



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eđitim Bilimleri Anabilim Dalı

Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**PISA 2009 OKUMA BECERİLERİ AÇIK UÇLU SORULARININ
PUANLANMASINDA GENELLENEBİLİRLİK KURAMINDAKİ
FARKLI DESENLERİN KARŞILAŞTIRILMASI**

Meral ALKAN

Doktora Tezi

Ankara, 2013

PISA 2009 OKUMA BECERİLERİ AÇIK UÇLU SORULARININ
PUANLANMASINDA GENELLENEBİLİRLİK KURAMINDAKİ FARKLI
DESENLERİN KARŞILAŞTIRILMASI

Meral ALKAN

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

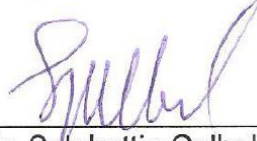
Ankara, 2013

KABUL VE ONAY

Meral ALKAN tarafından hazırlanan "PISA 2009 Okuma Becerileri Açık Uçlu Sorularının Puanlanmasında Genellenebilirlik Kuramındaki Farklı Desenlerin Karşılaştırılması" başlıklı bu çalışma, 24.01.2013 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



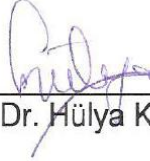
Prof. Dr. Şener Büyüköztürk (Başkan)



Prof. Dr. Selahattin Gelbal



Doç. Dr. Nuri Doğan (Danışman)



Doç. Dr. Hülya Kelecioğlu



Yrd. Doç. Dr. Burcu Atar

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

24.01.2013

Meral ALKAN

Kaan, Gökçe ve Hakan'a

TEŞEKKÜR

Çalışmalarımın her aşamasında beni yönlendiren, olur olmaz saatlerde gönderdiğim e-postalara anında yanıt veren, yapıcı önerileriyle bu çalışmanın tamamlanmasında değerli katkıları olan danışmanım Doç. Dr. Nuri DOĞAN'a,

Hacettepe Üniversitesinde yüksek lisans öğrencisi olarak başladığım ilk günden, doktora çalışmalarımı tamamladığım son güne kadar benden desteklerini hiç eksik etmeyen, her konuda önerileriyle yol gösteren, şu an bulunduğum aşamaya gelmemde büyük katkıları olan değerli hocalarım Prof. Dr. Selahattin GELBAL ve Doç. Dr. Hülya KELECİOĞLU'na,

Yüksek lisans tez çalışmalarım ve doktora tez çalışmalarım da yapıcı önerileriyle bana yol gösteren, her konuda destek olan, hem akademik yönden hem de insan ilişkileri açısından kendisinden çok şey öğrendiğim değerli hocam Prof. Dr. Şener BÜYÜKÖZTÜRK'e,

Yaptığım çalışmaya değerli katkılarıyla yön veren Yrd. Doç. Dr. Burcu ATAR'a,

Fikir ve önerileriyle çalışmalarımı yönlendiren ve destekleyen Yrd. Doç. Dr. Neşe GÜLER'e,

teşekkürlerimi sunarım.

ÖZET

ALKAN, Meral. *PISA 2009 Okuma Becerileri Açık Uçlu Sorularının Puanlanmasında Genellenebilirlik Kuramındaki Farklı Desenlerin Karşılaştırılması*, Doktora Tezi, Ankara, 2013.

Bu araştırmada, PISA 2009 Okuma Becerileri performansını değerlendirmede kullanılan açık uçlu soruların, birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla elde edilen farklı desenler Genellenebilirlik Kuramına göre karşılaştırılmıştır.

Bu çalışmanın örneklemini, PISA 2009 uygulamasına katılan 4996 öğrenciden, okuma becerileri alanında yer alan soruları yanıtlayan ve yanıtladıkları kitapçıklar birden fazla puanlayıcı tarafından puanlanan 886 öğrenci oluşturmaktadır. Araştırmada, iki farklı senaryo üzerinden genellenebilirlik kuramında kullanılmak üzere iki desen tasarlanmıştır. Bu desenlerden ilki, öğrenci (ö), soru (s) ve puanlayıcı (p) değişkenleri olmak üzere, öğrencilerin aynı beceriler konusunda puanlayıcıların her biri tarafından puanlandığı $\text{ö} \times \text{s} \times \text{p}$ çapraz desendir. İkinci desen ise, her bir puanlayıcının öğrencilerden sadece bir kısmını puanlamasıyla oluşan, öğrenci ve puanlayıcı değişkenlerinin yuvalanmış olduğu, soruların ise bu değişkenlerle çaprazlanmış olduğu $(\text{ö:p}) \times \text{s}$ desendir.

Analizler EduG 6 Programıyla gerçekleştirilmiştir. EduG 6 Programı, varyans analizine dayalı ve genellenebilirlik kuramı analizleri için geliştirilmiş bir programdır. Program, araştırmacının ölçmeye etki eden en büyük varyans kaynağını belirlemesine olanak sağlar ve Karar çalışmasıyla hata varyansına en çok katkısı olan örneklem deseninde yapılan değişikliğin etkilerini görmesine yardımcı olur.

$\text{Ö} \times \text{S} \times \text{P}$ ve $(\text{Ö:P}) \times \text{S}$ desenleri karşılaştırıldığında, $(\text{Ö:P}) \times \text{S}$ deseni ile kestirilen bağıl ve mutlak hata varyanslarının $\text{Ö} \times \text{S} \times \text{P}$ desenine göre daha

küçük olduğu, dolayısıyla G ve Phi katsayılarının daha büyük değerler aldığı görülmektedir. Bu iki desende yapılan Karar çalışmaları incelendiğinde, her iki desende de puanlayıcı sayısını artırmanın G ve Phi katsayılarında artış sağladığı görülmüştür. 2, 4, 5, 6 ve 7 nolu Kitapçıklarda, puanlayıcı sayısını yarıya indirerek ya da madde sayısını yarıya indirerek, çoğu zaman her ikisinin sayısını da yarıya indirerek kabul edilebilir düzeylerde G katsayısına ulaşmanın mümkün olduğu görülmektedir.

2, 4, 5, 6, 7 nolu Kitapçıklarda madde ve puanlayıcı sayısının indirilip, 8 ve 12 nolu Kitapçıklarda madde sayısı sabit tutularak, puanlayıcı sayısının artırılmasının zaman, işgücü ve ekonomi açısından uygun olacağı sonucuna ulaşılmıştır.

Anahtar Sözcükler

Genellenebilirlik Kuramı, Güvenirlilik, G çalışması, K çalışması, PISA 2009.

ABSTRACT

ALKAN, Meral. *Comparison of Different Designs In Scoring of PISA 2009 Reading Open Ended Items According to Generalizability Theory*, Ph.D. Dissertation, Ankara, 2013.

In this study, different designs which were formed jointly and alternatively assesment of students by more than one rater in scoring of PISA 2009 reading open ended items were compared according to Generalizability Theory.

886 students who answered reading items and scored by more than one rater were selected from 4996 students who participated PISA 2009. In the study, two different designs were formed and G and D studies were done according to Generalizability Theory. The first design is a crossed design $S \times I \times R$ (student x item x rater) which students answered all of the items and scored by all of the raters and the second design is a partially nested design $(S:R) \times I$ which students answered all of the items but they were nested in raters.

EduG 6 was used to carry out Generalizability Analysis. EduG is a program based on the Analysis of Variance (ANOVA) and designed to carry out Generalizability analysis. It enables you to identify which sources of variance have the greatest influence on your measurement observations, and, through D study allows you to see the potential effect of changing your sampling design to reduce the greatest contributions to measurement error.

When the results of Generalizability analysis according to $S \times I \times R$ and $(S:R) \times I$ designs compared, it was observed that variance rates that were estimated for variables in $(S:R) \times I$ design tends to be smaller and so G and Phi coefficients in this design tend to be higher. As a result of D studies, in both designs when the number of raters increases, G and Phi Coefficients increase as well. In Booklets 2, 4, 5, 6 and 7, reducing the number of raters by half or reducing the number of items by half or reducing the number of raters and items by half at the same time, still provide acceptable G and Ghi Coefficients.

In Booklets 2, 4, 5, 6 and 7, reducing the number of items and raters by half; in Booklets 8 and 12, keeping the number of items constant and increasing the number of raters, will be convenient in terms of time, labor and economy.

Key Words

Generalizability Theory, Reliability, G Study, D Study, PISA 2009.

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM	ii
TEŞEKKÜR	iv
ÖZET	v
ABSTRACT	vii
İÇİNDEKİLER	ix
KISALTMALAR	xii
TABLolar DİZİNİ	xiii
BÖLÜM I	1
GİRİŞ	1
1.1. PROBLEM DURUMU	1
1.1.1. Uluslararası Öğrenci Değerlendirme Programı (PISA).....	8
1.1.1.1. PISA açık uçlu sorularının puanlanması.....	11
1.1.1.2. Puanlama işlemlerinin temel çerçevesi	12
a) Uluslararası Puanlayıcı Eğitimi.....	12
b) Puanlayıcılar ve Puanlayıcı Kodları.....	12
c) Uluslararası Puanlama Servisi.....	13
d) Gizlilik Formları.....	13
e) Ulusal Puanlayıcı Eğitimi.....	14
1.1.2. Güvenirlilik.....	16
1.1.3. Genellenebilirlik Kuramı.....	17
1.1.3.1. Genellenebilirlik (G) çalışması.....	24
1.1.3.2. Karar (K) çalışması.....	24
1.2. PROBLEM CÜMLESİ	25
1.3. ALT PROBLEMLER	25
1.4. SAYILTILAR	26
1.5. SINIRLILIKLAR	27
1.6. ARAŞTIRMANIN AMACI VE ÖNEMİ	27
1.7. İLGİLİ ARAŞTIRMALAR	28
1.7.1. Yurtdışında Yapılan Çalışmalar.....	28

1.7.2. Yurtdışında Yapılan Çalışmalar.....	33
BÖLÜM II.....	41
YÖNTEM.....	41
2.1. ARAŞTIRMANIN TÜRÜ.....	41
2.2. EVREN VE ÖRNEKLEM.....	41
2.3. ARAŞTIRMA VERİLERİ.....	43
2.4. VERİLERİN ANALİZİ.....	44
BÖLÜM III.....	46
BULGULAR VE YORUMLAR.....	46
3.1. ÖĞRENCİ (Ö), SORU (S) VE PUANLAYICI (P) DEĞİŞKENLERİNİN ÇAPRAZ TASARLANDIĞI Ö X S X P DESENİNİN GENELLENEBİLİRLİK ÇALIŞMASI SONUÇLARI.....	46
3.1.1. Ö X S X P Deseninde G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	46
3.1.2. Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışması (K) Sonuçları.....	58
3.2. ÖĞRENCİ (Ö) ve PUANLAYICI (P) DEĞİŞKENLERİNİN YUVALANMIŞ, SORU (S) DEĞİŞKENİNİN İSE ÇAPRAZ TASARLANDIĞI (Ö:P) X S DESENİNİN GENELLENEBİLİRLİK ÇALIŞMASI SONUÇLARI.....	71
3.2.1. (Ö:P) X S Deseninde G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	72
3.2.2. (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar (K) Çalışması Sonuçları.....	74
3.3. Ö X S X P VE (Ö:P) X S DESENLERİNE AİT G ÇALIŞMASI SONUÇLARININ DEĞİŞİMİ.....	78
3.4. Ö X S X P ve (Ö : P) X S DESENLERİNDE PUANLAYICI ve ÖĞRENCİ SAYILARININ ARTTIRILIP AZALTILMASIYLA YAPILAN KARAR ÇALIŞMALARINDA ELDE EDİLEN SONUÇLARIN DEĞİŞİMİ.....	84

3.4.1. Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışmalarında Elde Edilen Bağıl ve Mutlak Hata Varyanslarının Değişimi.....	84
3.4.2. Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışmalarında Elde Edilen G ve Phi Katsayılarının Değişimi.....	89
3.5. HER BİR KİTAPÇIKTA KABUL EDİLEBİLİR DÜZEYDE BİR GENELLENEBİLİRLİK SEVİYESİ ELDE ETMEK İÇİN GEREKLİ MİNİMUM SORU SAYISI.....	92
BÖLÜM IV.....	100
SONUÇ VE ÖNERİLER.....	100
4.1. SONUÇLAR.....	100
4.2. ÖNERİLER.....	109
KAYNAKÇA.....	111
EKLER.....	118

KISALTMALAR

EARGED	: Eğitimi Araştırma ve Geliştirme Dairesi
ICILS	: Uluslararası Bilgisayar ve Bilgi Teknolojileri Okuryazarlığı (International Computer and Information Literacy Study)
MEB	: Milli Eğitim Bakanlığı
NUTS	: (Nomenclature D'unités Territoriales Statistiques) İstatistik Bölge Birimleri (İBB) Sınıflandırması
OECD	: Ekonomik İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Cooperation and Development)
PISA	: Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment)
TIMSS	: Uluslararası Matematik ve Fen Eğilimleri Araştırması
G Kuramı	: Genellenebilirlik Kuramı
G Katsayısı	: Genellenebilirlik Katsayısı
G Çalışması	: Genellenebilirlik Çalışması
K Çalışması	: Karar Çalışması
X	: Çaprazlanma
(:)	: Yuvalanma
ö	: Öğrenci
p	: Puanlayıcı
s	: Soru
ö x p	: Öğrenci x puanlayıcı ortak etkisi
ö x s	: Öğrenci x soru ortak etkisi
ö:p	: Öğrenci : puanlayıcı ana etkisi
ö x s x p	: Öğrenci x soru x puanlayıcı artık etkisi
p : ö x g	: Puanlayıcı : öğrenci x soru artık etkisi

TABLOLAR DİZİNİ

Tablo 1.	PISA Uygulamasında Yer Alan Soru Türleri.....	10
Tablo 2.	Kitapçık 2’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	47
Tablo 3.	Kitapçık 4’te, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	49
Tablo 4.	Kitapçık 5’te, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	50
Tablo 5.	Kitapçık 6’da, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	52
Tablo 6.	Kitapçık 7’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	53
Tablo 7.	Kitapçık 8’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	55
Tablo 8.	Kitapçık 12’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	57
Tablo 9.	Kitapçık 2’de, Ö X S X P Deseninde, Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	59
Tablo 10.	Kitapçık 2’de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları...60	
Tablo 11.	Kitapçık 4’te, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	61
Tablo 12.	Kitapçık 4’te, ÖXSXP Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	62
Tablo 13.	Kitapçık 5’te, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	62
Tablo 14.	Kitapçık 5’te, ÖXSXP Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	63

Tablo 15. Kitapçık 6'da, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	64
Tablo 16. Kitapçık 6'da, ÖXSXP Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	65
Tablo 17. Kitapçık 7'de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	66
Tablo 18. Kitapçık 7'de, ÖXSXP Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	67
Tablo 19. Kitapçık 8'de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	68
Tablo 20. Kitapçık 8'de, ÖXSXP Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	69
Tablo 21. Kitapçık 12'de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	70
Tablo 22. Kitapçık 12'de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	71
Tablo 23. Kitapçık 2'de, (Ö:P) X S Deseninde G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	72
Tablo 24. Kitapçık 8'de, (Ö:P) X S Deseninde G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	73
Tablo 25. Kitapçık 2'de, (Ö:P) X S Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	75
Tablo 26. Kitapçık 2'de, (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	76
Tablo 27. Kitapçık 8'de, (Ö:P) X S Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	77

Tablo 28. Kitapçık 8'de, (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	78
Tablo 29. Kitapçık 2'de, Ö X S X P ve (Ö:P) X S Desenlerine Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	79
Tablo 30. Kitapçık 8'de Ö X S X P ve (Ö:P) X S Desenlerine Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri.....	82
Tablo 31. Ö X S X P ve (Ö:P) X S Desenlerine Ait Genellenebilirlik Düzeyleri.....	84
Tablo 32. Kitapçık 2'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	85
Tablo 33. Kitapçık 8'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	88
Tablo 34. Kitapçık 2'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	90
Tablo 35. Kitapçık 8'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları.....	91
Tablo 36. Kitapçık 2'de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	93
Tablo 37. Kitapçık 4'te, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	94
Tablo 38. Kitapçık 5'te, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	95
Tablo 39. Kitapçık 6'da, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	96
Tablo 40. Kitapçık 7'de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik	

	Katsayısı İçin Gerekli Minimum Soru Sayısı.....	97
Tablo 41.	Kitapçık 8’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	98
Tablo 42.	Kitapçık 12’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı.....	99

BÖLÜM I

GİRİŞ

Bu bölümde, araştırmanın problem durumuna, problem cümlesine ve alt problemlere, araştırmanın amacı ve önemine, sayıtlara, sınırlılıklara, kısaltmalara ve ilgili araştırmalara yer verilmiştir.

1.1. PROBLEM DURUMU

21. yüzyıla girerken, yaşanan toplumsal, ekonomik ve teknolojik gelişmeler her alanda hızlı bir değişime sebep olmuştur. Toplumların bu değişime ayak uydurabilme isteği, eğitimin niteliği konusunu ön plana çıkarmıştır. Eğitimin niteliği, yeni nesillerin bu değişime ayak uyduracak yeni beceri ve yeterliklerle donatılmasında en önemli etkidir. Bilginin güç olarak kabul edildiği ve hızlı yayıldığı günümüzde, eleştirel düşünen, sorgulayan, kendi öğrenmelerinden sorumlu, yaratıcı, hayata hazır bireyler yetiştirmek eğitim sistemlerinin en önemli hedefi haline gelmiştir. Bu durum, eğitim uygulamalarını etkilediği kadar değerlendirme uygulamalarını da etkilemiştir. Eğer konu bilgisi tek başına yeterli bir kriter değilse, bireyin sorunun doğru yanıtını verilen seçenekler arasından seçmesi üzerine kurulu bilgi ve beceri testleri de tek başına yeterli değildir. Öğrenmeye getirilen bu anlayış testlerin de bireyin yanıtlarını yapılandıracağı formda düzenlenmesi gereğini ortaya çıkarmıştır. Öğretim programları öğrencilerde alt düzey düşünmeyi gerektiren bir anlayıştan, üst düzey düşünmeyi gerektiren bir anlayışa; öğretim yöntem ve teknikleri öğretmen merkezli bir yapıdan öğrenci merkezli bir yapıya; ölçme değerlendirme yaklaşımları ise bilgilerin ne derece kazanıldığını ölçen bir yapıdan, bilgilerin yeni durumlarda ya da gerçek yaşamda nasıl kullanılabileceğini ölçen bir yapıya dönüşmüştür. Bu durum, öğretim sürecinde yeni değerlendirme yaklaşımlarına yönelmenin tetikleyici gücü olmuştur (Biemer, 1993).

Ancak, bu gelişmelere rağmen, eğitim kurum ve kuruluşlarının bir çoğunda kullanılan ölçme araçları çoktan seçmeli testlerdir. Öğretmenlerin sınıf uygulamalarında, ulusal düzeyde gerçekleştirilen seviye belirleme sınavlarında, öğrenci seçme sınavlarında, özellikle çok sayıda öğrencinin katıldığı değerlendirmelerde çoktan seçmeli testlerinin diğer sınav türlerine oranla daha sık kullanıldığı görülmektedir. Bunun en önemli nedenleri arasında bu tür sınavların kalabalık gruplar için uygun olması, çok sayıda madde kullanarak pek çok davranışın ölçülebilmesi, test ve madde istatistiklerinin kolay hesaplanması ve kolay puanlanması sayılabilir.

Çoktan seçmeli sınavlarda, öğrencilerden kendi yanıtlarını oluşturmaları beklenmez. Öğrencilerin yapması gereken, soru kökünü okuyup soruya ilişkin olarak verilen dört ya da beş seçenektan kendilerince doğru olanı işaretlemeleridir. Her sorunun bir doğru yanıtı vardır. Bu nedenle 4 seçenekli bir testte 3 seçenek çeldirici görevi görür. Çeldiriciler, sorunun yanıtını bilmeyen ya da eksik veya yanlış bilgiye sahip öğrencilerin doğru yanıtı bulmalarını zorlaştırmak amacı taşımaktadır. Çoktan seçmeli sınavlarda cevaplama işlemi seçeneklerden birinin seçilmesiyle yapıldığından, cevapların puanlanmasında puanlayıcıdan puanlayıcıya farklılık göstermez. Çoktan seçmeli testler bu nedenle puanlamanın objektif yapılmasına olanak sağlar.

Çoktan seçmeli testlerin avantajları yanında sınırlılıkları da vardır. Bu testlerde tahminle doğru cevaba ulaşmak mümkündür. Doğru cevabı bilmeden, tahminle doğru cevaba ulaşmak şans başarısı olarak adlandırılır. Doğru cevabı bilmeyen bir öğrencinin 4 seçenek arasından doğru cevabı bulma ihtimali %25'tir. Bu nedenle çeldirici sayısının artırılması, öğrencinin doğru cevabı tahminle bulma ihtimalini düşürmektedir. Şans başarısı test puanlarındaki varyansı arttırmakta, dolayısıyla güvenilirlik ve geçerliğin düşmesine sebep olmaktadır. Şans başarısını önlemenin bir yolu, öğrencilerin sadece bildikleri soruları yanıtlamalarını sağlamak için, belirli miktarda yanlış sorunun doğru cevaplardan birinin iptaline dayanan uygulamadır. Örneğin SBS sınavlarında, her test için

yanlış cevap sayısının üçte biri, doğru cevap sayısından çıkarılarak geçerli cevaplara karşılık gelen ham puanlar bulunmaktadır (SBS Kılavuzu, 2012).

Çoktan seçmeli testlerde karşılaşılan bir diğer problem, bu testlerde kopya çekmenin açık uçlu sorulara oranla daha kolay olmasıdır. Bunu önlemek için paralel formların kullanılması ya da test ortamının teste uygun hale getirilmesi gibi çeşitli yollar kullanılabilir. Bunun yanında SBS kılavuzunda yer aldığı gibi:

Ayrıca sınavdan sonra cevap kâğıtlarının değerlendirmesi yapılırken aynı salonda sınava giren ve aynı tür soru kitapçığını kullanan öğrenciler arasında bilgisayar sistemi aracılığıyla ikili/toplu kopya taraması yapılır. İkili/toplu kopya taraması sonucunda kopya tespit edilen öğrencilerin ilgili testleri iptal edilir. Bu nedenle sınav esnasında adayların cevap kâğıtlarını hiçbir öğrencinin göremeyeceği şekilde önlerinde bulundurmaları ve kesinlikle herhangi bir yöntemle kopya çekme teşebbüsünde bulunmamaları gerekmektedir.

şeklinde uygulamalara ve test yönergesinde uyarılara yer verilebilir.

