmenliği	matematik	Eğitimde değerlendirme (dönemi)	ölçme ve (tez)	Matematik (ilköğretim)	eğitimi
Puanlayıcı 3	Matematik		Eğitimde değerlendirme	ölçme ve	Eğitimde değerlendirme (dönemi)	ölçme ve (ders)
Okuduğunu Anlama Uygulaması						
	Lisans		Yüksek Lisans		Doktora	
Puanlayıcı 1	Türkçe öğretmenliği		Türkçe Öğretmenliği			
Puanlayıcı 2*	İngilizce öğretmenliği		Eğitimde değerlendirme		ölçme ve	
Puanlayıcı 3**	Matematik		Eğitimde değerlendirme		ölçme ve (tez)	

*yüksek lisans tezinde Türkçe dil becerileri üzerine araştırma yapmıştır.

**yüksek lisans tezinde okuduğunu anlama üzerine araştırma yapmıştır.

Veri Toplama Aracı

Veri toplama aracı olarak “Matematik Okuryazarlığı Uygulaması”(Ek-1) ve “Okuduğunu Anlama Uygulaması” (Ek-2) kullanılmıştır. Burada bir sözel ve bir sayısal ders seçilmesine dikkat edilmiştir. Diğer yandan soruları puanlamak için araştırmacı tarafından hazırlanan DPA’na ilişkin uzman görüşü almak amacıyla “Uzman Görüş Formu” kullanılmıştır. Ayrıca uygulamalardan elde edilen verilerin puanlanmasında “DPA” (Ek-3) kullanılmıştır.

Matematik Okuryazarlığı Uygulaması

Matematik Okuryazarlığı Uygulaması, PISA uygulamasında kullanılan beş açık uçlu sorudan oluşan bir sınav şeklindedir. Sorular belirlenirken MEB tarafından yayınlanan 2003-2009 PISA uygulaması ulusal nihai raporlarından açıklanan sorular incelenmiştir. Açıklanan sorular içerisinde DPA hazırlamaya uygun olduğu belirlenen sorular incelenerek beş tanesi

seçilmiş ve form haline getirilmiştir. Bu uygulamada kullanılan sorular, PISA tarafından belirtilen matematiksel içeriğe göre 1. ve 2. soru değişim ve ilişkiler, 3. soru nicelik, 4. ve 5. soru belirsizlik konusuna yöneliktir. Kullanılan araçların güvenilirlik ve geçerliği için form haline getirilirken yönergesinin bulunmasına, soruların okunaklı ve uygun bir şekilde yazıldığına, soruların yanıtları için yeterli alanın bırakılıp bırakılmadığına dikkat edilmiştir. Cevaplama için yeterli sürenin verilmesi göz önüne alınmıştır. Ayrıca hazırlanan formun ilk sayfasında uygulamanın adı ve yönergenin bulunmasıyla birlikte yaş, cinsiyet gibi öğrencilere ait bazı betimsel bilgilerin istendiği bölüm bulunmaktadır. Hazırlanan formlar birkaç incelemeden sonra basıma hazır olduğuna karar verilerek çoğaltılmıştır.

Okuduğunu Anlama Uygulaması

Okuduğunu Anlama Uygulamasının hazırlanması için Matematik Okuryazarlığı Uygulamasına benzer bir yol izlenmiştir. Okuduğunu Anlama Uygulamasında PISA uygulamasında kullanılan altı soru kullanılmıştır. Sorular belirlenirken PISA tarafından açıklanan sorular incelenmiş ve DPA hazırlanması uygun olarak görülen sorulardan altı tanesi seçilmiştir. Belirlenen sorulardan üst düzey becerileri ölçme amacıyla yazılan sorulardan seçmeli tarzda yazılmış olanlar açık uçlu soru formatına çevrilmiştir. Bu uygulamada kullanılan sorular PISA tarafından belirlenen sorunun amacına göre bilgileri birleştirme ve yorumlama (soru 1, 2 ve 4) ile düşünme ve değerlendirme (soru 3, 5 ve 6) şeklindedir. Ayrıca yine forma son şekli verilirken yönergesinin bulunmasına, soruların okunaklı ve uygun bir şekilde yazıldığına, soruların yanıtları için yeterli alanın bırakılıp bırakılmadığına dikkat edilmiştir. Cevaplanma için yeterli sürenin verilmesi de dikkat edilen diğer bir noktadır. Ayrıca hazırlanan formun ilk sayfasında uygulamanın adı ve yönergenin bulunmasıyla birlikte yaş, cinsiyet gibi öğrencilere ait bazı betimsel bilgilerin istendiği bölüm bulunmaktadır. Hazırlanan formlar birkaç incelemeden sonra basıma hazır olduğuna karar verilerek çoğaltılmıştır.

Uzman Görüş Formu

Çalışma kapsamında soruları puanlamak için hazırlanan DPA'ların uygunluğunu belirlemek için uzman görüşüne başvurulmuştur ve bunun için araştırmacı tarafından hazırlanan uzman görüş formu oluşturulmuştur. Bu form genel olarak üç bölümden oluşmaktadır. Bölüm I de

kullanılan soruların genel içerikleri ile ilgili bilgilendirme için PISA madde içerikleri ile ilgili genel bilgiler yer almaktadır. Bölüm 2 de uzmanların özelliklerini betimleyebilmek için bazı kişisel bilgiler istenmiştir. Bölüm 3 de ise çalışma kapsamında kullanılan sorular, beklenen yanıtları ve bu soruları puanlamak için hazırlanan DPA'ları bulunmaktadır. Bölüm 1 deki açıklamalar ve Bölüm 3 deki soru yanıtları sadece DPA'ların uygunluğunu değerlendirmede kolaylık sağlaması için verilmiştir. Uzmanlara bu durum açıklama kısmında belirtilmiş ve Bölüm 1 ile ilgili herhangi bir işlem yapmalarına gerek olmadığı söylenmiştir.

Bölüm 3 de çalışmada kapsamında kullanılan PISA soruları iki başlık altında toplanmıştır: Matematik Okuryazarlığı ve Okuduğunu Anlama. Matematik Okuryazarlığı bölümünde PISA tarafından belirtilen matematiksel içeriğe göre 1. ve 2. soru değişim ve ilişkiler, 3. soru nicelik, 4. ve 5. soru belirsizlik konusuna yöneliktir. Bu nedenle soruları puanlandırmak için üç farklı DPA hazırlanmıştır. Okuduğunu anlama kısmında ise sorunun amacına göre bilgileri birleştirme ve yorumlama amacına yönelik üç soru (soru 1, 2 ve 4) ile düşünme ve değerlendirme amacına yönelik üç soru (soru 3, 5 ve 6) bulunmaktadır. Bu bölümde de sorunun amacı dikkate alınarak iki farklı DPA hazırlanmıştır. Her bir soru için PISA tarafından yayınlanan cevap anahtarı, ardından da araştırmacı tarafından yapılan çözümü bulunmaktadır. Çözümlerden sonra ise soruların içeriksel bilgisi ve beklenen cevaplardan hareket edilerek hazırlanan DPA'lar yer almaktadır. Uzmanlardan hazırlanan DPA'lardaki her bir kriterin uygunluğuna yönelik olarak aşağıda belirtilen sınıflandırmayı yapmaları beklenmiştir.

- I. “tamamen temsil ediyor” (tamamen uygun) = “gerekli/önemli”
- II. “oldukça temsil ediyor” (oldukça uygun) = “gerekli/önemli”
- III. “biraz temsil ediyor” (biraz uygun) = “yararlı ama önemli değil”
- IV. “hiç temsil etmiyor” (uygun değil) = “gereksiz”

Her DPA'nın yanında işaretlenmesi için gerekli bölüm ayrılmıştır. Ayrıca ölçütlere yönelik olarak III ve IV şeklinde görüş belirten uzmanların gerekçelerini açıklamaları istenmiştir.

Dereceli Puanlama Anahtarı (DPA)

DPA'ların hazırlanması aşamasında öncelikle kullanılan soruların cevaplanması araştırmacı tarafından yapılmış ardından MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü

tarafından hazırlanan PISA sorularının çözümleri incelenmiştir. Bu sayede sorulardan ne tür bir yanıtın beklendiği araştırmacı tarafından zihninde oluşturulmuştur. Ardından Matematik Okuryazarlığı Uygulamasında sorulan sorular için PISA tarafından belirlenen matematiksel içerik incelenmiş, soruların bulunduğu matematiksel içeriklerin (değişim ve ilişkiler, nicelik, belirsizlik) gerektirdiği becerilere ait açıklamalar ayrıntılı olarak gözden geçirildikten sonra her bir içeriğe yönelik olarak üç farklı DPA hazırlanmıştır. Benzer olarak Okuduğunu Anlama Uygulaması için PISA tarafından belirlenen soruların amaçları incelenmiş, soruların amacına (bilgileri birleştirme ve yorumlama; düşünme ve değerlendirme) ait açıklamalar ayrıntılı olarak gözden geçirilmiş ve soruların amacına göre gerektirdiği beceriler dikkate alınarak iki farklı DPA hazırlanmıştır.

Soruları puanlamak için kullanılacak olan araştırmacı tarafından geliştirilen DPA'ların uygunluğunu belirleyebilmek için uzman görüşü alınmış ve kapsam geçerlik oranı ve indeksi değerleri hesaplanmıştır. Kapsam geçerlik oranları (KGO) ve kapsam geçerlik indekslerinin (KGI) hesaplanmasının nedeni uzman görüşüne dayalı olarak toplanan nitel verilerdeki raporlama ve iletişim güçlüğü (Yurdugül ve Bayrak, 2012) ortadan kaldırmaktır. Bu amaçla kullanılan Davis ve Lawshe nin kapsam geçerlik oranları aşağıdaki şekilde açıklanabilir.

Davis Kapsam Geçerlik Oranı

Davis (1992)'ye göre kapsam geçerlik oranının hesaplanması için bir ölçek geliştirme çalışmasında konu alanı uzmanı, ilgili maddenin ölçülecek yapıyı temsil edip etmediğine ilişkin aşağıdaki şekilde görüş bildirir (Yurdugül, 2005; Yurdugül ve Bayrak, 2012):

- I. “tamamen temsil ediyor” (tamamen uygun)
- II. “oldukça temsil ediyor” (oldukça uygun)
- III. “biraz temsil ediyor” (biraz uygun)
- IV. “hiç temsil etmiyor” (uygun değil)

Kapsam geçerlik oranı denklem (1) kullanılarak elde edilir. Elde edilen kapsam geçerlik oranı 0,80 den büyük ise kapsam geçerliği yüksektir denir. Bu teknik için en az 3 en fazla 20 uzman önerilmektedir (Yurdugül, 2005).

$$\text{Kapsam Geçerlik Oranı (KMO)} = \frac{\text{I ve II şeklinde görüş bildiren uzman sayısı}}{\text{tüm uzman sayısı}} \quad \text{denklem (1)}$$

Lawshe Kapsam Geçerlik Oranı ve Kapsam Geçerlik İndeksi

Lawshe (1975)'nin yaklaşımı kapsam geçerlik oranı (madde bazında) ve kapsam geçerlik indeksini (ölçek bazında) dikkate alır. Lawshe'nin yaklaşımında geliştirilmesi amaçlanan ölçme aracındaki maddelerin ölçülmek istenilen özeliği ölçüp ölçmediğine ilişkin uzman görüşleri her bir madde için “gerekli/önemli”, “yararlı ama önemli değil” ve “gereksiz” şeklinde ölçeklendirilmektedir. Kapsam geçerlik oranı denklem (1) ile hesaplanır. Kapsam geçerlik indeksi ise ölçekte yer almasına karar verilen maddelerin kapsam geçerlik oranlarının ortalamasından elde edilir. Bir madde için uzmanların yarısı “Gerekli” dediyse $KMO = 0$; uzmanların yarısından fazlası gerekli dediyse $KMO > 0$ ve uzmanların yarısından azı gerekli dediyse $KMO < 0$ değerini alır (Yurdugül, 2005; Yurdugül ve Bayrak, 2012). Maddelerin kapsam geçerlik oranı ve indekslerinin uygun olup olmadığına Lawshe (1975) tarafından hazırlanan KGO ve KGİ için madde seçimine yönelik uzman sayısına göre minimum değerleri veren tablosuna (Tablo 11) bakılarak karar verilir.

$$KMO = \frac{N_G - \frac{N}{2}}{\frac{N}{2}} \text{ ya da } KMO = \frac{N_G}{\frac{N}{2}} - 1 \text{ denklem (2)}$$

N_G : Gerekli diyen uzmanların sayısı

N : Araştırmaya katılan uzmanların toplam sayısı

Tablo 11. $\alpha = 0,05$ Anlamlılık Düzeyinde KGO ve KGİ'ler için Uzman Sayısına Göre Minimum Değerler

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

Lawshe (1975)

Araştırma kapsamında hazırlanan DPA'ların uygunluğuna yönelik on uzmandan görüş alınmıştır. Toplanan görüşler sonucunda Davis ve Lawhse'ye göre hesaplanan kapsam geçerlik oranı ve kapsam geçerlik indeksleri Tablo 12'de verilmiştir.

Tablo 12. Kullanılan DPA'lara Yönelik Olarak Hesaplanan KGO (Davis-Lawhse) ve KGI'leri

	KGO(Davis)	KGO(Lawhse)	KGI
• <i>Matematik soru 1-2</i>			
İlişkileri belirleme	1,00	1,00	0,71
İşlemleri gerçekleştirme	0,80	0,60	
Sonuca ulaşma	0,80	0,60	
• <i>Matematik soru 3</i>			
Sayıları algılayabilme	0,60	0,40	0,80
Akıl yürütme / İlişkilendirme	1,00	1,00	
Nedenini açıklama	1,00	1,00	
• <i>Matematik soru 4-5</i>			
Grafiği okuma	1,00	1,00	0,80
Akıl yürütme / İlişkilendirme	0,90	0,80	
Neden belirtme	0,80	0,60	
• <i>Okuduğunu anlamam soru1-2-4</i>			
İlişkilendirme	0,90	0,80	0,94
Karşılaştırma	1,00	1,00	
Parça-bütün ilişkisini anlama	1,00	1,00	
• <i>Okuduğunu anlama 3-5-6</i>			
Yansıtma	1,00	1,00	1,00
Yorumlama	1,00	1,00	
Metinden bağımsız düşünme	1,00	1,00	

Davis'e göre kapsam geçerlik oranları incelendiğinde sadece matematik 3. soru için hazırlanan DPA'nın sayıları algılayabilme ölçütünde KGO değeri 0,80'nin altında kaldığı için kapsam olarak uygun olmadığı görülmektedir. Aynı ölçüt Lawshe'ye göre değerlendirildiğinde de aynı şekilde 0,62 nin altında olduğu görülmüş ve bu nedenle yeterli kapsamı sağlamadığı söylenebilir. Diğer yandan her bir soruya ait diğer ölçütler incelendiğinde Davis'e göre hesaplanan KGO değerlerinin 0,80 ve üzeri olduğu görülmüştür. Bu durumda tüm ölçütlerin yeterli kapsamı sağladığı söylenebilir. Diğer yandan Lawshe'ye göre KGO değerleri incelendiğinde üç ölçüt için (matematik 1. ve 2. soru işlemleri gerçekleştirme ile sonuca ulaşma; matematik 4. ve 5. soru neden belirtme) 0,62 değerinden düşük ancak çok yakın olduğu (0,60) olduğu görülmüştür. Ancak hazırlanan beş DPA'ya yönelik tek tek hesaplanan KGI değerleri incelendiğinde 0,62 den yüksek olduğu

için tüm DPA'ların kapsam olarak yeterli olduğu söylenebilir. Yine de uzman görüş formlarındaki dönütler incelenmiş, uygun görülen düzeltmeler yapılarak DPA'lara son şekli verilmiştir.

Veri Toplama Süreci

Veri toplama süreci genel olarak iki aşamadan oluşmaktadır. İlk aşamada, hazırlanan “Matematik Okuryazarlığı Uygulaması” ve “Okuduğunu Anlama Uygulaması” formları hazırlanıp, gerekli görüşmeler yapıldıktan sonra bir ortaokuldaki 8. sınıf öğrencilerine ve bir Anadolu Lisesi’ndeki 9. sınıf öğrencilerine uygulanmıştır. Öğrencilerden elde edilen veriler toplanıp, her bir öğrenci formu için kodlama işlemi yapılmıştır. Ardından ikinci aşama olan puanlama aşamasına geçilmiştir. Bu aşamada puanlayıcıların puanlama işlemini daha kolay yapabilmesi için her bir bireye ait her bir sorunun bulunduğu kitapçık setleri halinde fotokopi çekilmiştir. Diğer bir deyişle tüm öğrencilere ait 1. soruların yer aldığı tek bir kitapçık seti oluşturulmuştur. Bu işlem esnasında bağlantılı soruların aynı set içinde olacak şekilde ayarlanmasına dikkat edilmiştir. Ayrıca aynı sayfada bulunan sorularda aynı set içinde kalmıştır. Bu işlemin yapılmasının amacı puanlayıcıların bir soruya ait tüm cevapları arka arkaya puanlandırabilmesi ve puanlayıcılara puanlama işlemi esnasında kolaylık sağlanabilmesidir. Diğer yandan bu aşamada puanlayıcılar için genel bir yönerge oluşturulmuş ve bu yönergede çalışmanın amacı açıklanmış ve puanlama esnasında neler yapmaları gerektiği hakkında bilgi verilmiştir. Aynı zamanda her bir soru için sorunun özelliğinin, cevabının, soruya ait DPA’daki kriterlerin yer aldığı yönergeler oluşturulmuş ve her bir soruya ait kitapçık setinin içine DPA ile birlikte soruya ait yönergede yerleştirilmiştir (Yönerge örnekleri Ek-4 de verilmiştir). Tüm işlemlerden sonra daha önceden konuşulan ve puanlandırma konusunda onay alan puanlayıcılara öğrenci yanıtlarını içeren setler teslim edilmiş ve puanlandırma işlemi yapmaları istenmiştir. Dağıtım işlemi esnasında açıklamalar yapılmış ya da formları inceleyen puanlayıcılardan gelen sorular cevaplanmıştır. Yaklaşık bir hafta içerisinde tüm puanlayıcılar puanlandırma işlemini yaparak verileri teslim etmişlerdir.

Verilerin Analizi

Verilerin analizinde Windows için AgreeStat 2015.2 programı kullanılmıştır. Verilerin programa uygun düzenlenmesinde Excel 2013 programından yararlanılmıştır. AgreeStat

programı uygulamalar için görsel temel içeren kurulum gerektirmeyen bir programdır. Excel 2003, 2007, 2010 ve 2013 ile uyumlu çalışmaktadır. AgreeStat puanlayıcı güvenilirliği analizi için ideal bir programdır. İki veya daha fazla puanlayıcı için puanlayıcılar arası ve sınıf içi güvenilirlik katsayıları bu program sayesinde hesaplanabilir. Ayrıca AgreeStat programı tüm hesaplanan katsayılar için standart hata ve güven aralığı değerlerini hesaplar (AGREESTAT, 2015).

DPA'daki her bir ölçüte ait performans düzeyleri iki farklı ölçek türünde düşünülebilir. Puanlayıcılardan belirlenen performans düzeylerini birer sınıf olarak düşünüp puanlama yaptıkları kabul edilirse elde edilen puanların sınıflama ölçeğinde olduğu söylenebilir. Diğer yandan DPA'daki her bir performans düzeyi için aslında bir alt kategorinin bir üst kategoriden daha fazla bilgi taşıdığı dikkate alınırsa elde edilen puanlar sıralama ölçeği düzeyindedir. Bu çalışma kapsamında toplanan veriler sıralama ölçeğinde olduğu düşünülmüş, uzlaşma katsayıları hesaplanırken AgreeStat programında 'ordinal' seçeneğine göre işlemler yapılmıştır. Ayrıca sınıf içi korelasyon katsayılarını hesaplamak için her bir bireyin iki uygulamasına ait toplam puanlar oluşturulmuş ve ayrı ayrı analize alınmıştır. Bu durumda ise elde edilen toplam puanlar eşit aralıklı düzeyde olduğu için sınıf içi korelasyon katsayısının hesaplanması uygun bulunmuştur.

AgreeStat programı kayıp değerleri göz ardı etmemektedir. Onun yerine üst düzey matematiksel işlemler gerektiren güçlü katı (rigorous) istatistiksel yaklaşımlar kullanarak kayıp değerleri ele almaktadır. Kayıp değerlerin analiz edilebilmesi için kayıp değerlerin olduğu hücreye hiçbir değer girilmemeli, boş bırakılmalıdır. Bu durumda program kayıp değer olduğunu anlayarak analizleri bu duruma uygun istatistiksel teknikleri kullanarak yapmaktadır (AGREESTAT, 2015). AgreeStat programının kayıp veri yöntemine yönelik olarak kullandığı yaklaşım pairwise yöntemidir.*

Çalışma kapsamında iki ve üç puanlayıcı için uzlaşma katsayıları ve sınıf içi güvenilirlik katsayıları hesaplanmıştır. Uzlaşma katsayıları DPA'daki her bir ölçüt düzeyi için hesaplanmıştır ve analiz edilen veri sıralama ölçeğinde olduğu için ağırlıklandırılmış analiz yapılmıştır. İki puanlayıcı için hesaplanan puanlayıcılar arası güvenilirlik katsayıları Cohen's kappa, Gwet's AC₂, Scott's pi, Krippendorff's alpha, Brenann-Prediger ve uzlaşma yüzdesidir. Üç puanlayıcı için hesaplanan puanlayıcılar arası güvenilirlik katsayıları ise Cohen's kappa, Gwet's AC₂, Fleiss' kappa, Krippendorff's alpha, Brenann-Prediger ve

* AgreeStat destek (support@agreestat.com)'e atılan mail yoluyla bilgi alınmıştır.

uzlaşma yüzdesi şeklindedir. Sınıf içi korelasyon katsayı için uygun model belirlenirken her bir puanlayıcının her bir bireyi (hedefi) puanladığı, puanlayıcıların daha geniş bir evrenden değil de ilgilenilen konuyla ilgili olarak seçildiği (durum 3) ve puanlayan ile puanlanan grup arasında etkileşimin olmadığı model seçilmiştir. Bu duruma uygun olarak 2-yönlü karma ANOVA (2-way mixed ANOVA) modeli kullanılmıştır. Kayıp veri olmadığı duurm için varsayımlar incelenmiştir (Ek-6). Elde edilen güvenilirlik katsayıları, katsayı değerleri ve puanlayıcılar arası katsayılar için standart hataları, sınıf içi korelasyon katsayılarında ise hata varyansları dikkate alınarak karşılaştırmalar yapılmıştır. Karşılaştırma işlemi iki puanlayıcılı ve üç puanlıyıcılı durum için kayıp veri durumu olduğu ve olmadığı durumlar için ayrı ayrı yapılmıştır. Kayıp veri durumu incelenirken tüm veri üzerinden %5, %10 ve %15 kayıp veri olacak şekilde randum olarak belirlenen gözlemler silinmiştir ve her bir analiz durumu için tüm uzlaşma katsayısı değerleri hesaplanmıştır.