Sözü edilen sınırlılıklar, alınacak önlemlerle giderilebilir. Ancak çoktan seçmeli testlerin en büyük sınırlılığı performansı, yeteneği ya da beceriyi yeterli düzeyde ölçememesidir. Öğrenci davranışlarını değerlendirmek amacıyla kullanılan çoktan seçmeli, kısa yanıtlı, doğru yanlış, eşleştirmeli, boşluk tamamlamalı gibi klasik test yöntemleri, problem çözme, okuduğunu anlama, eleştirel düşünme, analitik düşünme, empati kurma, araştırma yapma, karar verme, toplumsal tarihin önemini anlama, yaratıcılık gibi üst düzey zihinsel süreçleri belirlemede yetersiz kalmaktadır (Kutlu, 2006).

Çoktan seçmeli testlerin bu sınırlılıkları, özellikle uluslararası düzeyde öğrencilerin davranışlarının ölçülmesinde başka yöntemlerin ön plana çıkması gereğini doğurmuştur. Bu yöntemlerden biri performans değerlendirmedir. Performans değerlendirmede kullanılan sorular, çoktan seçmeli maddeler yerine, açık uçlu sorulardan oluşur. Açık uçlu sorular, çoktan seçmeli testler ile

ölçmenin zor olduğu yüksek düzeydeki bilişsel beceriler için kolaylıkla kullanılabilirler. Üst düzey bilişsel becerileri kullanarak, öğrenciler sadece bilgiyi hatırlamaz, onun yerine elde edilen bilgiyi problem çözmede, analiz ve değerlendirme yapmada kullanırlar. Bu tür sorular öğrencilerin bir konuyu gerçekten anlayıp anlamadığı konusunda en fazla bilgi veren sorulardır. Sanders (1996), “İyi sorular farklı düşünme olasılıklarını tanımlarken farklı düşünme becerilerine dayanmaktadır ve dar anlamda ne öğrenildiğine odaklanmaktansa doğrudan öğrenmeye odaklanmaktadır” demektedir. Bu bağlamda, açık uçlu soruların öğrencilerin anlamalarına, mantık ve bilgiyi kullanma yeteneklerine odaklandığı ve öğrencilerin başarı düzeylerini çoktan seçmeli sorulara oranla daha iyi gösteren ve eğitim için daha iyi bir yol gösterici nitelik taşıdığı söylenebilir.

PIRLS, TIMSS, PISA, ICILS gibi bir çok uluslararası değerlendirme çalışmasında açık uçlu sorular kullanılmaktadır. Türkiye, öğrenci başarılarını ölçme ve değerlendirme çalışmalarını uluslararası boyutta sürdürmek, Türk öğrencilerin başarı düzeylerini, eğitim sistemini diğer ülkelerin verileriyle karşılaştırmak, güçlü ve iyileştirmeye açık yönleri belirlemek için uluslararası başarı belirleme çalışmalarına katılmaktadır. Bu çalışmalara PIRLS, TIMSS, ICILS, PISA çalışmaları örnek olarak verilebilir. Bütün bu değerlendirme çalışmalarında farklı soru türleri kullanılmaktadır.

Açık uçlu soruları yanıtlarken öğrencilerin kendi yanıtlarını oluşturmaları, öğrencilerin birbirlerinden farklı bireysel yanıtlar vermelerine ve soruları kendi bakış açılarıyla yanıtlamalarına imkân tanımaktadır. Dolayısıyla çok çeşitli yanıtlarla karşılaşmanın mümkün olduğu bu durumda puanlama yöntemleri ve puanlama güvenilirliği büyük önem taşımaktadır.

Öğrencilerin performanslarını güvenilir bir şekilde değerlendirmek için öğrenciden beklenen davranışlar önceden belirlenmeli ve bu davranışlara göre kontrol listesi veya dereceli puanlama anahtarı hazırlanmalıdır. Yapısal özellikleri bakımından iki tür dereceli puanlama anahtarı bulunmaktadır.

Bunlardan biri bütünsel dereceli puanlama anahtarı, diğeri de analitik dereceli puanlama anahtarıdır. Bütünsel puanlama anahtarında öğrencinin gösterdiği performansın bütününe tek bir puan verilmektedir ve her düzeyde performansın kalitesini belirleyen tanımlar bulunmaktadır (Kutlu, 2009). Analitik puanlama anahtarı ise, sadece belli parçaların değerlendirildiği, test edildiği yanıtın ya da yapılan işin aşama aşama puanlandığı durumlarda kullanılan puanlama anahtarıdır. Klein ve arkadaşları (1998), bütünsel puanlamanın analitik puanlama kadar güvenilir olduğunu bulmuşlardır. Bütünsel puanlama puanlayıcının daha az vaktini alır ancak verilen puanların doğrulanması daha zordur çünkü karara nasıl ulaşıldığını gösteren açık bir kanıt yoktur. Bu nedenle bütünsel puanlamanın, bütünün parçalara göre daha önem taşıdığı durumlarda kullanılması uygun olur. Analitik puanlama anahtarıyla yapılan puanlamalar performansın her bir alt boyutu için bilgi verir. Bu nedenle alt bileşenlerin önem taşıdığı durumlarda analitik puanlama yapılması daha uygun olur.

Performans değerlendirmede, puanlama sürecinde bir çok hata meydana gelebilir. Bu hata türleri üç başlık altında toplanabilir. Sabit hatalar, bir ölçmeden diğerine, miktarı değişmeyen hatalardır. Ölçmeye konu olan nesnelere üzerinde yapılan bütün ölçümlerde yapılan hata miktarı eşittir. Hata miktarı ölçmeden ölçmeye değişmediği için, burada ölçme sonuçlarına karışan hata sabit hatadır. Örnek olarak, yazılı sorularından 5. sorunun silik çıkması nedeniyle, hiç bir öğrenci tarafından yapılamaması gösterilebilir. Sistemik hatalar, ölçülen büyüklüğe, ölçmeceye veya ölçme koşullarına ve belli bir duruma bağlı olarak miktarı değişen hatalardır. Cetvelin 2 cm kısa ölçmesi nedeniyle, 10 cm yerine 8 cm, 20 cm yerine 16 cm ölçmesi bu duruma örnek olarak gösterilebilir. Gerek sistemik, gerekse sabit hatalarda ölçme sonuçlarına karışan hatanın miktarı, yönü ve kaynağı bellidir. Bu sebeple bu tür hataların ölçme sonuçlarına karışmasını engellemek ya da düzeltmek nispeten daha kolaydır. Tesadüfi hatalar, ölçme sonuçlarına geliş güzel karışan, kaynağı, miktarı ve yönü kesin olarak bilinmeyen hatalardır. Tesadüfi hatalar, ölçmenin yapıldığı ortam, ölçmeyi yapan kişi, ölçme aracı vb. gibi bir çok hata kaynağından gelen bir hata türüdür. Örneğin, dikkatsizce okunup puanlanan notlar, doğru puanlandığı halde yanlış

kayıt edilen notlar bu duruma bir örnek teşkil edebilir. Tesadüfi hatalar, ölçmelere tek yönlü olarak karışmazlar. Ölçme sonuçlarına bazen pozitif, bazen de negatif yönde karışabilirler ve kontrol edilemezler. Bu nedenle, ölçme sonuçlarına karışan hata ve hata kaynakları iyi tanımlanmalı ve hata miktarını kestirebilecek yöntemler bulunmalıdır (Turgut, 1992). Ölçme sonuçlarının doğruluğu, bu sonuçlara bağlı olarak verilecek kararları etkilediği için çok önemlidir. Güvenirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak tanımlanabilir (Baykul, 2000).

Değerlendirme sürecinde puanlayıcılar kullanıldığı zaman, değişkenlik kaynağı olabilmekte, ölçme hatalarına yol açabilmekte ve değerlendirmenin güvenilirliğini düşürebilmektedir. Performans değerlendirme çalışmalarında puanlayıcı güvenilirliği büyük önem taşımaktadır. Puanlayıcı güvenilirliği, birden çok puanlayıcının aynı özelliğe verdikleri puanların tutarlılığıdır. Puanlayıcılar performansın ölçülmesinde güvenilirliği etkileyen en önemli hata kaynaklarından biridir. Puanlamayı yapan kişiden kaynaklanan çeşitli puanlama yanlışlıkları söz konusudur. Scullen, Mount ve Goff (2000), puanlayıcı etkisini “öğrencinin gerçek performansı ile ilgili olmayan, performans değerlendirmede sistematik hataya sebep olan doğrudan puanlayıcıyla ilgili etkilerin geniş bir kategorisidir” şeklinde tanımlamaktadır. Puanlayıcı etkisinin, puanlayıcının motivasyon, endişe, başarı, öz yeterlik gibi psikolojik durumlarından (Alfallay, 2004; Bernardin & Villanova, 2005), kişisel özelliklerinden (Wexley & Youtz, 1985), önceki inanışlarından, puanlayıcının cinsiyeti, yaşı gibi demografik özelliklerinden, puanlayıcının puanlama deneyiminden (weigle, 1998) kaynaklandığı söylenebilir. Puanlayıcı kaynaklı hata türlerine bakıldığında, dört ana başlık altında toplamak mümkündür. (1) Cömertlik/Katılık: puanlayıcının öğrenciyi puanlarken genel olarak gerçek değerinin altında ya da üstünde puanlamasıdır. (2) Halo Etkisi: Puanlayıcının, bireyin tek bir kişilik özelliğini ya da davranışını temel alarak puanlama yapmasıdır. Puanlayıcının, bireyin bir alandaki özelliğinin etkisinde kalarak bunu kişinin tüm özelliklerine genellemesi ile ortaya çıkar. (3) Ortaya yönelme eğilimi, puanlayıcının uç değerlerdense puanlamada ortalama değerlere yönelmesidir. (4) Ranj Daralması:

Puanlayıcının puanlamada tüm puanları ölçeğin belirli bir aralığında veriyor olması durumudur.

Sözü edilen performans değerlendirmede puanlayıcı kaynaklı hataların güvenilirliği olumsuz yönde etkilememesi için çeşitli önlemler alınabilir. Bu önlemler kısaca şu şekilde özetlenebilir:

- Puanlayıcıların seçiminde dikkatli davranılmalıdır. Puanlayıcıların bu işi yapmak için motivasyonlarının yüksek olması gerekir. İlgi düzeyi düşük puanlayıcılar puanlama sürecini olumsuz etkiler.
- Puanlayıcılar için etkili bir eğitim programı düzenlenmelidir.
- Puanlayıcıların verdikleri puanların yapısı incelenmeli ve puanların güvenilirliği sorgulanmalıdır.
- İdeal puanlayıcı sayısı belirlenmelidir.
- Ölçek maddelerinin ideal sayısı belirlenmelidir. Ölçek maddelerinin sayısını artırmak puanlayıcının yükünü artırır, ancak olabildiğince çok sayıda madde kullanmak tercih edilmelidir. (Darst, Zakrajsek & Mancini (1989).

Puanlayıcıların ölçümlere karışan diğer değişkenlik kaynaklarıyla etkileşimleri de güvenilirlik için önemlidir (Brennan, 2001). Bu nedenle güvenilirlik belirlenirken birçok değişken kaynağından gelen hatalar da dikkate alınmalıdır. Güvenirliğin ölçülmesinde değişkenlik kaynaklarını ve bu kaynaklar arasındaki etkileşimleri de bir arada değerlendirebilen yöntemlerden biri de Genellenebilirlik Kuramı'dır. Genellenebilirlik kuramının aynı anda birden fazla hata kaynağını bir arada göz önünde bulundurmasından dolayı Türkiye'nin katıldığı uluslararası sınavları bir de bu açıdan değerlendirmekte fayda olduğu düşünülmektedir. Bu konuda yapılan çalışmalar oldukça azdır. Smith (1997) tarafından yapılan TIMSS açık uçlu sorularının Genellenebilirliği çalışmasında, açık uçlu sorularda puanlayıcı etkisine bakılmıştır. Bu çalışmada, İngilizce konuşulan 7 ülkenin herbirinden 50 kitapçık ile her madde için 150 öğrencinin yanıtlarını incelemiştir. Bütün maddelerde kabul edilebilir (0.80) bir Genellenebilirlik düzeyi için puanlayıcı

sayısını 5'ten 15'e yükseltmek gerektiğini ve puanlayıcı sayısının az olduğu ülkelerde bu durumun bir sorun yaratabileceği ifade edilmiştir. Sharma & Weathers (2003) tarafından yapılan bir çalışmada, uluslararası araştırma projelerinde kullanılan ölçeklerin katılan bütün ülkelere Genellenebilirliği araştırılmıştır. Ölçeğin bütün ülkelerde aynı anlamı taşıdığı ve belirli bir ülkeye özel olmadığı sonucuna ulaşılmıştır. Eğer istenilen Genellenebilirlik düzeyi 0.90 ise 17 maddeden 11'inin kullanılmasının bu düzey için yeterli olduğu sonucuna ulaşılmıştır. Örneklem büyüklüğü 290 kişidir ve yapılan çalışmada 140 kişinin istenilen düzeyde bir Genellenebilirlik düzeyi için yeterli olduğu sonucuna varılmıştır. Minimum sayıda madde kullanımı sınav/anket süresini düşürmekte, yorgunluk etkisini ortadan kaldırmakta ve maliyeti düşürmektedir.

Bu bağlamda, ülkemizin katıldığı uluslararası değerlendirme çalışmalarından PISA'da, gerekli minimum madde sayısını, gerekli minimum örneklem sayısını ya da gerekli minimum puanlayıcı sayısını bilmenin bu sınavlarda harcanan emeğin ve maliyetin azaltılmasına yardımcı olacağı düşünülmektedir. PISA uygulamasıyla ilgili detaylı bilgiye aşağıda yer verilmiştir.

1.1.1. Uluslararası Öğrenci Değerlendirme Programı (PISA)

PISA Projesi, 15 yaş grubu öğrencilerinin zorunlu eğitimin sonunda, bilgi toplumunda karşılaşılabilecekleri durumlara ne kadar hazırlıklı olarak yetiştirildiklerini belirlemek amacıyla Ekonomik İşbirliği ve Kalkınma Teşkilâtı - OECD (Organisation for Economic Co-Operation and Development) tarafından geliştirilmiştir (EARGED, 2005). PISA, öğrencilerin, matematik, fen ve okuma becerileri alanlarındaki bilgi ve becerilerinin değerlendirildiği uluslararası en büyük eğitim araştırmalarından biridir. Üç yılda bir yapılan bu araştırmayla, OECD üyesi ülkeler ve diğer katılımcı ülkelerdeki (dünya ekonomisinin yaklaşık olarak %90'ı) 15 yaş grubu öğrencilerin modern toplumdaki yerlerini alabilmeleri için gereken temel bilgi ve becerilere ne ölçüde sahip oldukları değerlendirilmektedir (EARGED, 2010).

PISA projesinde her dönem okuma becerileri, matematik ve fen okuryazarlığı alanlarından sadece birine temel alan olarak ağırlık verilmektedir. Bununla birlikte, diğer iki alan da yapılan değerlendirme kapsamına dâhil edilmektedir. Dokuz yıllık bir döngüde bu alanlardan her biri, bir kez temel alan olmaktadır. İlk defa 2000 yılında gerçekleştirilen PISA uygulamasında temel alan okuma becerileri, 2003'te matematik okuryazarlığı, 2006'da ise fen okuryazarlığı olmuştur. 2009 yılında ise dokuz yıllık yeni bir döngü başlamıştır. 2009 yılında ağırlıklı alan okuma becerileridir.

Dünya genelinde 15 yaş grubu bireyler bir okur olarak ne biliyor ve yeterlikleri nelerdir? Yazılı metinlerde ihtiyaçları olan bilgiyi bulabiliyor, yorumlayabiliyor ve kullanabiliyorlar mı? Kendi deneyim ve anlayışlarına göre eleştirel bir şekilde metin hakkında düşünebiliyorlar mı? Kişisel ilgileri ve istekleri doğrultusunda ya da daha farklı nedenlerden dolayı, farklı içerikli ve farklı amaçlar için yazılmış değişik metinleri okuyabiliyorlar mı? PISA 2009'daki okuma becerileri değerlendirmesi çerçevesinde bu sorulara yanıt aranmaktadır. PISA 2009'da okuma becerileri; kişisel hedefleri yakalama, belirli bir konuda kişinin sahip olduğu bilgiyi ve potansiyeli artırma, toplumda katılımcı bir birey olabilme ve yazılı metinleri anlama, kullanma, yansıtma olarak tanımlanmaktadır (OECD, 2009).

PISA 2009 uygulaması, ülkemizde 2009 yılının Nisan ayında gerçekleştirilmiştir. Bu uygulamada, 12 istatistikî bölge birimleri sınıflandırmasına (İBB, NUTS (Nomenclature D'unités Territoriales Statistiques)) ve okul türlerine göre tabakalandırılarak PISA uluslararası merkez tarafından seçkisiz yöntemle belirlenen toplam 170 okuldan 4996 öğrenci yer almıştır (EARGED, 2010).

PISA 2009'da 13 madde demeti (bundle) yer almaktadır. Bunlardan 7'si okuma becerileri, 3'ü Matematik ve 3'ü Fen maddelerine aittir. Bu madde demetleri belirli bir döngü düzenine göre 13 kitapçığa yerleştirilmiştir. Her bir kitapçıkta en az bir okuma becerileri madde demeti olmak kaydıyla, 4 madde demeti yer alır. Her bir madde demetinin yanıtlanma süresi 30 dakikadır. Dolayısıyla, her bir

öğrencinin değerlendirilme süresi 120 dakikadır. Her öğrenci seçkisiz yöntemle belirlenen 13 kitapçıktan birini yanıtlar (EARGED, 2010).

PISA uygulamasında farklı soru türleri yer almaktadır. Tablo 1’de soru türlerine ilişkin bilgi verilmiştir.

Tablo 1. PISA Uygulamasında Yer Alan Soru Türleri

<p>Kapalı Uçlu sorular (Tüm soruların %8’i)</p>	<p>Bu tür sorularda öğrencilerin önceden belirtilen formata uygun olarak kendi yanıtlarını oluşturmaları gerekmektedir. Öğrencilerin yanıtları veri giriş sistemine puanlayıcılar tarafından doğrudan yazılmakta ve sistem tarafından doğru (1) ya da yanlış (0) şeklinde puanlanmaktadır.</p>
<p>Çoktan Seçmeli Sorular (Tüm soruların %52’si)</p>	<p>Çoktan seçmeli sorularda, verilen dört ya da beş seçenekten birinin (çoktan seçmeli) ya da bir dizi seçenekten (“evet” ya da “hayır”, veya “katılıyorum” ya da “katılmıyorum”) bir ya da ikisinin seçilmesi gerekmektedir. Bu tür sorularda öğrencilerin işaretlediği seçenekler veri giriş sistemine puanlayıcılar tarafından girilmekte ve sistem tarafından doğru (1) ya da yanlış (0) şeklinde puanlanmaktadır.</p>
<p>Açık Uçlu Sorular (Tüm soruların %40’ı)</p>	<p>Açık uçlu sorularda, öğrencilerin kısa ya da uzun yanıtlar halinde kendi yanıtlarını oluşturmaları gerekmektedir. Bu tür sorular puanlama konusunda eğitim almış puanlayıcılar tarafından, tek bir puanlayıcının ya da 4 puanlayıcının kullanılan desenlere göre, bütünsel puanlama rehberi kullanılarak doğru (1) ya da yanlış (0) şeklinde, bazı sorularda ise doğru (2), kısmî doğru (1) ya da yanlış (0) şeklinde puanlanmaktadır.</p>

1.1.1.1. PISA Açık Uçlu Soruların Puanlanması

PISA 2009'da, Fen, Matematik ve Okuma becerilerinin değerlendirilmesine ilişkin olarak 13 kitapçık kullanılmıştır. Değerlendirmede yer alan Matematik, Fen ve Okuma becerilerine ait tüm bilişsel maddelerin toplamda %45'i, eğitim almış puanlayıcılar tarafından puanlanmıştır. Kitapçıklar, puanlayıcılara, her bir ülke için minimum örneklem sayısı olan 4500 kitapçıktan seçkisiz yöntemle seçilerek atanmıştır. 13 kitapçıktan her biri için ortalama 26 maddenin puanlanması gerekmektedir. PISA sonuçlarının karşılaştırılabilir olması için öğrenci yanıtlarının, puanlayıcıdan puanlayıcıya, ülkeden ülkeye hiçbir değişiklik göstermeden puanlanması hayati önem taşımaktadır. Puanlama için kabul edilebilir ve kabul edilemez örnek yanıtların da yer aldığı kapsamlı kriterler listesi uluslararası merkez tarafından hazırlanarak, her ülkenin proje yöneticisine Fen, Matematik ve Okuma becerileri puanlama rehberi olarak gönderilmiştir.

Açık uçlu soruların puanlanması işlemi için, öncelikle puanlama rehberinin çevirisinin ve adaptasyonu yapılmakta ve onay için uluslararası merkeze gönderilmektedir. Daha sonra puanlayıcılar seçilmekte ve puanlayıcılara puanlama eğitimi verilmektedir. Eğitim ve uygulama sırasında uygun yerel örneklere yer verilmektedir. Okullardan gelen kitapçıklar düzenlenmekte ve birden fazla puanlayıcı tarafından puanlanacak kitapçıkların seçimi yapılmaktadır. Öncelikle kitapçıklar uluslararası desene uygun olarak tek bir puanlayıcı tarafından puanlanmaktadır. Tek puanlayıcının kullanıldığı desene göre yapılan puanlama işlemi tamamlandıktan sonra seçilen alt grupların birden fazla puanlayıcı tarafından puanlanması işlemine geçilmektedir. 13 kitapçığın her birinden 100'er adet kitap, birden fazla puanlayıcının kullanıldığı desen için seçilmiştir. Bu kitapçıklardan 2, 4, 5, 6, 7, 8 ve 12 nolu kitapçıklar okuma becerilerinin birden fazla puanlayıcı tarafından puanlanmasında kullanılmıştır. Okuma becerileri soruları 16 puanlayıcı tarafından puanlanmıştır. Puanlayıcılar, bütün kısa cevaplı soruları ve açık uçlu soruları bağımsız olarak puanlamıştır. Kitapçık 2'de 6 soru, Kitapçık 4'te 5 soru, Kitapçık 6'da 7 soru, Kitapçık 7'de 7

soru, Kitapçık 8’de 8 soru ve Kitapçık 12’de 6 soru yer almaktadır. Birden çok puanlayıcı tarafından puanlanan kitapçıklardan bir örneklem kontrol için uluslararası merkeze gönderilmektedir.