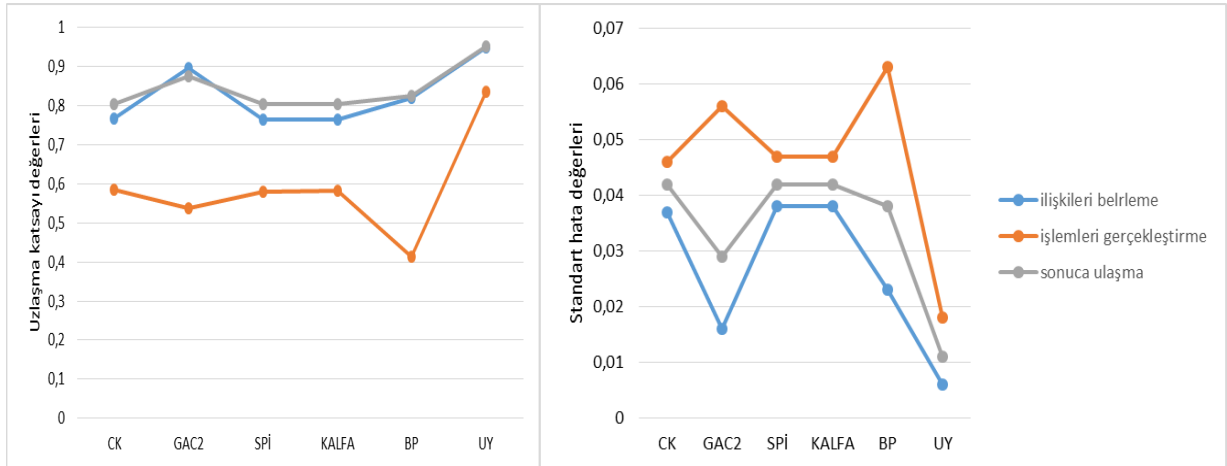
BÖLÜM 4

BULGULAR VE YORUM

Bu bölümde çalışmanın amacına yönelik elde edilen bulgular raporlanmış ve yorumlanmıştır. Anlaşılabilirlik açısından bulgular araştırma sorularına göre başlıklandırılarak özetlenmeye çalışılmıştır.

Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Karşılaştırılması

Kayıp verinin olmadığı durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için matematik okuryazarlığı uygulaması 1. sorudan elde edilen güvenilirlik değerleri ve bu değerlerin standart hataları Şekil 2’de verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5 Tablo 17’de verilmiştir.

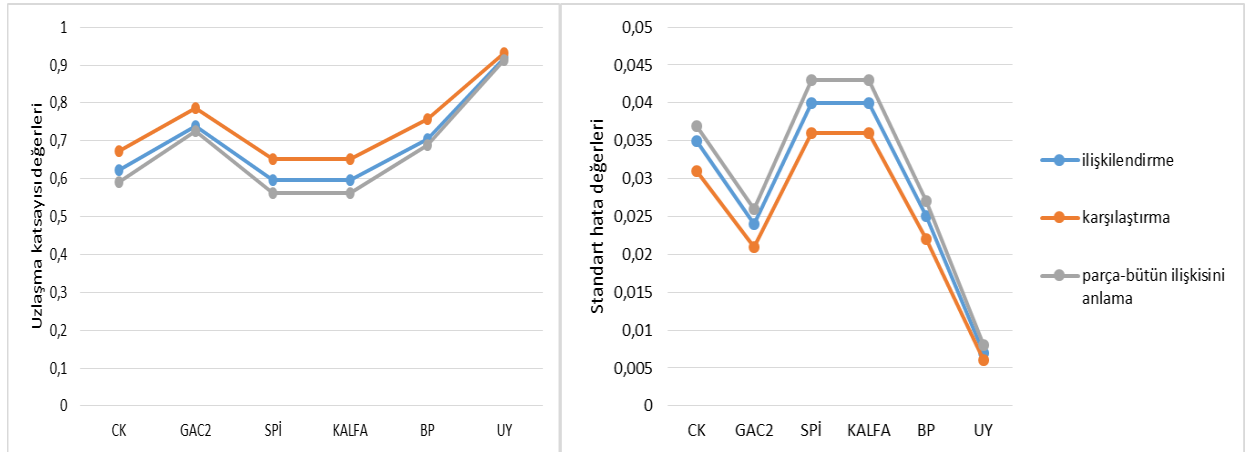


Şekil 2. İki puanlayıcı için matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC₂, SPİ = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma)

Şekil 2 incelendiğinde matematik okuryazarlığı soru 1'deki tüm kriterler için uzlaşma yüzdesinin en yüksek değere ve en düşük standart hataya sahip olduğu görülmektedir. Bu durumun nedeninin diğer uzlaşma katsayıları hesaplanırken puanlayıcılar arası uzlaşma için şans yoluyla beklenen uzlaşmaya yönelik hesaplama yapılarak bu etkinin ortadan kaldırılmaya çalışılması gösterilebilir. Uzlaşma yüzdesi hesaplanırken sadece puanlayıcılar arası uzlaşma oranı dikkate alınmakta ve uzlaşmanın şans yoluyla beklenen oranına yönelik hiçbir işlem yapılmamaktadır. Diğer yandan Gwet'in AC₂ ile Brennan-Prediger katsayılarının ve Cohen'nin kappa, Scott'nin pi ile Krippendorff alfa değerlerinin benzer sonuç verdiği görülmüştür. Bu beş uzlaşma katsayısının hesaplanması temel olarak puanlayıcıların uzlaştığı birimlerin oranından şansla uzlaşılan birimlerin oranı çıkarılıp elde edilen değer şansa uzlaşma olmama olasılığına bölümüne dayanmaktadır. Genelde tüm katsayılar için uzlaşılan birimlerin oranı benzer şekilde hesaplanırken, uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanış şekli farklılaşmaktadır. Gwet'in AC₂ ile Brennan-Prediger katsayıları için uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanış şekli benzer olduğu için bu iki katsayının benzer sonuçlar verdiği söylenebilir. Aynı şekilde Cohen'nin kappa, Scott'nin pi ile Krippendorff alfa katsayıların yakın sonuçlar vermesi bu üç katsayı için şans yoluyla beklendiği birimlerin oranının hesaplanış şeklinin benzer oluşundan kaynaklanabilir. Genel olarak güvenirliliğin 0,60'dan yüksek olduğu durumda Gwet'in AC₂ katsayısı en yüksek değere ve en düşük standart hataya sahip iken 0,60'dan düşük olduğu durum için ise Cohen'nin kappa, Scott'nin pi ile Krippendorff'un

alfasına göre daha düşük güvenilirlik değerine sahip ve daha yüksek standart hataya sahiptir. Soru 1 de ele alınan kriterler bazında güvenilirlik katsayıları incelendiğinde sonuca ulaşma kriterine yönelik yapılan puanlamalarda puanlayıcılar arası uzlaşmanın daha yüksek olduğu görülmüştür. Bunun nedenin ise sonuca ulaşma kriterinin daha net sınırlar içerek şekilde tanımlanabilmesinden kaynaklandığı düşünülmektedir. Puanlayıcılar arasında uzlaşmanın en düşük olduğu kriter ise işlemleri gerçekleştirme kriteridir. Bunun nedeninin ise farklı çözüm yolları gerektiren bu tür bir soruda puanlayıcıların zihinlerindeki işlem basamakları ve beklentilerinin farklılaşmasından kaynaklandığı söylenebilir.

Okuduğunu anlama kapsamında kayıp verinin olmadığı durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için 1. sorudan elde edilen güvenilirlik değerleri ve bu değerler ait standart hata değerleri Şekil 3'te verilmiştir. Ayrıca okuduğunu anlama uygulaması kapsamında iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5'te Tablo 18'de verilmiştir.



Şekil 3. İki puanlayıcı için Okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC2, SPI = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)

Şekil 3 incelendiğinde okuduğunu anlama soru 1'deki tüm kriterler için en yüksek değere ve en düşük standart hataya sahip uzlaşma katsayısının uzlaşma yüzdesi olduğu görülmektedir. Uzlaşma yüzdesinde şansa uzlaşma durumunun dikkate alınmaması ve bu duruma yönelik herhangi bir düzeltme kullanılmamasının uzlaşma yüzdesinin diğer katsayılardan daha yüksek güvenilirlik değerine sahip olmasına neden olduğu düşünülmektedir. Diğer yandan

Gwet'in AC_2 ile Brennan-Prediger katsayılarının ve Cohen'nin kappa, Scott'nin pi ile Krippendorff alfa değerlerinin benzer sonuç verdiği görülmüştür. Genel olarak, Gwet'in AC_2 katsayısı en yüksek değere ve en düşük standart hataya sahip olduğu tespit edilmiştir. Cohen'nin kappa, Scott'nin pi ile Krippendorff'un alfa değerleri ise diğer katsayılara göre daha düşük güvenilirlik değerine ve daha yüksek standart hataya sahiptir. Soru 1 de ele alınan kriterler bazında güvenilirlik katsayıları incelendiğinde genellikle tüm kriterler için benzer güvenilirlik sonuçları elde edildiği görülmekle birlikte en yüksekten en düşük güvenirlığe doğru karşılaştırma, ilişkilendirme, parça-bütün ilişkisini anlama şeklinde olduğu tespit edilmiştir.

Ek-5'de verilen Tablo 17 ve 18 incelendiğinde matematik okuryazarlığı için dördüncü sorunun üçüncü kriteri için Gwet'in AC_2 ve Brennan-Brediger katsayılarının; okuduğunu anlama için ikinci sorunun ikinci kriteri ve dördüncü sorunun birinci kriteri için Scott'nin pi ve Krippendorff'un alfa katsayılarının negatif değerler verdiği görülmektedir. Krippendorff (2004; s.222) Krippendorff'un alfa katsayısının örnekleme hatası ve sistematik uzlaşmama gibi nedenlerden dolayı 0 ile 1 aralığının dışına çıkabileceğini belirtmiştir. Bu nedenle hesaplanan uzlaşma katsayılarında negatif değerlerin çıkmış olabileceği söylenebilir. Ayrıca uzlaşma katsayılarında negatif değerler daha çok okuduğunu anlamaya ait sorular için görülmüştür. Bunun nedeninin ise okuduğunu anlama gibi sözel dersler için dereceli puanlama anahtarının hazırlanmasının zor olması, hazırlanan dereceli puanlama anahtarındaki açıklamalar ve soruların cevaplarının puanlayıcılar arasında farklı anlaşılabilme ihtimalinin daha yüksek olmasından kaynaklandığı söylenebilir.

Hem matematik okuryazarlığı hem de okuduğunu anlama için incelendiğinde kayıp verinin olmadığı iki puanlayıcılı durum için uzlaşma yüzdesinin her zaman en yüksek katsayı değerini aldığı ve en düşük standart hataya sahip olduğu görülmektedir. Uzlaşma yüzdesinden sonra genel olarak hemen her bir kriter için en az hataya sahip uzlaşma katsayısının Gwet's AC_2 olduğu belirlenmiştir. Aynı zamanda Gwet' AC_2 ile Brennan-Prediger uzlaşma katsayılarının kendi aralarında ve Cohen's kappa, Scott's pi ile Krippendorff alfa uzlaşma katsayılarının kendi aralarında yakın değerlere sahip olduğu görülmüştür. Uzlaşma katsayılarının değerleri dikkate alındığı uzlaşma yüzdesinin her zaman en yüksek değere sahip olması nedeniyle çok cömert bir katsayı olduğu yorumu yapılabilir. Cohen's kappa, Scott's pi ve Krippendorff alfa değerleri ise çoğu durumda en katı değerleri veren katsayı değerleri olarak karşımıza çıkmış olduğu düşünülebilir. Bu

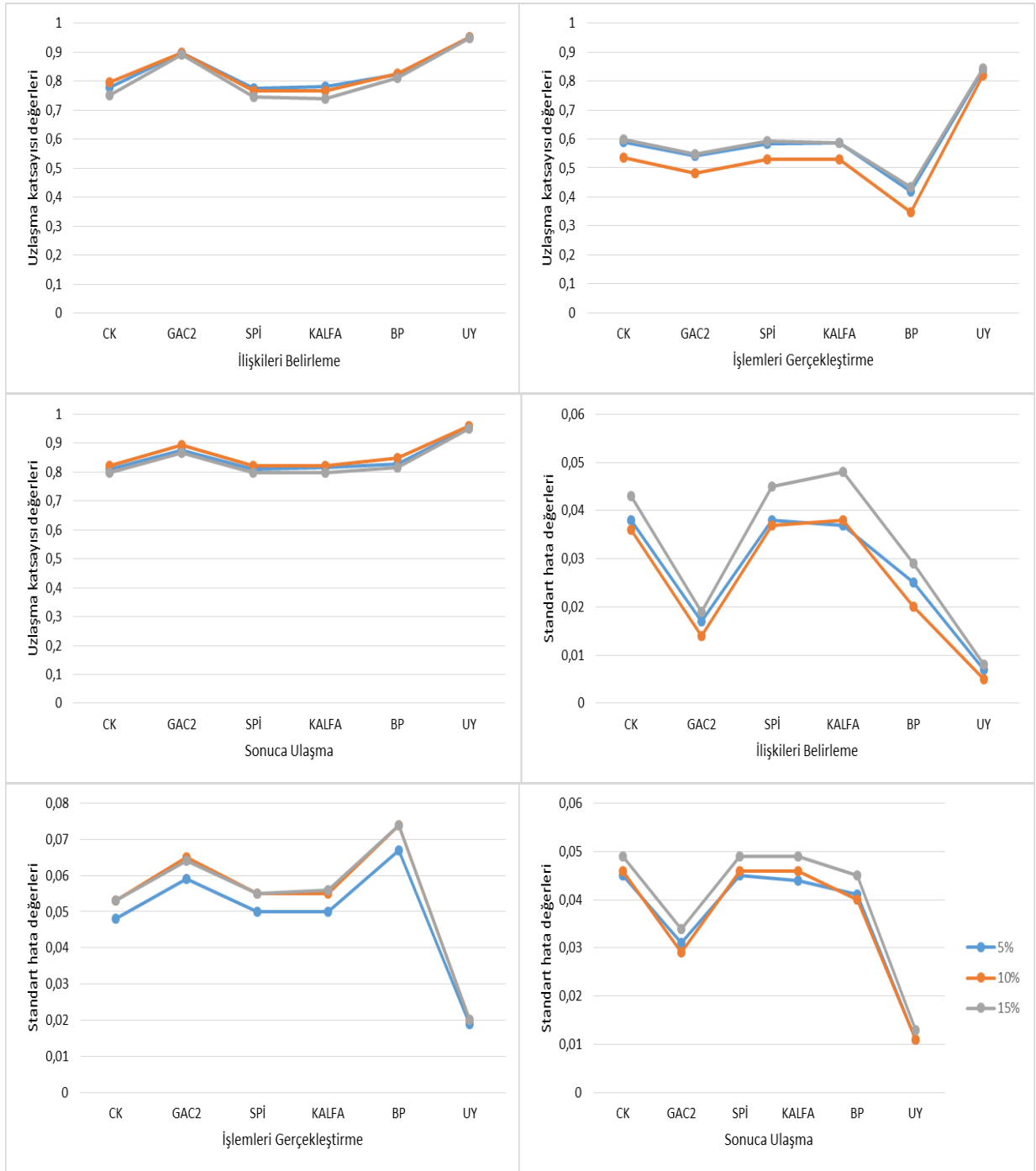
durumda uzlaşma yüzdesinin cömert bir katsayı olduğu düşünülürse bazı durumlarda uzlaşma miktarını olduğundan fazla tahmin edebilir olabileceği yorumu yapılabilir. Benzer şekilde Cohen's kappa, Scott's pi ile Krippendorff alfa değerlerinin daha katı uzlaşma değerlerini verdiği düşünülürse uzlaşma miktarını olduğundan düşük olarak gösterebileceği fikri ortaya çıkabilir.

Kılıç (2009) adlandırma ölçekte değerlendirme yapan iki değerlendirici durumunda kappa ve Krippendorff alfa benzer sonuçlar verdiği ile adlandırma ölçekte değerlendirme yapan çoklu değerlendirici durumunda Fleiss kappa ve Krippendorff alfa benzer sonuçlar verdiği belirtmiştir. Aynı şekilde Bıkmaz (2011) kappa ve Krippendorff alfa tekniğinin benzer sonuçlar verdiği bulgusuna ulaşmıştır. Gwet (2002) de pi ve kappa istatistiklerinin çoğu zaman benzer sonuçlar ürettiğini ifade etmiştir. Viera ve Garrett (2005) göre kappa istatistiği uzlaşmanın basit bir ham oranının hesaplamasından daha fazla bilgi vermektedir.

Diğer taraftan bakıldığında Gwet's AC_2 ile Brennan-Prediger uzlaşma katsayıları Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayılarına göre daha yüksek değerlere sahiptir. Bu durum genel olarak uzlaşma yüzdesinden en yüksek güvenilirlik değerlerini gösterdiğini ve bu nedenle güvenilirliği olduğundan yüksek gösterebileceği; Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayılarının ise en düşük güvenilirlik değerlerine sahip olduğu ve bu nedenle güvenilirliği olduğundan düşük gösterebileceği düşünülmektedir. Gwet's AC_2 ile Brennan-Prediger uzlaşma katsayılarının ise güvenilirliği ne yüksek ne de düşük gösterdiği, genel olarak diğer katsayıların arasında değer aldığı söylenebilir. Bu nedenle puanlama güvenilirliğinde Gwet's AC_2 ile Brennan-Prediger uzlaşma katsayılarının kullanımı önerilebilir. Wongpakaran vd. (2013) Gwet's AC_1 katsayısının Cohen's kappaya göre daha kararlı bir puanlayıcılar arası güvenilirlik katsayısı olduğunu göstermişlerdir. Gwet (2002) Pi ve Kappa istatistiğinin tahmin edilemeyen davranışlarının şanla uzlaşma olasılığının hesaplanmasındaki bir yanlış yöntemden kaynaklandığını göstererek AC_1 istatistiğinin daha güvenilir bir yaklaşım olduğunu belirtmişlerdir.

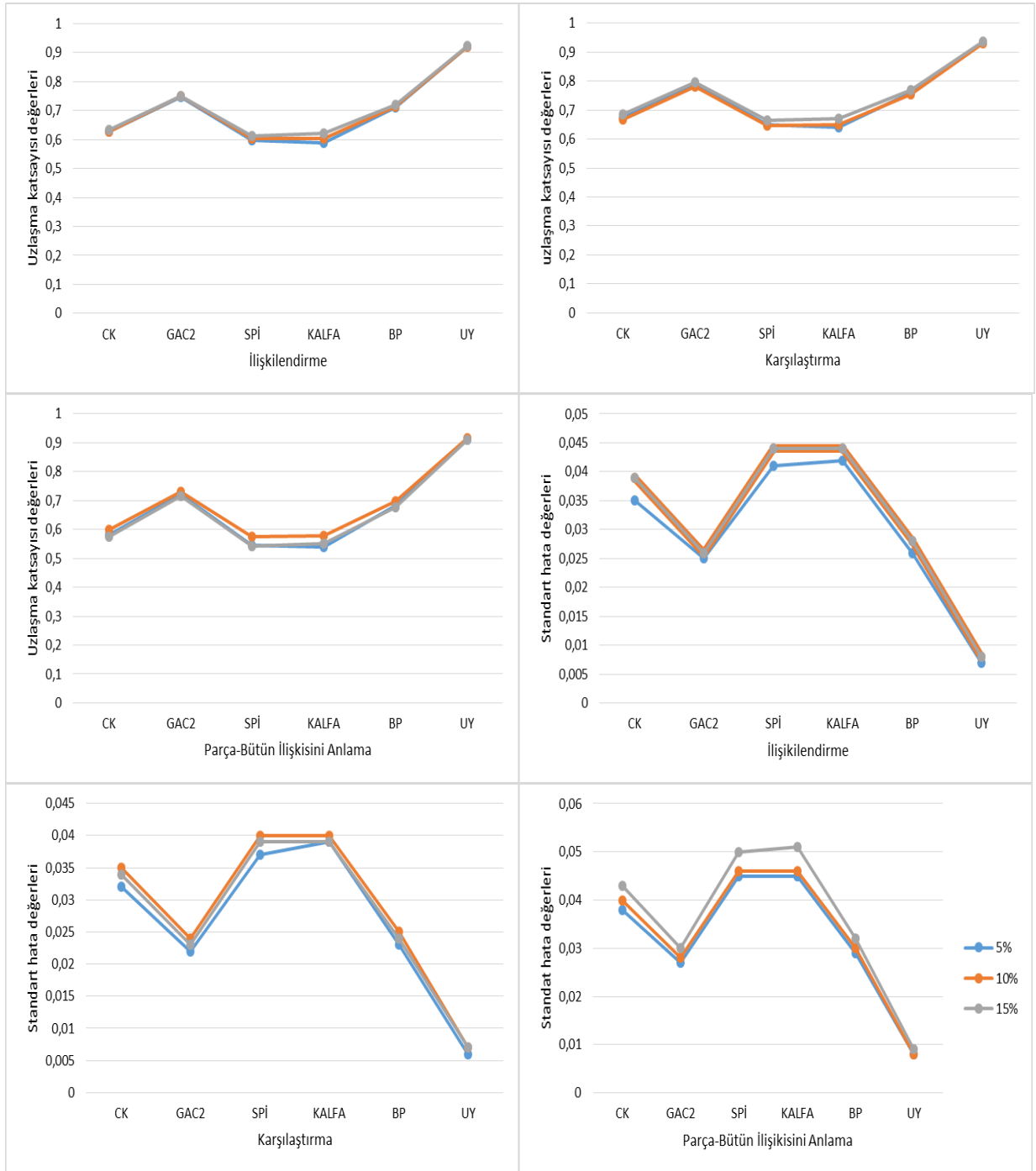
Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Kayıp Veri Oranı Dikkate Alınarak Karşılaştırılması

Kayıp verinin olduğu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için matematik okuryazarlığı uygulaması 1. sorudan elde edilen güvenilirlik değerleri ve bu değerlerin standart hataları Şekil 4'te verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında Kayıp verinin olduğu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5 Tablo 19, 20 ve 21'de verilmiştir.