Test kitapçıklarında her bir sorunun kod numarası, sorunun sağ köşesinde yer almaktadır. Örneğin R412S1, 412 kodlu okuma (Reading) ünitesinin 1. sorusu anlamına gelmektedir. Tek puanlayıcının kullanıldığı puanlama deseninde puanlar, soru kod numarasının hemen yanında yer alan puan kodlarının daire içine alınması ile kitabın üzerinde verilir. Birden fazla puanlayıcının kullanıldığı puanlama deseninde ise ilk üç puanlama, hazırlanmış puan formları üzerinde yapılırken, son puanlayıcı kitap üzerinde puanlama yapmakta ve böylece puanlayıcıların diğer puanlayıcıların verdiği puanı görmemeleri ve bilmemeleri sağlanmaktadır. Puanlama sorumlusu, her bir puanlama formunda öğrenci kodunun ve puanlayıcı kodunun doğru olarak yazıldığını kontrol etmektedir. Çünkü bu kodlar verinin kalitesi açısından çok önemlidir. Puanlamalar, madde madde yapılmaktadır. Okuma ünitelerinde 0, 1 ve 9, sırasıyla yanlış, doğru ve boş anlamına gelmektedir. Bazı açık uçlu sorularda 2 tam puan, 1 kısmi puan anlamına gelmektedir.

1.1.1.2. Puanlama İşlemlerinin Temel Çerçevesi

a) *Uluslararası Puanlayıcı Eğitimi*

Katılımcı her ülkeden temsilcilerin, biri pilot uygulamadan önce, diğeri esas uygulamadan önce olmak üzere uluslararası iki puanlayıcı eğitimine katılmaları gerekmektedir.

b) *Puanlayıcılar ve Puanlayıcı Kodları*

Ülke temsilcileri puanlayıcıları belirlemekle sorumludur. Puanlayıcıların üst düzey akademik niteliklerinin olması gerekmez, ancak ortaöğretim Matematik, Fen ve Türkçe/Türk Dili Edebiyatı öğretmenleri olmaları ve bu yaş çocuklarının

kendilerini ifade etme biçimlerine alışık olmaları beklenmektedir. Puanlama işlemleri yaklaşık 2 ay sürmektedir ve puanlayıcıların bu süre boyunca kendilerini bu işe adanmaları beklenmektedir.

Puanlayıcı sayısı birden fazla puanlayıcı tarafından yapılan puanlama desenine göre değişmektedir. Esas uygulamada okuma becerileri için tavsiye edilen toplam puanlayıcı sayısı 16'dır. Bu sayı 3 aylık test döneminde, 4500 – 6000 öğrenciyi değerlendirmeye tabi tutan ülkeler için yeterli bir sayıdır. Daha geniş öğrenci grubu için, ülkeler farklı desenlerin bir kombinasyonunu oluşturabilirler ve bu desen uluslararası merkez tarafından kabul edilirse uygulanabilir. Birden fazla puanlayıcının kullanıldığı desende, her bir alan için minimum 4 puanlayıcıya ihtiyaç vardır. Puanlama döneminin uzunluğu nedeniyle, yedek 2 puanlayıcının bulunmasında yarar görülmektedir.

Uluslararası merkez tarafından belirlendiği şekilde, puanlayıcılara 3 basamaklı bir kod numarası verilir. Okuma becerilerini puanlayacak olan 16 puanlayıcıya 201, 202, 203, 204..... kodları verilmiştir. Kod numaraları uygulanan desene göre kitapçıkların puanlayıcılara dağıtımını sağlamak ve birden çok puanlayıcının kullanıldığı desenlerde puanlayıcı tutarlılığını takip etmek içindir.

c) *Uluslararası Puanlama Servisi*

Ulusal düzeyde belirli maddelerle ilgili çözümlenemeyen herhangi bir konu ortaya çıktığında danışmak üzere uluslararası puanlama servisi oluşturulmuştur. Sorulan sorulara verilen yanıtlar web sitesinde yayınlanmakta ve yanıtlara bütün katılımcı ülke temsilcileri tarafından ulaşılabilir.

d) *Gizlilik Formları*

PISA dokümanlarını görmeden önce, aday puanlayıcıların gizlilik sözleşmesi imzalaması gerekmektedir. Bu sözleşmeyle puanlayıcılar bu soruları puanlayıcı grubu dışına çıkarmayacağını beyan etmektedirler.

e) *Ulusal Puanlayıcı Eğitimi*

PISA uygulamasında puanlayıcı olarak görev alacak herkesin, PISA ile ilgili puanlama konusunda tecrübesi olup olmadığına ve pilot uygulamada puanlayıcı eğitimi alıp almadığına bakılmaksızın belirli bir puanlama eğitimine katılması gerekmektedir. Proje yöneticilerine yardımcı olmak amacıyla, puanlama rehberine ek olarak eğitim materyalleri hazırlanmıştır.

Kitapçıkların puanlanması, kitapçıklarda yer alan farklı ünitelere (cluster) göre yapılmaktadır. Bu, puanlama işleminin aynı ünitenin içinde yer alan bütün maddelerin ayrı ayrı kodlanması ve bütün kitaplarda bu ünitelerin tamamlandıktan sonra yeni bir üniteye geçilmesi anlamına gelmektedir. Her üniteyle ilgili yapılan eğitim çalışması sonrası, doğru puanlamaları uluslararası merkez tarafından yapılmış örnek maddeler üzerinde uygulama yapılmaktadır. Puanlayıcılar sonuçları hep birlikte gözden geçirmekte, tartışmaya zaman ayrılmakta ve daha önceden verilen puanlamanın nedenini açıklığa kavuşturmaktadırlar. Bu aşamadan sonra puanlayıcılar yerel örneklerin üzerinde çalışmaktadır. Eğitimin başında, puanlayıcı eğitimi alanlardan puanlama rehberinin uygulanmasında yüksek düzeyde tutarlılık beklendiği, verilen cevapların puanlama rehberine uygun olarak yapılıp yapılmadığının çok önemli olduğu söylenir.

PISA uygulamasında bütünsel dereceli puanlama anahtarı kullanılmıştır. PISA uygulamasında yer almış sorulardan biri ve sorunun yanıtına ilişkin puanlama anahtarı aşağıda yer almaktadır. PISA Okuma Becerileri değerlendirilmesinde kullanılan soru ve puanlama rehberine ilişkin örnekler (EARGED, 2010), Ek.1'de paylaşılmıştır.

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.

Önceki sayfada yer alan "Kan Araniyor" adlı metin bir Fransız internet sitesinden alınmıştır. Aşağıdaki sorulara yanıt vermek için bu metinden yararlanınız.

Soru 8: KAN ARANIYOR

R429Q08 - 0 1 9

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

.....

.....

KAN ARANIYOR PUANLAMA 8

SORUNUN AMACI:

Bilgileri birleştirme ve yorumlama: Bir yorum geliştirme
Kısa bir metinde bağlantılar kurarak bir sonuca ulaşma.

Tam Puan

Son kan bağışından bu yana yeterli zaman geçmesi gerektiğini vurgular.

- Onun son kan vermesinden bu yana 8 hafta geçip geçmediğine bağlı.
- Yeterince süre geçtiyse kan verebilir aksi takdirde veremez.

1.1.2. Güvenirlik

Ölçme çalışmalarında amaç, olabildiğince gerçek puanlara yakın gözlenen puanlar elde etmektir. Gerçek puanlara yakın ölçme sonuçları, ölçmelerdeki hata puanlarının azlığı ölçüsünde gerçekleşir (Baykul, 2000). Ölçme sonuçları bulundurdukları tesadüfi hataların azlığı ölçüsünde güvenilirdir. Bu sebeple güvenirlik “ölçme sonuçlarının tesadüfi hatalardan arınlık derecesi” (Turgut, 1992, s. 27) olarak da tanımlanır. Aynı araçla ve aynı bireyler üzerinde yapılan birden çok ölçmede aynı sonuçların alınması beklenir. Bu sebeple, güvenirlik, aynı bireyler üzerinde yapılan, bir niteliğe ait ölçmelerin benzer şartlarda tekrar elde edilebilirliği (Croker ve Algina, 1986, s. 105) olarak da tanımlanır. Ölçme aracının önemli teknik özelliklerinden biri olan güvenirlik, ölçme aracının ölçtüğü özelliği ya da özellikleri, ne derecede bir kararlılıkla ölçülere yansıtılabileceğini gösterir (Özçelik, 1992). Ölçümlerimizin güvenirliliğinin yüksek olması demek ölçümlerdeki tesadüfi hataların az olması demektir. Benzer şekilde, ölçümlerde ölçme hatası arttıkça veya ölçümlerin gerçeği yansıtma düzeyi azaldıkça güvenirlik de azalır.

Ölçme yaparken ölçülen özelliğin miktarı hakkında doğru bilgi edinebilmek için ve bu ölçme sonuçlarına dayanarak doğru kararlar verebilmek için, mümkün olduğunca hatasız ölçme elde etmek istenir. Ancak gerek fiziksel bilimlerdeki ölçümlerde, gerekse eğitim ve psikoloji gibi sosyal bilimlerde ölçme sonuçlarına bir hata karışır. Bu hatanın miktarı genelde bilinemez. Eğitim ve psikolojideki

değişkenlerin yapısı göz önüne alındığında, bu alanlarda yapılacak ölçümlerde hata olma olasılığı daha fazladır. Eğitim uygulamalı bir bilim dalıdır ve eğitim süreci içerisinde birçok ölçme durumu yer alır. Eğitimdeki ölçmelerde hatalardan sakınılmak isteniyorsa, hata kaynakları ve özellikle bu kaynakların ölçme sonuçlarını nasıl etkilediği bilinmelidir.

Ölçme sonuçlarına dayanarak bir karara varılması amaçlandığından, ölçmelerde ne kadar hata olabileceğini kestirmek zorunlu bir hal alır. Çünkü verilecek kararların isabetlilik derecesi yapılan ölçümlerin doğruluğuna bağlıdır. Bu sebeple ölçme sonuçlarına karışan hata ve hata kaynakları iyi tanımlanmalı ve hata miktarını kestirebilecek yöntemler bulunmalıdır (Turgut, 1992). Güvenirliğin hesaplanmasında kullanılacak yöntemlerden biri Genellenebilirlik kuramıdır.

1.1.3. Genellenebilirlik Kuramı

Genellenebilirlik kuramı, davranış ölçmede güvenirliliğin değerlendirilmesini, güvenilir gözlemlerin tasarlanmasını, araştırılmasını ve kavramlaştırılmasını sağlayan, temelinde varyans analizine (ANOVA) dayalı olan istatistiksel bir kuramdır (Cronbach, Rajaratnam Gleser, 1963; Cronbach, Gleser, Rajaratnam, 1972).

Genellenebilirlik kuramının amacı, ölçme sonuçlarını farklı varyans kaynaklarına ayırarak, yorumlayarak ve tanımlayarak, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına doğrulukla genellenmesini sağlamaktır (Brennan, 2001).

Genellenebilirlik kuramında aynı anda birden fazla hata kaynağı bir arada göz önünde bulundurulurken tek bir güvenirlilik değerine ulaşmak mümkün olmaktadır. Shavelson ve Webb'e (1991) göre, Genellenebilirlik kuramı, çoklu varyans kaynaklarını tek bir analizde ele alır ve her bir varyans kaynağının büyüklüğünün belirlenmesini sağlar. Birçok olası test formundan biri, olası

birçok durumdan birinde, olası birçok test uygulayıcısının biri tarafından uygulanabilir. Bu seçimlerin herbiri potansiyel bir hata kaynağıdır. G kuramı ölçme ve ölçme desenini geliştirmek için bu hata kaynaklarının her birini dikkate alır. Klasik test kuramına dayalı geleneksel güvenilirlik yöntemi, ölçmelerde bir kerede sadece bir hata kaynağını dikkate alır. Örneğin test-tekrar test yönteminde sadece test durumu hata kaynağı olarak ele alınır. Paralel formlarda güvenilirlik hesaplanmasında sadece test formları hata kaynağı olarak kabul edilir. İç tutarlılıkta ise sadece maddeler hata kaynağı olarak ele alınır. Bu nedenle Klasik Test Kuramı çok sınırlı bilgi vermektedir. Güvenirliğin bir türü (örn. Test-tekrar test), diğer bir tür güvenilirlik (iç tutarlılık) konusunda yorumlar yaparken kullanılamaz. Farklı türde güvenilirlik katsayıları mevcut olsa da, güvenilir ölçümler elde etmek için kaç test formuna, kaç maddeye ve duruma ihtiyaç olduğunu birleştirilmiş bilgilerle belirlemek oldukça zordur (Webb, Rowley, & Shavelson, 1988).

Genellenebilirlik kuramında, puanlayıcı, madde ya da zaman gibi hata kaynaklarına değişkenlik kaynağı (facet) adı verilir. Değişkenlik kaynağı, ölçme hatasının olası kaynağı olarak tanımlanabilir. Değişkenlik kaynağıyla ilişkili bir varyans istenen bir varyans olmayıp bu tür bir varyansın olabildiğince küçük olması istenir (Alharby, 2006). Değişkenlik kaynaklarının düzeyleri koşullar (conditions) olarak adlandırılır. Örneğin, "puanlayıcılar" bir değişkenlik kaynağı ise birinci puanlayıcı, ikinci puanlayıcı, üçüncü puanlayıcı vb.nin her biri bir koşuldur. Genellikle var olan bir değişkenlik kaynağının olası koşullarının sonsuz büyüklükte olduğu varsayılır. Araştırmada yer alan gözlemlerin yapıldığı örneklemin yerine geçebilecek olası gözlemlerin tümüne "*kabul edilebilir gözlemlerin evreni (the universe of admissible observation)*" adı verilir. "*Genellenebilirlik evreni (the universe of generalization)*" ise araştırmacının genellemek istediği koşulların tümüdür. Pek çok ölçme durumunda bireyler ya da öğrenciler ölçmenin objesi (the object of measurement) olarak ele alınırlar. Bireyler arası farklılıklar doğaldır ve sistemattir. Bu sebeple bireylere bağlı varyans istenilen bir durum olduğu için bireyler değişkenlik kaynağı olarak düşünülmez. Genellenebilirlik kuramı, belirli bir ölçme sonucuna ya da gözlenen

puana değil ölçme sonuçlarının belirli bir örneklemeden çok daha geniş olan evrenine nasıl genellenebileceğine odaklanmaktadır. Genellenebilirlik evreni, ölçülmek istenilen özelliğe ilişkin tüm ölçme sonuçlarını kapsar. Bir puanın kullanılabilirliği, o puanın genellenebilirlik evreni adı verilen daha geniş durumlara doğru bir şekilde nasıl genellenebileceğine bağlıdır. Genellenebilirlik kuramı, “gözlenen puanlar, bireylere ilişkin davranışların, belirlenen evrendeki tüm durumlara doğru bir şekilde genellenmesine ne ölçüde izin vermektedir” sorusuna cevap arar. Bir bireyden elde edilebilecek ideal ölçme sonucu, bireyin, tüm kabul edilebilir gözlemleri üzerinden elde edilecek olan puanlarının ortalamasıdır ve bu değere “evren puanı” adı verilir. (Shavelson, Webb ve Rowley, 1989).

Genellenebilirlik kuramına bağlı yapılan çalışmalarda ölçme sonuçları çapraz (crossed) ya da yuvalanmış (nested) olabilmektedir. Tamamıyla çapraz desenlerde, ölçme objesi olan bireyler tüm değişkenlik kaynaklarının tüm durumlarında puanlanmıştır. Örneğin, tüm öğrenciler (b) testte yer alan tüm maddeleri (m) cevaplandırmış ve iki farklı puanlayıcı (p) tüm öğrencilerin tüm cevaplarını puanlamışsa, öğrenciler, maddeler ve puanlayıcılara göre tümüyle çaprazlanmış bir desen (b x m x p) söz konusu olacaktır. Yuvalanmış desende ise, bireyler değişkenlik kaynaklarının sadece bir durumu için puanlanmışken diğer durumlarda puanları bulunmamaktadır. Örneğin, bir testte yer alan her bir maddeyi (m) farklı bir öğrenci (b) cevapladığında ve her bir öğrencinin cevabını farklı bir puanlayıcının (p) puanladığı durumdaki desenler yuvalanmış desen (b:m:p) olarak tanımlanır. Genellenebilirlik kuramında yapılan çalışmalarda ölçme sonuçlarının bir kısmının çapraz, bir kısmının ise yuvalanmış olduğu karışık (mixed) desenler de mümkündür (Brennan, 2001). Ancak tüm olası değişkenlik kaynaklarına ilişkin hatanın kestirime imkân verdiğinden dolayı G çalışmalarında tümüyle çapraz desenler daha çok tercih edilmektedir (Musquash ve O’Connor, 2006). Cronbach (1972), bütün varyans bileşenlerini hesaplamaya olanak tanımasından dolayı çapraz desen kullanımını önermektedir. Ancak yuvalanmış desenin bazı varyans bileşenlerinin

hesaplanmasında serbestlik derecesini artırmasından dolayı yararlı olacağını ifade etmiştir.

Genellenebilirlik kuramında söz konusu olan deęişkenlik kaynakları sabit ya da tesadüfi olarak tanımlanabilmektedir. Bir deęişkenlik kaynağında yer alan tüm durumlar, o deęişkenlik kaynağında yer alabilecek olası tüm dięer durumlar ile deęiştirilebilir olma özelliğine sahipse, bu deęişkenlik kaynağı tesadüfi olarak tanımlanır (Kieffer, 1998). Örneęin, yapılan bir matematik sınavında yer alan maddeler yine aynı alanda yapılacak bir sınavda olabilecek dięer maddeler ile deęiştirilebilir nitelikteyse, bu durumda çalışmada kullanılan maddeler tesadüfi olarak ele alınır. Tesadüfi durumların söz konusu olduęu deęişkenlik kaynaklarına baęlı yapılan çalışmalar, araştırmacıya o deęişkenlik kaynağı için tüm durumların yer aldıęı evrene genelleme yapabilme olasılıęı sağlar. Aksine, araştırmacı yaptıęı çalışmada yer alan deęişkenlik kaynağına baęlı sadece belirli durumlarla ilgileniyorsa, dięer durumlara genelleme yapmak gibi bir amacı yoksa, bu durumda ele alınan deęişkenlik kaynağı sabit olarak tanımlanır (Crocker ve Algina, 1986). Sabit deęişkenlik kaynaklarının bulunduęu çalışmalarda araştırmacının genelleme yapması doęru olmayacaktır (Kieffer, 1998).

Tüm öğrencilerin (b) testte yer alan tüm maddeleri (m) cevaplandırdıęı ve iki farklı puanlayıcı (p) tarafından tüm öğrencilerin tüm cevaplarının puanlandıęı durumda, öğrenciler, maddeler ve puanlayıcıların tümüyle çaprazlandıęı bir desen ($b \times m \times p$) elde etmiş oluruz. Genellenebilirlik kuramında, madde ve puanlayıcı deęişkeni gibi 2 deęişkenin yer aldıęı, bu deęişkenlerin sonsuz büyüklükte bir evrenden rastlantısal olarak seçildięi ve deęişkenlerin çapraz olduęu bir desene ait gözlenen puanı şu şekilde tanımlayabiliriz:

$$X_{bmp} =$$

μ	genel ortalama
$+ \mu_b - \mu$	birey etkisi
$+ \mu_m - \mu$	madde etkisi
$+ \mu_p - \mu$	puanlayıcı etkisi
$+ \mu_{bm} - \mu_b - \mu_m + \mu$	birey x madde etkisi
$+ \mu_{bp} - \mu_b - \mu_p + \mu$	birey x puanlayıcı etkisi
$+ \mu_{mp} - \mu_m - \mu_p + \mu$	madde x puanlayıcı etkisi
$+ X_{bmp} - \mu_b - \mu_m - \mu_p + \mu_{bm} + \mu_{bp} + \mu_{mp} - \mu$	artık etkisi

G kuramında önemli konulardan biri G çalışmasında kullanılacak desene karar vermektir. Çünkü G çalışmasında belirli bir desen için toplanan veri, D çalışması desenlerinin herbirinde gözlenen puan varyansını hesaplamada kullanılamaz (Crocker and Algina, 1986). Bütün varyans bileşenlerini hesaplamaya olanak tanimasından dolayı çapraz desen kullanımını önerilir (Cronbach,1972).

G kuramı, bir tek güvenilirlik katsayısı yerine G katsayılarından bahseder ve evren tanımının değişmesiyle katsayının değerinin de değişeceğini öngörür. Bir başka deyişle, Genellenebilirlik kuramı hem bireylerin performanslarına dayalı bağıl kararlar hem de bireylerin performanslarıyla ilgili mutlak kararlar alınmasına ilişkin iki farklı güvenilirlik katsayısının hesaplanmasına olanak tanır. Bağıl değerlendirmeler için genellenebilirlik katsayısı (G), mutlak değerlendirmeler için ise güvenilirlik katsayısı (Phi) hesaplanır (Shavelson & Webb, 1991; Brennan, 2001). Bu durum, Genellenebilirlik kuramında hem bağıl hem de mutlak olmak üzere iki ayrı anlamda karar vermeyi mümkün hale getirir. Bu bağlamda, G kuramında, karar vericiler bir tek test puanından farklı şekillerde yararlanabilirler (Shavelson,1991).

Bağıl karar için hesaplanan G katsayısı her bir ölçme objesinin, değişkenlik kaynağındaki aldığı ham puanın ne kadar yüksek olduğuna değil, diğer ölçme objelerinin puanlarının sıralaması arasındaki yerine bağlı olarak hesaplanır. Mutlak karar için hesaplanan G katsayısı ise çok daha katı bir değer olup, hem ölçmenin objesine ilişkin (genellikle bireylerdir) puanların sıralamasındaki

tutarlılığın derecesini hem de ham puanların tutarlılığın derecesini ortaya koyar. Belirli bir kesme puanının üzerindeki puanın önem taşıdığı performans ölçümlerinde (örneğin, ehliyet sınavlarında, uzmanlık sınavlarında vb.) mutlak G katsayısı tercih edilebilir (Lee ve Frisbie, 1999; Brennan, 1992). Elde edilen puanların puan sıralamasındaki yerinin önem taşıdığı durumlarda bağıl G katsayısını kullanmak uygun olacaktır. Bağıl ve mutlak kararlar için hesaplanan G katsayılarındaki karışıklığı ortadan kaldırmak üzere, bağıl kararlar için hesaplanan değere G katsayısı, mutlak kararlar için hesaplanan değere Phi katsayısı ya da güvenilirlik (dependability) katsayısı adı verilmektedir. Brennan (2001), G ve Phi katsayılarının yeterlik ölçütlerinin isteğe bağlı olarak değiştiğini ancak bazı araştırmacıların G ve Phi katsayılarının 0,80'den büyük olması durumunda "yüksek" olarak değerlendirilebileceğini ifade etmektedir. Shavelson ve Webb (1991), 0.80 ve üzeri Genellenebilirlik katsayılarının anlamlı olduğunu söylemektedir.

G katsayısını hesaplamak için öncelikle Bağıl Hata Varyansını hesaplamak gerekir. Bağıl değerlendirmede etkili olan 3 hata kaynağı vardır. Bunlar sırasıyla; birey madde ortak etkisi, birey puanlayıcı ortak etkisi ve artık varyanstır.

Bağıl hata, gözlenen sapma puanı ile evren sapma puanı arasındaki farktır. Bağıl hata varyansı, σ^2_{δ} şeklinde gösterilir ve klasik test kuramındaki hata varyansına karşılık gelir (Brennan, 2001). Bağıl hata varyansında, öğrencilerin başarısı içinde buldukları gruba göre değerlendirilir.