Şekil 4. İki puanlayıcı için kayıp veri olduğu durumda matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Cohen'nin kappa, GAC2 = Gwet'in AC2, SPI = Scott'nin pi, KALFA = Krippendorff'un alfa, BP = Brennan-Prediger, UY = Uzlaşma Yüzdesi)

Kayıp verinin olduđu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değışimleri görmek için okuduđunu anlama uygulaması 1. sorudan elde edilen güvenilirlik değeri ve bu değeriin standart hataları Şekil 5’te verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında Kayıp verinin olduđu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değeri ve bu değeriin standart hataları tüm sorular için Ek-5 Tablo 22, 23 ve 24’te verilmiştir.



Şekil 5. İki puanlayıcı için kayıp veri olduğu durumda okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger’ın kappa, GAC2 = Gwet’in AC2, FK = Fleiss’nin kappa, KALFA = Krippendorff’un alfa, BP = Brennan-Prediger, UY = Uzlaşma yüzdesi)

Şekil 4 ve 5'e göre hem matematik okuryazarlığı hem de okuduğunu anlama için incelendiğinde kayıp verinin olduğu durumda üç puanlayıcı durum için uzlaşma katsayıları arasındaki ilişki kayıp veri %5, %10 ve %15 olduğu durumlarda kayıp veri olmayan durumdakine benzer sonuçlar gösterdiği görülmektedir. Hem matematik okuryazarlığı hem de okuduğunu anlama için incelendiğinde kayıp verinin olduğu durumda iki puanlayıcı durum için veri setinde %15'e kadar kayıp veri bulunduğu uzlaşma katsayılarından en yüksek değere ve en düşük standart hataya sahip olanın iki puanlayıcı kayıp veri bulunmadığı duruma benzer olarak uzlaşma yüzdesi olduğu görülmektedir. Gwet's AC₂ ve Brennan-Prediger uzlaşma katsayılarının Cohen's kappa, Scott's pi ve Krippendorff alfa uzlaşma katsayılarına göre çoğu zaman daha yüksek değer aldığı söylenebilir. Ayrıca uzlaşma yüzdesinden sonra görece en az standart hata değerine sahip uzlaşma katsayısını Gwet's AC₂ olduğu görülmüştür. Ayrıca kayıp değer miktarındaki artışa göre uzlaşma katsayıları incelendiğinde kayıp veri miktarının yaklaşık %5 ve %10 olduğu durumlar için tüm uzlaşma katsayılarının değerlerinin neredeyse aynı olduğu, standart hatalarındaki değişimin de yok denecek kadar düşük olduğu söylenebilir.

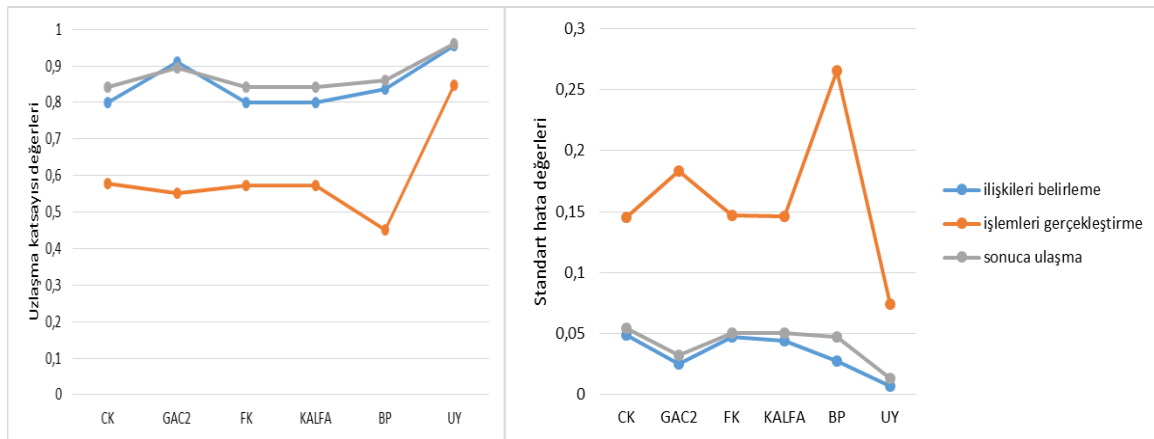
Kayıp veri miktarının %15 olduğu durumla kayıp veri miktarının %5 ve %10 olduğu durumlarla karşılaştırıldığında ise %15 olan durumda, uzlaşma katsayılarının bir miktar düştüğü, standart hatalarda bir miktar artış olduğu görülmekle birlikte bu artışın yapılacak yorumları etkileyebilecek düzeyde olmadığı söylenebilir. Bu durumda uzlaşma katsayılarında %15'e kadar kayıp verinin olmasının yaklaşık 250 kişilik bir veri grubunda sorun yaratmadığı söylenebilir. Ayrıca kayıp veriyi egale etmek için yapılan matematiksel hesaplamaların yaklaşık 250 kişilik bir veri grubunda %15'e kadar kayıp verinin bulunması durumunda iyi bir şekilde çalıştığı söylenebilir. Bu durum daha küçük ve daha büyük örneklemelerde araştırılabilir.

Izquierdo ve Pedrero (2014) genel olarak güvenilirlik katsayıları kestirimlerinde kayıp yanıtların toplam yüzdesinin %10 aşmadığı durumlarda makul bir şekilde iyi sonuçlar verdiğini belirtmiştir. Bal (2003) düşük hacimli örneklemelerde kayıp değer atama yöntemlerinin (listwise, pairwise, ortalama, regresyon ve EM) tutarsız sonuçlar verdiğini, büyük hacimli örneklemelerde ise EM atama yönteminin tam veri setine oldukça yakın değerler verdiğini belirtmektedir. Ayrıca Bal (2003) özellikle 200 birim ve üzerindeki veri setleri için %5-%20 arasındaki eksik veri yapılarında EM yönteminin diğer yöntemlere göre daha iyi sonuçlar verdiğini ileri sürmüştür. Kılıç (2009) sürekli ölçekte değerlendirme yapan

çoklu değerlendirici için eksik veri olmadığı ve olduğu durumlar için ICC ve Krippendorff alfa ölçütlerinin benzer sonuçlar verdiği belirtilmiştir.

Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının İncelenmesi

Kayıp verinin olmadığı durumda üç puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için matematik okuryazarlığı uygulaması 1. sorudan elde edilen güvenilirlik değerleri ve bu değerlerin standart hataları Şekil 6'da verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5 Tablo 25'te verilmiştir.

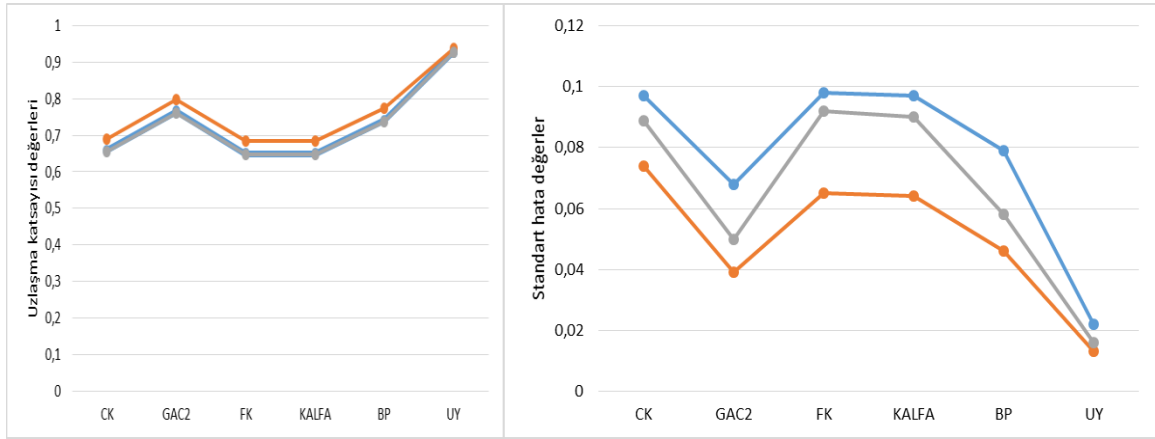


Şekil 6. Üç puanlayıcı için matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC₂, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)

Şekil 6 incelendiğinde matematik okuryazarlığı soru 1'deki tüm kriterler için uzlaşma yüzdesinin en yüksek değere ve en düşük standart hataya sahip olduğu görülmektedir. Bu durumun nedeninin diğer uzlaşma katsayıları hesaplanırken puanlayıcılar arası uzlaşma için şans yoluyla beklenen uzlaşmaya yönelik hesaplama yapılarak bu etkinin ortadan kaldırılmaya çalışılması gösterilebilir. Uzlaşma yüzdesi hesaplanırken sadece puanlayıcılar arası uzlaşma oranı dikkate alınmakta ve uzlaşmanın şans yoluyla beklenen oranına yönelik hiçbir işlem yapılmamaktadır. Diğer yandan Gwet'in AC₂ ile Brennan-Prediger

katsayılarının ve Conger'in kappa, Fleiss'nin kappa ile Krippendorff alfa değerlerinin benzer sonuç verdiği görülmüştür. Bu beş uzlaşma katsayısının hesaplanması temel olarak puanlayıcıların uzlaştığı birimlerin oranından şansla uzlaşılan birimlerin oranı çıkarılıp elde edilen değer in şansla uzlaşma olmama olasılığına bölümüne dayanmaktadır. Genelde tüm katsayılar için uzlaşılan birimlerin oranı benzer şekilde hesaplanırken, uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanış şekli farklılaşmaktadır. Gwet'in AC₂ ile Brennan-Prediger katsayıları için uzlaşmanın şans yoluyla beklendiği birimlerin oranının hesaplanış şekli benzer olduğu için bu iki katsayının benzer sonuçlar verdiği söylenebilir. Aynı şekilde Conger'in kappa, Fleiss'nin kappa ile Krippendorff alfa katsayıların yakın sonuçlar vermesi bu üç katsayı için şans yoluyla beklendiği birimlerin oranının hesaplanış şeklinin benzer oluşundan kaynaklanabilir. Genel olarak güvenilirliğin 0,60'dan yüksek olduğu durumda Gwet'in AC₂ katsayısı en yüksek değere ve en düşük standart hataya sahip iken 0,60'dan düşük olduğu durum için ise Cohen'nin kappa, Scott'nin pi ile Krippendorff'un alfasına göre daha düşük güvenilirlik değerine sahip ve daha yüksek standart hataya sahiptir. Soru 1 de ele alınan kriterler bazında güvenilirlik katsayıları incelendiğinde sonuca ulaşma kriterine yönelik yapılan puanlamalarda puanlayıcılar arası uzlaşmanın daha yüksek olduğu görülmüştür. Bunun nedenin ise sonuca ulaşma kriterinin daha net sınırlar içerecek şekilde tanımlanabilmesinden kaynaklandığı düşünülmektedir. Puanlayıcılar arasında uzlaşmanın en düşük olduğu kriter ise işlemleri gerçekleştirme kriteridir. Bunun nedeninin ise farklı çözüm yolları gerektiren bu tür bir soruda puanlayıcıların zihinlerindeki işlem basamakları ve beklentilerinin farklılaşmasından kaynaklandığı söylenebilir.

Kayıp verinin olmadığı durumda üç puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için matematik okuryazarlığı uygulaması 1. sorudan elde edilen güvenilirlik değerleri ve bu değerlerin standart hataları Şekil 7'de verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5 Tablo 26'da verilmiştir.



Şekil 7. Üç puanlayıcı için okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC₂, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan- Prediger, UY = Uzlaşma yüzdesi)

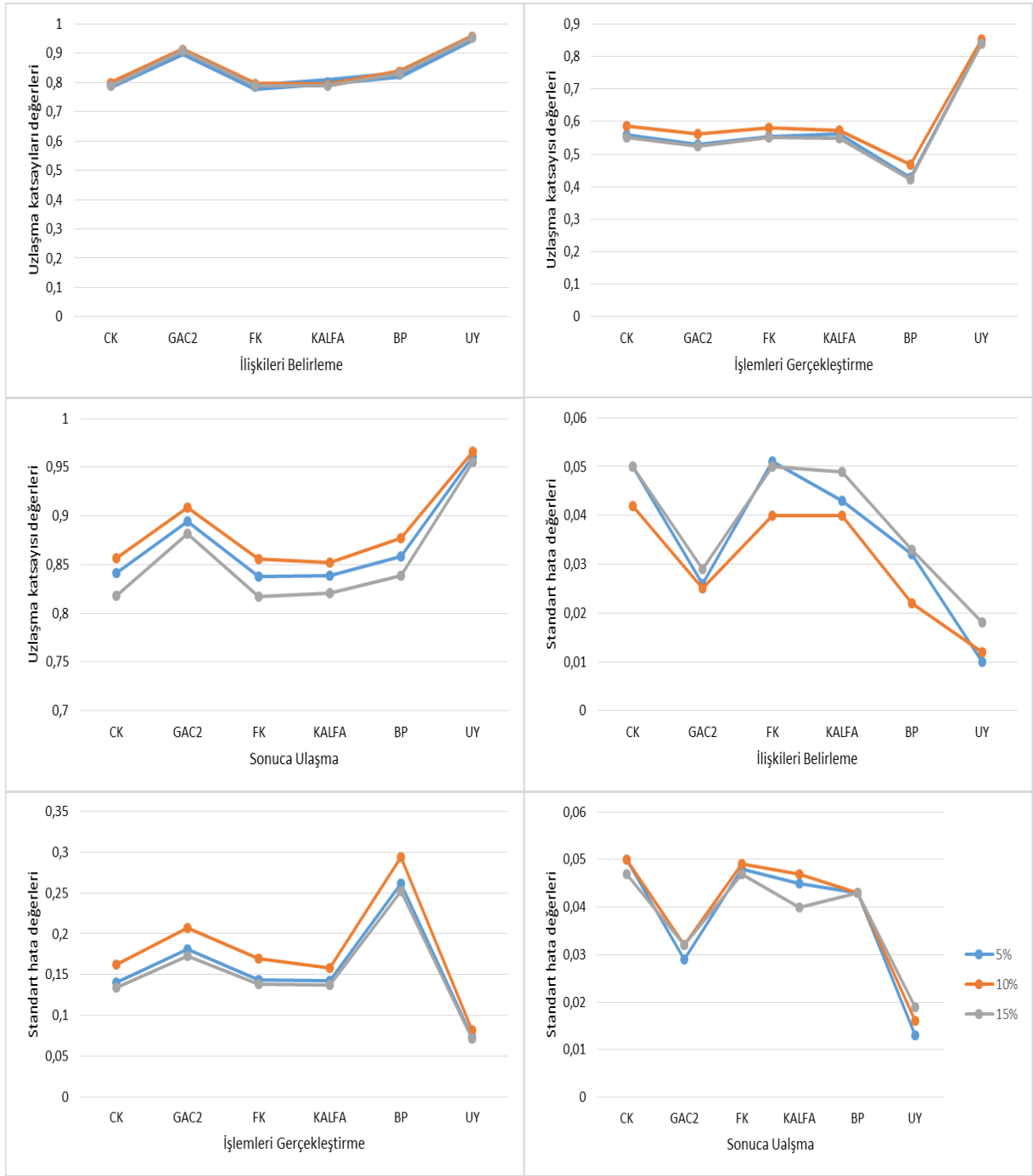
Şekil 7 incelendiğinde okuduğunu anlama soru 1'deki tüm kriterler için en yüksek değere ve en düşük standart hataya sahip uzlaşma katsayısının uzlaşma yüzdesi olduğu görülmektedir. Uzlaşma yüzdesinde şansla uzlaşma durumunun dikkate alınmaması ve bu duruma yönelik herhangi bir düzeltme kullanılmamasının uzlaşma yüzdesinin diğer katsayılardan daha yüksek güvenilirlik değerine sahip olmasına neden olduğu düşünülmektedir. Diğer yandan Gwet'in AC₂ ile Brennan-Prediger katsayılarının ve Conger'in kappa, Fleiss'nin kappa ile Krippendorff alfa değerlerinin benzer sonuç verdiği görülmüştür. Genel olarak, Gwet'in AC₂ katsayısı en yüksek değere ve en düşük standart hataya sahip olduğu tespit edilmiştir. Conger'in kappa, Fleiss'nin kappa ile Krippendorff'un alfa değerleri ise diğer katsayılara göre daha düşük güvenilirlik değerine ve daha yüksek standart hataya sahiptir. Soru 1 de ele alınan kriterler bazında güvenilirlik katsayıları incelendiğinde genellikle tüm kriterler için benzer güvenilirlik sonuçları elde edildiği görülmekle birlikte en yüksekten en düşük güvenirlığe doğru karşılaştırma, ilişkilendirme, parça-bütün ilişkisini anlama şeklinde olduğu tespit edilmiştir.

Hem matematik okuryazarlığı hem de okuduğunu anlama için incelendiğinde kayıp verinin olmadığı üç puanlayıcı durum için en yüksek katsayı değerini alan ve en düşük standart hata değerine sahip olan uzlaşma katsayısının uzlaşma yüzdesi olduğu görülmektedir. Uzlaşma katsayılarından Gwet's AC₂'nin hemen hemen her bir kriterde genel olarak uzlaşma katsayısından sonra en yüksek uzlaşma düzeyini veren ve en az standart hataya sahip olan katsayı olduğu tespit edilmiştir. Diğer yandan, Gwet' AC₂ ile Brennan-Prediger uzlaşma

katsayılarının kendi aralarında ve Conger's kappa, Fleiss' kapa ile Krippendorff alfa uzlaşma katsayılarının kendi aralarında yakın değerlere sahip olduğu belirlenmiştir. En düşük uzlaşma katsayısı değerleri ise Fleiss' kappa katsayısında gözlemlenmiştir. Uzlaşma katsayılarının değerleri dikkate alındığı uzlaşma yüzdesinin her zaman en yüksek değere sahip olması nedeniyle çok cömert bir katsayı olduğu yorumu yapılabilir. Conger's kappa, Scott's pi ve Krippendorff alfa değerleri ise çoğu durumda en katı değerleri veren katsayı değerleri olarak karşımıza çıktığı düşünülebilir. Bu durumda uzlaşma yüzdesinin cömert bir katsayı olduğu düşünülürse bazı durumlarda uzlaşma miktarını olduğundan fazla tahmin edebilir olabileceği şeklinde yorum yapılabilir. Benzer şekilde Conger's kappa, Fleiss' kappa ile Krippendorff alfa değerlerinin daha katı uzlaşma değerlerini verdiği düşünülürse uzlaşma miktarını olduğundan düşük olarak gösterebileceği fikri ortaya çıkabilir. Üç puanlayıcı için elde edilen bulgular iki puanlayıcı için hesaplanan uzlaşma katsayılarından elde edilen bulgular ile benzerlik göstermektedir.

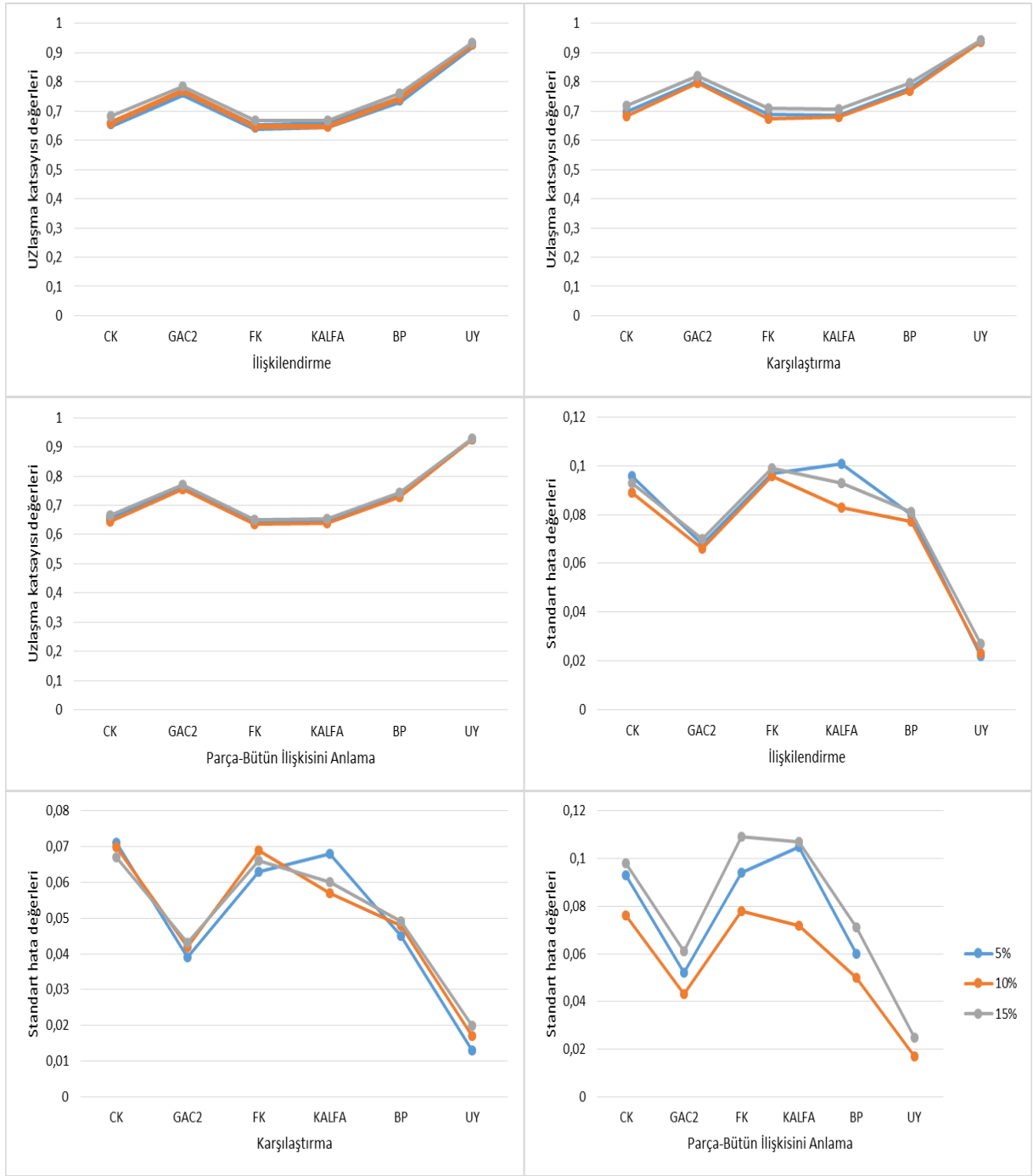
Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Uzlaşma Katsayılarının Kayıp Veri Oranı Dikkate Alınarak İncelenmesi

Kayıp verinin olduğu durumda üç puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değişimleri görmek için matematik okuryazarlığı uygulaması 1. sorudan elde edilen güvenilirlik değerleri ve bu değerlerin standart hataları Şekil 8'de verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında Kayıp verinin olduğu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değerler ve bu değerlerin standart hataları tüm sorular için Ek-5 Tablo 27, 28 ve 29'da verilmiştir.