Bağıl hata varyansı ve G katsayısını hesaplama formülleri aşağıda verilmiştir:

$$\sigma^2_{\delta} = \frac{\sigma^2_{os}}{n_s} + \frac{\sigma^2_{op}}{n_p} + \frac{\sigma^2_{osp,e}}{n_s n_p}$$

$$G = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_\delta^2}$$

Mutlak hata, bireyin gözlenen puanı ve evren puanı arasındaki farktır. Mutlak hata varyansı, σ^2_Δ şeklinde gösterilir ve öğrencilerdeki değişim gruba bağlı olmadan mutlak bir ölçüte göre belirlenir.

Mutlak Hata Varyansını ve Phi Katsayısını hesaplama formülleri aşağıda verilmiştir:

$$\sigma^2_\Delta = \frac{\sigma_s^2}{n_s} + \frac{\sigma_p^2}{n_p} + \frac{\sigma_{os}^2}{n_s} + \frac{\sigma_{op}^2}{n_p} + \frac{\sigma_{sp}^2}{n_s n_p} + \frac{\sigma_{osp,e}^2}{n_s n_p}$$

$$\Phi = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_\Delta^2}$$

Mutlak değerlendirmede, belirli bir test formunun güçlüğü, o formu yanıtlayan kişinin gözlenen puanını etkileyecektir, dolayısıyla evren puanı da bundan etkilenir. Oysaki bağıl değerlendirmede, bu formun etkisi formu yanıtlayan herkes için sabittir. Bu nedenle formu yanıtlayanların sıralamasında bir farka neden olmamaktadır (Erlish & Shavelson, 1976 b). Dolayısıyla madde ana etkisi bağıl hata kararlarında yer almazken, mutlak performansı etkilediği için ölçme hatası tanımı içinde yer alır. Bu durum, Phi katsayısının G katsayısından daha küçük bir değer almasına sebep olmaktadır.

Genellenebilirlik kuramı, belirli bir amaca bağlı olarak, ölçme hatasının en aza indirgenebileceği ölçmelerin düzenlenmesine yani Karar (K) çalışmalarına imkân sağlar. Genellenebilirlik kuramına başlarken G ve K çalışmaları arasındaki ayrımları anlamamız gerekir. G: Genellenebilme ve K: Karar verme çalışmalarıdır. Araştırmacı öncelikli olarak hangi örneklemelerin evren için

genellenebileceğini düşünmelidir. G çalışmasının amacı, K çalışmasının planlanabilmesine yardımcı olmaktır. K çalışması, veriler toplandıktan sonra özel amaçlı karar verme konusu ile ilgilenir. Deney gruplarını karşılaştırır ya da daha çok değişkenler arası ilişkileri inceler.

1.1.3.1 Genellenebilirlik (G) Çalışması

Genellenebilirlik çalışması, örneklemden elde edilen ölçümlerin evrene genellenebilirlik derecesiyle ilgilidir. Amacı, evrensel puanlarla gözlenen puanlar arasındaki ilişkileri incelemektir. Zamana karşı tepkilerin kararlılığıyla, bir ölçme aracının iki ya da daha çok formundaki puanların eşdeğerliliğiyle ya da alt test puanlarının birbiriyle ilgili çalışmalar Genellenebilirlik çalışması kapsamına girer. Genellenebilirlik çalışması değişkenlik kaynaklarıyla ilgili bilgi sağlamak için tasarlanır.

1.1.3.2. Karar (K) Çalışması

Karar çalışmaları Genellenebilirlik çalışması tarafından genellenen katsayıların kısmi kararlar ve kısmi faktör grupları için bağımsız ölçümlerden, karar vericinin hangisini genellenebilmek istediğine göre tasarlanır. Araştırmada kullanılan değişkenlere ait değerler değiştirilerek hatayı en aza indirmek ve güvenilirliği maksimum hale çıkarmak için çeşitli tasarımlar yapmak mümkündür.

Bu çalışmada, PISA 2009 okuma becerileri sorularının birden fazla puanlayıcı tarafından puanlanmasına ilişkin farklı desenler ele alınmıştır. PISA çalışmasında sadece okuma becerilerinin puanlanmasında 16 puanlayıcıya ihtiyaç vardır. Bu puanlayıcıların okulların açık olduğu bir dönemde, okullarda görev yapan öğretmenlerden seçilmesi ve puanlama işleminin yaklaşık 2 ay sürmesi puanlama sürecinde sorunlar yaratmaktadır. Gerek puanlayıcı sayısının fazlalığı gerekse sürenin uzunluğu puanlayıcı sayısını ve süreyi düşürebilecek çözümler üretmeyi gerektirmiştir. Bu amaçla farklı desenler denenmiş, en uygun desen, en uygun puanlayıcı ve madde sayısı ve kabul

edilebilir düzeyde bir Genellenebilirlik için gerekli örneklem sayısı belirlenmeye çalışılmıştır. Puanlayıcılar performansın ölçülmesinde güvenilirliği etkileyen en önemli hata kaynaklarından biridir. Puanlamayı yapan kişiden kaynaklanan çeşitli puanlama yanlışlıkları söz konusudur. Puanlayıcı güvenilirliği ve puanlamaya karışan hata kaynakları ortaya çıkarıldığında, bu tür hata kaynaklarının güvenilirliği olumsuz yönde etkilememesi için çeşitli önlemler almak mümkün olabilir. En uygun soru sayısını, en uygun puanlayıcı sayısını belirlemek ve kabul edilebilir düzeyde bir Genellenebilirlik için minimum örneklem sayısını belirlemek, minimum sayıda madde kullanımını sağlayarak sınav/anket süresini düşürecek, yorgunluk etkisini ortadan kaldıracak ve maliyeti düşürecektir.

1.2. PROBLEM CÜMLESİ

PISA 2009 Okuma becerileri performansının ölçülmesinde kullanılan maddelerin bütünsel bir dereceli puanlama anahtarıyla birden fazla puanlayıcı tarafından, çapraz (ÖXSXP) ve yuvalanmış [(Ö:P)XS] desene göre puanlanması ile elde edilen puanların Genellenebilirlik (G) ve Karar (K) çalışmaları sonuçları nasıldır?

1.3. ALT PROBLEMLER

1. Öğrenci (ö), PISA okuma becerilerindeki performansı değerlendirmede kullanılan sorular (s) ve puanlayıcı (p) değişkenlerinin çapraz tasarlandığı Ö X S X P deseninin genellenebilirlik kuramı sonuçları nasıldır?
 - 1.1 Ö X S X P deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?
 - 1.2 Ö X S X P deseninde puanlayıcı ve öğrenci sayılarının arttırılıp azaltılmasıyla yapılan Karar çalışması (K) sonuçları nasıldır?

2. Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, soru (s) değişkeninin ise çapraz tasarlandığı (Ö:P) X S deseninin Genellenebilirlik kuramı sonuçları nasıldır?
 - 2.1. (Ö:P) X S deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?
 - 2.2. (Ö:P) X S deseninde puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla yapılan Karar çalışması (K) sonuçları nasıldır?
3. Ö X S X P ve (Ö:P) X S desenlerinden elde edilen G çalışması sonuçlarının değişimi nasıldır?
4. Ö X S X P ve (Ö:P) X S desenlerinde puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla yapılan Karar çalışmalarında elde edilen sonuçların değişimi nasıldır?
 - 4.1. Ö X S X P ve (Ö:P) X S desenlerinde puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla yapılan Karar çalışmalarında elde edilen bağıl ve mutlak hata varyanslarının değişimi nasıldır?
 - 4.2. Ö X S X P ve (Ö:P) X S desenlerinde puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla yapılan Karar çalışmalarında elde edilen G ve Phi sayılarının değişimi nasıldır?
5. Her bir kitapçıkta kabul edilebilir düzeyde bir Genellenebilirlik seviyesini elde etmek için gerekli minimum soru sayısı nedir?

1.4. SAYILTILAR

PISA 2009 okuma becerilerini ölçen sorulara verilen yanıtları puanlayan dört puanlayıcının puanları, birbirlerinden bağımsız olarak verdiği kabul edilmiştir.

1.5. SINIRLILIKLAR

Bu araştırma, PISA uygulamasında yer alan okuma becerileri alan sorularını yanıtlamış ve verdikleri yanıtlar birden fazla puanlayıcı tarafından değerlendirilmiş olan 886 öğrenciyle sınırlı tutulmuştur.

1.6. ARAŞTIRMANIN AMACI VE ÖNEMİ

Bu araştırmanın amacı, Genellenebilirlik kuramına göre PISA 2009 Okuma Becerileri öğrenci performanslarının puanlanmasında, öğrencilerin birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerden elde edilen G çalışması ve bu desenlerle yapılan Karar çalışması sonuçlarını karşılaştırmaktır.

Türkiye, 2000’li yılların başından itibaren, TIMSS, PIRLS, PISA gibi uluslararası düzeyde karşılaştırmalı değerlendirme çalışmaların içinde yer almaktadır. Bu çalışmaların en önemli özelliklerinden biri, Türkiye’de uygulanan büyük ölçekli değerlendirme çalışmalarından farklı olarak açık uçlu sorulara yer veriliyor olmasıdır. Bu durum beraberinde, açık uçlu soruların puanlayıcılar tarafından dereceli puanlama anahtarı kullanarak puanlanmasını getirmektedir. Puanlayıcılar tarafından yapılan puanlamanın objektif olması gerekir. Bir puanlamanın objektif olması, puanlayıcıların aynı sonuçta birleşmesi, bütün puanlayıcıların aynı yanıtı aynı puanı vermesi anlamına gelmektedir. Ölçme sonuçlarının objektifliği artıkça güvenilirliği de artar (Turgut, 1992).

Açık uçlu soruların puanlanmasında, öğrenci sayısına ve puanlanacak madde sayısına bağlı olarak, görev yapacak puanlayıcı sayıları artmakta ve puanlama süresi uzamaktadır. PISA 2009 uygulamasında, sadece okuma becerileri kısmının puanlanmasında 16 Türkçe/ Türk Dili Edebiyatı öğretmenine ihtiyaç duyulmuştur. Bu çalışmalar için okullarda görev yapan öğretmenlerin görevlendirilmesi, puanlama işlemlerinin okulların açık olduğu döneme denk gelmesinden dolayı sorun olmaktadır. PISA uygulamasında, ülkeler farklı

desenlerin bir kombinasyonunu oluşturabilirler ve bu desen konsorsiyum tarafından kabul edilirse, bu desen uygulanabilir. Bu nedenle, aynı sınav için oluşturulmuş farklı desenlerden hangisinin kullanılmasının daha uygun olduğunun belirlenmesinin ve her bir kitapçıkta istenilen Genellenebilirlik düzeyini sağlayan minimum madde sayısının belirlenmesinin, zaman, maliyet ve iş gücü açısından bu tür sınavlara katkı sağlaması beklenmektedir. Bu çalışmada zaman, maliyet ve iş gücü açısından çözümler üretebilecek sonuçlar elde edilebileceği düşünülmektedir.

Türkiye'nin yer aldığı uluslararası öğrenci değerlendirme programlarına ilişkin gerçek verilerin kullanıldığı, Genellenebilirlik kuramına dayalı çok az sayıda çalışma vardır. TIMSS, PIRLS, ICILS gibi uluslararası öğrenci değerlendirme çalışmalarında açık uçlu soruların puanlayıcı güvenilirliklerinin belirlenmesi ve karar çalışmalarının yapılması açısından bu çalışmanın örnek teşkil edeceği ve sonraki çalışmalara ışık tutacağı düşünülmektedir.

ÖSYM (Ölçme, Seçme ve Yerleştirme Merkezi) tarafından gerçekleştirilen seçme ve yerleştirme sınavlarında açık uçlu soruların kullanımına yönelik çalışmalar başlatılmıştır. Ancak bu tür sorulara ilişkin güvenilirlik ve karar çalışmaları bulunmadığından ve katılımın çok yüksek olduğu bu sınavlarda kaç maddenin yeterli olacağı, kaç puanlayıcının nasıl bir desene puanlama yapması gerektiği gibi konular büyük önem taşımaktadır. Bu nedenle bu çalışmanın yol gösterici olacağı düşünülmektedir.

1.7. İLGİLİ ARAŞTIRMALAR

1.7.1 Yurtiçinde Yapılan Çalışmalar

Atılgan ve Tezbaşaran (2005), aynı değişkenlik kaynaklarının farklı düzeyleri için alternatif K Çalışmaları senaryolarıyla elde edilen G ve Phi katsayıları ile aynı değişkenlik kaynaklarının aynı düzeyleriyle gerçekte uygulanması sonucunda elde edilen G ve Phi katsayılarının tutarlılıklarını karşılaştırmıştır.

Araştırmada müzik öğretmenliği programına öğrenci seçmek amacı ile yapılan özel yetenek seçme sınavlarının ardışık iki yıl verileri kullanılmıştır. Bu sınavlarda ortak boyut olan çift ses, ezgi ve tartım boyutları bir yıl üç puanlayıcı diğer yıl ise dört puanlayıcı tarafından puanlanmıştır. Araştırmada kullanılan özel yetenek sınavlarının alt boyutlardan oluşması nedeniyle, birey (b), görev (g) ve puanlayıcı (p) değişkenlik kaynağı (facets) olmak üzere, G-kuramının çok değişkenli ($b \times g \times p$) deseni kullanılmıştır. Üç puanlayıcı sınav durumuyla dört puanlayıcı sınav durumunun G ve Phi katsayıları hesaplanarak, senaryolarla elde edilen katsayılarla gerçekleşenler karşılaştırılmıştır. Sonuç olarak; puanlayıcı sayısının artırılması senaryosu ile kestirilen G ve Phi katsayılarının gerçekten büyük, puanlayıcı sayısının azaltılması yoluyla kestirilen G ve Phi katsayılarının ise gerçekte olduğundan küçük çıktığı görülmüştür.

Kan (2005), yazılı yoklama sınavlarının puanlanmasında puanlama cetveli ve yanıt anahtarı kullanımının ve puanlamalar arasında geçen zamanın, farklı öğretmenler tarafından yapılan puanlamaların güvenilirliğine etkisini incelemiştir. Bu çalışmada, a) farklı öğretmenlerin puanlama cetveli ve yanıt anahtarı kullanarak ve kullanmadan verdiği yazılı yoklama puanlarının karşılaştırılması sonucunda ve b) farklı öğretmenlerin farklı zamanlarda, puanlama cetveli ve yanıt anahtarı kullanarak ve kullanmadan verdiği yazılı yoklama puanlarının karşılaştırılması sonucunda puanların ortalamaları arasında anlamlı bir fark olduğu ve puanlamalar arasında geçen zamanın puanların tutarlılığı üzerinde etkili bir rol oynadığı tespit edilmiştir.

Güler (2008), Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modelinin karşılaştırılmasını yaptığı çalışmada, TIMSS 1999'da yer alan açık uçlu sorulardan 24'ü 203 öğrenciye uygulanmış ve öğrencilerin verdikleri yanıtlar bütünsel puanlama rehberine dayalı olarak puanlanmıştır. Elde edilen puanların güvenilirliği farklı kuramlara göre hesaplanmıştır. Klasik test kuramında Cronbach alfa güvenirlik katsayısı, puanlayıcılar arası uyumun belirlenmesinde Kendall'in uyum katsayısı, puanlayıcılar arası korelasyon katsayısı hesaplanmış ve puanlayıcıların verdikleri puanların ortalamasında bir fark olup olmadığını

belirlemek için F testi yapılmıştır. Genellenebilirlik kuramında ise, birey x görev x puanlayıcı çapraz deseni kullanılarak Genellenebilirlik katsayıları hesaplanmıştır. Çok değişkenli Rasch Modeli ile birey, puanlayıcı ve madde boyutlarına ilişkin güvenilirlik hesaplanmıştır. Klasik test kuramına göre puanların iç tutarlılığı 0,92 olarak hesaplanmıştır. Kendal'in uyum katsayısı 0,52 bulunmuş ancak puanlayıcılar arası korelasyon katsayıları 0,90 - 0,97 aralığında hesaplanmıştır. Puanlayıcılar arasında anlamlı ilişki bulunmuş ancak yapılan F testinde puanların ortalamaları arasında farklılık olduğu belirlenmiştir. Genellenebilirlik kuramına göre Genellenebilirlik katsayısı 0,92 ve Güvenirlik katsayısı 0,90 bulunmuştur. Matematik başarısını ölçmek için kullanılan ölçme araçlarının güvenilir sonuçlar verdiği ve puanlayıcıların birbiriyle tutarlı puanlama yaptığı ifade edilmiştir. Güvenirliğin belirlenmesinde en az iki kuramdan yararlanılmasının daha uygun olacağı sonucuna varılmıştır.

Atılgan (2008), İnönü üniversitesinde Müzik öğrencilerinin seçiminde kullanılan özel yetenek seçme sınavına ilişkin güvenilirlik analizi yapmıştır. Çalışmada, 2003-2004 öğretim yılında sınava giren 249 öğrencinin sınav sonuçları kullanılmıştır. Bu çalışmanın üç temel amacı vardır. Bunlar : 1. Özel yetenek seçme sınavının güvenilirliğini değerlendirmek, 2. Değerlendirme için en uygun görev sayısını belirlemek, 3. Değerlendirme için en uygun puanlayıcı sayısını belirlemek. Çalışmada, öğrenci x görev x puanlayıcı çapraz deseni kullanılmıştır. Özel yetenek seçme sınavı 3 alt bölümden oluşmaktadır. Bu bölümler: Çözümleme, şarkı söyleme ve çalmadır. Herbir alt birimde birçok görev bulunmaktadır. Puanlayıcı sayısı gibi öğrencinin başarısını etkileyecek birçok değişken vardır. Dolayısıyla görev ve puanlayıcı gibi değişkenlik kaynaklarının birden fazla olmasından dolayı güvenilirlik analizi Genellenebilirlik kuramıyla yapılmıştır. Genellenebilirlik çalışması sonrası, testin uzunluğunun değiştirilmesi ve puanlayıcı sayısının azaltılması için Karar çalışması yapılmıştır. Çalışmada, bağıl ve mutlak değerlendirme için G ve Phi katsayıları hesaplanmıştır. Ayrıca puanlayıcı sayısını bir arttırıp bir azaltarak ve görev sayısını 2 arttırıp 2 azaltarak karar çalışması yapılmıştır.

Çözümleme bölümünde, G katsayısına etki açısından, görev sayısındaki artışın, puanlayıcı sayısındaki artıştan daha etkili olduğu görülmüştür. Şarkı söyleme bölümünde, G katsayısı için, puanlayıcı sayısını arttırmanın görev sayısını arttırmaktan daha etkili olduğu bulunmuştur. Ancak Phi katsayısının artışında, şarkı söyleme görev sayısının arttırılması, puanlayıcı sayısının arttırılmasından daha etkilidir. Çalma bölümünde, G ve Phi katsayıları için puanlayıcı sayısının arttırılması, görev sayısının arttırılmasından daha etkili olmuştur. Testin bütününde, G katsayısı için, puanlayıcı sayısının ve görev sayısının arttırılması daha yararlı görülmektedir. Sonuç olarak, hem alt bölümler hem de testin bütününü için, G katsayısını dikkate aldığımızda görev ve puanlayıcı sayısındaki artışın çok etkili olmadığı dolayısıyla 3 puanlayıcı ve daha uzun bir test uygulamasının daha ekonomik olacağı sonucuna varılmıştır.

Nalbantoğlu (2009), Genellenabilirlik kuramına göre performans puanlamada öğrencilerin birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerden elde edilen G ve K çalışmaları sonuçlarının karşılaştırıldığı çalışmada, çalışmanın örneklemini, 2007 – 2008 öğretim yılında Hacettepe Üniversitesi Tıp fakültesi 3. sınıf öğrencilerinden 48 öğrenci oluşturmaktadır. Araştırmada, 3 puanlayıcı aynı iletişim becerileri formuyla, 15 görev doğrultusunda öğrencileri puanlamışlardır. Çalışmada Genellenebilirlik kuramına göre öğrenci x görev x puanlayıcı ve (öğrenci:puanlayıcı) x görev desenleri olmak üzere iki desen kullanılmıştır. Kullanılan iki desende değişkenler için kestirilen varyans değerlerinin birbirleriyle paralellik gösterdiği görülmüştür. Ancak (öğrenci:puanlayıcı) x görev deseninde, G ve Phi katsayılarının daha büyük çıkma eğiliminde olduğu görülmüştür. Bu nedenle, çok sayıda öğrencinin bulunduğu performans sınavlarında puanlayıcılar arası tutarlık sağlandığında öğrencilerin puanlayıcıların hepsi tarafından tek tek puanlanması yerine puanlayıcıların belli sayıdaki öğrencileri dönüşümlü olarak puanlanmasının zaman, iş gücü ve ekonomiklik açısından daha uygun olduğu sonucuna varılmıştır.

Güler (2009), Genellenabilirlik Kuramı ve SPSS ile GENOVA programlarıyla hesaplanan G ve K çalışmalarına ilişkin sonuçların karşılaştırılmasının yapıldığı çalışmada, GENOVA programının anlaşılmasının ve kullanılmasının karmaşık oluşunun Genellenebilirlik kuramıyla yapılan çalışmaların oldukça az olmasını açıklayan sebeplerden biri olarak göstermiştir. Musquash ve O'Conner tarafından geliştirilen Genellenebilirlik kuramının kullanılmasına izin veren SPSS programının oldukça pratik bir alternatif olduğunu ifade etmiştir. İki programa ilişkin değerlerin neredeyse eşit değerler olması nedeniyle, Musquash ve O'Conner tarafından geliştirilen programın kullanılması tavsiye edilmiştir.

Güler (2011), rasgele veriler üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramına göre güvenilirliğin karşılaştırılması çalışmada, 125 öğrenciye 18 satır ve 4 sütun içeren tablolar dağıtılmış ve öğrencilerin bu satırları 0 ve 5 arasında tamamen rasgele doldurmaları istenmiştir. Bu şekilde 125 öğrencinin 18 madde üzerinden 4 puanlayıcı tarafından puanlandığı varsayılan veri elde edilmiştir. Dolayısıyla bu veri rastlantısal olarak oluşturulmuş bir veridir ve güvenilirliği düşüktür. Rastlantısal olarak oluşturulmuş bu veride, Genellenebilirlik Kuramına ve Klasik Test Kuramına göre güvenilirlik hesaplanmıştır. Değişkenlik kaynağının maddeler olduğu tek değişken kaynaklı çapraz desen (b x m) için hesaplanan G katsayısı ile Cronbach α değerleri her bir puanlayıcı için ayrı ayrı hesaplanmış ve çok düşük değerler elde edilmiştir. Değişkenlik kaynağının maddeler ve puanlayıcılar olduğu tümüyle çapraz desen (b x m x p) için Genellenebilirlik kuramına dayalı G katsayısı ve Phi katsayısı sırasıyla 0.46 ve 0.47 olarak hesaplanmıştır. Tamamen gelişigüzel oluşturulan (ölçme hatasının yüksek, güvenilirliğin düşük olmasının beklendiği) ölçme sonuçlarına dayandırılmış ve her iki kurama göre elde edilen değerler, verilerin güvenilir olmadığını teyid ederek, negatif bir durum için de alanyazını destekler sonuçlar vermiştir.

Yelboğa (2012), Genellenebilirlik Kuramına göre iş performansı ölçeklerinde güvenilirlik çalışmada, 2005 yılında bir finans şirketinde çalışan 170 kişinin 3 yönetici tarafından 32 görev üzerinden değerlendirildiği verileri kullanmıştır.

Çalışmada, Yelboğa tarafından geliştirilen iş performans değerlendirme ölçeği kullanılmıştır. Ölçekte 4 alt ölçek bulunmaktadır. Bu alt ölçeklerin ilkinde 9 madde, ikinci ve üçüncüsünde 10'ar madde ve dördüncüsünde 3 madde olmak üzere toplam 32 madde vardır. Puanlamada her madde 5'li likert tipi ölçekler 1=yetersiz, 5= mükemmel olmak üzere değerlendirilmiştir. Genellenebilirlik analizinde, kişi x görev x puanlayıcı çapraz deseninde G katsayısı 0,89 ve Phi katsayısı 0,87 olarak hesaplanmıştır. Yapılan karar çalışmasında puanlayıcı sayısındaki artışın G ve Phi katsayılarında artışa sebep olduğu, 4 puanlayıcı sonrası puanlayıcı sayısını artırmanın bir fark yaratmadığı görülmüştür. Genellenebilirlik kuramının, iş performansı ölçeğinde çoklu hata kaynaklarını ele almada en uygun yöntem olduğu sonucuna varılmıştır.

Yurtiçinde yapılan çalışmalar incelendiğinde, çalışmaların büyük bir çoğunluğunda farklı desenlerin karşılaştırıldığı, farklı senaryolara dayanan K çalışmaları yapıldığı ve bu çalışmalarda en uygun madde sayısı ve en uygun puanlayıcı sayısına odaklandığı görülmektedir. Ayrıca Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modelinin karşılaştırıldığı çalışmalar mevcuttur. Yapılan çalışmalarda, Genellenebilirlik kuramının, birden fazla değişkenlik kaynağının olduğu durumlarda değişkenlik kaynaklarını birlikte ele alıyor olmasından ve Karar çalışması yapılmasına olanak sağlamasından dolayı güvenilirlik hesaplamalarında tercih edilen bir yaklaşım olduğu görülmektedir.

1.7.2. Yurtdışında Yapılan Çalışmalar

Smith (1997), TIMSS açık uçlu sorularının puanlanmasının Genellenebilirliği çalışmasında, madde türünün bir fonksiyonu olarak, puanlayıcı etkisinin açık uçlu sorularda ülke ve öğrenci düzeyinde, hata varyans kaynaklarına katkılarını araştırmıştır. İngilizce konuşulan 7 ülkenin herbirinden 50 kitapçık ile her madde için 350 öğrencinin yanıtlarını incelemiştir. Genellenebilirlik katsayısı, TIMSS açık uçlu sorularında, ülkeler arasında güvenilirliğin yüksek düzeyde olduğunu göstermektedir. Bir bireye ait belirli bir maddenin puanlanmasıyla ilgili Genellenebilirliğe bakıldığında, bazı maddeler için Genellenebilirliğin

değişkenlik gösterdiği görülmüştür. Ancak TIMSS ülke bazında ortalamalara dayandığı için, bu durumun sorun yaratmayacağı ifade edilmiştir. TIMSS'e katılan 21 ülkeden 39 puanlayıcı bu çalışmada görev almıştır. 14 Matematik ve 17 Fen maddesi olmak üzere toplam 31 madde puanlanmıştır. Bu maddeler TIMSS'de yer alan toplam Matematik ve Fen açık uçlu sorularının yarısından fazlasını oluşturmaktadır. Genellenebilirlik deseni, öğrencilerin ülkede yuvalandığı ((öğrenci:ülke) x puanlayıcı) desendir. Araştırmada örneklem büyüklüğü ve puanlayıcı sayısının etkisi araştırılmıştır. Analizler sonucunda ülke düzeyinde ortalama puanlar ve öğrenci düzeyindeki puanların her ikisi için Genellenebilirlik katsayıları hesaplanmıştır. Ülke düzeyi ortalamaların Genellenebilirliği oldukça yüksek çıkmıştır. Sadece 3 fen ve 1 matematik sorusunda Genellenebilirlik katsayısı 0.7'nin altında yer almıştır. Öğrenci sayısı ve puanlayıcı sayısı artırıldığında hata varyansı düşmekte dolayısıyla G katsayısı artmaktadır. Fakat 500-1000 öğrencinin bulunduğu uygulamalarda, Genellenebilirlik katsayısı, öğrenci sayısının artırılıp azaltılmasındansa, puanlayıcı sayısından daha çok etkilenmektedir. Öğrenci sayısı 200'den 1000'e çıkarıldığında Genellenebilirlik katsayısı 0,75'ten 0,84'e çıkmaktadır. Puanlayıcı sayısındaki artışın etkisi maddeden maddeye farklılık göstermektedir. Puanlayıcı sayısını 5'ten 15'e yükseltmek Genellenebilirlik katsayısını bütün maddeler için 0,7'nin üzerine çıkarmıştır. Bu durum puanlayıcı sayısının az olduğu ülkelerde bir sorun olabileceğini göstermektedir.

Mcbee ve Barnes (1998), 8. Sınıf Matematik performansının değerlendirilmesi çalışmasında, performansın zaman içerisinde değişimini ve görev benzerliklerinin değerlendirme sonuçlarının Genellenebilirliğe etkisini araştırmıştır. Bu çalışmada, 8. Sınıf öğrencilerine, dört performans değerlendirme görevi uygulanmıştır. Görevler program hedeflerine uygun olarak hazırlanmıştır. Bu görevlerden karmaşık problem çözme alanını ölçmek için geliştirilen ikisi oldukça benzerdir. İlk uygulama Nisan ayının ilk haftası, ikinci uygulama ilk uygulamadan 3 hafta sonra gerçekleştirilmiştir. Bütünsel Puanlama anahtarı kullanılmıştır. Puanlama sırasında puanlayıcılar performansın hangi uygulamaya (birinci ya da ikinci) ait olduğunu bilmeden puanlama yapmışlardır.

Genellenebilirlik deseni öğrenci x görev x puanlayıcı x uygulama çapraz desendir. Bütün değişkenlik kaynakları rastlantısal olarak alınmıştır. Benzer görevler ve 2 puanlayıcı ile Genellenebilirlik katsayısı 0,66 olarak hesaplanmıştır. 0.80 düzeyinde Genellenebilirlik için 20 göreve ihtiyaç vardır. Daha genel 4 görev için Genellenebilirlik katsayısı 0.39'da kalmaktadır. 32 görev ve 2 puanlayıcı kullanıldığında 0,72'ye ulaşmaktadır. Bu çalışmada görev seçiminin temel hata kaynağı olduğu görülmüştür. Bütün görevler üzerinden değerlendirme yapıldığında görevler arası tutarlık çok düşüktür ve mantıklı sayıda görevle bile kabul edilebilir düzeyde bir Genellenebilirliğe ulaşmamaktadır. Test birbirine benzer görevlerden oluşturulduğunda Genellenebilirlik düzeyi artmaktadır ancak bu da yapının temsilinde problem yaratmaktadır. Araştırma, kabul edilebilir düzeyde Genellenebilirlik için, sadece benzer görevler kullanılmış olsa bile, görev sayısının oldukça yüksek olması gerektiğini ortaya çıkarmıştır. Performans değerlendirme ile eğitim arasında bağ kurmanın, daha üst düzey düşünme becerilerinin karmaşıklığı nedeniyle, çoktan seçmeli maddeler ve eğitim arasında bağ kurmaktan daha zor olduğu sonucuna ulaşılmıştır.

Lee (1999), daha önce yapılan çalışmalarda madde bazlı güvenilirlik yöntemleri kullanılarak yapılan hesaplamalarda, alt testlerden oluşan test puanlarının güvenilirliğinin gerçekte olduğundan daha yüksek bulunduğu ortaya konulduğunu ifade etmiştir. Bu çalışmanın amacı alt testlerden oluşan bir testten elde edilen puanların güvenilirliğini hesaplamada Genellenebilirlik kuramını kullanmanın uygun olup olmayacağını araştırmaktır. G kuramında alttestlerin ve alttestlerdeki madde sayısının değiştirilmesinin etkisi karar çalışmaları yapılarak araştırılabilir. Çalışmada öğrencilerin çapraz, maddelerin pasajda yuvalandığı öğrenci x (madde : pasaj) deseni kullanılmıştır. Genellenebilirlik katsayısının pasaj sayısını artırmaktan, pasajdaki madde sayısını artırmaya göre, daha büyük oranda etkilendiği görülmüştür. Cronbach alpha ve Genellenebilirlik katsayısı arasındaki fark pasajda yer alan madde sayısını artırmakla artmaktadır. Bu nedenle pasaj sayısını arttırmanın daha etkili olduğu düşünülmüştür. Ancak gerçek durumda, pasaj sayısını arttırırken sınav

süresinde meydana gelecek değişiklik dikkate alınarak karar verilmesi önerilmiştir.

Tobar, Stegner & Kane (1999), ruh hali profili (POMS)'nin Genellenebilirliğini ele aldıkları çalışmada, madde ve alt ölçek sayılarının Genellenebilirlik katsayısına etkilerini incelemişlerdir. Ruh hali profili, 6 alt ölçekten ve 65 sorudan oluşmaktadır. Alt ölçeklerde madde sayısı 7-15 arasında değişmektedir ve 0 (hiç) – 4 (aşırı) arasında puanlanmaktadır. Bu çalışmaya 40'ı kadın 40'ı erkek olmak üzere 80 üniversite öğrencisi katılmıştır. Hesaplamalar 58 madde üzerinden yapılmıştır. Alt ölçeklerdeki madde sayısının 7-15 arasında değişmesi G kuramına göre dengesiz (unbalanced) bir desen oluşturacağı için, her bir altölçek 7 maddeye düşürülmüş, Genellenebilirlik analizleri 42 madde üzerinden yapılmıştır. Dengeli (balanced) ve dengesiz (unbalanced) desenin farkını görmek için hesaplamalar iki desen içinde gerçekleştirilmiş ve bu durumun varyans bileşenlerine etkisi olmadığı görülmüştür. İlk desen öğrenci x madde desenedir. Alt ölçeklerin birinde (confusion) Genellenebilirlik Katsayısı, 7 madde için 0.74, 8 madde için 0.76 ve 10 madde için 0.80 olarak hesaplanmıştır. Öğrenci x (madde:altölçek) deseni, alt ölçek rastlantısal kabul edildiğinde, Genellenebilirlik düzeyi 6 altölçek ve herbiri için 7 maddede 0.90, 12 altölçek ve herbiri için 7 maddede 0,94 ve 6 altölçek ve herbiri için 14 maddede 0.91 olmaktadır. Oysa alt ölçekler sabit kabul edildiğinde Genellenebilirlik katsayısı 0.96 olarak hesaplanmaktadır. Altölçeklerdeki madde sayısı 7'ye indirilebilir, ya da bu altölçeklerin ötesine genelleme yapma niyeti yoksa, alt ölçekler sabit alınarak 0,96 düzeyinde Genellenebilirlik elde edilebilir sonucuna ulaşmışlardır.

Goodwin (2001), puanlayıcılar arası uyum ve güvenilirlik çalışmasında kullanılan 3 yaklaşımı karşılaştırmıştır: a) uyum yüzdesi ve kappa b) basit korelasyon yöntemleri c) Genellenebilirlik (G) kuramı teknikleri. Çalışmada hipotetik veriler kullanılmıştır ve veri sayısı 10 alınarak çok küçük sayıda tutulmuştur. Çalışmada 10 öğrenci, 2 puanlayıcı tarafından, 6 farklı günde fiziksel aktivitelerinin kalitesi bakımından puanlanmıştır. Her puanlayıcı 7 puan üzerinden (1-en düşük, 7-en yüksek) değerlendirme yapmıştır. Uyum yüzdesine

2 şekilde bakılmıştır: 1. Tam uyum, 2. ± 1 puanın kabul edildiği durum. Tam uyumda uyum yüzdesi %20 iken, ± 1 puanın kabul edildiği durumda %80 olmaktadır. Cohen Kappa uyum katsayısı öğrencilerin her bir gün için hesaplanan iki puanı dikkate alındığında 0,51, toplam puanlar dikkate alındığında 0.90 olarak hesaplanmıştır. G kuramına göre hesaplama yapıldığında puanlayıcı x gün x öğrenci çapraz deseninde 2 puanlayıcı 6 gün için G katsayısı 0.88 olarak hesaplanmıştır. Yapılan karar çalışmasında, 1 puanlayıcı 3 gözlem günü için G katsayısı 0,68 olarak hesaplanmıştır. Farklı yaklaşımların avantajları ve dezavantajları karşılaştırması sonucunda, Genellenebilirlik kuramı tekniklerinin en kapsamlı, en esnek ve bir çalışmada farklı kaynakların neden olduğu ölçme hatalarının izole edilmesine imkan tanıyan teknikler olduğu vurgulanmıştır. G kuramının puanların genellenebilirliği ya da güvenilirliği konusunda en çok bilgi veren yaklaşım olduğu ifade edilmiştir.

Sharma & Weathers (2003), uluslararası araştırma projelerinde kullanılan ölçeklerin genellenebilirliğini araştırdıkları çalışmalarında, yapının araştırmaya katılan bütün ülkelere genellenebilir olmasının bu tür araştırmalar için en önemli nitelik olduğunu ifade etmişlerdir. Önemli bir diğer özellik ise, ölçeğin bütün ülkelere genellenebilmesi için kaç maddeye ve kaç kişiye ihtiyaç duyulduğu şeklinde ifade edilmiştir. Araştırmada doğrulayıcı faktör analizi ve Genellenebilirlik kuramı bir arada kullanılmıştır. Ülkeler arasında ölçmenin değişmezliğine kanıt sağlamak için, ölçmenin değişmezliği ve ölçeğin ülkelere genellenebilmesi için kaç kişi ve kaç maddeye ihtiyaç duyulduğunu hesaplamada iki tekniğin tamamlayıcı teknikler olarak ele alınması önerilmiştir.

Örnekleme için Amerika'dan 71, Fransa'dan 70, Japonya'dan 76 ve Almanya'dan 73 kişi seçilmiş ve 17 madde kullanılmıştır. Genellenebilirlik deseni kişilerin ülkede yuvalandığı (Kişi:Ülke) x Madde desenidir. Hesaplamalar bu örnekleme göre yapılmış, daha sonra eşit örnekleme büyüklüğü varsayımı ihlal edilerek rastlantısal olarak seçilen 40, 50, 60 ya da 70 kişi ülkeler arası dengesiz (unbalanced) deseni oluşturmuştur. Varyans bileşenleri PROC varcomp SAS ile hesaplanmıştır. Yapılan faktör analizinde ilk faktör tarafından açıklanan varyans

yüzdesi farklı örneklem için %43.2 - %57.3 aralığında değişmektedir. İkinci faktör tarafından açıklanan varyans yüzdesi %6.6 - %9.7 aralığındadır ve ilk eigen değerinden sonra önemli bir düşüş söz konusudur. Bu nedenle tek faktörlü model kabul edilmiştir. 17 madde de, birleştirilmiş ve ayrı ayrı ele alınmış örneklem için 1. Faktörde kabul edilebilir düzeyde yüklenmiştir. Faktör yapısının ülkeler arasında aynı olduğu görülmüştür. G kuramına göre, ülkeler rastlantısal faktörler olarak ele alınmıştır. Ülkeler %7.01 ile oldukça küçük varyans değerine sahiptir ve bu durumda veri toplanan ülkeler arasında büyük fark olmadığı ifade edilmiştir. Varyansın %35'i ülkeler içindeki kişilerden kaynaklanmaktadır. Yani maddelere verilen yanıtlar kişiler arasında farklılaşmaktadır. Madde x ülke etkileşimi büyük (%5.51) değildir, ortalamalar farklı olsa da yanıt desenleri ülkeler arasında benzerlik göstermektedir, yani ölçek belirli bir ülkeye özel değildir ve bütün ülkelerde aynı anlamı taşımaktadır.

Ülkelerin sabit ya da rastlantısal olarak alınmasının bir fark yaratmadığı görülmüştür. Eğer istenilen Genellenebilirlik düzeyi 0.90 ise 11 maddenin bu düzey için yeterli olduğu ($G=0.905$), 17 maddenin 11'i rastlantısal olarak seçildiğinde, istenilen düzeyde Genellenebilirliğe sahip daha kısa bir ölçeğin hazırlanabileceği ifade edilmiştir. Eğer ülkeler sabit olarak alınırsa, Amerika için 8 madde, Fransa için 4 madde, Japonya için 15 madde ve Almanya için 11 maddeye ihtiyaç vardır.

Madde sayısının belirlenmesinden sonra verilmesi gereken ikinci karar gerekli örneklem büyüklüğünü belirlemektir. Ülkeler rastlantısal olarak alındığında 140 kişi 0,90 Genellenebilirlik katsayısı için yeterlidir. Belirlenen örneklem bundan daha büyük bir örnektir. Ülkeler sabit olarak alındığında, 15 madde kullanımında 0,90'dan daha büyük bir Genellenebilirlik katsayısı için her ülkeden 100'er kişi olmalıdır sonucuna ulaşılmıştır. Gerekli minimum madde sayısını ve kişi sayısını bilmek, uluslararası sınavlarda harcanan emeğin ve maliyetin azaltılmasına yardımcı olur.

Schoonen (2005), yazma puanlarının genellenebilirliği: bir yapısal eşitlik modeli uygulaması adlı çalışmasında, kompozisyon yazma sınavlarında sonucu etkileyen değişkenlik kaynaklarını ele almıştır. Hollanda'da gerçekleştirilen bu çalışmada, 22 okul ve 442 kişilik bir gruptan 89 öğrenci rastlantısal olarak seçilmiştir. 6. sınıf öğrencisi 89 kişi, dörder makale yazmış ve herbiri 5 puanlayıcı tarafından "içerik ve düzen" ve "dilin kullanımı" boyutlarında iki ayrı puanlama metodu (bütünsel ve analitik) kullanılarak puanlanmıştır. Yazarın yazma yeteneği dışında, görevin konusu, puanlanan nitelikler ("içerik ve düzen" ve "dilin kullanımı"), bu niteliklerin puanlanma şekli (bütünsel ya da analitik) yazarın puanında etkili olabilmektedir. Bu çalışmada bu değişkenlik kaynakları varyans analizi tekniği kullanılarak Genellenebilirlik çalışması yapılmıştır. Yapısal Eşitlik Modeli yazma puanlarında varyans bileşenlerini hesaplamak için kullanılmıştır. 4 görev x 2 nitelik ("içerik ve düzen" ve "dilin kullanımı") x 2 puanlama süreci (bütünsel ve analitik) x 5 puanlayıcı deseni oluşturulmuştur. Burada 4 değişkenlik kaynağından söz edilmektedir. Sonuç olarak, yazma puanlarının yazarın yeteneğinden çok, değişkenlik kaynaklarından etkilendiği bulunmuştur. Görev ve puanlayıcıların puanlardaki değişimde çok etkili olduğu ortaya çıkmıştır. Bütünsel puanlama, analitik puanlamadan daha genellenebilir bulunmuştur. Dilin kullanımına ilişkin puanlar, içerik ve düzen puanlarından daha genellenebilir bulunmuştur. Bu bulgular ışığında, puanların genellenebilirliğine bakarken görev ve puanlayıcı sayısı yanında puanlama işlemlerinin ve puanlanacak niteliklerin de dikkate alınması gerektiği ifade edilmiştir.

Lee (2005), yaptığı çalışmada TOEFL testinde görev ve puanlayıcı sayısındaki değişimin Genellenebilirliğe etkisini araştırmış ve maksimum güvenilirlik için en uygun görev sayısını belirlemeye çalışmıştır. TOEFL 2000, konuşma çerçevesinde 3 temel konuşma görevi yer almaktadır: bağımsız konuşma, birleştirilmiş konuşma görevleri (dinleme – konuşma ve okuma – konuşma). Bağımsız konuşmada kişinin, kişisel deneyimleri ya da genel bilgisini kullanarak konuşma görevlerini yanıtlaması gerekmektedir. Birleştirilmiş görevlerde kişinin önce akademik bir konuşmayı ya da metni anlaması ve sonrasında sözlü

yanıtlar vermesi gerekmektedir. 13 görevin bir kısmına ya da bütününe yanıt veren 478 öğrenciden, 11 görevin tamamına yanıt veren 261 kişi örnekleme oluşturmaktadır. Ortalama yaş 24'tür. Bağımsız ve birleştirilmiş görevler için 2 puanlama rehberi oluşturulmuştur. Her öğrenci 2 puanlayıcı tarafından 1-5 puan aralığında puanlanmıştır. Puanlayıcılar bütün görevleri puanlamıştır. Ancak puanlayıcılar öğrencilerle yuvalanmıştır. Puanlayıcılar önce bir maddeye ait yanıtların tümünü tamamlamış sonra bir sonraki maddeye geçmiştir. (puanlayıcı:öğrenci) x görev ve öğrenci x görev x puanlayıcı ve öğrenci x görev desenleri kullanılmıştır. İlk iki desen için görev sayıları 1-12 arasında değiştirilerek karar çalışması yapılmıştır. Görev sayısını artırmanın puanlayıcı sayısını arttırmaktan daha etkili olduğu görülmüştür. Bu nedenle performans değerlendirmede daha az puanlayıcı kullanmanın kabul edilebilir bir genellenebilirlik düzeyi için uygun olduğu sonucuna varılmıştır.

Yurtdışında yapılan çalışmalar incelendiğinde, uluslararası araştırma projelerinde kullanılan ölçeklerin Genellenebilirliğinin araştırıldığı çalışmalar dikkat çekmektedir. Bu çalışmalarda farklı senaryolara ilişkin K çalışmaları yapılmış, farklı desenler karşılaştırılmış, en uygun madde sayısı, en uygun puanlayıcı sayısı ve uygun örneklem büyüklüğüne ilişkin çalışmalara yer verilmiştir. Genellenebilirlik kuramının birden fazla değişkenlik kaynağının olduğu durumlarda değişkenlik kaynaklarını birlikte ele alıyor olmasından ve Karar çalışması yapılmasına olanak sağlamasından dolayı güvenilirlik hesaplamalarında tercih edilen bir yaklaşım olduğu ve karar çalışmalarının gelecekte yapılacak uygulamalarda değişkenliği düzeltmeye yönelik olduğu görülmektedir.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, evren ve örneklem, araştırma verilerinin toplanma şekli ve verilerin analizine yer verilmiştir.