Şekil 8. Üç puanlayıcı için kayıp veri olduğu durumda matematik okuryazarlığı uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan-Prediger, UY = Uzlaşma Yüzdesi)

Kayıp verinin olduđu durumda üç puanlayıcı için hesaplanan farklı uzlaşma katsayıları arasındaki değışimleri görmek için okuduđunu anlama uygulaması 1. sorudan elde edilen güvenilirlik değeri ve bu değeriin standart hataları Şekil 9’da verilmiştir. Ayrıca matematik okuryazarlığı uygulamasında Kayıp verinin olduđu durumda iki puanlayıcı için hesaplanan farklı uzlaşma katsayılarına ait değeri ve bu değeriin standart hataları tüm sorular için Ek-5 Tablo 30, 31 ve 32’de verilmiştir.



Şekil 9. Üç puanlayıcı için kayıp veri olduğu durumda okuduğunu anlama uygulaması soru 1 için kriterlere göre uzlaşma katsayıları ve standart hatalar (CK = Conger'in kappa, GAC2 = Gwet'in AC2, FK = Fleiss'nin kappa, KALFA = Krippendorff'un alfa, BP = Brennan-Prediger, UY = Uzlaşma yüzdesi)

Şekil 8 ve 9'a göre hem matematik okuryazarlığı hem de okuduğunu anlama için incelendiğinde kayıp verinin olduğu durumda üç puanlayıcı durum için uzlaşma katsayıları arasındaki ilişki kayıp veri %5, %10 ve %15 olduğu durumlarda kayıp veri olmayan

durumdakine benzer sonuçlar vermektedir. Veri setinde %15'e kadar kayıp veri bulunduğu durum için uzlaşma katsayılarından en yüksek değere ve en düşük standart hataya sahip olanın üç puanlayıcı kayıp veri bulunmadığı duruma benzer olarak uzlaşma yüzdesidir. Gwet's AC₂ ve Brennan-Prediger uzlaşma katsayılarının Conger's kappa, Fleiss' kappa ve Krippendorff alfa uzlaşma katsayılarına göre çoğu zaman daha yüksek değer aldığı söylenebilir. Diğer yandan Gwet's AC₂'nin uzlaşma yüzdesinden sonra nispeten en az standart hata değerine sahip uzlaşma katsayısı olduğu görülmüştür. Ayrıca kayıp değer miktarındaki artışa göre uzlaşma katsayıları incelendiğinde kayıp veri miktarının yaklaşık %5 ve %10 olduğu durumlar için tüm uzlaşma katsayılarının değerlerinin neredeyse aynı olduğu, standart hatalarındaki değişimin de yok denecek kadar az olduğu söylenebilir. Kayıp veri miktarının %15 olduğu durumla kayıp veri miktarının %5 ve %10 olduğu durum karşılaştırıldığında ise %15 olan durumda uzlaşma katsayılarının bir miktar düştüğü, standart hatalarda bir miktar artış olduğu görülmekle birlikte bu artışın yapılacak yorumları etkileyebilecek düzeyde olmadığı düşünülmektedir. Bu durumda uzlaşma katsayılarında %15'e kadar kayıp verinin olmasının yaklaşık 250 kişilik bir veri grubunda sorun yaratmadığı söylenebilir. Ayrıca kayıp veriyi egale etmek için yapılan matematiksel hesaplamaların yaklaşık 250 kişilik bir veri grubunda %15'e kadar kayıp verinin bulunması durumunda iyi bir şekilde çalıştığı söylenebilir. Bu durum daha küçük ve daha büyük örneklemelerde araştırılabilir. Aynı zamanda bu durum ikinci alt probleme göre bulunan kayıp verinin bulunduğu iki puanlayıcılı durum ile benzerlik göstermektedir.

Matematik Okurvazarlığı ve Okuduğunu Anlama Uygulamalarında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerlerinin Kayıp Veri Oranına Göre İncelenmesi

İki puanlayıcı için kayıp verinin olduğu ve olmadığı durumlar için hesaplanan sınıf içi korelasyon değerleri Tablo 13 ve 14'te verilmiştir. Tablolarda kayıp verinin bulunduğu durumun %5, %10 ve %15 olduğu duruma göre hesaplanan sınıf içi korelasyon değerlerine yer verilmiştir.

Tablo 13. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri

	Hedef sayısı	Ölçümlerin sayısı	Hata varyansı	ICC(3,1)	Güven aralığı
Kayıp veri yok	228	456	6,647	0,901	[0,873;0,922]
Kayıp veri %5	228	432	6,483	0,903	[0,872;0,922]
Kayıp veri %10	228	410	6,808	0,900	[0,863;0,916]
Kayıp veri %15	228	386	7,361	0,888	[0,843;0,905]

Tablo 14. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri

	Hedef sayısı	Ölçümlerin sayısı	Hata varyansı	ICC(3,1)	Güven aralığı
Kayıp veri yok	230	460	23,047	0,681	[0,604;0,744]
Kayıp veri %5	230	436	24,128	0,665	[0,584;0,731]
Kayıp veri %10	230	414	23,833	0,669	[0,589;0,736]
Kayıp veri %15	230	393	21,987	0,669	[0,587;0,736]

Tablolar incelendiğinde kayıp veri miktarının %5 ve %10 olduğu durum ile kayıp verinin olmadığı durum için hesaplanan sınıf içi korelasyon değerlerinin birbirine çok yakın olduğu ve neredeyse fark olmadığı söylenebilir. Ancak kayıp veri miktarı %15 olduğu durum için nispeten farklılaşma olduğu görülmektedir. Ayrıca hata varyansları incelendiğinde kayıp verinin olmadığı ve kayıp verinin %5 ile %10 olduğu durumlar için hata varyans değerleri birbirine çok yakinken kayıp veri miktarının %15 olduğu durum için hata varyansında da bir miktar artış olduğu görülmüştür ancak bu artışında çok düşük düzeyde olduğu söylenebilir. Bu durumda iki puanlayıcı için hesaplanan sınıf içi korelasyon katsayısında yaklaşık 250 kişilik bir veri grubunda %15 civarında kayıp veri olduğu veri seti için kayıp veri değerlerini manipüle edebildiği söylenebilir.

Matematik Okuryazarlığı ve Okuduğunu Anlama Uygulamalarında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerlerinin Kayıp Veri Oranına Göre İncelenmesi

Üç puanlayıcı için kayıp verinin olduğu ve olmadığı durumlar için hesaplanan sınıf içi korelasyon değerleri Tablo 15 ve 16'da verilmiştir. Tablolarda kayıp verinin bulunduğu durumun %5, %10 ve %15 olduğu duruma göre hesaplanan sınıf içi korelasyon değerlerine yer verilmiştir.

Tablo 15. Matematik Okuryazarlığı Uygulamasında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri

	Hedef sayısı	Ölçümlerin sayısı	Hata varyansı	ICC(3,1)	Güven aralığı
Kayıp veri yok	228	684	6,908	0,888	[0,862;0,909]
Kayıp veri %5	228	648	6,721	0,890	[0,863;0,910]
Kayıp veri %10	228	615	7,272	0,882	[0,850;0,901]
Kayıp veri %15	228	579	7,219	0,878	[0,840;0,894]

Tablo 16. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Sınıf İçi Korelasyon Değerleri

	Hedef sayısı	Ölçümlerin sayısı	Hata varyansı	ICC(3,1)	Güven aralığı
Kayıp veri yok	230	690	18,181	0,756	[0,708;0,799]
Kayıp veri %5	230	654	18,832	0,751	[0,702;0,795]
Kayıp veri %10	230	621	18,308	0,758	[0,709;0,801]
Kayıp veri %15	230	588	18,160	0,747	[0,695;0,792]

Tablolar incelendiğinde iki puanlayıcının olduğu duruma benzer olarak kayıp veri miktarının %5 ve %10 olduğu durum ile kayıp verinin olmadığı durum için hesaplanan sınıf içi korelasyon değerlerinin birbirine çok yakın olduğu ve neredeyse fark olmadığı söylenebilir. Ancak kayıp veri miktarı %15 olduğu durum için nispeten farklılaşma olduğu görülmektedir. Ayrıca hata varyansları incelendiğinde kayıp verinin olmadığı ve kayıp verinin %5 ile %10 olduğu durumlar için hata varyans değerleri birbirine çok yakınken kayıp veri miktarının %15 olduğu durum için hata varyansında da bir miktar artış olduğu görülmüştür ancak bu artışta çok düşük düzeyde olduğu söylenebilir. Bu durumda üç puanlayıcı için hesaplanan sınıf içi korelasyon katsayısında yaklaşık 250 kişilik bir veri grubunda %15 civarında kayıp veri olduğu veri seti için kayıp veri değerlerini manipüle edebildiği söylenebilir.

Puanlayıcı Sayısının Artmasının Puanlama Güvenirliğine Etkisi

Puanlayıcı sayısının artmasının puanlayıcı güvenirliğine etkisini incelemek için iki puanlayıcı (Tablo 10-17) ve üç puanlayıcı (Tablo 18-25) için kayıp verinin olmadığı ve olduğu durumlara ait uzlaşma katsayıları karşılaştırılarak incelenmiştir. Yapılan inceleme sonucunda bazı durumlarda üç puanlayıcı için hesaplanan uzlaşma katsayıları iki puanlayıcı için hesaplanan uzlaşma katsayısına göre düşüş gösterse de genel eğilim artış şeklinde olduğu söylenebilir. Diğer bir deyişle puanlayıcı sayısı ikiden üçe çıktığında yani arttığında uzlaşma katsayısı değerleri artmaktadır. Aynı zamanda tablolardaki standart hata değerleri

incelendiğinde üç puanlayıcılı durum için iki puanlayıcılı duruma göre standart hata değerlerinin daha fazla olduğu söylenebilir. Bunun nedeni ise üç puanlayıcılı durumda hem puanlayıcı hem de hedeflerden gelen hata dikkate alınırken iki puanlayıcılı durum için sadece hedeflerden gelen hata dikkate alınması olabilir. Diğer yandan sınıf içi korelasyon değerleri için iki puanlayıcı (Tablo 26-27) ve üç puanlayıcı (Tablo 28-29) karşılaştırıldığında pek fazla bir değişimin olmadığı görülmektedir. Matematik okuryazarlığı uygulaması için neredeyse hiçbir değişim olmadığı söylenebilirken, okuduğunu anlama için üç puanlayıcının olduğu durumdaki sınıf içi korelasyon değerlerinin iki puanlayıcının olduğu duruma göre nispeten daha yüksek olduğu söylenebilir. Bu durum uzlaşma katsayılarındakine benzerdir. Diğer yandan hata varyansı değerleri de göreceli olarak üç puanlayıcının olduğu durumda iki puanlayıcının olduğu duruma göre bir miktar fazladır. Genel olarak puanlayıcı sayısının artmasının puanlama güvenilirliğini arttırdığı şeklinde bir yorumlanabilir.

Özder (2012) beş puanlayıcının bulunmasının puanlama güvenilirliğini arttırdığı sonucu ortaya konmuştur. Diğer yandan Bıkmaz (2011) ise kappa, log-lineer analiz ve Krippendorff alfa tekniği için yapılan hesaplamalarda puanlayıcı sayısındaki artışın güvenilirlik düzeyine düşüşe neden olduğu ve en yüksek güvenilirlik değerlerinin iki puanlayıcının bulunduğu durumda elde edildiğini tespit etmiştir. Kanık vd. (2010) değerlendirici sayısının en az beş ve tanı kategorisi sayısının ona çıkması durumunda Krippendorff alfa katsayısının beklenen değerden daha büyük tahminler yaptığını gözlemlemiştir. Aktaş (2013) genellenebilirlik kuramı için puanlayıcı sayısının beş ve kategori sayısının iki olduğu durumda güvenilirlik kestirimlerinin en yüksek değeri verdiğini belirtmiştir. Çalışma kapsamında klasik test kuramına göre hesaplanan uzlaşma katsayılarında üç puanlayıcı için standart hata değerlerinin iki puanlayıcı için olanlardan daha yüksek olduğu tespit edilmiştir. Diğer yandan Aktaş (2013) genellenebilirlik kuramında elde edilen standart hata değerlerinin puanlayıcı sayısı arttıkça azaldığını tespit etmiş ve aynı zamanda en düşük standart hata değerine on puanlayıcının olduğu durumda ulaşıldığı bulgusuna ulaşmıştır.

KTK'ya Göre Puanlama Güvenirliği İçin Hesaplanan Güvenirlik Katsayılarının Sayısal ve Sözel Ders Olma Durumuna Göre Kıyaslanması

Puanlama güvenilirliği için hesaplanan uzlaşma katsayıları (Tablo 10-25) matematik okuryazarlığı uygulaması (sayısal ders olarak) ile okuduğunu anlama uygulamasından (sözel ders olarak) elde edilen değerler üzerinden incelendiğinde genel olarak matematik

okuryazarlığında puanlayıcıların uzlaşmasının daha yüksek olduğu ve elde edilen verilerin daha güvenilir olduğu söylenebilir. Sınıf içi korelasyon değerleri (Tablo 26-29) içinde aynı durum geçerlidir, yani matematik okuryazarlığında puanlayıcıların yaptığı değerlendirme üzerinden hesaplanan sınıf içi korelasyon değerleri okuduğunu anlama uygulamasına göre daha yüksektir. Bu durumda puanlama güvenilirliğinde sayısal derslerde yapılan işlemlerin daha güvenilir sonuçlar verdiği söylenebilir. Pantzare (2015) matematik dersinde yapılan uygulamalara yönelik değerlendirmelerin daha güvenilir olarak çıkmasının nedeni çoğu matematik maddesinin çözümünün belirli bir mantık izlediğini ve bu durumun açık puanlama DPA'larının hazırlanmasını kolaylaştırdığını ve bu nedenle güvenilir puanlama olasılığının arttırdığı şeklinde yorumlamıştır. Bu durumun ayrıca sözel derslerin yoruma açık olmasından dolayı puanlayıcıların soruları farklı yorumlamalarından kaynaklanmış olabilir. Murphy (1982) ve Newton (1996) farklı okul konuları ile ilgili yapılan puanlayıcı güvenilirliği çalışmalarında matematikle ilgili yapılan değerlendirmelerin daha güvenilir olarak bulunduğunu göstermiştir (Murphy, 1982 ve Newton, 1996'dan aktaran Pantzare, 2015).

BÖLÜM 5

SONUÇ VE ÖNERİLER

Sonuçlar

Uzlaşma katsayıları arasındaki ilişki hem iki puanlayıcı hem de üç puanlayıcının bulunduğu durumlar için kayıp veri durumu olup olmasına göre benzer sonuçlar göstermektedir. Her zaman en yüksek uzlaşma değerine ve en düşük hataya sahip olan uzlaşma katsayı uzlaşma yüzdesidir. Bunun nedeni ise uzlaşma yüzdesi hesaplanırken şans yoluyla uzlaşmaya yönelik herhangi bir düzeltme yapılmamasıdır. Genel olarak Gwet's AC₂ uzlaşma katsayısının uzlaşma yüzdesinden sonra en yüksek değere sahiptir. Ayrıca Gwet's AC₂ ile Brennan-Prediger uzlaşma katsayılarının genel olarak birbirine çok benzer sonuçlar vermektedir. Aynı şekilde Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayıları birbirlerine çok yakın değerler vermektedir. Bu katsayıların birbirine yakın değerler vermesinin formüllerindeki benzerliklerden dolayı olduğu ortaya konulmuştur. Yapılan uygulama sonucunda ise Gwet's AC₂ ile Brennan-Prediger uzlaşma katsayılarının genel olarak Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayılarından daha yüksek değerler aldığı tespit edilmiştir.

Kayıp değer miktarına göre uzlaşma katsayılarına bakıldığında ise hem uzlaşma katsayılarında hem de sınıf içi korelasyon katsayısında kayıp veri olmadığı durum ile kayıp verinin %5 ve %10 olduğu durumlarda çok benzer sonuçlar göstermektedir. Kayıp verinin %15 olduğu durum için nispeten farklılıklar görülse de bu farklılıklar elde edilen değerler üzerinden yapılan yorumu etkileyecek kadar büyük değildir. Bu durumda uzlaşma yüzdesi, Gwet's AC₂, Brennan-Prediger, Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayıları ve sınıf içi korelasyon katsayılarını kayıp veri ile baş

etme yöntemi olarak pairwise kullandığında 250 kişilik bir hedef grup için yaklaşık %15'e kadar olan kayıp verinin üstesinden gelebilmektedir.

İki puanlayıcı ve üç puanlayıcının bulunduğu durum için uzlaşma katsayısı ve sınıf içi korelasyon değerleri karşılaştırıldığında üç puanlayıcının bulunduğu durum için değerler genel olarak daha yüksektir. Diğer bir deyişle, puanlayıcı sayısı artıkça puanlama güvenilirliği ait katsayı değerleri artmaktadır.

Sayısal (matematik okuryazarlığı) ve sözel ders (okuduğunu anlama) açısından puanlayıcı güvenilirliği katsayıları incelendiğinde ise genel olarak matematik okuryazarlığı uygulaması için hesaplanan puanlayıcılar arası güvenirlilik katsayısı değerleri daha yüksektir. Bu durumun sayısal derslerdeki sorulara verilen cevapların kişiden kişiye fazla değişmeyen net yanıtlarının olmasından dolayı sözel derslere göre daha kolay puanlandırılabilirdiği sonucunu desteklemektedir.

Öneriler

Çalışma kapsamında elde edilen buğular ışığında uygulayıcılara ve araştırmacılara aşağıdaki önerilerde bulunulmuştur:

- Puanlama güvenilirliğine ait uzlaşma katsayıları birbirinden farklı değerler sunduğu için ihtiyaç elde edilen verilerin güvenilirliği belirlenirken en az iki uzlaşma katsayısı raporlanabilir. Özellikle seçilen uzlaşma katsayıları birbirinden farklılaşan katsayılardan seçilmesi önerilmektedir. Örneğin Gwet's AC₂ ile Brennan-Prediger katsayısı benzer sonuçlar verdiği için bunlardan birinin; Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa ve Krippendorff alfa uzlaşma katsayıları benzer sonuçlar verdiği için bunlardan birinin seçilmesi uygun görülmektedir.
- Sözel derslere ait puanlama güvenilirliği katsayıları hesaplanırken DPA'nın hazırlanmasına dikkat edilmesi, puanlayıcılara puanlama konusunda konuya özgü eğitimler verilmesi önerilmektedir.
- Üçten fazla puanlayıcının olduğu durum için araştırma tekrarlanabilir, en uygun uzlaşma değerinin ne kadar puanlayıcı ile elde edilebileceği araştırılabilir.
- Kayıp veri durumu için daha küçük ya da daha büyük örneklemeler açısından nasıl bir değişim olabileceği incelenebilir.

- Bu çalışma kapsamında kayıp veri oranı dikkate alınmıştır, kayıp veri başa çıkma yöntemlerine yönelik bir karşılaştırma yapılmamıştır. Puanlama güvenirliğinde kayıp veri ile başa çıkma yöntemlerinin karşılaştırmasını yapan bir çalışma yapılabilir.
- Soru düzeylerine göre puanlama güvenirliği katsayıları arasındaki ilişkinin nasıl olduğu yönünde bir inceleme yapılmamıştır, bu yönde çalışmalar yapılması alan yazına katkı yapabilir.

KAYNAKLAR

- AGREESTAT (2015). *AgreeStat 2015.1 for Excel Windows/Mac User's Guide*. Advanced Analytics, LLC: USA.
- Aktaş, M. (2013). *Aynı performans görevinin farklı sayıda puanlayıcılar tarafından üç farklı teknikle puanlanmasından elde edilen puanların güvenilirliklerinin genellenebilirlik kuramına göre incelenmesi*. Yüksek Lisans Tezi. Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Mersin.
- Atılgan, H. (2005). Genellenebilirlik kuramı ve puanlayıcılar arası güvenirlilik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama*, (4) 7. 95-108
- Bal, C. (2003). *Çok gruplu verisetlerinde eksik gözlem sorununun çözümlenmesi ve sağlık alanında bir uygulama*. Doktora Tezi. Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Eskişehir.
- Baştürk, R., & Işıkoğlu, N. (2008). Okul öncesi eğitim kurumlarının işlevsel kalitelerinin çok-yüzeyli Rasch ölçme modeli ile analizi. *Kurum ve Uygulamada Eğitim Bilimleri*, 8(1),7-32.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. 2. Baskı. Pegem Akademi: Ankara.
- Bıkmaz, Ö. (2011). *Üst düzey zihinsel özelliklerin ölçülmesinde puanlayıcılar arası güvenirlilik belirleme tekniklerinin karşılaştırılması*. Yüksek Lisans Tezi. Hacettepe Üniversitesi Sosyal bilimler Enstitüsü, Ankara.

- Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., & Hickmott, J. (2009). Examining design and inter-rater reliability of rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation*.
- Büyükökçü, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenirliliğin 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. (2012). *Bilimsel araştırma yöntemleri*. 11. Baskı. Pegem Akademi: Ankara.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. XX(1).
- Cohen, R. J., & Swerdlik, M. E. (2013). *Psikolojik Test ve Değerleme Testlere ve Ölçmeye Giriş*. (G. Gözen, Çev). Ezel Tavşancıl (Çev. Ed.). Ankara: Nobel.
- Crocker, L., & Algina, J. (2006). *Introduction to classical & modern test theory*. United States of America.
- Dates, B. G., & King, J. E. (2008). SPSS algorithms for bootstrapping and Jackknifing generalized measures of agreement. *Annual meeting of the Southwest Educational Research Association, New Orleans, LA, February 6-9*. http://www.ccitonline.org/jking/homepage/bootstrap_paper.doc adresinden 19.07.2015 tarihinde indirilmiştir.
- Demir, E., & Özbaşı, D. (2013). Çoktan Seçmeli Testlerde Kayıp Veri Sorunu: SBS Örneği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 76-101.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382.
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve rasch modeli üzerine bir araştırma*. Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N., Kaya Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik Kuramı*. Ankara: Pegem Akademi.
- Gwet, K. (2001). *Handbook of inter-rater reliability*. Gaithersburg: STATAXIS.

- Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods For Inter-Rater Reliability Assessment*. No. 1.
- Gwet, K. L. (2008a). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 61, 29-48.
- Gwet, K. L. (2008b). Variance estimation of nominal-scale inter-rater reliability with random selection of raters. *Psychometrika*. 73(3). 407-430. DOI: 10.1007/s11336-007-9054-8.
- Gwet, K. L. (2011). On the krippendorff's alpha coefficient. *Manuscript Submitted for Publication in Communication Methods and Measures*.
- Gwet, K. L. (2012). *Handbook of inter-rater reliability*. 3rd ed. Gaithersburg: Advanced Analytics, LLC.
- Hambleton, R. K. & Swaminathan, H. (1991). *Item response theory principles and applications*. Springer Science & Business Media, LLC
- Izquierdo, M.C., & Pedrero, E.F. (2014). Estimating the reliability coefficient of tests in presence of missing values. *Psicothema*. Vol. 26. No.4. 516-523. DOI: 10.7334/psicothema2014.98.
- Kan, A. (2014). Eğitimde ölçme ve değerlendirme. H. Atılgan (Ed). *Ölçme araçlarında bulunması gereken nitelikler* içinde (s.23-80). 7. Baskı. Ankara: Anı Yayıncılık.
- Kanık, E. A., Orekici Temel, G., & Ersöz Kaya, İ. (2010). Fleiss kappa ve Krippendorff alpha uyum katsayılarının örneklem genişliği, değerlendirici sayısı ve kullanılan ölçeğin kategori sayısından etkilenme durumları üzerine bir benzetim çalışması. *Türkiye Klinikleri J Biostat*, 2(2), 74-81.
- Kılıç, S. (2009). *Ölçümlerin uyumluluğu ve tıptaki uygulamaları*. Yüksek Lisans Tezi. Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü, Adana.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2010). *Öğrenci başarısının belirlenmesi performansa ve portfolyoya dayalı durum belirleme*. 3. Baskı. Pegem Akademi: Ankara.
- Krippendorff, K. (2004). *Content analysis: an introduction to its methodology*. 2nd ed. USA: Sage Publications.

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Laxshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Linacre, J. M. (1993). Generalizability theory and many facet Rasch measurement. *Annual Meeting Of The American Educational Research Association*. (April, 13, 1993), (ED 364 573). Atlanta Georgia.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46.
- Moskal, B. M, & Leydens, J. A. (2000). Scoring rubric development, validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10).
- Mushquash, C. & O'Connor, B. P. (2006). SPSS and SAS Programs for Generalizability Theory Analysis. *Behavior Research Methods*. 38(3),542-547.
- Newell, J. A., Dahm, K. D., & Newell, H. L. (2002). Rubric development and inter-rater reliability issues in assessing learning outcomes. *Proceeding of the 2002 American Society for Engineering Education Annual Conference & Exposition*.
- Özçelik, D. A. (2013a). *Test hazırlama kılavuzu*. 5. Baskı. Pegem Akademi: Ankara.
- Özçelik, D. A. (2013b). *Okullarda ölçme ve değerlendirme öğretmen el kitabı*. 2. Baskı. Pegem Akademi: Ankara.
- Özder, H. (2012). Öğretmen adaylarının seçiminde uygulanan mülakat sınavının geçerlik ve güvenirliği: KKTC örneği. *Eğitim ve Bilim*, 37(166).
- Pantzare, A. L. (2015). Interrater reliability in large-scale assessments-can teachers score national test reliably without external controls?. *Practical Assessment, Research & Evaluation*, 20(9).
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*. 35(4) , 435-448.
- Semerci, Ç., Semerci, N., & Duman, B. (2013). Yüksek lisans öğrencilerinin seminer sunu performanslarının çok-yüzeyle rasch modeli ile analizi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 25, 7-22.

- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*. 86(2), 420-428.
- Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. (2013). A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Medical Research Methodology*. 13 (61), <http://www.biomedcentral.com/1471-2288/13/61>.
- Tan, Ş. (2013). *Öğretimde ölçme ve değerlendirme KPSS el kitabı*. 8. Baskı. Ankara: Pegem Akademi.
- Tekindal, (1998). *Klasik yazılı sınavla ve çok sorulu testlerle elde edilen ölçümlerin güvenirlik ve geçerliği*. IV. Ulusal Sınıf Öğretmenliği Sempozyumu, Pamukkale Üniversitesi, Denizli.
- Turgut, F. M., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. 4. Baskı. Pegem Akademi: Ankara.
- Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*. 37(5), 360-3.
- Yelboğa, A. (2008). Güvenirliğin Değerlendirilmesinde Genellenebilirlik Kuramı'nın Kullanılması: Endüstri ve Örgüt Psikolojisinde Bir Uygulama. *Psikoloji Çalışmaları Dergisi*, 28, 35-54.
- Yelboğa, A, & Tavşancıl, E. (2010). Klasik test ve genellenebilirlik kuramına göre güvenirliğin bir iş performansı ölçeği üzerinden incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10 (3), 1825-1854.
- Yıldıztekin, B. (2014). *Klasik test kuramı ve genellenebilirlik kuramından puanlayıcılar arası tutarlılığın farklı yöntemlere göre karşılaştırılması*. Yüksek Lisans Tezi. Hacettepe Üniversitesi Eğitim Bilimler Enstitüsü, Ankara.
- Yurdugül, H. (2005). *Ölçek Geliştirme Çalışmalarında Kapsam Geçerliği için Kapsam Geçerlik İndekslerinin Kullanılması*. XIV. Ulusal Eğitim Bilimleri Kongresi Pamukkale Üniversitesi Eğitim Fakültesi, 28-30 Eylül, Denizli <http://www.ua.gov.tr/basin-odasi/yayinlar> adresinden 18.06.2015 tarihinde indirilmiştir.

- Yurdugöl, H., Bayrak, F. (2012). Ölçek Geliştirme Çalışmalarında Kapsam Geçerlik Ölçüleri: Kapsam Geçerlik İndeksi ve Kappa İstatistiğinin Karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi. Özel Sayı 2*. 264-271.
- Yüzüak, A. V., Yüzüak, B., & Kaptan, F (2015). Performans görevinin akran gruplar ve öğretmen yaklaşımları doğrultusunda çok-yüzeyli Rasch ölçeği modeli ile analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 1-11.

EK-1:

MATEMATİK OKURYAZARLIĞI UYGULAMA FORMU

**MATEMATİK OKURYAZARLIĞI
UYGULAMASI**

AD: _____

SOYAD: _____

SINIF: _____

OKUL: _____

YAŞ: _____

CİNSİYET: ☐ K ☐ E

AÇIKLAMA

- Bu uygulama sizin matematik okuryazarlık becerinizi ölçmeye yönelik olarak hazırlanmıştır.
- Uygulama içerisinde beş tane açık uçlu soru bulunmaktadır.
- Çözümleri soruların hemen altında yer alan boşluklara yapabilirsiniz.
- Soruların cevaplarını yazarken istediğiniz renkte ve türde (kurşun, tükenmez vb.) kalem kullanabilirsiniz.
- Uygulama süresi 1 ders saatidir.
- Tüm soruları yanıtlamanız beklenmektedir.
- Bu sorulara vereceğiniz yanıtlar bilimsel çalışmalar için kullanılacaktır.

Başarılar Dilerim.



Resimde, yürüyen bir erkeğin ayak izleri görülüyor. Adım uzunluğunu gösteren “p”, ardışık iki ayak izinin topukları arasındaki mesafedir.

Erkekler için, “n” ile “p” arasındaki ilişki yaklaşık olarak $\frac{n}{p} = 140$ formülü ile gösterilmektedir.

Burada;

n= bir dakikadaki adım sayısını,

p= metre cinsinden adım uzunluğunu göstermektedir.

SORU 1:

Dakikada 70 adım atarak yürüyen Hakkı’ya bu formül uygulandığında, Hakkı’nın bir adım uzunluğu ne olur? İşleminizi gösteriniz.

SORU 2:

Burak, adım uzunluğunun 0,80 metre olduğunu biliyor. Formül Burak'ın yürüyüşüne uygulanabilir.

Burak'ın bir dakikadaki yürüme hızını metre olarak ve bir saatteki yürüme hızını kilometre olarak hesaplayınız. İşlemlerinizi gösteriniz.

Singapur'dan Mei-Ling karşılıklı öğrenci değişim programından yararlanarak 3 ay süreyle Güney Afrika'ya gitmek için hazırlık yapıyordu. Onun, bir miktar Singapur Dolarını (SGD) Güney Afrika para birimi olan "Rand" (GAR) çevirmesi gerekti.

SORU 3:

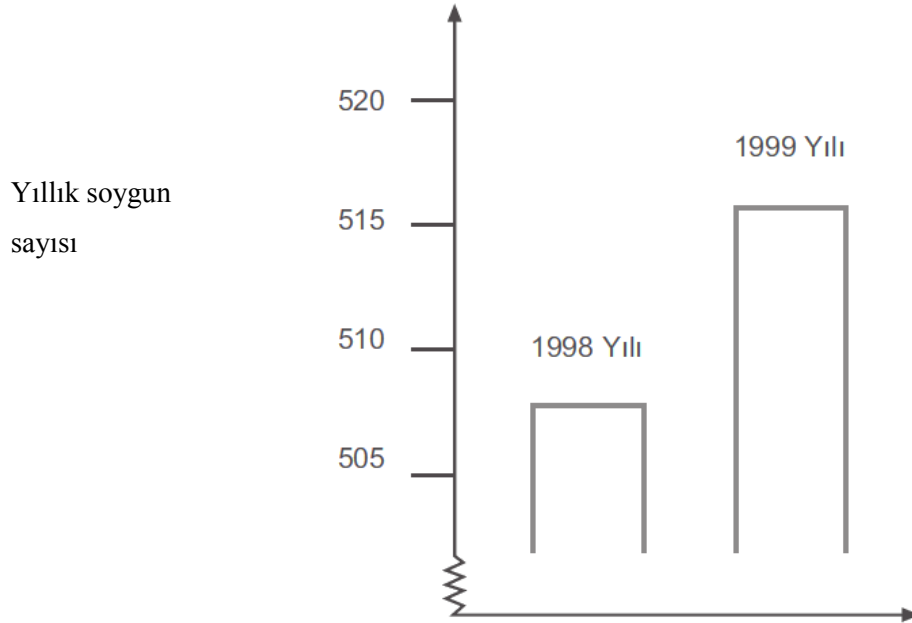
Bu üç ay süresince döviz kuru oranı bir SGD için 4,2 den 4,0 GAR'a değişmiştir.

Mei-Ling Güney Afrika randını yeniden Singapur dolarına çevirdiğinde, döviz kurunun 4,2 GAR yerine 4,0 GAR olması Mei-Ling'in yararına mı olmuştur? Yanıtınızı destekleyecek bir açıklama yazınız.

SORU 4:

Bir televizyon muhabiri, bu grafiđi gösterdi ve řöyle dedi:

Bu grafik 1998 yılından 1999'a kadar soygunların sayısında çok büyük bir artış olduğunu göstermektedir.

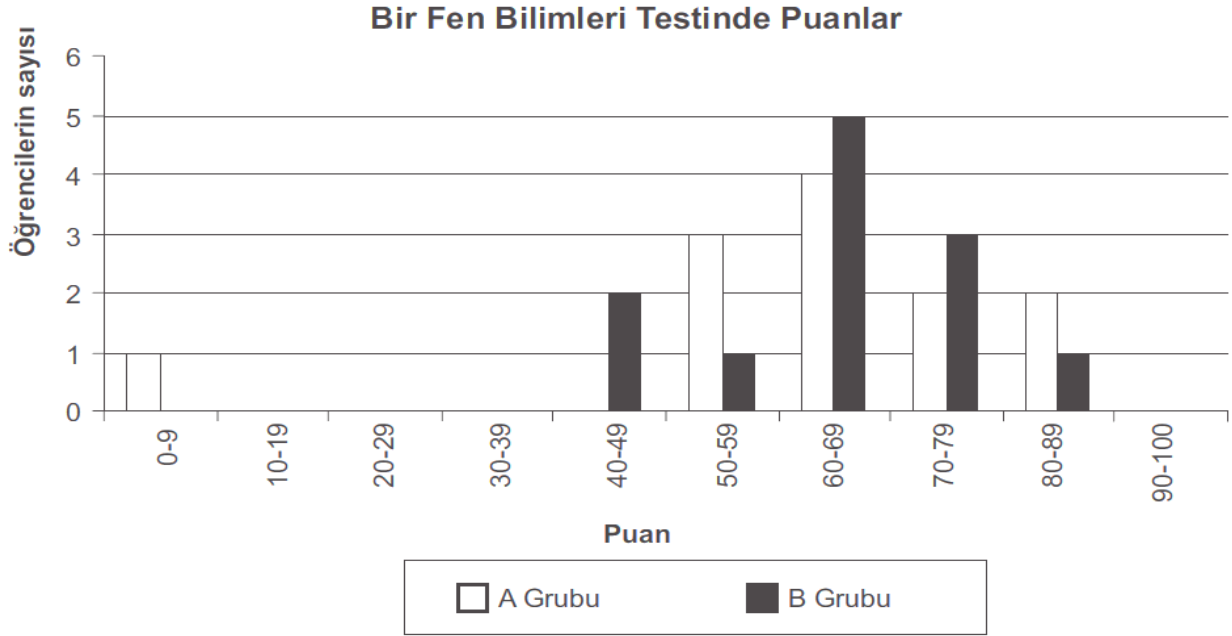


Muhabirin sözlerinin grafiđin kabul edilebilir bir yorumu olduğunu düşünüyor musunuz? Yanıtınızı desteklemek için bir açıklama yapınız.

SORU 5:

Aşağıdaki grafik, A Grubu ve B Grubu olarak adlandırılan iki grubun bir fen bilimleri testinde aldıkları puanları göstermektedir.

A Grubu için ortalama 62,0 ve B Grubu için ortalama 64,5'tir. Puanları, 50 ya da daha fazla olan öğrenciler, bu testten geçerler.



Bir öğretmen, grafiğe bakarak bu testte B Grubunun A Grubundan daha başarılı olduğunu ileri sürmektedir.

A Grubundaki öğrenciler, öğretmenleriyle aynı düşüncede değildir. Onlar, B grubundaki öğrencilerin, daha başarılı sayılmamaları gerektiği konusunda öğretmenlerini inandırmaya çalışıyorlar.

Grafiği kullanarak A grubundaki öğrencilerin kullanabileceği matematiksel bir dayanak veriniz.

EK-2:

OKUDUĞUNU ANLAMA UYGULAMA FORMU

**OKUDUĞUNU ANLAMA
UYGULAMASI**

AD: _____

SOYAD: _____

SINIF: _____

OKUL: _____

YAŞ: _____

CİNSİYET: ☐ K ☐ E

Haftada okudunuz kitap sayfa sayısı: _____

Ayda okuduğunuz kitap sayısı: _____

Yılda okuduğunuz kitap sayısı: _____

AÇIKLAMA

- Bu uygulama sizin okuduğunuzu anlama becerinizi ölçmeye yönelik olarak hazırlanmıştır.
- Uygulama içerisinde altı tane açık uçlu soru bulunmaktadır.
- Çözümleri soruların hemen altında yer alan boşluklara yapabilirsiniz.
- Soruların cevaplarını yazarken istediğiniz renkte ve türde (kurşun, tükenmez vb.) kalem kullanabilirsiniz.
- Uygulama süresi 1 ders saatidir.
- Tüm soruları yanıtlamanız beklenmektedir.
- Bu sorulara vereceğiniz yanıtlar bilimsel çalışmalar için kullanılacaktır.

Başarılar Dilerim.

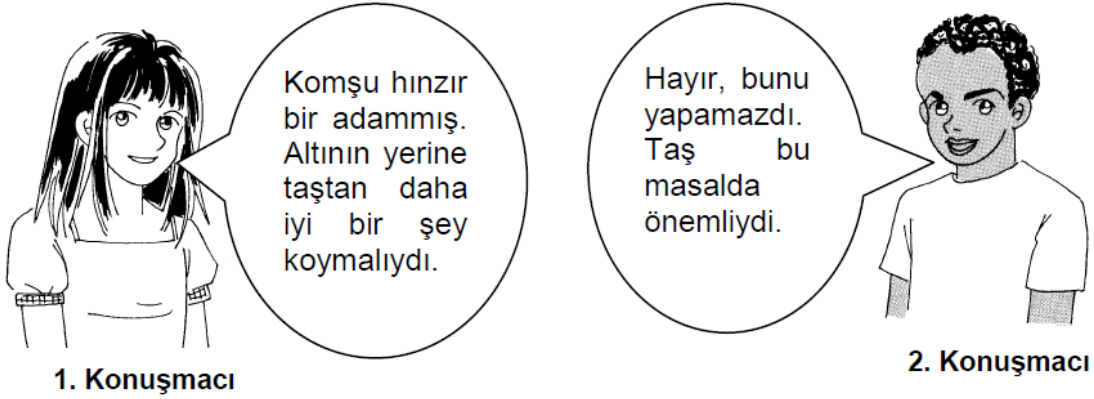
CİMRİ VE ALTINI

Bir Ezop masalı

Cimri bir adam, bütün mal varlığını satıp bir külçe altın almış ve bu altını eski bir duvarın dibinde açtığı bir çukura gömmüş. Adam, her gün gidip altınına bakarmış. İşçilerinden biri, cimrinin bu yeri sık sık ziyaret ettiğini fark etmiş ve davranışlarını izlemeye karar vermiş. İşçi kısa bir süre sonra, gizli hazinenin sırrını keşfetmiş, burayı kazmış ve çıkan altını çalmış. Cimri, ertesi gün dönünce altının yerinde yeller estiğini görmüş, ağlayarak saçını başını yolmaya başlamış. Onu böyle perişan gören komşusu nedenini öğrenince şöyle demiş: "Üzme kendini bu kadar, git bir taş al, aynı çukura koy ve altının hâlâ orada olduğunu hayal et. Çünkü altın çukurda iken kullanmayı hiç düşünmediğine göre, altın nasıl olsa senin değildi. Taş da aynı işi görecektir."

SORU 1:

Aşağıda "Cimri ve Altını" masalını okuyan iki kişi arasında geçen konuşmanın bir bölümü yer almaktadır.



2. Konuşmacı kendi görüşünü desteklemek için ne söyleyebilirdi?

KAN ARANIYOR



Kan bağıışı gereklidir.

İnsan kanının yerini tam olarak tutabilecek başka bir madde yoktur. Kan bağıışı, bu nedenle çok önemlidir ve yaşamları kurtarmada onun yeri doldurulamaz.

Fransa'da, her yıl, 500,000 hasta kan naklinden yararlanmaktadır.

Kan almak için kullanılan araçlar sterildir ve tek kullanımlıktır. (şırıngalar, tüpler, kan torbaları).

Kan vermenizde herhangi bir tehlike yoktur.

Kan bağıışı:

Kan bağıışı en çok bilinen bağıış şeklidir ve süresi 45 dakika ile 1 saat arasında değişmektedir.

450-ml'lik torbaların yanı sıra testlerin ve kontrollerin yapılması için birkaç küçük numune de alınır.

- Bir erkek yılda beş defa, bir kadın üç defa kan verebilir.
- Kan bağıışı yapanlar 18 ile 65 yaş arasında olabilir.

Her bağıış arasında geçmesi gereken zorunlu süre 8 haftadır.

SORU 2:

Son on iki ayda iki defa kan veren, on sekiz yaşındaki bir kadın tekrar kan vermek istiyor. "Kan Aranıyor" adlı metne göre, onun kan vermesine hangi koşulda izin verilir?

SORU 3:

Metinde: “Kan almak için kullanılan araçlar sterildir ve tek kullanımlıktır...” denilmektedir.

Bu metinde bu bilgi neden yer almaktadır?

CEP TELEFONU GÜVENLİĞİ

Cep telefonları tehlikeli midir?

Önemli Nokta

Cep telefonlarının yol açtığı sağlık problemleriyle ilgili birbiriyle çelişen raporlar 1990ların sonunda ortaya çıkmıştır.

Önemli Nokta

Cep telefonlarının etkilerini araştırmak için yapılan bilimsel araştırmalara şimdiye kadar milyonlarca YTL yatırım yapılmıştır.

Evet	Hayır
1. Cep telefonlarından yayılan radyo dalgaları, vücut dokularında ısı artışına neden olabilir, bu nedenle, zararlı etkileri olabilir.	Radyo dalgaları vücuda ısı kaynaklı zarar verecek kadar güçlü değildir.
2. Cep telefonları tarafından yaratılan manyetik alanlar, vücut hücrelerinin çalışma biçimini etkileyebilir.	Manyetik alanlar son derece küçüktür ve bu nedenle vücudumuzdaki hücreleri etkileme ihtimali yoktur.
3. Cep telefonu ile uzun konuşmalar yapan insanlar bazen yorgunluk, baş ağrısı ve konsantrasyon kaybından şikayetçi olurlar.	Bu etkiler laboratuvar koşulları altında hiç gözlemlenmedi ve modern yaşam tarzlarındaki diğer nedenlerden kaynaklanıyor olabilirler.
4. Cep telefonu kullanıcılarının beyinlerinin cep telefonuna maruz kalan bölümünde kanser gelişmesi riski 2,5 kat daha fazladır.	Araştırmacılar bu artışın nedeninin cep telefonu kullanımıyla ilişkisinin kesin olmadığını kabul etmektedirler.
5. Uluslar arası Kanser Araştırmaları Derneği, çocukluk yıllarında gelişen kanser ile yüksek gerilim hatları arasında bir ilişki olduğunu ortaya çıkarmıştır. Yüksek gerilim hatları da cep telefonları gibi radyasyon yaymaktadır.	Yüksek gerilim hatları tarafından üretilen radyasyon, cep telefonlarından gelen enerjiden çok daha fazla enerjiye sahip, farklı bir tür radyasyondur.
6. Cep telefonlarındakine benzeyen radyo frekans dalgaları, iplik kurtlarının gen dizilimini değiştirmiştir.	İplik kurtları insan değildir. Bu nedenle, bizim beyin hücrelerimizin aynı şekilde tepki vereceği kesin değildir.