2.1. ARAŞTIRMANIN TÜRÜ

Araştırma Genellenabilirlik kuramı ile PISA 2009 okuma becerileri puanlamasına ait özelliklerin belirlenmesi yönüyle durum belirleme çalışmasıdır ve betimsel bir araştırma niteliği taşımaktadır. Betimsel araştırma, var olan bir olayı, birey ya da grubun özelliklerini var olduğu şekilde tanımlayan, var olan durumu nicel ya da nitel yönden betimleyen bir araştırma türüdür (Karasar, 1998).

2.2. EVREN VE ÖRNEKLEM

Ülke sonuçlarının karşılaştırılabilmesi için PISA'da, karşılaştırılabilir bir hedef kitlenin değerlendirilmesine önem verilmektedir. Uygulamaların yapıldığı ülkelerde, okul öncesi eğitimlerin yapı ve kapsamı, okula başlama yaşı ve eğitim sitemlerinin yapısına ilişkin farklılıklar, öğrencilerin sınıf düzeylerinin ülkeler arası karşılaştırma yapabilecek şekilde tanımlanmasına izin vermemektedir. Eğitimdeki performansla ilgili yapılan uluslararası karşılaştırmaların geçerli olabilmesi için, örneklem, belirli bir yaş grubuna göre belirlenmektedir. PISA öğrenci evreni, okul türüne bakılmaksızın okullarda öğrenim gören, değerlendirmenin yapılacağı tarih itibariyle yaşları 15 yıl 3 ay ve 16 yıl 2 ay arasında değişen, en az altı yıllık örgün eğitimi tamamlamış öğrencilerden oluşmaktadır. 15 yaş grubunun kullanılması, öğrenci başarılarının bütün ülkelerdeki zorunlu eğitimin bitiminde veya zorunlu eğitim tamamlanmadan önce karşılaştırılabilmesini sağlamaktadır (EARGED, 2010)

PISA uygulamasında öncelikle, ulusal merkezler tarafından tabakalı örneklem yöntemine göre örneklemin çatısı oluşturulur. Bu çatıda, öğrencilerin ve okulların hangi kriterlere göre örnekleme dâhil edileceği (bölge, program türü, okul türü vb.) belirlenir. Oluşturulan örneklem dosyasında, belirlenen tabakalara göre okul ve öğrenci sayıları yer almaktadır. Daha sonra ulusal merkezler bu dosyayı uluslararası merkeze gönderirler. Belirlenen tabakalara ve öğrencilerin temsil edilme oranlarına göre okullar belirlenir ve ülkelere gönderilir. Ülkeler belirlenen okullardaki 7. sınıf ve üzeri sınıflarda okuyan tüm 15 yaş grubu öğrencilerin arasından seçkisiz yöntemle 35 öğrenciyi PISA'ya katılmak üzere seçer. 35'ten daha az sayıda öğrencisi olan okullardaki tüm 15 yaş grubu öğrenciler seçilir. Bu liste kontrol edilmek üzere okullara gönderilir ve okullardan onay alındıktan sonra PISA uygulamasına katılacak okul ve öğrenciler kesinleşmiş olur (EARGED, 2010).

PISA 2009 uygulamasının Türkiye örneklemini, 12 İstatistik Bölge Birimleri Sınıflandırmasına (İBBS, NUTS) ve okul türlerine göre tabakalandırılarak PISA uluslararası merkez tarafından seçkisiz yöntemle belirlenen toplam 170 okuldan 4996 öğrenci oluşturmaktadır. Bu çalışmada, PISA 2009 uygulamasında okuma becerileri alanında yer alan soruları yanıtlayan ve yanıtladıkları kitapçıklar çoklu puanlama için seçilen 886, öğrenci alt örneklemini oluşturmaktadır.

Puanlayıcı evrenini, daha önce uluslar arası projelerde puanlayıcı olarak yer alan ya da 15 yaş öğrencilerine (7. sınıftan itibaren 15 yaş öğrencilerinin bulunduğu seviyede) öğretmenlik yapmakta olan Türkçe ya da Edebiyat öğretmenleri oluşturmaktadır. Puanlama için, bu evrenden Ankara'da görev yapan 16 Türkçe/Edebiyat öğretmeni puanlayıcı olarak seçilmiştir. Bu puanlayıcılar 4'erli gruplar halinde puanlama yapmıştır.

Soru evrenini PISA sınavlarında sorulabilecek sorular oluşturmaktadır. Bu çalışmada PISA 2009 uygulamasında yer alan 13 kitapçıktan, okuma becerileri sorularının bulunduğu ve birden fazla puanlayıcı tarafından puanlama yapılan 7 kitapçıkta yer alan 44 madde test edilmiştir. Kitapçık 2'de 6 soru, Kitapçık 4'te 5

soru, Kitapçık 6'da 7 soru, Kitapçık 7'de 7 soru, Kitapçık 8'de 8 soru ve Kitapçık 12'de 6 soru yer almaktadır. 2, 8 ve 12 nolu kitapçıklarda yer alan soruların puanlanmasında tam puan 2, kısmi puan 1 ve yanlış/ilgisiz yanıt 0 olarak değerlendirilmektedir. 4, 5, 6, 7. kitapçıklarda puanlama tam puan 1, yanlış/ilgisiz yanıt 0 olarak değerlendirilmektedir.

2.3. ARAŞTIRMA VERİLERİ

PISA uygulamasında 13 kitapçık kullanılmaktadır ve uygulamaya katılan okulların herbirinden 15 yaş grubunda yer alan 35 öğrenci seçilmekte ve bu öğrencilere kitapçıklar rastlantısal olarak dağıtılmaktadır. Uygulamada ilk öğrencinin hangi kitapçığı alacağı rastlantısal olarak belirlenmekte ve listede yer alan öğrencilerin alacağı kitapçıklar bu sırayı takip edecek şekilde düzenlenmektedir. Örneğin listedeki ilk öğrencinin sınav kitapçığının 9 olduğunu varsayalım. Bir sonraki öğrenci 10. kitapçığı alacaktır ve devamında 11,12,13,1,2... şeklinde sıralama devam etmektedir.

PISA 2009 test kitapçıklarında her bir sorunun kod numarası, sorunun sağ köşesinde yer almaktadır. Örneğin R412S1, 412 kodlu okuma (Reading) ünitesinin 1. sorusu anlamına gelmektedir. Tek puanlayıcı deseninin kullanıldığı kitapçıklarda, puanlar, soru kod numarasının hemen yanında yer alan puan kodlarının daire içine alınması ile kitabın üzerinde verilir. Birden fazla puanlayıcının (dört puanlayıcı) yer aldığı desende ise ilk üç puanlama, hazırlanmış puan formları üzerinde verilirken, son puanlayıcı kitap üzerinde puanlama yapmakta ve böylece puanlayıcıların diğer puanlayıcıların verdiği puanı görmemeleri ve bilmemeleri sağlanmaktadır. Bütün kitapçıklar bu desene göre 4 puanlayıcı tarafından puanlanmış, puanlama işlemi için toplam 16 puanlayıcı görev yapmıştır. Puanlama yapılırken bütün kitapçıklarda aynı soruya ait yanıtların tamamı puanlanmadan, bir sonraki soruya geçilmemiştir.

Puanlama işlemine başlamadan önce, örnek sorular üzerinde cevap anahtarlarını kullanarak bir eğitim yapılmış ve puanlama işlemine bu eğitimden

sonra geçilmiştir. Bu puanlama işleminde elde edilen veriler çalışmamızın temel verilerini oluşturmaktadır. Veriler EARGED tarafından sağlanmıştır.

2.4 VERİLERİN ANALİZİ

Verilerin analizi alt problemlere göre Genellenebilirlik kuramıyla belirlenen desenler doğrultusunda yapılmıştır. Genellenebilirlik kuramı ile desenlere ait varyans bileşenlerinin kestirilmesi, değişkenlerin toplam varyansı açıklama oranlarının hesaplanması, her bir desen için karar çalışmalarını yapılmasında EduG 6 programı kullanılmıştır. EduG 6 Programı, Genellenebilirlik kuramı analizleri için geliştirilmiş olup, araştırmacının tanımlamış olduğu değişkenlik kaynakları ve bu değişkenlik kaynaklarıyla oluşturduğu desen için G ve K çalışması yapılmasına olanak sağlamaktadır.

PISA okuma becerileri verilerine ilişkin olarak, araştırmacı tarafından iki farklı senaryo üzerinden Genellenebilirlik kuramında kullanılmak üzere iki desen tasarlanmıştır. Bu desenlerden ilki, öğrenci (ö), soru (s) ve puanlayıcı (p) değişkenleri olmak üzere, öğrencilerin aynı beceriler konusunda puanlayıcıların her biri tarafından puanlandığı $\text{ö} \times \text{s} \times \text{p}$ çapraz desendir. İkinci desen ise, her bir puanlayıcının öğrencilerden sadece bir kısmını puanlamasıyla oluşan, öğrenci ve puanlayıcı değişkenlerinin yuvalanmış olduğu, soruların ise bu değişkenlerle çaprazlanmış olduğu $(\text{ö:p}) \times \text{s}$ desendir.

Çapraz desende, 2, 4, 5, 6, 7, 8 ve 12 nolu kitapçıkların herbirinden 100'er kitapçığa ait veriler kullanılmış ve her bir kitapçık için Genellenebilirlik ve Karar çalışmaları yapılmıştır. Genellenebilirlik ve Karar çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri, G ve Phi katsayıları her bir kitapçık için ayrı ayrı verilmiştir.

Yuvalanmış desende, 2 ve 8 nolu kitapçıklar için, PISA 2009'dan 100'er kitapçık ve PISA 2006'dan 93'er kitapçığa ait veriler kullanılmış ve her bir kitapçık için Genellenebilirlik ve Karar Çalışması yapılmıştır. Genellenebilirlik ve

Karar çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri, G ve Phi katsayıları herbir kitapçık için ayrı ayrı verilmiştir.

Karar çalışmalarında, 4 olan puanlayıcı sayısı artırıp azaltılarak 2-6 puanlayıcıya ilişkin Genellenebilirlik ve Phi katsayılarının değişimine bakılmıştır. Öğrenci sayısının etkisini araştırmak için her iki desende de öğrenci sayıları aşamalı olarak 10'ar 10'ar düşürülerek yarıya indirilmiş ve Genellenebilirlik ve Phi katsayılarının değişimine bakılmıştır. Öğrenci sayısının yarıya indirildiği durumda elde edilen değerlere tablolarda yer verilmiştir. Her iki desene ilişkin olarak karşılaştırma yaparken aynı sayıda öğrencinin kullanıldığı durumlar karşılaştırılmıştır.

En son aşamada, soru sayıları aşamalı olarak azaltılmış ve kabul edilebilir bir Genellenebilirlik düzeyi elde etmek için gerekli minimum soru sayısı belirlenmiştir.

BÖLÜM III

BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın alt problemlerinin her biri için istatistiksel çözümlenmelerin sonucunda elde edilen bulgulara ve bulgulara ilişkin yorumlara yer verilmiştir.

3.1. ÖĞRENCİ (Ö), SORU (S) VE PUANLAYICI (P) DEĞİŞKENLERİNİN ÇAPRAZ TASARLANDIĞI ÖXSXP DESENİNİN GENELLENEBİLİRLİK ÇALIŞMASI SONUÇLARI

Bu bölümde, Ö X S X P deseninin Genellenebilirlik çalışması sonuçları ve bu desende puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla gerçekleştirilen Karar çalışması sonuçları iki alt başlık halinde verilmiştir.

3.1.1. Ö X S X P Desende, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Okuma becerilerinin yer aldığı ve birden fazla puanlayıcı tarafından puanlanan kitapçıklar 2, 4, 5, 6, 7, 8 ve 12 nolu kitapçıklardır. Bu kitapçıkların herbirinden 100'er adet kullanılmıştır. Bu desende yer alan bütün değişkenler (öğrenci, soru ve puanlayıcı) çaprazlanmıştır. Ö x S X P çaprazlanmış desen ile Genellenebilirlik çalışması yapılmıştır. Genellenebilirlik çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri her bir kitapçık için ayrı ayrı verilmiştir.

Kitapçık 2 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 2'de verilmiştir.

Tablo 2. Kitapçık 2’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	267.59333	99	2.70296	0.10577	16.0
Soru	137.14333	5	27.42867	-0.00095*	0.0
Puanlayıcı	39.53333	3	13.17778	-0.02470*	0.0
OS	60.85667	495	0.12294	-0.03699*	0.0
OP	92.80000	297	0.31246	0.00692	1.0
SP	419.35667	15	27.95711	0.27686	41.9
OSP, e	402.31000	1485	0.27092	0.27092	41.0
Toplam	1419.59333	2399			100%

*-Varyans bileşenleri 0 olarak kabul edilmiştir (Brennan, 2011)
o: öğrenci, s: soru, p: puanlayıcı

Tablo 2’de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %16’sını açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranının %16 olması, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Tabloda yer alan varyans bileşenlerinden bir kısmı negatif değer almıştır. Negatif varyans kavramsal olarak mümkün olmadığından Negatif değer alan varyans bileşenleri 0 kabul edilmektedir (Brennan 2001). Buna ilişkin iki yaklaşımdan birini kullanmak uygun olur. Bu yaklaşımlardan ilki negatif değeri 0 kabul ederek, negatif varyans bileşeni diğer varyans bileşenleri için beklenen kareler ortalamaları denkleminde yer alıyorsa, negatif değer yerine 0 değeri kullanılarak hesaplama yapılmasıdır (Cronbach, 1972). Diğer yaklaşım ise bu değeri 0 kabul etmek ancak denkleminde negatif değeri kullanmaktır (Brennan, 1983). Bu araştırmada, ilk yaklaşım dikkate alınmıştır. Varyans bileşenlerinin negatif değer alması örneklem hatasından ya da modelin yanlış belirlenmesinden kaynaklanmaktadır (Shavelson & Webb, 1991). Calkins (1978) ve Leone & Nelson (1966) yaptıkları çalışmalarda değişkenlik

kaynaklarının düzeyi az sayıda ise (örn. beş) varyans bileşenlerinin negatif çıkabileceğini bulmuşlardır.

Soru değişkeni ve puanlayıcı değişkeni için kestirilen varyans bileşenlerinin, toplam varyansı açıklama yüzdeleri %0'dır. Bu durumda, bu kitapçıkta yer alan 6 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediği, puanlayıcılar arasında farklılık bulunmadığı, birbirleri yerine kullanılabileceği ve puanlayıcıların birbirleriyle tutarlı puanlama yaptıkları söylenebilir.

Öğrenci ve Soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Öğrenci ve Soru etkileşimi öğrencilerin her bir soru için performanslarında meydana gelen değişikliğin göstergesidir. Bu durumda öğrenci performansının sorudan soruya değişmediğini söylemek mümkündür. Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesine bakıldığında, %1'lik bu değer çok küçük bir değer olduğu, dolayısıyla öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine değişiklik göstermediği söylenebilir.

Soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %41,9 ile oldukça yüksek bir değerdir. Bu durum bazı sorulara verilen puanların bir puanlayıcıdan diğerine değişiklik gösterdiği şeklinde yorumlanabilir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %41'dir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 4 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 3'te verilmiştir.

Tablo 3. Kitapçık 4'te, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	172.16950	99	1.73909	0.09485	27.7
Soru	3.48700	4	0.87175	-0.00738	0.0
Puanlayıcı	0.02150	3	0.00717	-0.00755	0.0
OS	18.01300	396	0.04549	-0.04112	0.0
OP	1.92850	297	0.00649	-0.04070	0.0
SP	47.84100	12	3.98675	0.03777	11.0
OSP, e	249.45900	1188	0.20998	0.20998	61.3
Toplam	492.91950	1999			100%

Tablo 3'te verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %27'sini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı öğrencilerin okuma becerileri bakımından farklılık gösterip göstermediğinin bir göstergesidir. Öğrencilerin okuma becerileri açısından farklılaştığı söylenebilir.

Soru değişkeni için kestirilen varyans bileşeninin ve puanlayıcı değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama yüzdeleri %0'dır. Buradan 4. Kitapçıkta yer alan 5 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediği, puanlayıcılar arasında farklılık bulunmadığı ve puanlayıcıların birbirleriyle tutarlı puanlama yaptıkları anlaşılmaktadır.

Öğrenci ve Soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Öğrenci ve Soru etkileşimi öğrencilerin her bir soruda performanslarında meydana gelen değişikliğin göstergesidir. Öğrenci performansı soru değiştiğinde aynı kalmaktadır, öğrencilerin performanslarında sorudan soruya herhangi bir farklılaşma görülmemektedir.

Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesine bakıldığında, %0 olduğu görülmektedir. Bu değer öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine değişiklik göstermediğini göstermektedir.

Soru ve puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %11'dir. Bu durumda bazı sorulara verilen puanlar bir puanlayıcıdan diğerine farklılık göstermektedir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %61,3'tür. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşenin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 5 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 4'te verilmiştir.

Tablo 4. Kitapçık 5'te, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	133.84200	99	1.35194	0.07283	21.9
Soru	9.05200	4	2.26300	-0.02357	0.0
Puanlayıcı	0.01000	3	0.00333	-0.02332	0.0
OS	12.84800	396	0.03244	-0.02776	0.0
OP	1.89000	297	0.00636	-0.02743	0.0
SP	141.62000	12	11.80167	0.11658	35.0
OSP, e	170.48000	1188	0.14350	0.14350	43.1
Toplam	469.74200	1999			100%

Tablo 4'te, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşenin

toplam varyansın %21,9'unu açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Soru değişkeni için kestirilen varyans bileşeninin ve puanlayıcı değişkeni için kestirilen varyans bileşeninin, toplam varyansı açıklama yüzdeleri %0'dır. Bu durum, 5. kitapçıkta yer alan 5 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediğinin ve puanlayıcılar arasında farklılık bulunmadığının ve puanlayıcıların birbirleriyle tutarlı puanlama yaptıklarının bir göstergesidir.

Öğrenci ve Soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Öğrenci ve Soru etkileşimi öğrencilerin her bir soru da performanslarında meydana gelen değişikliğin göstergesidir. Bu durumda öğrenci performansının sorudan soruya değişmediğini söylemek mümkündür.

Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Bu değere bakarak öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine değişiklik göstermediği söylenebilir.

Soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %35'dir. Bu oldukça yüksek bir değerdir. Buna göre bazı sorulara verilen puanların bir puanlayıcıdan diğerine değişiklik gösterdiği söylenebilir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %43,1'dir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 6 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 5'te verilmiştir.

Tablo 5. Kitapçık 6'da, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	133.71429	99	1.35065	0.05163	15.4
Soru	5.36214	6	0.89369	-0.02410	0.0
Puanlayıcı	0.01714	3	0.00571	-0.01494	0.0
OS	46.28071	594	0.07791	-0.02554	0.0
OP	2.12571	297	0.00716	-0.02470	0.0
SP	191.47786	18	10.63766	0.10458	31.1
OSP, e	320.87929	1782	0.18007	0.18007	53.5
Toplam	699.85714	2799			100%

Tablo 5'te, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %15,4'ünü açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Soru değişkeni için kestirilen varyans bileşeninin, Puanlayıcı değişkeni için kestirilen varyans bileşeninin ve Öğrenci ve Soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdeleri %0'dır. Bu durum, 6. kitapçıkta yer alan soruların zorluk düzeyi bakımından birbirinden farklılık göstermediğinin, puanlayıcılar arasında farklılık bulunmadığının ve puanlayıcıların birbirleriyle tutarlı puanlama yaptıklarının ve öğrenci performansının sorudan soruya değişmediğinin bir göstergesidir.

Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Bu değer performanslarının puanlanması sırasında puanların bir puanlayıcıdan diğerine değişiklik göstermediğini göstermektedir.

Soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %31,1'dir. Bu oldukça yüksek bir değerdir. Bu durumda bazı sorulara verilen puanların bir puanlayıcıdan diğerine değişiklik gösterdiği söylenebilir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %53,5'dir. Üç değişkenlik kaynağının etkileşiminden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 7 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 6'da verilmiştir.

Tablo 6. Kitapçık 7'de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	305.90143	99	3.08991	0.11586	20.6
Soru	34.06214	6	5.67702	-0.03023	0.0
Puanlayıcı	0.02714	3	0.00905	-0.02522	0.0
OS	64.72357	594	0.10896	-0.04009	0.0
OP	1.83000	297	0.00616	-0.03760	0.0
SP	322.69786	18	17.92766	0.17658	31.4
OSP, e	479.94500	1782	0.26933	0.26933	47.9
Toplam	1209.18714	2799			100%

Tablo 6'da, Kitapçık 7 için G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdeleri yer almaktadır. Bu tabloya bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %20,6'sını açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı ne kadar yüksekse, öğrencilerin okuma becerileri bakımından o kadar farklılaştığı söylenebilir.

Soru deęiřkeni iin kestirilen varyans bileřeni, toplam varyansı aıklama yzdesi %0'dır. Bu durumda, 7 kitapıkta yer alan 7 sorunun zorluk dzeyi bakımından birbirinden farklılık gstermedięi sylenebilir.

Puanlayıcı deęiřkeni iin kestirilen varyans bileřenin toplam varyansı aıklama yzdesi %0'dır. Bu durumda, puanlayıcılar arasında farklılık bulunmamaktadır ve puanlayıcıların birbirleriyle tutarlı puanlama yaptıkları sylenebilir.

Öęrenci ve Soru ortak etkisine ait varyans bileřeninin toplam varyansı aıklama yzdesi %0'dır. Öęrenci ve Soru etkileřimi öęrencilerin her bir soruda performanslarında meydana gelen deęiřiklięin gstergesidir. Bu durumda öęrenci performansının sorudan soruya deęiřmedięini sylemek mmkündür. Öęrenci ve Puanlayıcı ortak etkisine ait varyans bileřenin toplam varyansı aıklama yzdesi %0'dır. Dolayısıyla bu deęer öęrencilerin performansları puanlanırken puanların bir puanlayıcıdan dięerine deęiřiklik gstermedięini gstermektedir.

Soru ve puanlayıcı ortak etkisine ait varyans bileřeninin toplam varyansı aıklama yzdesi %31,4'dür. Bu durum bazı sorulara verilen puanların bir puanlayıcıdan dięerine deęiřiklik gsterdięinin bir kanıtıdır. Puanlayıcılar bazı sorulara cmert davranırken bazı sorulara cimri davranmış olabilirler.

Öęrenci Soru Puanlayıcı ortak etkisine ait varyans bileřenin toplam varyansı aıklama yzdesi %47,9'dur. Ü deęiřkenlik kaynaęının etkileřimden elde edilen varyans bileřeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öęrenci, soru, puanlayıcı etkileřiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceęinin bir gstergesi olabilir.

Kitapık 8 iin, Ö X S X P desenine ait G alıřması sonuçları Tablo 7'de verilmiřtir.

Tablo 7. Kitapçık 8’de, Ö X S X P Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	398.14344	99	4.02165	0.09072	19.8
Soru	14.25469	7	2.03638	0.00039	0.1
Puanlayıcı	60.97844	3	20.32615	0.02181	4.8
OS	114.08906	693	0.16463	-0.01120	0.0
OP	345.49031	297	1.16327	0.11923	26.0
SP	40.41406	21	1.92448	0.01715	3.7
OSP, e	435.36719	2079	0.20941	0.20941	45.7
Toplam	1408.73719	3199			100%

Tablo 7’de, 8. Kitapçık için G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdeleri yer almaktadır. Tabloda öğrenci değişkenine ait varyans bileşeninin toplam varyansın %19,8’ini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranının %19,8 olması, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Soru değişkeni için kestirilen varyans bileşeni, toplam varyansı açıklama yüzdesi %0,1’dir. Bu durum, 8. kitapçıkta yer alan 8 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediğinin bir kanıtıdır.

Puanlayıcı değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama yüzdesi %4,8’dir. Bu durum puanlayıcılar arasında bir farklılık bulunduğunun ve puanlayıcıların bazı sorularda birbirleriyle tutarlı puanlama yapmadıklarının bir göstergesi olabilir. 2, 8 ve 12 nolu kitapçıklarda yer alan soruların puanlanmasında tam puan 2, kısmi puan 1 ve yanlış/ilgisiz yanıt 0 olarak değerlendirilmektedir. 4, 5, 6, 7 nolu kitapçıklarda ise puanlama tam puan 1, yanlış/ilgisiz yanıt 0 olarak değerlendirilmektedir. Puanlamada kategoriler artıkça puanlayıcılar arasında tutarlılığın bu durumdan etkilendiği düşünülmektedir.

Öğrenci ve Soru ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %0'dır. Öğrenci ve Soru etkileşimi öğrencilerin her bir soruda performanslarında meydana gelen değişikliğin göstergesidir. Bu durumda öğrenci performansının sorudan soruya değişmediğini söylemek mümkündür.

Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi, %26'dır. Bu değer oldukça yüksek bir değerdir. Bu durumda öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine değişiklik gösterdiğini söyleyebiliriz.

Soru ve puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %3,7'dir. Bu yüzde yüksek olmamakla birlikte, bazı sorulara verilen puanların bir puanlayıcıdan diğerine değişiklik gösterdiğinin bir göstergesi olarak yorumlanabilir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %45,7'dir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşenin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 12 için, Ö X S X P desenine ait G çalışması sonuçları Tablo 8'de verilmiştir.

Tablo 8. Kitapçık 12'de Ö X S X Deseninde, G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	147.87458	99	1.49368	0.05200	19.9
Soru	0.62708	5	0.12542	-0.00343	0.0
Puanlayıcı	33.37458	3	11.12486	0.01563	6.0
OS	36.99792	495	0.07474	-0.01912	0.0
OP	95.66708	297	0.32211	0.02848	10.9
SP	23.63792	15	1.57586	0.01425	5.4
OSP, e	224.57042	1485	0.15123	0.15123	57.8
Toplam	562.74958	2399			100%

Tablo 8'de, 12. Kitapçığa ait G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdeleri verilmektedir. Öğrenci değişkenine ait varyans bileşeninin toplam varyansın %19,9'unu açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesi olarak yorumlanabilir.

Soru değişkeni için kestirilen varyans bileşeninin, toplam varyansı açıklama yüzdesi %0'dır. Bu durum, 12. kitapçıkta yer alan 6 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediğinin bir göstergesidir.

Puanlayıcı değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama yüzdesi %6'dır. Bu durum puanlayıcılar arasında bir farklılık bulunduğunun ve puanlayıcıların birbirleriyle tutarlı puanlama yapmadıklarının bir göstergesi olarak yorumlanabilir. 8. Kitapçıkta olduğu gibi, 12. kitapçıkta da soruların puanlanmasında tam puan 2, kısmi puan 1 olarak değerlendirilmektedir. Puanlamanın 1 tam puan üzerinden yapıldığı diğer kitapçıklarla karşılaştırıldığında, puanlamada kategoriler artıkça puanlayıcılar arasında tutarlılığın bu durumdan etkilendiği görülmektedir.

Öğrenci ve Soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %0'dır. Öğrenci ve Soru etkileşimi öğrencilerin her bir soruda

performanslarında meydana gelen deęişiklięin göstergesidir. Bu durumda öğrenci performansının sorudan soruya deęişmedięini söylemek mümkündür. Öğrenci ve Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesine bakıldığında, %10,9 olduęu görülmektedir. Bu deęer öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan dięerine deęişiklik gösterdięi şeklinde yorumlanabilir.

Soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %5,4'tür. Bu durum, bazı sorulara verilen puanların bir puanlayıcıdan dięerine deęişiklik gösterdięi şeklinde yorumlanabilir. Puanlayıcılar bazı sorulara cömert davranırken bazı sorulara cimri davranmış olabilirler.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %57,8'dir. Üç deęişkenlik kaynaęının etkileşimden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceęinin bir göstergesi olabilir.

3.1.2. Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışması (K) Sonuçları

PISA 2009 Okuma becerilerine ilişkin çoklu puanlama yapılan 7 kitapçıkta (2, 4, 5, 6, 7, 8, 12 nolu kitapçıklar) herbirinden 100 kitapçık 4 puanlayıcı tarafından deęerlendirilmiştir. Kitapçıklarda puanlanan soru sayısı birbirinden farklılık göstermektedir. 2. kitapçıkta 6 soru, 4. kitapçıkta 5 soru, 5. kitapçıkta 5 soru, 6. kitapçıkta 7 soru, 7. kitapçıkta 7 soru, 8. kitapçıkta 8 soru, 12. Kitapçıkta 6 soru yer almaktadır. Bu kitapçıklarda tüm deęişkenlerin çaprazlandığı Ö X S X P deseninde görevler ölçmenin nesnesi olarak alınmış ve öğrenci ve puanlayıcı sayılarının artırılıp azaltılmasıyla oluşturulan senaryolar için Genellenebilirlik katsayıları kestirilmiştir. Her bir kitapçık için kestirilen Genellenebilirlik katsayısı, Phi katsayısı, baęıl ve mutlak hata varyansları aşıęıda verilmiştir.

Kitapçık 2 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 9 ve Tablo 10'da verilmiştir.

Tablo 9. Kitapçık 2'de, Ö X S X P Deseninde Puanlayıcı Sayılarının Artırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1200		1800		2400		3000		3600	
G Katsayısı	0.80		0.86		0.89		0.91		0.92	
Phi Katsayısı	0.68		0.76		0.81		0.84		0.87	
Bağıl Hata V.	0.02604		0.01736		0.01302		0.01042		0.00868	
Ölçm. Std. H.	0.16136		0.13175		0.11410		0.10206		0.09316	
Mutlak H. V.	0.04911		0.03274		0.02455		0.01964		0.01637	
Ölç. Std. H..	0.22161		0.18094		0.15670		0.14016		0.12795	

Kitapçık 2'de yer alan 6 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,89, Phi katsayısı 0,81 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda, G katsayısı 0,91, Phi katsayısı 0,84 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda, G katsayısı 0.92, Phi katsayısı 0,86 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3'e düşürüldüğünde G katsayısı 0,85'e, Phi katsayısı 0,76'ya, puanlayıcı sayısı 2'ye düşürüldüğünde G katsayısı 0,80'e, Phi katsayısı 0,68'e düşmektedir. Bu durumda Genellenebilirlik katsayısı kabul edilebilir düzeyde kalırken, Phi katsayısı kabul edilebilir düzeyin altına düşmektedir.

Tablo 10. Kitapçık 2'de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	480		720		960		1200		1440	
G Katsayısı	0.80		0.86		0.89		0.91		0.92	
Phi Katsayısı	0.71		0.79		0.83		0.86		0.88	
Bağıl Hata V.	0.03017		0.02012		0.01509		0.01207		0.01006	
Ölçm. Std. H.	0.17371		0.14183		0.12283		0.10986		0.10029	
Mutlak H. V.	0.04886		0.03260		0.02446		0.01958		0.01633	
Ölç. Std. H.	0.22105		0.18054		0.15641		0.13994		0.12779	

Öğrenci sayıları 100'den 40 kadar aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları arttırılıp azaltılmıştır. Öğrenci sayısı 100'den 40'a düşürüldüğünde Genellenebilirlik katsayısında büyük bir düşüş olmadığını, 4 puanlayıcı için 0,89'dan 0,88'e; 2 puanlayıcı için 0,80'den 0,79'a düştüğünü; 3, 5 ve 6 puanlayıcı için neredeyse sabit kaldığını söyleyebiliriz. Öğrenci sayısı 100'den 40'a düşürüldüğünde, Phi katsayısı 4, 5, 6 puanlayıcı için 0,80'nin üzerinde bir değer alırken 2 ve 3 puanlayıcı için 0,80 düzeyinin altında yer almaktadır.

Kitapçık 4 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 11 ve Tablo 12'de verilmiştir.

Tablo 11. Kitapçık 4'te, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzey	Evren	Düzey	Evren	Düzey	Evren	Düzey	Evren	Düzey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	5	INF	5	INF	5	INF	5	INF	5	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1000		1500		2000		2500		3000	
G Katsayısı	0.82		0.87		0.90		0.92		0.93	
Phi Katsayısı	0.79		0.85		0.88		0.91		0.92	
Bağıl Hata V.	0.02100		0.01400		0.01050		0.00840		0.00700	
Ölçm. Std. H.	0.14491		0.11832		0.10247		0.09165		0.08366	
Mutlak H. V.	0.02478		0.01652		0.01239		0.00991		0.00826	
Ölç. Std. H.	0.15740		0.12852		0.11130		0.09955		0.09088	

Kitapçık 4'de yer alan 5 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,90 Phi katsayısı 0,88 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda G katsayısı 0,92, Phi katsayısı 0,91 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda, G katsayısı 0.93, Phi katsayısı 0,92 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3'e düşürüldüğünde G katsayısı 0,87'ye Phi katsayısı 0,85'e, puanlayıcı sayısı 2'ye düşürüldüğünde G katsayısı 0,81'e Phi katsayısı 0,79'a düşmektedir. Puanlayıcı sayısı düştüğünde G ve Phi katsayılarında önemli bir düşüş meydana gelmektedir ancak bu durumda bile Genellenebilirlik ve Phi katsayıları kabul edilebilir düzeydedir.

Tablo 12. Kitapçık 4'te, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	5	INF	5	INF	5	INF	5	INF	5	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	400		600		800		1000		1200	
G Katsayısı	0.82		0.87		0.90		0.92		0.93	
Phi Katsayısı	0.80		0.86		0.89		0.91		0.92	
Bağıl Hata V.	0.02128		0.01418		0.01064		0.00851		0.00709	
Ölçm. Std. H.	0.14586		0.11909		0.10314		0.09225		0.08421	
Mutlak H. V.	0.02444		0.01629		0.01222		0.00978		0.00815	
Ölç. Std. H.	0.15632		0.12764		0.11054		0.09887		0.09025	

Kitapçık 4 için, öğrenci sayıları 100'den 40'a aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları azaltılıp artırılmıştır. Öğrenci sayısı 100'den 40'a düşürüldüğünde G ve Phi katsayılarında büyük bir düşüş olmadığı 6, 5, 4, 3, ve 2 puanlayıcı kullanılan her bir desende neredeyse sabit kaldığı görülmüştür. Kitapçık 5 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 13 ve Tablo 14'te verilmiştir.

Tablo 13. Kitapçık 5'te, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	5	INF	5	INF	5	INF	5	INF	5	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1000		1500		2000		2500		3000	
G Katsayısı	0.84		0.88		0.91		0.93		0.94	
Phi Katsayısı	0.73686		0.80771		0.84850		0.87501		0.89363	
Bağıl Hata V.	0.01435		0.00957		0.00718		0.00574		0.00478	
Ölçm. Std. H.	0.11979		0.09781		0.08471		0.07576		0.06916	
Mutlak H. V.	0.02601		0.01734		0.01300		0.01040		0.00867	
Ölç. Std. H.	0.16127		0.13168		0.11404		0.10200		0.09311	

Kitapçık 5’de yer alan 5 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,91, Phi katsayısı 0,85 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda G katsayısı 0,92, Phi katsayısı 0,88 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda G katsayısı 0.93, Phi katsayısı 0,89 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3’e düşürüldüğünde, G katsayısı 0,88’e, Phi katsayısı 0,81’e, puanlayıcı sayısı 2’ye düşürüldüğünde G katsayısı 0,83’e Phi katsayısı 0,74’e düşmektedir. Bu durumda Genellenebilirlik katsayısı kabul edilebilir düzeyde kalırken, Phi katsayısı kabul edilebilir düzeyin altına düşmektedir.

Tablo 14. Kitapçık 5’te, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	5	INF	5	INF	5	INF	5	INF	5	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem		400		600		800		1000		1200
G Katsayısı		0.83		0.88		0.91		0.92		0.93
Phi Katsayısı		0.75		0.82		0.86		0.88		0.90
Bağıl Hata V.		0.01669		0.01113		0.00835		0.00668		0.00556
Ölçm. Std. H.		0.12921		0.10550		0.09136		0.08172		0.07460
Mutlak H. V.		0.02683		0.01789		0.01342		0.01073		0.00894
Ölç. Std. H.		0.16381		0.13375		0.11583		0.10360		0.09458

Kitapçık 6 için, öğrenci sayıları 100’den 40’a aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları azaltılıp artırılmıştır. Öğrenci sayısı 100’den 40’a düşürüldüğünde Genellenebilirlik ve Phi katsayılarında büyük bir düşüş olmadığı, Genellenebilirlik katsayısının 4 puanlayıcı için 0,91’den 0,90’a; 2 puanlayıcı için 0,83’den 0,82’ye; 3 puanlayıcı için 0.88’den 0,87’ye düştüğü görülmüştür, 5 ve 6 puanlayıcı için G katsayısının neredeyse sabit kaldığını

söyleyebiliriz. Phi katsayısının 2 puanlayıcı için 0,80'in altında kalırken, 3, 4, 5 ve 6 puanlayıcı için 0,80'in üzerinde bir değer aldığı görülmektedir.

Kitapçık 6 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 15 ve Tablo 16'da verilmiştir.

Tablo 15. Kitapçık 6'da, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzye	Evren	Düzye	Evren	Düzye	Evren	Düzye	Evren	Düzye	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	7	INF	7	INF	7	INF	7	INF	7	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1400		2100		2800		3500		4200	
G Katsayısı	0.80		0.86		0.89		0.91		0.92	
Phi Katsayısı	0.72		0.79		0.84		0.86		0.88	
Bağıl Hata V.	0.01286		0.00857		0.00643		0.00514		0.00429	
Ölçm. Std. H.	0.11341		0.09260		0.08019		0.07173		0.06548	
Mutlak H. V.	0.02033		0.01355		0.01017		0.00813		0.00678	
Ölç. Std. H.	0.14259		0.11642		0.10083		0.09018		0.08232	

Kitapçık 6'da yer alan 7 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,89, Phi katsayısı 0,84 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda, G katsayısı 0,91, Phi katsayısı 0,86 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda, G katsayısı 0.92, Phi katsayısı 0,88 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3'e düşürüldüğünde G katsayısı 0,85'e, Phi katsayısı 0,79'a, puanlayıcı sayısı 2'ye düşürüldüğünde G katsayısı 0,80'e, Phi katsayısı 0,72'ye düşmektedir. Bu durumda Genellenebilirlik katsayısı kabul edilebilir düzeyde kalırken Phi katsayısı 2 puanlayıcı için kabul edilebilir düzeyin altında kalmaktadır.

Tablo 16. Kitapçık 6'da, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	7	INF	7	INF	7	INF	7	INF	7	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	560		840		1120		1400		1680	
G Katsayısı	0.84		0.89		0.91		0.93		0.94	
Phi Katsayısı	0.76		0.82		0.86		0.89		0.90	
Bağıl Hata V.	0.01156		0.00771		0.00578		0.00463		0.00385	
Ölçm. Std. H.	0.10753		0.08780		0.07604		0.06801		0.06208	
Mutlak H. V.	0.01947		0.01298		0.00974		0.00779		0.00649	
Ölç. Std. H.	0.13954		0.11394		0.09867		0.08826		0.08057	

Kitapçık 6 için, öğrenci sayıları 100'den 40'a aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları artırılıp azaltılmıştır. Öğrenci sayısı 100'den 40'a düşürüldüğünde Genellenebilirlik katsayılarında bir düşüş görülmediği, 4 puanlayıcı için 0,89'dan 0,91'e; 2 puanlayıcı için 0,80'den 0,83'e; 3 puanlayıcı için 0,86'dan 0,88'e, 5 puanlayıcı için 0,91'den 0,93'e ve 6 puanlayıcı için 0,92'den 0,94 yükseldiği görülmüştür. Phi katsayısının ise 2 puanlayıcı için 0,80 seviyesinin altında kaldığı, 3, 4, 5, ve 6 puanlayıcı için 0,80 değerinden daha büyük bir değer aldığı görülmüştür.

Kitapçık 7 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 17 ve Tablo 18'de verilmiştir.

Tablo 17. Kitapçık 7’de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	7	INF	7	INF	7	INF	7	INF	7	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1400		2100		2800		3500		4200	
G Katsayısı	0.86		0.90		0.92		0.94		0.95	
Phi Katsayısı	0.78		0.85		0.88		0.90		0.92	
Bağıl Hata V.	0.01924		0.01283		0.00962		0.00770		0.00641	
Ölçm. Std. H.	0.13870		0.11325		0.09808		0.08772		0.08008	
Mutlak H. V.	0.03185		0.02123		0.01593		0.01274		0.01062	
Ölç. Std. H.	0.17847		0.14572		0.12620		0.11287		0.10304	

Kitapçık 7’de yer alan 7 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,92, Phi katsayısı 0,88 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda G katsayısı 0,94, Phi katsayısı 0,90 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda G katsayısının 0.94, Phi katsayısı 0,92 olarak hesaplanmaktadır. Puanlayıcı sayısı 4’ten 5’e çıkarıldığında G katsayısında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3’e düşürüldüğünde, G katsayısı 0,90’a, Phi katsayısı 0,85’e, puanlayıcı sayısı 2’ye düşürüldüğünde G katsayısı 0,86’ya, Phi katsayısı 0,78’e düşmektedir. Bu durumda Genellenebilirlik katsayısı kabul edilebilir düzeyde kalırken Phi katsayısı kabul edilebilir düzeyin altına düşmektedir.

Tablo 18. Kitapçık 7'de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	7	INF	7	INF	7	INF	7	INF	7	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	560		840		1120		1400		1680	
G Katsayısı	0.86		0.91		0.93		0.94		0.95	
Phi Katsayısı	0.79		0.85		0.88		0.90		0.92	
Bağıl Hata V.	0.01824		0.01216		0.00912		0.00730		0.00608	
Ölçm. Std. H.	0.13505		0.11027		0.09550		0.08542		0.07797	
Mutlak H. V.	0.03169		0.02113		0.01584		0.01268		0.01056	
Ölç. Std. H.	0.17802		0.14535		0.12588		0.11259		0.10278	

Kitapçık 7 için, öğrenci sayıları 100'den 40'a aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları artırılıp azaltılmıştır. Öğrenci sayısı 100'den 40'a düşürüldüğünde Genellenabilirlik ve Phi katsayılarında büyük bir düşüş olmadığı görülmektedir. Öğrenci sayısı 100'den 40'a düşürüldüğünde bile G katsayısı 4 puanlayıcı için 0,93; 2 puanlayıcı için 0,86; 3 puanlayıcı için 0,90; 5 puanlayıcı için 0,94 ve 6 puanlayıcı için 0,95'dir. Phi katsayısı ise 2 puanlayıcı için 0,79 iken, 3 puanlayıcı için 0,85, 4 puanlayıcı için 0,88, 5 puanlayıcı için 0,90 ve 6 puanlayıcı için 0,92 değerini almaktadır.

Kitapçık 8 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 19 ve Tablo 20'de verilmiştir.

Tablo 19. Kitapçık 8’de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	8	INF	8	INF	8	INF	8	INF	8	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1600		2400		3200		4000		4800	
G Katsayısı	0.56		0.65		0.71		0.76		0.79	
Phi Katsayısı	0.52		0.62		0.68		0.73		0.76	
Bağıl Hata V.	0.07270		0.04847		0.03635		0.02908		0.02423	
Ölçm. Std. H.	0.26964		0.22016		0.19066		0.17053		0.15568	
Mutlak H. V.	0.08473		0.05650		0.04239		0.03392		0.02828	
Ölç. Std. H.	0.29108		0.23770		0.20589		0.18418		0.16815	

Kitapçık 8’de yer alan 8 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,71, Phi katsayısı 0,68 olarak hesaplanmıştır. G ve Phi katsayıları için bu düzey kabul edilebilir düzey olan %80’in oldukça altındadır. Genellenebilirlik katsayısı, öğrenci sayısı sabit tutulup, puanlayıcı sayısı 2’ye indirildiğinde 0,56’ya, Phi katsayısı 0,52’ye kadar düşmektedir. Puanlayıcı sayısı 5 olduğunda G katsayısı 0,76, Phi katsayısı 0,73 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda G katsayısı 0.79, Phi katsayısı 0,76 olarak hesaplanmaktadır. Puanlayıcı sayısı arttırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak kabul edilebilir düzeyde bir G katsayısı için en az 6 puanlayıcıya ihtiyaç vardır.

Tablo 20. Kitapçık 8'de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	8	INF	8	INF	8	INF	8	INF	8	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	640		960		1280		1600		1920	
G Katsayısı	0.54		0.64		0.70		0.75		0.78	
Phi Katsayısı	0.48		0.58		0.65		0.69		0.73	
Bağıl Hata V.	0.06891		0.04594		0.03446		0.02757		0.02297	
Ölçm. Std. H.	0.26251		0.21434		0.18563		0.16603		0.15156	
Mutlak H. V.	0.08862		0.05915		0.04442		0.03558		0.02968	
Ölç. Std. H.	0.29770		0.24321		0.21075		0.18862		0.17228	

Kitapçık 8 için, öğrenci sayıları 100'den 40'a kadar aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları azaltılıp artırılmıştır. Öğrenci sayısı 100'den 40'a düşürüldüğünde, Genellebilirlik katsayısı 4 puanlayıcı için 0,71'den 0,70'e, 3 puanlayıcı için 0,65'ten 0,64'e, 2 puanlayıcı için 0,56'dan 0,54'e düşmektedir. Puanlayıcı sayısı 5 olduğunda 0,75'den 0,74'e, puanlayıcı sayısı 6 olduğunda 0,79'dan 0,78 düşmektedir. Phi katsayısı ise 4 puanlayıcı için 0,68'den 0,65'e, 3 puanlayıcı için 0,62'den 0,58'e ve 2 puanlayıcı için 0,52'den 0,48'e düşmektedir. Öğrenci sayısı düşürüldüğünde, G ve Phi katsayılarında bir düşüş meydana gelmektedir. Ancak bu düşüş önemli bir düşüş değildir. Burada asıl önemli olan, 6 puanlayıcı kullanıldığında bile kabul edilebilir düzeyde G ve Phi katsayıları elde etmede bir sıkıntı yaşanıyor olmasıdır. Burada iki ayrı sorun söz konusu olabilir. Birincisi bu kitapçıkta kısmi puanlama söz konusudur, tam puan 2, kısmi puan 1 üzerinden değerlendirilmektedir. Bu durumda puanlama kategorisi artmıştır ve puanlama kategorisinin artmış olması puanlamayı zorlaştıran bir durumdur. Bu kitapçıkta çoklu puanlama yapılan soru sayısı diğer kitapçıklardan fazladır. Dolayısıyla soruların aynı grup içinde birbiriyle bağlantılı sorular olması söz konusudur. Puanlayıcılar herbir soru için puanlama kılavuzuna bakmak yerine ilişkili soruların yanıtlarında kendi içinde bir gruptandırmaya gitmiş olabilirler.