Önemli Nokta

Cep telefonu kullanıcılarının sayısının ne kadar çok olduğu düşünülürse, telefonların yaratacağı küçük bir yan etki bile büyük bir halk sağlığı sorunu yaratabilir.

Önemli Nokta

2000 yılında yayınlanan Stewart Raporu (İngiltere kaynaklı bir rapor) cep telefonları kaynaklı bilinen bir sağlık problemi olmadığını buldu fakat daha fazla araştırma yapılana kadar özellikle gençlerin önlem almalarını önerdi. 2004 yılında yayınlanan bir sonraki rapor, önceki rapor sonuçlarını desteklemiştir.

Eğer bir cep telefonu kullanıyorsanız ...

Şunları Yapınız

Aramaları kısa tutunuz.

Cep telefonunuzu bekleme konumundayken vücudunuzdan uzakta tutunuz.

Konuşma süresi uzun olan cep telefonları satın alın. Bu daha verimli olur ve daha az radyasyon yayar.

Şunları Yapmayınız

Telefon baz istasyonu ile iletişim için daha fazla güce ihtiyaç duyacağı ve bu nedenle daha fazla radyo dalgası yayacağı için çekim gücü zayıf olduğunda cep telefonunuzu kullanmayınız.

"SAR" değeri¹ yüksek olan cep telefonu satın almayınız. SAR değerinin yüksek olması, telefonun daha fazla radyasyon yaydığı anlamına gelir.

Bağımsız bir kuruluş tarafından test edilmeyen sözde koruyucu cihazları satın almayın.

SORU 4:

Metinde yer alan **önemli noktaların** amacı nedir?

SORU 5:

“Bir şeyin kesinlikle başka bir şeye sebep olduğunu kanıtlamak zordur.”

Verilen bu bilginin **Cep telefonları tehlikeli midir?** adlı tablonun 4. maddesinde yer alan **Evet** ve **Hayır** ifadeleriyle ilişkisi nedir?

SORU 6:

Tablonun **Hayır** sütununda yer alan 3. maddeye bakınız. Burada sözü edilen “diğer nedenlerden” biri ne olabilir? Cevabınızı bir neden yazarak açıklayınız.

EK-3:

DERECELİ PUANLAMA ANAHTARLARI

Matematik Okuryazarlığı Soru 1-2 Dereceli Puanlama Anahtarı

Ölçütler	Performans Düzeyleri				Başarı Puanı
	0	1	2	3	
İlişkileri Belirleme	Cevap yok ya da tamamen ilgisiz.	Soruda verilen ilişkinin adım ve dakika arasındaki ilişki olduğu anlayamamış.	Soruda verilen ilişkinin adım ve dakika arasındaki ilişki olduğunu anlamış ancak ilişkinin nasıl kullanılacağını kavrayamamış.	Soruda verilen ilişkinin adım ve dakika arasındaki ilişki olduğunu anlamış ve ilişkinin nasıl kullanılacağını kavramış.	
İşlemleri Gerçekleştirme	Cevap yok ya da tamamen ilgisiz.	İşlemleri yaparken hangisinin ne ile ilgili olduğu anlaşılmıyor, çok karmaşık bir sıra ile yazmış.	İşlemleri yaparken hangi işlemin hangi amaca hizmet ettiği anlaşılmasına rağmen çok karmaşık bir sıra ile yazmış.	İşlemleri yaparken hangi işlemin hangi amaca hizmet ettiği anlaşılıyor ve sıralı bir düzen içerisinde yazmış.	
Sonuca Ulaşma	Cevap yok ya da tamamen ilgisiz.	Yaptığı işlemleri tamamlayamamış, sonuca ulaşamamış.	Yaptığı işlemleri tamamlamış ancak yanlış sonuca ulaşmış.	Yaptığı işlemleri tamamlamış ve doğru sonuca ulaşmış.	

Matematik Okuryazarlığı Soru 3 Dereceli Puanlama Anahtarı

	Performans Düzeyleri				Başarı Puanı
Ölçütler	0	1	2	3	
Sayıların ifade ettiği değeri anlama	Cevap yok ya da tamamen ilgisiz.	Hangi sayının hangi para birimine yönelik olduğunu anlasa da bunu işlemlerinde ve açıklamalarında gösterememiştir.	Hangi sayının hangi para birimine yönelik olduğunu anlasa da bunu işlemlerinde ya da açıklamalarında gösterememiştir.	Hangi sayının hangi para birimine yönelik olduğunu anlamış ve bunu işlemlerinde ya da açıklamalarında doğru bir şekilde ortaya koymuştur.	
İlişkilendirme	Cevap yok ya da tamamen ilgisiz.	Para birimleri arasındaki ilişkiyi gözlemleyememiş.	Para birimleri arasındaki ilişkiyi gözlemlemiş ancak yeterince ifade edememiş.	Para birimleri arasındaki ilişkiyi doğru bir şekilde kurmuştur.	
Nedenini Açıklama	Cevap yok ya da tamamen ilgisiz.	Yaptığı açıklamayı doğru bir şekilde para birimleriyle kıyaslayarak açıklayamamıştır.	Açıklama yaparken para birimlerinin dönüşümlerini kıyaslasa da doğru bir şekilde ifade edememiştir.	Açıklama yaparken para birimlerinin dönüşümlerini kıyaslayarak da doğru şekilde ifade etmiştir.	

Matematik Okuryazarlığı Soru 4-5 Dereceli Puanlama Anahtarı

	Performans Düzeyleri				Başarı Puanı
Ölçütler	0	1	2	3	
Grafiği Okuma	Cevap yok ya da tamamen ilgisiz.	Grafikte verilmek istenen bilginin ne olduğunu anlayamamıştır.	Grafikle verilmek istenen bilgiyi anlamasına rağmen cevap verirken doğru bir şekilde ifade edememiştir.	Grafikle verilmek istenen bilgiyi tam olarak anlamış ve cevabında tam olarak ifade etmiştir.	
İlişkilendirme	Cevap yok ya da tamamen ilgisiz.	Verilen açıklama ve bilgiler arasındaki ilişkiyi kuramamıştır.	Verilen açıklama ve bilgiler arasındaki ilişkiyi anlamış gözükmesine rağmen tam anlamıyla ifade edememiştir.	Verilen açıklama ve bilgiler arasındaki ilişkiyi tam ve doğru bir şekilde kurmuş ve açıklamıştır.	
Neden Belirtme	Cevap yok ya da tamamen ilgisiz.	Belirttiği görüşe yönelik açıklama yapamamıştır.	Belirttiği görüşe yönelik açıklamalarını verilen açıklama ve bilgilere dayandıramamıştır.	Belirttiği görüşe yönelik açıklamalarını verilen açıklama ve bilgilere dayandırarak tam ve doğru bir şekilde yapmıştır.	

Okuduğunu Anlama Soru 1-2-4 Dereceli Puanlama Anahtarı

Ölçütler	Performans Düzeyleri				Başarı Puanı
	0	1	2	3	
İlişkilendirme	Cevap yok ya da tamamen ilgisiz.	Soruda açıklaması beklenen durum ile metin arasında ilişki kuramamıştır.	Soruda açıklaması beklenen durum ile metni ilişkilendirmesine rağmen bu ilişkiye yönelik açıklamayı uygun olarak yanıtında ifade edememiştir.	Soruda açıklaması beklenen durum ile metni ilişkilendirmiş ve bu ilişkiye yönelik açıklamasını uygun bir şekilde yapmıştır.	
Kıyaslama	Cevap yok ya da tamamen ilgisiz.	Metinde yer alan bilgi ile soruda yer alan bilgiyi kıyaslayamamıştır.	Metinde yer alan bilgi ile soruda yer alan bilgiyi kıyaslamaya çalışmış ancak tam olarak açıklayamamıştır.	Metinde yer alan bilgi ile soruda yer alan bilgiyi kıyaslamış ve tam olarak açıklamıştır.	
Parça-Bütün ilişkisini Anlama	Cevap yok ya da tamamen ilgisiz.	Verilen durumun metnin bütünündeki yerini anlamamıştır.	Verilen durumun metnin bütünündeki yerini anlamasına rağmen doğru bir şekilde ifade edememiştir.	Verilen durumun metnin bütünündeki yerini belirlemiş ve doğru bir şekilde ifade etmiştir.	

Okuduğunu Anlama Soru 3-5-6 Dereceli Puanlama Anahtarı

	Performans Düzeyleri				Başarı Puanı
Ölçütler	0	1	2	3	
Yansıtma	Cevap yok ya da tamamen ilgisiz.	Kendi düşünce ve deneyimleri ile metinde verilen arasında bağlantı kuramamıştır.	Kendi düşünce ve deneyimleri ile metinde verilenler arasında bağlantı kurmasına rağmen açıklamasında durumlar arasındaki ilişkilendirmeyi yapamamıştır.	Kendi düşünce ve deneyimleri ile metinde verilenler arasında bağlantı kurmuş ve açıklamalarında durumlar arasındaki ilişkiyi belirtmiştir.	
Yorumlama	Cevap yok ya da tamamen ilgisiz.	Daha önceki bilgi ve deneyimlerine dayalı olarak metinde verilenler hakkında değerlendirmelerde bulunmamıştır.	Daha önceki bilgi ve deneyimlerine dayalı olarak metinde verilenler hakkında yorumlar yapmaya çalışmış ama bir bütünsellik oluşturamamıştır.	Daha önceki bilgi ve deneyimlerini kullanarak metinde verilenler hakkında doğru bir şekilde yorumlarda bulunmuştur.	
Metinden Bağımsız Düşünme	Cevap yok ya da tamamen ilgisiz.	İstenilen durumu açıklamada sadece metinde verilen bilgilere bağlı kalmıştır.	İstenilen duruma ait yapılan açıklamalarını kendi görüşleriyle desteklemeye çalışmasına rağmen daha metin dışına tam anlamıyla çıkamamıştır.	Metinden bağımsız ve nesnel bir şekilde düşünerek istenilen durum hakkında fikir yürütmüştür.	

EK-4:

YÖNERGE ÖRNEKLERİ

MATEMATİK OKURYAZARLIĞI SORU 1 YÖNERGE ÖRNEĞİ

Değerli Puanlayıcı,

Matematik okuryazarlığı kapsamında kullanılan 4.soru belirsizlik konusunu yönelik olarak hazırlanmış bir sorudur. Bu sorunun olası cevapları (bu ve benzeri cevaplar beklenmektedir) şu şekildedir:

Hayır, kabul edilmez. Grafiğin sadece **küçük bir parçasının** gösterilmiş olması gerçeği üzerinde odaklanır.

- Kabul edilemez. Grafiğin tamamı gösterilmeli.
- Grafiğin kabul edilebilir bir yorumu olduğunu düşünmüyorum, çünkü eğer grafiğin tamamı gösterilseydi soygunlarda sadece küçük bir artış olduğunu görebilirdiniz.
- Hayır, çünkü grafiğin en üst kısmını kullanmış ve eğer grafiğin 0-520 arasındaki tamamına bakarsanız, çok fazla artmamıştır.
- Hayır, çünkü grafik büyük bir artış varmış gibi gösteriyor fakat, sayılara baktığımızda çok artış yok.

Hayır, düşünmüyorum. Çünkü grafik sıfırdan başlamıyor. 500'den başlayıp beşer birim artıyor. 1998 yılının soygun sayısı 508 iken 1999'un yaklaşık 515'tir. Yani aradaki fark 7'dir. Yüzde olarak hesaplarsak $\frac{7}{508} \times 100 = 1,37$ dir. Yani yaklaşık iki katı değil %1,37 si kadar artmıştır.

Öğrenci yanıtlarını değerlendirmeniz için hazırlanmış olan dereceli puanlama anahtarında bulunan ölçütler ve açıklamaları aşağıda verilmiştir:

Grafiği Okuma: Bu kriter ile öğrencilerinin grafikte anlatılmak istenen bilgiyi anlamaları ve anladıkları tam ve doğru açıklamalar ile ifade etmeleri beklenmektedir. *Bu soru için grafiğin sıfır noktasından başlamadığını fark etmek önemlidir.*

İlişkilendirme: Bu kriterde öğrencilerden soruda verilen açıklama ile grafikte verilen bilgi arasındaki ilişkiyi kurmaları beklenmektedir.

Nedenini Açıklama: Bu kriterde öğrencinin ileri sürdüğü görüşe yönelik olarak belirttikleri nedenleri doğru bir şekilde açıklayıp açıklamadıkları, açıklamalarını soruda verilen bilgilere dayandırıp dayandıramadığı incelenmeye çalışılmıştır.

DİKKAT: Lütfen her bir öğrenciye ait sayfayı okurken öğrenci numarasını kontrol ediniz ve puanlayıcı değerlendirme formunda öğrenciye ait kısma değerlendirmenizi yazdığınızdan emin olunuz. Öğrenci numarası sayfanın ya da sorunun sağ üst köşesinde bulunmaktadır.

NOT: LÜTFEN PUANLAMA YAPARKEN GENEL YÖNERGEDEKİ AÇIKLAMALARI DİKKATE ALINIZ. PUANLAMA YAPMADAN ÖNCE DERECELİ PUANLAMA

ANAHTARINI AYRINTILI OLARAK İNCELEYİNİZ. HER BİR BİREYE AİT YANITLARI DEĞERLENDİRİRKEN DERECELİ PUANLAMA ANAHTARINIZ HEY YANINIZDA BULUNSUN.

NOT: GENEL YÖNERGEDE AÇIKLANAN KAYIP VERİ DURUMUNU OKUMA İŞLEMLERİNİZ ESNASINDA LÜTFEN UNUTMAYINIZ.

OKUDUĞUNU ANLAMA SORU 1 YÖNERGE ÖRNEĞİ

Değerli Puanlayıcı,

Okuduğunu anlama kapsamında kullanılan 1.soru birleştirme ve yorumlama amacına yönelik olarak hazırlanmış bir sorudur. Bu sorunun olası cevapları (bu ve benzeri cevaplar beklenmektedir) şu şekildedir:

Masalda verilen altının yararsız ya da değersiz bir şeyle değiştirilmesi mesajını fark eder.

•Özellikle anlatılmak istenen altının yerine değersiz bir şeyin konulması.

•Taş bu masalda önemli, çünkü altın, adam için ne ifade ediyorsa taş da onu ifade etmektedir.

•Eğer onun yerine taştan daha iyi bir şey konulursa, anlatılmak istenen asıl konu kaybolur. Bu yüzden gömülen şeyin gerçekten işe yaramayan bir şey olması gerekir.

•Bir taş yararsızdır, fakat cimri için altın da öyle!

•Taştan daha iyi bir şey onun kullanabileceği bir şeydir. O altını kullanmadı, komşunun özellikle

Altın genel anlamda değerli bir taş olmasına rağmen bu hikâyede adam için sadece gömülü duran, değeri olmayan bir şeyi ifade ediyordu. Yani onun için sadece gömdüğü yerdeki anlamı önemliydi. Bu nedenle gömülü yerde ha taş olmuş ha altın hiç fark eden bir şey olmayacaktı adam için.

Öğrenci yanıtlarını değerlendirmeniz için hazırlanmış olan dereceli puanlama anahtarında bulunan ölçütler ve açıklamaları aşağıda verilmiştir:

İlişkilendirme: Bu kriter ile öğrencilerden soruda öğrencilerden açıklanması beklenen kısım ile metinde verilen bilgileri ilişkilendirmesi ve bu açıklamaya yönelik olarak uygun açıklamalar yapabilmesi beklenmektedir. *Bu soru için metindeki açıklamalara dayalı olarak 2. konuşmacı tarafından ileri sürülen yanıtı destekleyecek açıklamalar yapması beklenmektedir.*

Kıyaslama: Bu kriter ile soruda açıklanması istenen bilgi ile metinde verilen bilginin benzer ve farklı yönlerini görerek, birbirine göre ne anlama geldiğini anlayıp anlamadığını ve açıklamalarında uygun doğru ve tam bir şekilde ifade edip etmediğinin belirlemek amaçlanmıştır.

Parça-Bütün İlişkisini Anlama: Bu kriterde öğrencinin metnin bir parçası olarak verilen bilginin metnin tamamındaki yerini anlayıp, bunu tam ve doğru bir şekilde ifade etmiş olması beklenmektedir.

DİKKAT: Lütfen her bir öğrenciye ait sayfayı okurken öğrenci numarasını kontrol ediniz ve puanlayıcı değerlendirme formunda öğrenciye ait kısma değerlendirmenizi yazdığınızdan emin olunuz. Öğrenci numarası sayfanın ya da sorunun sağ üst köşesinde bulunmaktadır.

NOT: LÜTFEN PUANLAMA YAPARKEN GENEL YÖNERGEDEKİ AÇIKLAMALARI DİKKATE ALINIZ. PUANLAMA YAPMADAN ÖNCE DERCELİ PUANLAMA ANAHTARINI AYRINTILI OLARAK İNCELEYİNİZ. HER BİR BİREYE AİT YANITLARI DEĞERLENDİRİRKEN DERCELİ PUANLAMA ANAHTARINIZ HEPSİNİZDE BULUNSUN.

NOT: GENEL YÖNERGEDE AÇIKLANAN KAYIP VERİ DURUMUNU OKUMA İŞLEMLERİNİZ ESNASINDA LÜTFEN UNUTMAYINIZ.

EK-5:

ARAŞTIRMA BULGULARINA AİT TABLOLAR

Tablo 17. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,767 (0,037)	0,896 (0,016)	0,764 (0,038)	0,765 (0,038)	0,819 (0,023)	0,950 (0,006)
İşlemleri gerçekleştirme	0,585 (0,046)	0,538 (0,056)	0,581 (0,047)	0,582 (0,047)	0,414 (0,063)	0,837 (0,018)
Sonuca ulaşma	0,804 (0,042)	0,875 (0,029)	0,804 (0,042)	0,804 (0,042)	0,825 (0,038)	0,951 (0,011)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,833 (0,031)	0,926 (0,014)	0,833 (0,031)	0,834 (0,031)	0,902 (0,018)	0,923 (0,005)
İşlemleri gerçekleştirme	0,613 (0,049)	0,661 (0,052)	0,612 (0,049)	0,613 (0,049)	0,525 (0,062)	0,868 (0,017)
Sonuca ulaşma	0,780 (0,037)	0,897 (0,018)	0,780 (0,037)	0,780 (0,034)	0,833 (0,028)	0,954 (0,008)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,617 (0,046)	0,775 (0,031)	0,591 (0,054)	0,592 (0,054)	0,719 (0,035)	0,922 (0,010)
Akıl yürütme / İlişkilendirme	0,67 (0,044)	0,725 (0,038)	0,626 (0,051)	0,627 (0,01)	0,696 (0,039)	0,916 (0,011)
Nedenini açıklama	0,657 (0,044)	0,811 (0,025)	0,641 (0,049)	0,641 (0,049)	0,768 (0,028)	0,936 (0,008)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,576 (0,047)	0,720 (0,032)	0,570 (0,049)	0,571 (0,049)	0,665 (0,034)	0,907 (0,010)
Akıl yürütme / İlişkilendirme	0,705 (0,034)	0,780 (0,022)	0,696 (0,036)	0,697 (0,036)	0,761 (0,023)	0,934 (0,006)
Neden belirtme	0,215 (0,036)	-0,044 (0,076)	0,044 (0,065)	0,046 (0,065)	-0,189 (0,090)	0,670 (0,025)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,710 (0,034)	0,692 (0,034)	0,709 (0,035)	0,709 (0,035)	0,679 (0,034)	0,911 (0,009)
Akıl yürütme / İlişkilendirme	0,816 (0,024)	0,809 (0,023)	0,816 (0,024)	0,816 (0,024)	0,808 (0,023)	0,947 (0,006)
Neden belirtme	0,719 (0,031)	0,683 (0,033)	0,714 (0,033)	0,714 (0,033)	0,660 (0,033)	0,905 (0,009)

Tablo 18. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,623 (0,035)	0,739 (0,024)	0,596 (0,040)	0,597 (0,040)	0,704 (0,025)	0,918 (0,007)
Karşılaştırma	0,674 (0,031)	0,786 (0,021)	0,653 (0,036)	0,653 (0,036)	0,758 (0,022)	0,933 (0,006)
Parça-bütün ilişkisini anlama	0,591 (0,037)	0,725 (0,026)	0,561 (0,043)	0,562 (0,043)	0,689 (0,027)	0,914 (0,008)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,170 (0,063)	0,581 (0,034)	0,098 (0,076)	0,100 (0,076)	0,463 (0,036)	0,851 (0,010)
Karşılaştırma	0,121 (0,045)	0,378 (0,032)	-0,041 (0,070)	-0,039 (0,070)	0,214 (0,045)	0,782 (0,013)
Parça-bütün ilişkisini anlama	0,113 (0,054)	0,421 (0,035)	-0,003 (0,073)	0,008 (0,074)	0,303 (0,043)	0,806 (0,012)
• Okuduğunu anlama soru 3						
Yansıtma	0,263 (0,057)	0,602 (0,033)	0,227 (0,066)	0,229 (0,066)	0,517 (0,038)	0,866 (0,011)
Yorumlama	0,340 (0,059)	0,638 (0,033)	0,337 (0,060)	0,338 (0,060)	0,550 (0,040)	0,875 (0,011)
Metinden bağımsız düşünme	0,572 (0,050)	0,758 (0,030)	0,569 (0,050)	0,570 (0,050)	0,697 (0,035)	0,916 (0,010)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,191 (0,049)	0,741 (0,026)	-0,030 (0,081)	-0,028 (0,081)	0,602 (0,033)	0,889 (0,009)
Karşılaştırma	0,397 (0,083)	0,935 (0,012)	0,363 (0,094)	0,365 (0,093)	0,842 (0,025)	0,956 (0,007)
Parça-bütün ilişkisini anlama	0,271 (0,074)	0,901 (0,016)	0,203 (0,089)	0,205 (0,088)	0,779 (0,029)	0,939 (0,008)
• Okuduğunu anlama soru 5						
Yansıtma	0,302 (0,055)	0,690 (0,033)	0,279 (0,060)	0,281 (0,060)	0,597 (0,037)	0,888 (0,010)
Yorumlama	0,319 (0,052)	0,707 (0,035)	0,252 (0,067)	0,259 (0,067)	0,597 (0,040)	0,888 (0,111)
Metinden bağımsız düşünme	0,393 (0,048)	0,632 (0,041)	0,307 (0,066)	0,309 (0,066)	0,546 (0,044)	0,874 (0,012)
• Okuduğunu anlama soru 6						
Yansıtma	0,599 (0,046)	0,663 (0,037)	0,590 (0,048)	0,591 (0,048)	0,577 (0,042)	0,883 (0,012)
Yorumlama	0,440 (0,048)	0,406 (0,048)	0,398 (0,057)	0,400 (0,057)	0,343 (0,050)	0,817 (0,014)
Metinden bağımsız düşünme	0,404 (0,047)	0,356 (0,046)	0,352 (0,058)	0,353 (0,058)	0,275 (0,052)	0,799 (0,014)

Tablo 19. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,778 (0,038)	0,898 (0,017)	0,776 (0,038)	0,783 (0,037)	0,822 (0,025)	0,951 (0,007)
İşlemleri gerçekleştirme	0,590 (0,048)	0,543 (0,059)	0,585 (0,050)	0,587 (0,050)	0,420 (0,067)	0,839 (0,019)
Sonuca ulaşma	0,811 (0,045)	0,877 (0,031)	0,811 (0,045)	0,816 (0,044)	0,829 (0,041)	0,953 (0,011)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,850 (0,031)	0,933 (0,014)	0,850 (0,031)	0,849 (0,031)	0,912 (0,018)	0,975 (0,005)
İşlemleri gerçekleştirme	0,617 (0,052)	0,663 (0,054)	0,615 (0,051)	0,617 (0,051)	0,522 (0,066)	0,867 (0,018)
Sonuca ulaşma	0,772 (0,049)	0,893 (0,020)	0,771 (0,041)	0,771 (0,041)	0,825 (0,031)	0,952 (0,008)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,619 (0,049)	0,769 (0,034)	0,595 (0,057)	0,597 (0,057)	0,715 (0,038)	0,921 (0,011)
Akıl yürütme / İlişkilendirme	0,645 (0,048)	0,717 (0,041)	0,626 (0,054)	0,628 (0,054)	0,690 (0,043)	0,914 (0,012)
Nedenini açıklama	0,669 (0,046)	0,817 (0,027)	0,655 (0,051)	0,654 (0,051)	0,776 (0,030)	0,938 (0,008)
• <i>Matematik soru 4</i>						
Grafiği okuma	0,570 (0,049)	0,705 (0,035)	0,565 (0,051)	0,573 (0,050)	0,653 (0,037)	0,904 (0,010)
Akıl yürütme / İlişkilendirme	0,700 (0,036)	0,774 (0,024)	0,692 (0,038)	0,694 (0,039)	0,756 (0,025)	0,932 (0,007)
Neden belirtme	0,216 (0,042)	-0,049 (0,080)	0,047 (0,069)	0,054 (0,069)	-0,192 (0,095)	0,669 (0,026)
• <i>Matematik soru 5</i>						
Grafiği okuma	0,697 (0,037)	0,683 (0,036)	0,695 (0,038)	0,695 (0,038)	0,668 (0,036)	0,908 (0,010)
Akıl yürütme / İlişkilendirme	0,812 (0,025)	0,805 (0,024)	0,812 (0,025)	0,812 (0,026)	0,805 (0,024)	0,946 (0,007)
Neden belirtme	0,719 (0,032)	0,682 (0,034)	0,714 (0,034)	0,714 (0,034)	0,662 (0,034)	0,906 (0,010)

Tablo 20. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,797 (0,036)	0,899 (0,014)	0,767 (0,037)	0,768 (0,038)	0,826 (0,020)	0,952 (0,005)
İşlemleri gerçekleştirme	0,535 (0,053)	0,482 (0,065)	0,529 (0,055)	0,530 (0,055)	0,348 (0,074)	0,819 (0,020)
Sonuca ulaşma	0,822 (0,046)	0,893 (0,029)	0,822 (0,046)	0,822 (0,046)	0,848 (0,040)	0,958 (0,011)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,814 (0,036)	0,917 (0,016)	0,814 (0,037)	0,813 (0,037)	0,890 (0,021)	0,970 (0,006)
İşlemleri gerçekleştirme	0,598 (0,056)	0,649 (0,058)	0,596 (0,056)	0,589 (0,057)	0,499 (0,072)	0,861 (0,020)
Sonuca ulaşma	0,766 (0,043)	0,891 (0,020)	0,765 (0,043)	0,765 (0,043)	0,824 (0,032)	0,951 (0,009)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,670 (0,048)	0,802 (0,030)	0,650 (0,054)	0,655 (0,054)	0,755 (0,034)	0,932 (0,010)
Akıl yürütme / İlişkilendirme	0,691 (0,045)	0,758 (0,037)	0,675 (0,050)	0,684 (0,049)	0,733 (0,038)	0,926 (0,011)
Nedenini açıklama	0,690 (0,047)	0,829 (0,026)	0,677 (0,052)	0,688 (0,051)	0,790 (0,030)	0,942 (0,008)
• <i>Matematik soru 4</i>						
Grafiği okuma	0,576 (0,051)	0,710 (0,037)	0,571 (0,054)	0,583 (0,052)	0,658 (0,039)	0,905 (0,011)
Akıl yürütme / İlişkilendirme	0,714 (0,036)	0,784 (0,025)	0,705 (0,039)	0,712 (0,038)	0,766 (0,026)	0,935 (0,007)
Neden belirtme	0,223 (0,045)	-0,053 (0,084)	0,046 (0,073)	0,044 (0,073)	-0,196 (0,100)	0,668 (0,028)
• <i>Matematik soru 5</i>						
Grafiği okuma	0,704 (0,039)	0,691 (0,038)	0,703 (0,039)	0,701 (0,040)	0,679 (0,038)	0,911 (0,011)
Akıl yürütme / İlişkilendirme	0,806 (0,028)	0,800 (0,026)	0,805 (0,028)	0,803 (0,029)	0,800 (0,026)	0,944 (0,007)
Neden belirtme	0,696 (0,036)	0,656 (0,038)	0,687 (0,039)	0,688 (0,039)	0,633 (0,039)	0,898 (0,011)

Tablo 21. Matematik Okuryazarlık Uygulamasında İki Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,751 (0,043)	0,892 (0,019)	0,747 (0,045)	0,741 (0,048)	0,811 (0,029)	0,948 (0,008)
İşlemleri gerçekleştirme	0,598 (0,053)	0,547 (0,064)	0,593 (0,055)	0,586 (0,056)	0,433 (0,074)	0,843 (0,020)
Sonuca ulaşma	0,797 (0,049)	0,867 (0,034)	0,797 (0,049)	0,797 (0,049)	0,816 (0,045)	0,949 (0,013)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,842 (0,033)	0,928 (0,015)	0,842 (0,033)	0,845 (0,033)	0,904 (0,020)	0,973 (0,005)
İşlemleri gerçekleştirme	0,659 (0,052)	0,703 (0,054)	0,659 (0,052)	0,669 (0,051)	0,583 (0,065)	0,884 (0,018)
Sonuca ulaşma	0,810 (0,040)	0,910 (0,019)	0,810 (0,040)	0,814 (0,035)	0,854 (0,030)	0,959 (0,008)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,603 (0,055)	0,774 (0,036)	0,575 (0,063)	0,571 (0,065)	0,713 (0,041)	0,920 (0,011)
Akıl yürütme / İlişkilendirme	0,626 (0,051)	0,716 (0,041)	0,604 (0,058)	0,611 (0,058)	0,685 (0,043)	0,912 (0,012)
Nedenini açıklama	0,643 (0,052)	0,801 (0,030)	0,624 (0,058)	0,628 (0,059)	0,757 (0,035)	0,932 (0,010)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,553 (0,055)	0,696 (0,039)	0,544 (0,058)	0,552 (0,058)	0,637 (0,042)	0,899 (0,012)
Akıl yürütme / İlişkilendirme	0,708 (0,038)	0,780 (0,026)	0,703 (0,040)	0,701 (0,042)	0,763 (0,027)	0,934 (0,008)
Neden belirtme	0,158 (0,055)	-0,066 (0,095)	0,014 (0,081)	0,036 (0,078)	-0,195 (0,109)	0,668 (0,030)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,703 (0,037)	0,688 (0,036)	0,699 (0,039)	0,690 (0,041)	0,677 (0,036)	0,910 (0,010)
Akıl yürütme / İlişkilendirme	0,826 (0,026)	0,826 (0,023)	0,826 (0,026)	0,819 (0,028)	0,826 (0,023)	0,952 (0,006)
Neden belirtme	0,688 (0,038)	0,655 (0,040)	0,682 (0,040)	0,676 (0,043)	0,632 (0,040)	0,898 (0,011)

Tablo 22. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,626 (0,035)	0,746 (0,025)	0,598 (0,041)	0,589 (0,042)	0,710 (0,026)	0,919 (0,007)
Karşılaştırma	0,674 (0,032)	0,790 (0,022)	0,650 (0,037)	0,640 (0,039)	0,760 (0,023)	0,933 (0,006)
Parça-bütün ilişkisini anlama	0,580 (0,038)	0,723 (0,027)	0,546 (0,045)	0,539 (0,045)	0,684 (0,029)	0,912 (0,008)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,172 (0,068)	0,580 (0,036)	0,094 (0,082)	0,113 (0,081)	0,460 (0,039)	0,850 (0,011)
Karşılaştırma	0,117 (0,049)	0,370 (0,034)	-0,052 (0,074)	-0,042 (0,074)	0,204 (0,048)	0,779 (0,013)
Parça-bütün ilişkisini anlama	0,097 (0,058)	0,406 (0,037)	-0,025 (0,077)	-0,009 (0,077)	0,285 (0,045)	0,802 (0,013)
• Okuduğunu anlama soru 3						
Yansıtma	0,227 (0,062)	0,583 (0,035)	0,191 (0,070)	0,199 (0,070)	0,496 (0,041)	0,860 (0,011)
Yorumlama	0,296 (0,065)	0,618 (0,036)	0,294 (0,065)	0,298 (0,065)	0,527 (0,043)	0,868 (0,012)
Metinden bağımsız düşünme	0,568 (0,053)	0,752 (0,032)	0,564 (0,054)	0,570 (0,054)	0,693 (0,038)	0,915 (0,011)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,183 (0,054)	0,741 (0,028)	-0,033 (0,088)	-0,003 (0,086)	0,600 (0,036)	0,889 (0,010)
Karşılaştırma	0,378 (0,091)	0,933 (0,013)	0,344 (0,102)	0,377 (0,097)	0,836 (0,028)	0,954 (0,008)
Parça-bütün ilişkisini anlama	0,244 (0,081)	0,897 (0,017)	0,173 (0,097)	0,210 (0,092)	0,768 (0,032)	0,936 (0,009)
• Okuduğunu anlama soru 5						
Yansıtma	0,272 (0,057)	0,683 (0,035)	0,248 (0,063)	0,250 (0,061)	0,585 (0,039)	0,885 (0,011)
Yorumlama	0,273 (0,055)	0,698 (0,038)	0,198 (0,070)	0,197 (0,069)	0,577 (0,044)	0,882 (0,012)
Metinden bağımsız düşünme	0,338 (0,051)	0,621 (0,044)	0,243 (0,070)	0,245 (0,070)	0,523 (0,047)	0,868 (0,01305)
• Okuduğunu anlama soru 6						
Yansıtma	0,343 (0,057)	0,709 (0,033)	0,323 (0,061)	0,317 (0,061)	0,621 (0,037)	0,895 (0,010)
Yorumlama	0,354 (0,056)	0,724 (0,036)	0,295 (0,069)	0,287 (0,069)	0,617 (0,041)	0,894 (0,012)
Metinden bağımsız düşünme	0,389 (0,051)	0,633 (0,043)	0,303 (0,069)	0,309 (0,068)	0,548 (0,045)	0,874 (0,013)

Tablo 23. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,626 (0,039)	0,749 (0,026)	0,604 (0,044)	0,604 (0,044)	0,714 (0,028)	0,921 (0,008)
Karşılaştırma	0,666 (0,035)	0,780 (0,024)	0,647 (0,040)	0,650 (0,040)	0,753 (0,025)	0,931 (0,007)
Parça-bütün ilişkisini anlama	0,600 (0,040)	0,731 (0,028)	0,574 (0,046)	0,578 (0,046)	0,697 (0,030)	0,916 (0,008)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,167 (0,071)	0,587 (0,038)	0,082 (0,086)	0,072 (0,091)	0,464 (0,041)	0,851 (0,011)
Karşılaştırma	0,134 (0,045)	0,380 (0,034)	-0,046 (0,077)	-0,044 (0,079)	0,221 (0,049)	0,784 (0,013)
Parça-bütün ilişkisini anlama	0,112 (0,057)	0,421 (0,037)	-0,020 (0,080)	-0,023 (0,083)	0,302 (0,045)	0,806 (0,012)
• Okuduğunu anlama soru 3						
Yansıtma	0,259 (0,065)	0,591 (0,036)	0,223 (0,074)	0,243 (0,072)	0,505 (0,043)	0,862 (0,012)
Yorumlama	0,333 (0,067)	0,629 (0,037)	0,329 (0,068)	0,338 (0,067)	0,542 (0,045)	0,873 (0,012)
Metinden bağımsız düşünme	0,578 (0,057)	0,760 (0,035)	0,576 (0,058)	0,586 (0,057)	0,699 (0,041)	0,917 (0,011)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,169 (0,058)	0,734 (0,030)	-0,056 (0,089)	-0,046 (0,088)	0,591 (0,038)	0,887 (0,011)
Karşılaştırma	0,349 (0,092)	0,929 (0,014)	0,304 (0,104)	0,291 (0,104)	0,829 (0,030)	0,952 (0,008)
Parça-bütün ilişkisini anlama	0,223 (0,080)	0,894 (0,019)	0,139 (0,095)	0,122 (0,092)	0,763 (0,035)	0,934 (0,010)
• Okuduğunu anlama soru 5						
Yansıtma	0,352 (0,057)	0,715 (0,031)	0,329 (0,063)	0,336 (0,062)	0,628 (0,035)	0,897 (0,010)
Yorumlama	0,346 (0,057)	0,711 (0,036)	0,275 (0,073)	0,290 (0,071)	0,607 (0,042)	0,891 (0,012)
Metinden bağımsız düşünme	0,373 (0,057)	0,608 (0,049)	0,278 (0,077)	0,294 (0,076)	0,520 (0,052)	0,867 (0,015)
• Okuduğunu anlama soru 6						
Yansıtma	0,591 (0,048)	0,667 (0,037)	0,577 (0,051)	0,554 (0,056)	0,578 (0,042)	0,883 (0,012)
Yorumlama	0,418 (0,051)	0,383 (0,050)	0,360 (0,062)	0,334 (0,067)	0,315 (0,054)	0,810 (0,015)
Metinden bağımsız düşünme	0,377 (0,049)	0,330 (0,047)	0,306 (0,064)	0,283 (0,069)	0,233 (0,055)	0,787 (0,015)

Tablo 24. Okuduğunu Anlama Uygulamasında İki Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Cohen'in Kappa (SH)	Gwet'in AC ₂ (SH)	Scott'nin Pi (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,634 (0,039)	0,751 (0,026)	0,611 (0,044)	0,622 (0,044)	0,719 (0,028)	0,922 (0,008)
Karşılaştırma	0,684 (0,034)	0,797 (0,023)	0,665 (0,039)	0,670 (0,039)	0,770 (0,024)	0,936 (0,007)
Parça-bütün ilişkisini anlama	0,576 (0,043)	0,715 (0,030)	0,543 (0,050)	0,551 (0,051)	0,678 (0,032)	0,910 (0,009)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,201 (0,064)	0,614 (0,034)	0,131 (0,077)	0,109 (0,082)	0,498 (0,037)	0,861 (0,010)
Karşılaştırma	0,135 (0,048)	0,397 (0,035)	-0,028 (0,075)	-0,041 (0,078)	0,236 (0,050)	0,788 (0,014)
Parça-bütün ilişkisini anlama	0,070 (0,060)	0,411 (0,040)	-0,052 (0,081)	-0,046 (0,081)	0,288 (0,048)	0,802 (0,013)
• Okuduğunu anlama soru 3						
Yansıtma	0,267 (0,064)	0,599 (0,038)	0,225 (0,075)	0,239 (0,074)	0,514 (0,044)	0,865 (0,012)
Yorumlama	0,326 (0,069)	0,625 (0,040)	0,321 (0,071)	0,333 (0,070)	0,536 (0,048)	0,871 (0,013)
Metinden bağımsız düşünme	0,562 (0,056)	0,753 (0,033)	0,560 (0,057)	0,567 (0,057)	0,693 (0,039)	0,915 (0,011)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,179 (0,060)	0,740 (0,030)	-0,029 (0,090)	-0,03322 (0,090)	0,600 (0,039)	0,889 (0,011)
Karşılaştırma	0,364 (0,102)	0,931 (0,015)	0,331 (0,113)	0,338 (0,112)	0,832 (0,032)	0,953 (0,009)
Parça-bütün ilişkisini anlama	0,237 (0,090)	0,894 (0,020)	0,165 (0,106)	0,174 (0,103)	0,763 (0,037)	0,934 (0,010)
• Okuduğunu anlama soru 5						
Yansıtma	0,264 (0,067)	0,682 (0,039)	0,240 (0,073)	0,255 (0,070)	0,583 (0,044)	0,884 (0,012)
Yorumlama	0,280 (0,063)	0,695 (0,042)	0,215 (0,079)	0,213 (0,076)	0,578 (0,049)	0,883 (0,014)
Metinden bağımsız düşünme	0,325 (0,061)	0,584 (0,053)	0,230 (0,082)	0,242 (0,080)	0,490 (0,057)	0,858 (0,016)
• Okuduğunu anlama soru 6						
Yansıtma	0,085 (0,077)	0,242 (0,083)	0,050 (0,084)	0,056 (0,082)	0,044 (0,094)	0,734 (0,026)
Yorumlama	0,443 (0,054)	0,403 (0,055)	0,385 (0,066)	0,387 (0,067)	0,332 (0,060)	0,814 (0,017)
Metinden bağımsız düşünme	0,399 (0,051)	0,350 (0,051)	0,329 (0,066)	0,329 (0,068)	0,257 (0,058)	0,794 (0,016)

Tablo 25. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,800 (0,049)	0,910 (0,025)	0,799 (0,047)	0,799 (0,434)	0,837 (0,027)	0,955 (0,007)
İşlemleri gerçekleştirme	0,577 (0,145)	0,551 (0,183)	0,573 (0,147)	0,573 (0,146)	0,451 (0,266)	0,848 (0,074)
Sonuca ulaşma	0,841 (0,054)	0,896 (0,032)	0,841 (0,050)	0,842 (0,050)	0,861 (0,047)	0,961 (0,013)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,783 (0,076)	0,860 (0,071)	0,783 (0,057)	0,783 (0,057)	0,834 (0,073)	0,954 (0,020)
İşlemleri gerçekleştirme	0,597 (0,160)	0,685 (0,149)	0,596 (0,141)	0,596 (0,141)	0,593 (0,195)	0,887 (0,054)
Sonuca ulaşma	0,631 (0,214)	0,840 (0,057)	0,626 (0,183)	0,626 (0,183)	0,775 (0,065)	0,938 (0,018)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,699 (0,134)	0,843 (0,088)	0,691 (0,124)	0,692 (0,123)	0,796 (0,099)	0,943 (0,027)
Akıl yürütme / İlişkilendirme	0,614 (0,187)	0,762 (0,166)	0,598 (0,198)	0,599 (0,197)	0,713 (0,171)	0,920 (0,047)
Nedenini açıklama	0,650 (0,141)	0,835 (0,101)	0,637 (0,121)	0,638 (0,119)	0,786 (0,104)	0,941 (0,029)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,457 (0,164)	0,602 (0,216)	0,416 (0,221)	0,417 (0,219)	0,558 (0,210)	0,877 (0,058)
Akıl yürütme / İlişkilendirme	0,507 (0,233)	0,717 (0,079)	0,476 (0,292)	0,477 (0,291)	0,650 (0,126)	0,903 (0,035)
Neden belirtme	0,246 (0,116)	0,201 (0,378)	0,152 (0,258)	0,153 (0,257)	0,187 (0,412)	0,774 (0,114)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,731 (0,105)	0,723 (0,129)	0,730 (0,101)	0,730 (0,100)	0,717 (0,136)	0,923 (0,038)
Akıl yürütme / İlişkilendirme	0,789 (0,051)	0,807 (0,035)	0,788 (0,039)	0,788 (0,038)	0,804 (0,032)	0,945 (0,009)
Neden belirtme	0,740 (0,060)	0,742 (0,094)	0,738 (0,048)	0,739 (0,046)	0,735 (0,100)	0,926 (0,028)

Tablo 26. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,659 (0,097)	0,766 (0,068)	0,650 (0,098)	0,650 (0,097)	0,740 (0,079)	0,928 (0,022)
Karşılaştırma	0,691 (0,074)	0,799 (0,039)	0,684 (0,065)	0,685 (0,064)	0,775 (0,046)	0,938 (0,013)
Parça-bütün ilişkisini anlama	0,656 (0,089)	0,762 (0,050)	0,648 (0,092)	0,648 (0,090)	0,737 (0,058)	0,927 (0,016)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,389 (0,427)	0,690 (0,211)	0,368 (0,468)	0,369 (0,466)	0,592 (0,246)	0,887 (0,068)
Karşılaştırma	0,303 (0,438)	0,554 (0,335)	0,247 (0,549)	0,248 (0,547)	0,412 (0,391)	0,837 (0,109)
Parça-bütün ilişkisini anlama	0,336 (0,445)	0,606 (0,307)	0,297 (0,523)	0,298 (0,520)	0,490 (0,337)	0,858 (0,094)
• Okuduğunu anlama soru 3						
Yansıtma	0,405 (0,259)	0,682 (0,144)	0,390 (0,275)	0,391 (0,273)	0,603 (0,162)	0,890 (0,045)
Yorumlama	0,454 (0,167)	0,691 (0,070)	0,447 (0,155)	0,448 (0,154)	0,616 (0,079)	0,893 (0,022)
Metinden bağımsız düşünme	0,534 (0,117)	0,742 (0,029)	0,532 (0,103)	0,533 (0,102)	0,672 (0,037)	0,909 (0,010)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,277 (0,191)	0,815 (0,096)	0,209 (0,251)	0,210 (0,248)	0,702 (0,120)	0,917 (0,033)
Karşılaştırma	0,418 (0,178)	0,932 (0,010)	0,405 (0,097)	0,405 (0,094)	0,842 (0,020)	0,956 (0,005)
Parça-bütün ilişkisini anlama	0,300 (0,198)	0,914 (0,027)	0,277 (0,097)	0,278 (0,093)	0,803 (0,049)	0,945 (0,013)
• Okuduğunu anlama soru 5						
Yansıtma	0,457 (0,197)	0,743 (0,114)	0,442 (0,159)	0,443 (0,158)	0,670 (0,110)	0,908 (0,031)
Yorumlama	0,463 (0,242)	0,799 (0,126)	0,440 (0,240)	0,441 (0,239)	0,714 (0,157)	0,920 (0,044)
Metinden bağımsız düşünme	0,475 (0,174)	0,775 (0,149)	0,446 (0,169)	0,446 (0,167)	0,692 (0,157)	0,914 (0,044)
• Okuduğunu anlama soru 6						
Yansıtma	0,498 (0,148)	0,480 (0,259)	0,475 (0,171)	0,476 (0,169)	0,448 (0,289)	0,847 (0,080)
Yorumlama	0,467 (0,216)	0,428 (0,357)	0,434 (0,254)	0,435 (0,252)	0,417 (0,381)	0,838 (0,106)
Metinden bağımsız düşünme	0,443 (0,230)	0,407 (0,386)	0,405 (0,276)	0,406 (0,274)	0,375 (0,415)	0,826 (0,115)

Tablo 27. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,791 (0,050)	0,905 (0,026)	0,785 (0,051)	0,801 (0,043)	0,827 (0,032)	0,952 (0,010)
İşlemleri gerçekleştirme	0,560 (0,140)	0,529 (0,181)	0,553 (0,144)	0,563 (0,142)	0,428 (0,262)	0,841 (0,073)
Sonuca ulaşma	0,841 (0,050)	0,894 (0,029)	0,838 (0,048)	0,839 (0,045)	0,858 (0,043)	0,961 (0,013)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,790 (0,078)	0,866 (0,072)	0,792 (0,063)	0,789 (0,063)	0,840 (0,076)	0,956 (0,021)
İşlemleri gerçekleştirme	0,595 (0,156)	0,680 (0,151)	0,591 (0,144)	0,592 (0,137)	0,585 (0,196)	0,885 (0,054)
Sonuca ulaşma	0,634 (0,204)	0,841 (0,055)	0,631 (0,178)	0,627 (0,177)	0,776 (0,060)	0,938 (0,017)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,685 (0,132)	0,834 (0,092)	0,675 (0,125)	0,682 (0,120)	0,784 (0,102)	0,940 (0,029)
Akıl yürütme / İlişkilendirme	0,614 (0,197)	0,756 (0,177)	0,596 (0,216)	0,601 (0,209)	0,707 (0,185)	0,919 (0,052)
Nedenini açıklama	0,650 (0,142)	0,834 (0,103)	0,637 (0,131)	0,643 (0,122)	0,784 (0,108)	0,940 (0,030)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,456 (0,160)	0,598 (0,218)	0,414 (0,208)	0,419 (0,216)	0,553 (0,209)	0,876 (0,059)
Akıl yürütme / İlişkilendirme	0,502 (0,231)	0,716 (0,070)	0,470 (0,293)	0,474 (0,278)	0,647 (0,120)	0,902 (0,034)
Neden belirtme	0,241 (0,107)	0,194 (0,378)	0,148 (0,249)	0,153 (0,231)	0,179 (0,409)	0,772 (0,114)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,716 (0,110)	0,709 (0,133)	0,717 (0,106)	0,720 (0,106)	0,702 (0,142)	0,917 (0,040)
Akıl yürütme / İlişkilendirme	0,785 (0,051)	0,804 (0,039)	0,784 (0,041)	0,785 (0,038)	0,800 (0,037)	0,944 (0,013)
Neden belirtme	0,744 (0,060)	0,745 (0,095)	0,740 (0,051)	0,740 (0,047)	0,738 (0,100)	0,927 (0,029)

Tablo 28. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,797 (0,042)	0,910 (0,025)	0,796 (0,040)	0,793 (0,040)	0,837 (0,022)	0,955 (0,012)
İşlemleri gerçekleştirme	0,585 (0,162)	0,561 (0,207)	0,580 (0,170)	0,573 (0,158)	0,467 (0,294)	0,852 (0,082)
Sonuca ulaşma	0,857 (0,050)	0,909 (0,032)	0,856 (0,049)	0,852 (0,047)	0,877 (0,043)	0,966 (0,016)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,784 (0,059)	0,863 (0,062)	0,783 (0,044)	0,783 (0,039)	0,836 (0,063)	0,954 (0,021)
İşlemleri gerçekleştirme	0,614 (0,170)	0,697 (0,162)	0,615 (0,161)	0,600 (0,158)	0,605 (0,215)	0,890 (0,061)
Sonuca ulaşma	0,623 (0,219)	0,840 (0,062)	0,621 (0,196)	0,610 (0,203)	0,775 (0,071)	0,937 (0,023)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,698 (0,107)	0,842 (0,072)	0,686 (0,094)	0,705 (0,096)	0,793 (0,079)	0,943 (0,023)
Akıl yürütme / İlişkilendirme	0,628 (0,162)	0,769 (0,144)	0,608 (0,167)	0,625 (0,174)	0,721 (0,146)	0,923 (0,041)
Nedenini açıklama	0,672 (0,115)	0,847 (0,087)	0,655 (0,098)	0,666 (0,100)	0,800 (0,089)	0,944 (0,026)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,473 (0,143)	0,614 (0,194)	0,430 (0,191)	0,433 (0,191)	0,570 (0,188)	0,881 (0,053)
Akıl yürütme / İlişkilendirme	0,515 (0,243)	0,723 (0,084)	0,479 (0,309)	0,481 (0,306)	0,657 (0,131)	0,905 (0,038)
Neden belirtme	0,247 (0,115)	0,202 (0,385)	0,157 (0,262)	0,149 (0,266)	0,185 (0,419)	0,774 (0,117)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,743 (0,100)	0,740 (0,124)	0,745 (0,100)	0,734 (0,093)	0,734 (0,130)	0,926 (0,037)
Akıl yürütme / İlişkilendirme	0,779 (0,050)	0,801 (0,025)	0,780 (0,041)	0,777 (0,041)	0,797 (0,023)	0,944 (0,012)
Neden belirtme	0,734 (0,055)	0,738 (0,096)	0,732 (0,051)	0,726 (0,042)	0,731 (0,102)	0,925 (0,030)

Tablo 29. Matematik Okuryazarlık Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• <i>Matematik soru 1</i>						
İlişkileri belirleme	0,790 (0,050)	0,908 (0,029)	0,789 (0,050)	0,788 (0,049)	0,832 (0,033)	0,953 (0,018)
İşlemleri gerçekleştirme	0,551 (0,134)	0,525 (0,173)	0,552 (0,138)	0,548 (0,137)	0,422 (0,252)	0,839 (0,071)
Sonuca ulaşma	0,818 (0,047)	0,882 (0,032)	0,817 (0,047)	0,821 (0,040)	0,839 (0,043)	0,955 (0,019)
• <i>Matematik soru 2</i>						
İlişkileri belirleme	0,791 (0,072)	0,865 (0,072)	0,791 (0,061)	0,788 (0,061)	0,840 (0,075)	0,955 (0,026)
İşlemleri gerçekleştirme	0,642 (0,155)	0,723 (0,147)	0,631 (0,140)	0,632 (0,154)	0,641 (0,182)	0,900 (0,053)
Sonuca ulaşma	0,669 (0,208)	0,856 (0,067)	0,661 (0,195)	0,658 (0,202)	0,797 (0,079)	0,943 (0,027)
• <i>Matematik soru 3</i>						
Sayıların ifade ettiği değeri anlama	0,708 (0,145)	0,852 (0,102)	0,705 (0,151)	0,686 (0,143)	0,806 (0,118)	0,946 (0,038)
Akıl yürütme / İlişkilendirme	0,609 (0,085)	0,764 (0,036)	0,592 (0,228)	0,585 (0,212)	0,713 (0,194)	0,920 (0,057)
Nedenini açıklama	0,648 (0,137)	0,832 (0,117)	0,639 (0,143)	0,624 (0,121)	0,783 (0,122)	0,940 (0,039)
• <i>Matematik soru 4</i>						
Grafîği okuma	0,455 (0,069)	0,597 (0,214)	0,413 (0,206)	0,422 (0,200)	0,551 (0,212)	0,875 (0,061)
Akıl yürütme / İlişkilendirme	0,487 (0,223)	0,703 (0,066)	0,460 (0,282)	0,474 (0,274)	0,635 (0,115)	0,899 (0,036)
Neden belirtme	0,230 (0,150)	0,215 (0,400)	0,149 (0,268)	0,154 (0,257)	0,203 (0,428)	0,778 (0,120)
• <i>Matematik soru 5</i>						
Grafîği okuma	0,731 (0,119)	0,730 (0,147)	0,735 (0,121)	0,720 (0,116)	0,724 (0,153)	0,923 (0,045)
Akıl yürütme / İlişkilendirme	0,789 (0,053)	0,817 (0,042)	0,79199 (0,05078)	0,788 (0,040)	0,812 (0,040)	0,948 (0,018)
Neden belirtme	0,734 (0,070)	0,745 (0,117)	0,73408 (0,06517)	0,728 (0,065)	0,737 (0,121)	0,927 (0,037)

Tablo 30. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %5 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,655 (0,096)	0,763 (0,068)	0,645 (0,097)	0,650 (0,101)	0,737 (0,080)	0,927 (0,022)
Karşılaştırma	0,697 (0,071)	0,803 (0,039)	0,689 (0,063)	0,687 (0,068)	0,779 (0,045)	0,939 (0,013)
Parça-bütün ilişkisini anlama	0,657 (0,093)	0,763 (0,052)	0,646 (0,094)	0,646 (0,105)	0,737 (0,060)	0,927 (0,017)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,402 (0,441)	0,696 (0,219)	0,378 (0,491)	0,388 (0,476)	0,599 (0,262)	0,889 (0,073)
Karşılaştırma	0,319 (0,459)	0,561 (0,345)	0,259 (0,577)	0,261 (0,570)	0,421 (0,414)	0,839 (0,115)
Parça-bütün ilişkisini anlama	0,351 (0,469)	0,611 (0,322)	0,308 (0,557)	0,307 (0,539)	0,495 (0,360)	0,860 (0,100)
• Okuduğunu anlama soru 3						
Yansıtma	0,384 (0,268)	0,669 (0,147)	0,366 (0,285)	0,369 (0,284)	0,589 (0,165)	0,886 (0,046)
Yorumlama	0,429 (0,173)	0,677 (0,071)	0,420 (0,156)	0,422 (0,156)	0,601 (0,082)	0,889 (0,023)
Metinden bağımsız düşünme	0,537 (0,114)	0,742 (0,029)	0,533 (0,095)	0,529 (0,101)	0,673 (0,039)	0,909 (0,012)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,284 (0,193)	0,819 (0,093)	0,206 (0,253)	0,213 (0,221)	0,705 (0,117)	0,918 (0,033)
Karşılaştırma	0,432 (0,200)	0,934 (0,013)	0,408 (0,136)	0,415 (0,110)	0,844 (0,024)	0,9567 (0,010)
Parça-bütün ilişkisini anlama	0,319 (0,213)	0,916 (0,029)	0,280 (0,118)	0,282 (0,085)	0,805 (0,051)	0,946 (0,016)
• Okuduğunu anlama soru 5						
Yansıtma	0,455 (0,203)	0,741 (0,115)	0,441 (0,171)	0,430 (0,166)	0,668 (0,113)	0,908 (0,032)
Yorumlama	0,450 (0,264)	0,798 (0,131)	0,429 (0,272)	0,420 (0,271)	0,710 (0,168)	0,920 (0,047)
Metinden bağımsız düşünme	0,445 (0,198)	0,774 (0,159)	0,419 (0,210)	0,414 (0,199)	0,685 (0,174)	0,912 (0,050)
• Okuduğunu anlama soru 6						
Yansıtma	0,477 (0,172)	0,759 (0,106)	0,461 (0,126)	0,464 (0,140)	0,685 (0,096)	0,913 (0,027)
Yorumlama	0,487 (0,221)	0,80769 (0,11798)	0,464 (0,212)	0,466 (0,226)	0,725 (0,145)	0,923 (0,041)
Metinden bağımsız düşünme	0,473 (0,178)	0,77501 (0,15143)	0,442 (0,183)	0,452 (0,176)	0,692 (0,161)	0,914 (0,045)

Tablo 31. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %10 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,658 (0,089)	0,770 (0,066)	0,647 (0,096)	0,648 (0,083)	0,744 (0,077)	0,929 (0,023)
Karşılaştırma	0,684 (0,070)	0,796 (0,042)	0,675 (0,069)	0,680 (0,057)	0,770 (0,048)	0,936 (0,017)
Parça-bütün ilişkisini anlama	0,645 (0,076)	0,756 (0,043)	0,635 (0,078)	0,640 (0,072)	0,729 (0,050)	0,925 (0,017)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,337 (0,412)	0,663 (0,210)	0,325 (0,450)	0,325 (0,450)	0,561 (0,239)	0,878 (0,067)
Karşılaştırma	0,277 (0,438)	0,533 (0,346)	0,223 (0,563)	0,213 (0,558)	0,389 (0,399)	0,830 (0,111)
Parça-bütün ilişkisini anlama	0,294 (0,434)	0,579 (0,311)	0,256 (0,525)	0,251 (0,520)	0,459 (0,335)	0,850 (0,094)
• Okuduğunu anlama soru 3						
Yansıtma	0,399 (0,259)	0,683 (0,149)	0,375 (0,282)	0,387 (0,276)	0,599 (0,167)	0,889 (0,047)
Yorumlama	0,431 (0,159)	0,680 (0,070)	0,419 (0,149)	0,429 (0,147)	0,602 (0,077)	0,890 (0,023)
Metinden bağımsız düşünme	0,544 (0,125)	0,751 (0,034)	0,542 (0,107)	0,534 (0,121)	0,680 (0,045)	0,911 (0,015)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,251 (0,217)	0,803 (0,097)	0,197 (0,293)	0,184 (0,236)	0,688 (0,126)	0,913 (0,036)
Karşılaştırma	0,386 (0,211)	0,926 (0,016)	0,395 (0,096)	0,361 (0,134)	0,832 (0,023)	0,953 (0,012)
Parça-bütün ilişkisini anlama	0,262 (0,245)	0,906 (0,035)	0,260 (0,149)	0,229 (0,114)	0,789 (0,063)	0,941 (0,020)
• Okuduğunu anlama soru 5						
Yansıtma	0,472 (0,161)	0,749 (0,118)	0,459 (0,132)	0,453 (0,117)	0,677 (0,109)	0,910 (0,032)
Yorumlama	0,47160 (0,22251)	0,798 (0,126)	0,447 (0,245)	0,438 (0,213)	0,714 (0,158)	0,920 (0,045)
Metinden bağımsız düşünme	0,46370 (0,18089)	0,765 (0,160)	0,430 (0,201)	0,424 (0,172)	0,680 (0,172)	0,911 (0,049)
• Okuduğunu anlama soru 6						
Yansıtma	0,51393 (0,13469)	0,500 (0,251)	0,499 (0,166)	0,476 (0,160)	0,465 (0,287)	0,851 (0,080)
Yorumlama	0,47866 (0,20895)	0,440 (0,360)	0,449 (0,248)	0,430 (0,269)	0,428 (0,390)	0,841 (0,109)
Metinden bağımsız düşünme	0,44606 (0,22944)	0,410 (0,398)	0,410 (0,282)	0,394 (0,294)	0,373 (0,432)	0,826 (0,120)

Tablo 32. Okuduğunu Anlama Uygulamasında Üç Puanlayıcı İçin Kayıp Veri %15 Olduğunda Hesaplanan Puanlayıcılar Arası Güvenirlilik Katsayıları

	Conger'in Kappa (SH)	Gwet'in AC ₂ (SH)	Fleiss'nin Kappa (SH)	Krippendorff'un Alpha (SH)	Brennan- Prediger (SH)	Uzlaşma Yüzdesi (SH)
• Okuduğunu anlama soru 1						
İlişkilendirme	0,682 (0,093)	0,785 (0,070)	0,668 (0,099)	0,669 (0,093)	0,759 (0,081)	0,933 (0,027)
Karşılaştırma	0,720 (0,067)	0,819 (0,043)	0,711 (0,066)	0,707 (0,060)	0,796 (0,049)	0,943 (0,020)
Parça-bütün ilişisini anlama	0,667 (0,098)	0,769 (0,061)	0,652 (0,109)	0,653 (0,107)	0,742 (0,071)	0,928 (0,025)
• Okuduğunu anlama soru 2						
İlişkilendirme	0,667 (0,098)	0,769 (0,189)	0,409 (0,430)	0,396 (0,431)	0,621 (0,218)	0,895 (0,062)
Karşılaştırma	0,321 (0,416)	0,570 (0,318)	0,269 (0,520)	0,259 (0,526)	0,430 (0,370)	0,842 (0,104)
Parça-bütün ilişisini anlama	0,350 (0,400)	0,623 (0,304)	0,316 (0,529)	0,305 (0,519)	0,507 (0,336)	0,863 (0,094)
• Okuduğunu anlama soru 3						
Yansıtma	0,372 (0,270)	0,664 (0,153)	0,362 (0,307)	0,357 (0,275)	0,580 (0,172)	0,883 (0,051)
Yorumlama	0,451 (0,171)	0,688 (0,073)	0,442 (0,172)	0,436 (0,153)	0,617 (0,084)	0,893 (0,029)
Metinden bağımsız düşünme	0,519 (0,158)	0,737 (0,044)	0,519 (0,150)	0,512 (0,147)	0,666 (0,055)	0,907 (0,023)
• Okuduğunu anlama soru 4						
İlişkilendirme	0,312 (0,205)	0,822 (0,095)	0,246 (0,248)	0,243 (0,258)	0,715 (0,119)	0,921 (0,037)
Karşılaştırma	0,442 (0,195)	0,935 (0,021)	0,419 (0,093)	0,414 (0,102)	0,847 (0,026)	0,957 (0,019)
Parça-bütün ilişisini anlama	0,288 (0,228)	0,910 (0,033)	0,249 (0,099)	0,265 (0,090)	0,795 (0,052)	0,943 (0,023)
• Okuduğunu anlama soru 5						
Yansıtma	0,436 (0,198)	0,729 (0,112)	0,403 (0,167)	0,424 (0,167)	0,649 (0,104)	0,902 (0,031)
Yorumlama	0,463 (0,222)	0,798 (0,126)	0,438 (0,222)	0,424 (0,233)	0,712 (0,154)	0,920 (0,044)
Metinden bağımsız düşünme	0,456 (0,195)	0,764 (0,175)	0,421 (0,215)	0,417 (0,209)	0,675 (0,187)	0,910 (0,053)
• Okuduğunu anlama soru 6						
Yansıtma	0,489 (0,147)	0,470 (0,269)	0,454 (0,184)	0,463 (0,186)	0,434 (0,303)	0,843 (0,085)
Yorumlama	0,480 (0,219)	0,438 (0,380)	0,434 (0,283)	0,435 (0,269)	0,422 (0,406)	0,839 (0,114)
Metinden bağımsız düşünme	0,447 (0,232)	0,403 (0,410)	0,393 (0,303)	0,399 (0,290)	0,371 (0,441)	0,825 (0,123)

EK-6:

SINIF İÇİ KORELASYON İÇİN VARSAYIM İNCELEMESİ

Tablo 33. Puanlayıcılara göre matematik okuryazarlığı toplam puanına ait betimsel istatistik değerleri

	Ortalama	Standart sapma	Çarpıklık (SH)	Basıklık (SH)	Kolmogorov-Smirnov (p)
Puanlayıcı 1	21,01	8,29	-0,248 (0,161)	-0,401 (0,321)	0,052 (0,200)
Puanlayıcı 2	23,73	8,10	-0,289 (0,161)	-0,503 (0,321)	0,062 (0,033)
Puanlayıcı 3	18,83	7,08	-0,254 (0,161)	-0,138 (0,321)	0,059 (0,053)

Tablo 34. Puanlayıcılara göre okuduğunu anlama toplam puanına ait betimsel istatistik değerleri

	Ortalama	Standart sapma	Çarpıklık (SH)	Basıklık (SH)	Kolmogorov-Smirnov (p)
Puanlayıcı 1	29,23	9,46	-0,804 (0,160)	0,739 (0,320)	0,128 (0,000)
Puanlayıcı 2	29,50	7,41	-0,116 (0,160)	1,171 (0,320)	0,071 (0,006)
Puanlayıcı 3	29,44	8,92	-0,193 (0,160)	0,708 (0,320)	0,063 (0,027)

Tablo 35. Varyansların homojenliği için Levene Testi

<i>Matematik okuryazarlığı toplam puanı için</i>		
	Levene istatistiği	p*
Puanlayıcı 1-2	0,041	0,839
Puanlayıcı 1-3	6,483	0,011
Puanlayıcı 2-3	5,717	0,017
Puanlayıcı 1-2-3	3,962	0,019
<i>Okuduğunu anlama toplam puanı için</i>		
	Levene istatistiği	p*
Puanlayıcı 1-2	9,328	0,002
Puanlayıcı 1-3	0,282	0,596
Puanlayıcı 2-3	6,873	0,009
Puanlayıcı 1-2-3	5,193	0,006
*p < 0,001		



GAZİ GELECEKTİR...