Dolayısıyla getirilen genel bir yorum, puanlama kılavuzundan sapmalara sebep olmuş olabilir.

Kitapçık 12 için, Ö X S X P desenine ait K çalışması sonuçları Tablo 21 ve Tablo 22’de verilmiştir.

Tablo 21. Kitapçık 12’de, Ö X S X P Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1200		1800		2400		3000		3600	
G Katsayısı	0.66		0.74		0.79		0.83		0.85	
Phi Katsayısı	0.59		0.69		0.74		0.78		0.81	
Bağıl Hata V.	0.02684		0.01790		0.01342		0.01074		0.00895	
Ölçm. Std. H.	0.16384		0.13377		0.11585		0.10362		0.09459	
Mutlak H. V.	0.03584		0.02390		0.01792		0.01434		0.01195	
Ölç. Std. H.	0.18933		0.15459		0.13387		0.11974		0.10931	

Kitapçık 12’de yer alan 6 soru doğrultusunda, 100 öğrenci 4 puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,79, Phi katsayısı 0,74 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 olduğunda G katsayısı 0,83, Phi katsayısı 0,78 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda G katsayısı 0.85, Phi katsayısı 0,81 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir. Ancak bu artış büyük bir artış değildir. Puanlayıcı sayısı 3’e düşürüldüğünde, G katsayısı 0,74’e, Phi katsayısı 0,69’a, puanlayıcı sayısı 2’ye düşürüldüğünde, G katsayısı 0,65’e, Phi katsayısı 0,59’a düşmektedir. Kabul edilebilir düzeyde bir Genellenebilirlik katsayısı için puanlayıcı sayısının en az 4 olması gerekmektedir.

Tablo 22. Kitapçık 12’de, Ö X S X P Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	40	INF	40	INF	40	INF	40	INF	40	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
Puanlayıcı	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	480		720		960		1200		1440	
G Katsayısı	0.64		0.73		0.78		0.82		0.84	
Phi Katsayısı	0.60		0.69		0.75		0.79		0.82	
Bağıl Hata V.	0.03551		0.02367		0.01775		0.01420		0.01184	
Ölçm. Std. H.	0.18843		0.15385		0.13324		0.11917		0.10879	
Mutlak H. V.	0.04306		0.02870		0.02153		0.01722		0.01435	
Ölç. Std. H.	0.20750		0.16942		0.14672		0.13123		0.11980	

Kitapçık 12 için, öğrenci sayıları 100’den 40’a aşamalı olarak düşürülmüş ve buna paralel olarak puanlayıcı sayıları artırılıp azaltılmıştır. Öğrenci sayısı 100’den 40’a düşürüldüğünde, 4 puanlayıcı için Genellebilirlik katsayısı 0,78’dir. Genellebilirlik katsayısı 3 puanlayıcı için 0,73, 2 puanlayıcı için 0,64’tür. Genellebilirlik katsayısı 5 puanlayıcı için 0,81 ve 6 puanlayıcı için 0,84’dür. 100 öğrenci 4 puanlayıcı için Genellebilirlik katsayısı kabul edilebilir bir düzey olan 0,80 iken, öğrenci sayısı 40’a düşürüldüğünde Genellebilirlik katsayısı da 0,78’e düşmektedir. Phi katsayısı ise öğrenci sayısı 40’a düşürüldüğünde aynı düzeyde kalmakta, 5 ve 6 puanlayıcı için kabul edilebilir düzey olan 0,80’in üzerinde bir değer almaktadır.

3.2. ÖĞRENCİ (Ö) ve PUANLAYICI (P) DEĞİŞKENLERİNİN YUVALANMIŞ, SORU (S) DEĞİŞKENİNİN İSE ÇAPRAZ TASARLANDIĞI (Ö:P)XS DESENİNİN GENELLENEBİLİRLİK ÇALIŞMASI SONUÇLARI

PISA 2009’da okuma becerilerine ilişkin soruların bir kısmı daha önceki uygulamalarda kullanılan sorulardır. Bu sorular yıllar içinde okuma becerilerinin gelişimini izlemek için kullanılan sorulardır ve bu nedenle hiçbir değişikliğe uğramadan kullanılmakta ve puanlama kriterleri hiçbir şekilde değişmemektedir.

PISA 2009 uygulamasında yer alan 2. ve 8. Kitapçıklar 2006 uygulamasında kullanılan sorulardan oluşmaktadır. Bu nedenle (Ö : P) X S desenini kullanırken 2006 ve 2009 verileri birleştirilerek öğrenci değişkeniyle puanlayıcı değişkeninin yuvalandığı soru değişkeninin çapraz olduğu desen oluşturulmuştur. 2. kitapçıkta yer alan toplam 6 soru için 2006 yılından 93, öğrenci 2009 yılından 93 öğrenci olmak üzere toplam 186 öğrenci, 4 puanlayıcı tarafından puanlanmıştır. Aynı şekilde, 8. kitapçıkta yer alan toplam 8 soru için 2006 yılından 93 öğrenci 2009 yılından 93 öğrenci olmak üzere toplam 186 öğrenci 4 puanlayıcı tarafından puanlanmıştır. Bu desende sorular ölçmenin nesnesi olarak alınmış ve öğrenci ve puanlayıcı sayılarının artırılıp azaltılmasıyla oluşturulan senaryolar için Genellenebilirlik katsayıları kestirilmiştir.

Bu bölümde, (Ö:P) X S deseninin Genellenebilirlik çalışması sonuçları ve bu desende puanlayıcı ve öğrenci sayılarının arttırılıp azaltılmasıyla gerçekleştirilen Karar çalışması sonuçları iki alt başlık halinde verilmiştir.

3.2.1. (Ö:P) X S Desende G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Kitapçık 2 için, (Ö:P) X S desenine ait G çalışması sonuçları Tablo 23'te verilmiştir.

Tablo 23. Kitapçık 2'de, (Ö:P) X S Desende G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	520.26725	185	2.81226	0.11064	16.8
Soru	200.43660	5	40.08732	0.05369	8.2
P:O	281.29167	558	0.50411	0.00197	0.3
OS	134.27173	925	0.14516	-0.08678	0.0
SP:O, e	1373.45833	2790	0.49228	0.49228	74.7
Toplam	2509.72558	4463			100%

Tablo 23'te, Kitapçık 2 için, (Ö:P) X S desenine ait G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdeleri verilmiştir. G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %16,8'ini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranının %16,8 olması, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Soru değişkeni için kestirilen varyans bileşeninin, toplam varyansı açıklama yüzdesi %8,2'dir. Bu durum, bu kitapçıkta yer alan 6 sorunun zorluk düzeyi bakımından birbirinden farklılaştığının bir göstergesidir.

Bu çalışmada öğrenci değişkeniyle puanlayıcı değişkeni yuvalanmıştır. (Ö:P) için kestirilen varyans değeri toplam varyansın %0,3'ünü açıklamaktadır. Bu değer küçük olması öğrenci puanlayıcı etkileşiminin farklılaşmadığı, puanlayıcıların yaptığı puanlamanın bir öğrenciden diğerine farklılaşmadığı şeklinde yorumlanabilir.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi %74,7'dir. Artık varyans bileşenin yüksek çıkması öğrenci görev ortak etkileşimi, soru puanlayıcı ortak etkileşiminin yüksek olması ve/veya tesadüfi hata kaynaklarının yüksek olması olarak yorumlanabilir. Kitapçık 8 için, (Ö:P) X S desenine ait G çalışması sonuçları Tablo 24'te verilmiştir.

Tablo 24. Kitapçık 8'de (Ö:P) X S Deseninde G Çalışmasına Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	780.85417	185	4.22083	0.07946	15.5
Soru	31.11492	7	4.44499	0.00572	1.1
P:O	966.25000	558	1.73163	0.18648	36.5
OS	241.38508	1295	0.18640	-0.01336	0.0
SP:O, e	936.75000	3906	0.23982	0.23982	46.9
Toplam	2956.35417	5951			100%

Tablo 24'te, Kitapçık 8 için, (Ö:P) X S desenine ait G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdeleri verilmiştir. G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin toplam varyansın %15,5'ini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranı, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir. Soru değişkeni için kestirilen varyans bileşeninin, toplam varyansı açıklama yüzdesi %1,1'dir. Bu durum, 8. kitapçıkta yer alan 8 sorunun zorluk düzeyi bakımından birbirinden farklılık göstermediği şeklinde yorumlanabilir.

Bu çalışmada öğrenci değişkeniyle puanlayıcı değişkeni yuvalanmıştır. (Ö:P) için kestirilen varyans değeri toplam varyansın %36,5'ini açıklamaktadır. Bu değer büyük olması öğrenci puanlayıcı etkileşiminin farklılaştığı, puanlayıcıların yaptığı puanlamanın bir öğrenciden diğerine farklılık gösterdiği şeklinde yorumlanabilir.

Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi %46,9'dur. Artık varyans bileşeninin yüksek çıkması öğrenci görev ortak etkileşimi, soru puanlayıcı ortak etkileşiminin yüksek olması ve/veya tesadüfi hata kaynaklarının yüksek olması olarak yorumlanabilir.

3.2.2. (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar (K) Çalışması Sonuçları

Kitapçık 2 için, (Ö:P) X S desenine ait K çalışması sonuçları Tablo 25 ve Tablo 26'da verilmiştir.

Tablo 25. Kitapçık 2’de, (Ö:P) X S Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	186	INF	186	INF	186	INF	186	INF	186	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
P:O	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	2232		3348		4464		5580		6696	
G Katsayısı	0.98		0.98		0.99		0.99		0.99	
Phi Katsayısı	0.97		0.97		0.98		0.98		0.98	
Bağıl Hata V.	0.00132		0.00088		0.00066		0.00053		0.00044	
Ölçm. Std. H.	0.03638		0.02970		0.02572		0.02301		0.02100	
Mutlak H. V.	0.00192		0.00148		0.00126		0.00113		0.00104	
Ölç. Std. H.	0.04386		0.03848		0.03548		0.03356		0.03221	

Kitapçık 2’de yer alan 6 soru doğrultusunda, 186 öğrenci iki grup halinde 4’er puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,99, Phi katsayısı 0,98 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 5 ve 6’ya çıkarıldığında G ve Phi katsayıları sabit kalmaktadır. Hesaplanan G katsayısı çok yüksek olduğundan puanlayıcı sayısı artırıldığında G katsayılarında meydana gelen artış çok küçüktür. Puanlayıcı sayısı 3’e ve 2’ye düşürüldüğünde de G ve Phi katsayıları neredeyse sabit kalmaktadır.

Tablo 26. Kitapçık 2’de, (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	6	INF	6	INF	6	INF	6	INF	6	INF
P:O	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1200		1800		2400		3000		3600	
G Katsayısı	0.95		0.96		0.97		0.98		0.98	
Phi Katsayısı	0.92		0.94		0.95		0.95		0.95	
Bağıl Hata V.	0.00218		0.00145		0.00109		0.00087		0.00073	
Ölçm. Std. H.	0.04665		0.03809		0.03298		0.02950		0.02693	
Mutlak H. V.	0.00336		0.00261		0.00223		0.00201		0.00186	
Ölç. Std. H.	0.05793		0.05105		0.04724		0.04480		0.04309	

Kitapçık 2 için, öğrenci sayısı 186’dan 100’e kadar aşamalı olarak düşürülmüştür ve buna paralel olarak puanlayıcı sayıları azaltılıp artırılmıştır. Öğrenci sayısı 186’dan 100’e düşürüldüğünde Genellenebilirlik katsayısının 4 puanlayıcı için 0,99’dan 0,97’ye; 3 puanlayıcı için 0,98’den 0,96’ya, 2 puanlayıcı için 0,98’den 0,95’e düştüğü görülmektedir. 5 ve 6 puanlayıcı için 0,99’dan 0,98’e düşmektedir. Phi katsayısı ise 4 puanlayıcı için 0,98’den 0,95’e, 3 puanlayıcı için 0,97’den 0,94’e ve 2 puanlayıcı için 0,97’den 0,92’ye düşmektedir. 5 ve 6 puanlayıcı için Phi katsayısı sabit kalmaktadır.

Kitapçık 8 için, (Ö:P) X S desenine ait G çalışması sonuçları Tablo 27 ve Tablo 28’de verilmiştir.

Tablo 27. Kitapçık 8’de, (Ö:P) X S Deseninde Puanlayıcı Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	186	INF	186	INF	186	INF	186	INF	186	INF
Soru	8	INF	8	INF	8	INF	8	INF	8	INF
P:O	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	2976		4464		5952		7440		8928	
G Katsayısı	0.90		0.93		0.95		0.96		0.96	
Phi Katsayısı	0.78		0.83		0.85		0.87		0.88	
Bağıl Hata V.	0.00064		0.00043		0.00032		0.00026		0.00021	
Ölçm. Std. H.	0.02539		0.02073		0.01795		0.01606		0.01466	
Mutlak H. V.	0.00157		0.00119		0.00100		0.00089		0.00081	
Ölç. Std. H.	0.03966		0.03451		0.03163		0.02976		0.02845	

Kitapçık 8’de yer alan 8 soru doğrultusunda, 186 öğrenci iki grup halinde 4’er puanlayıcı tarafından birbirlerinden bağımsız olarak puanlanmıştır. Buna göre G katsayısı 0,95, Phi katsayısı 0,85 olarak hesaplanmıştır. Öğrenci sayısı sabit tutulup, puanlayıcı sayısı 3’e indirildiğinde G katsayısı 0,93, Phi katsayısı 0,83, puanlayıcı sayısı 2’ye düşürüldüğünde G katsayısı 0,90, Phi katsayısı 0,78 olarak hesaplanmaktadır. Puanlayıcı sayısı 5 olduğunda G katsayısı 0,96, Phi katsayısı 0,87 olarak hesaplanmaktadır. Puanlayıcı sayısı 6 olduğunda G katsayısı 0,96, Phi katsayısı 0,88 olarak hesaplanmaktadır. Puanlayıcı sayısı artırıldığında G ve Phi katsayılarında bir artış meydana gelmektedir.

Tablo 28. Kitapçık 8’de, (Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

	Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren	Düzyey	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	8	INF	8	INF	8	INF	8	INF	8	INF
P:O	2	INF	3	INF	4	INF	5	INF	6	INF
Gözlem	1600		2400		3200		4000		4800	
G Katsayısı	0.80		0.86		0.89		0.91		0.92	
Phi Katsayısı	0.61		0.68		0.72		0.74		0.76	
Bağıl Hata V.	0.00121		0.00081		0.00061		0.00049		0.00040	
Ölçm. Std. H.	0.03484		0.02845		0.02464		0.02204		0.02012	
Mutlak H. V.	0.00308		0.00230		0.00191		0.00168		0.00153	
Ölç. Std. H.	0.05548		0.04797		0.04374		0.04099		0.03905	

Kitapçık 8 için, öğrenci sayısı 186’dan 100’e kadar aşamalı olarak düşürülmüştür ve buna paralel olarak puanlayıcı sayıları azaltılıp artırılmıştır. Öğrenci sayısı 186’dan 100’e düşürüldüğünde Genellenebilirlik katsayısının 4 puanlayıcı için 0,95’den 0,88’e; 3 puanlayıcı için 0,93’ten 0,85’e, 2 puanlayıcı için 0,90’dan 0,80’e düştüğü görülmektedir. G katsayısı 5 puanlayıcı için 0,96’dan 0,91’e, 6 puanlayıcı için 0,96’dan 0,92’ye düşmektedir. Phi katsayısı ise 4 puanlayıcı için 0,85’ten 0,72’ye, 3 puanlayıcı için 0,83’ten 0,68’e, 2 puanlayıcı için 0,78’den 0,61’e düşmektedir. Öğrenci sayısı 100’e indirilip puanlayıcı sayısı 2’ye indirildiğinde G katsayısı kabul edilebilir düzeyde kalırken 2, 3, 4, 5 ve 6 puanlayıcı için Phi katsayısı kabul edilebilir düzeyin altında bir değer almaktadır.

3.3. Ö X S X P ve (Ö:P) X S DESENLERİNE AİT G ÇALIŞMASI SONUÇLARININ DEĞİŞİMİ

Kitapçık 2 için, Ö X S X P ve (Ö:P) X S desenlerine ait G çalışması sonuçları Tablo 29’da verilmiştir.

Tablo 29. Kitapçık 2’de, Ö X S X P ve (Ö:P) X S Desenlerine Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Çapraz Desen					Varyans Kaynağı	Yuvalanmış Desen				
	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%		Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	267.59	99	2.70	0.10577	16.0	Öğrenci	287.68	99	2.90	0.11076	18.5
Soru	137.14	5	27.42	-0.00095	0.0	Soru	78.79	5	15.75	0.03899	6.5
Puanlayıcı	39.53	3	13.17	-0.02470	0.0	P:O	156.45	300	0.52	0.01439	2.4
OS	60.85	495	0.122	-0.03699	0.0	OS	79.91	495	0.16	-0.06844	0.0
OP	92.80	297	0.312	0.00692	1.0						
SP	419.35	15	27.95	0.27686	41.9						
OSP, e	402.31	1485	0.27	0.27092	41.0	SP:O, e	652.79	1500	0.43	0.43519	72.6
Toplam	1419.59	2399			100%	Toplam	1255.64	2399			100%

Tablo 29'da, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelere bakıldığında öğrenci değişkenine ait varyans bileşeninin çapraz desen için toplam varyansın %16'sını, yuvalanmış desen için %18,5'ini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeni, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir. Bu durum her iki desende de benzerdir.

Sorulara ait varyans bileşenlerinin toplam varyansı açıklama oranına bakıldığında, çapraz desende %0 iken, yuvalanmış desende %6,5 olarak hesaplandığı görülmektedir. Bu durumda soruların çapraz desende zorluk kolaylık açısından farklılaşmadığı görülürken, yuvalanmış desende farklılaştığı görülmektedir. Yuvalanmış Desende varyans bileşeninin daha büyük olması görevlerin daha iyi ayırt edildiğinin göstergesidir.

Her iki desende kestirilen puanlayıcı ana etkisine ait varyans bileşenleri incelendiğinde, her iki desende de puanlayıcı etkisine ait varyans bileşeninin oldukça küçük olduğu, dolayısıyla puanlayıcıların her iki desende de öğrencileri tutarlı bir şekilde puanladıkları söylenebilir.

Tablo 29'da, öğrenci puanlayıcı varyans bileşeni incelendiğinde öğrencilerin durumlarının bir puanlayıcıdan diğerine değişmediği görülmektedir, her iki desende de öğrenci soru ortak etkisi %0'dır. Bu durum öğrencilerin performanslarının bir görevden diğerine farklılaşmadığını göstermektedir.

Soru puanlayıcı ortak varyans bileşeni incelendiğinde, varyans bileşeninin oldukça yüksek olduğu, bu durumda puanlayıcıların puanlamada sorudan soruya farklılaştığı söylenebilir.

Her iki desende de, artık varyans yüksek çıkmıştır, ancak yuvalanmış desende daha yüksektir. Soruların zorluk derecesinin farklılaşması ve Puanlayıcı ile öğrenci arasındaki etkileşim bu durumda etkili olduğu düşünülmektedir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşeninin sıfır (0)

olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Kitapçık 8 için, Ö X S X P ve (Ö:P) X S desenlerine ait G çalışması sonuçları Tablo 30'da verilmiştir.

Tablo 30. Kitapçık 8’de, Ö X S X P ve (Ö:P) X S Desenlerine Ait Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Çapraz Desen					Varyans Kaynağı	Yuvalanmış Desen				
	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%		Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	%
Öğrenci	398.14	99	4.02	0.09072	19.8	Öğrenci	434.02	99	4.38	0.07488	13.7
Soru	14.25	7	2.03	0.00039	0.1	Soru	14.99	7	2.14	0.00484	0.9
P	60.97	3	20.32	0.02181	4.8	P:O	607.90	300	2.02	0.22295	40.9
OS	114.08	693	0.16	-0.01120	0.0	OS	141.53	693	0.20	-0.00964	0.0
OP	345.49	297	1.16	0.11923	26.0						
SP	40.41	21	1.92	0.01715	3.7						
OSP, e	435.36	2079	0.20	0.20941	45.7	SP:O, e	509.84	2100	0.24	0.24278	44.5
Toplam	1408.73	3199			100%	Toplam	1708.30	3199			100%

Tablo 30'da, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında öğrenci değişkenine ait varyans bileşeninin çapraz desen için toplam varyansın %19,8'ini, yuvalanmış desen için %13,7'sini açıkladığı görülmektedir. Öğrencilere ait varyans bileşeni, öğrencilerin okuma becerileri bakımından farklılaştığının bir göstergesidir.

Sorulara ait varyans bileşenlerinin toplam varyansı açıklama oranına bakıldığında çapraz desende %0,1 iken, yuvalanmış desende %0,9 olarak hesaplandığı görülmektedir. Bu durumda soruların her iki desende de zorluk kolaylık açısından farklılaşmadığı görülmektedir.

Puanlayıcı ana etkisine ait varyans bileşenleri incelendiğinde, her iki desende de puanlayıcı etkisine ait varyans bileşeninin oldukça yüksek olduğu, dolayısıyla puanlayıcıların her iki desende de öğrencileri puanlarken tutarlı davranmadığı görülmektedir. Tabloda öğrenci puanlayıcı varyans bileşeni incelendiğinde bu değer yüksek olduğu ve öğrencilerin puanlarının bir puanlayıcıdan diğerine değişiklik gösterdiğini söyleyebiliriz.

Çapraz desende, soru puanlayıcı ortak varyans bileşeni incelendiğinde, varyans bileşeninin küçük bir değer olduğu ve puanlayıcıların puanlamada sorudan soruya farklılaşmadığı görülmektedir.

Artık varyans her iki desende de yaklaşık %45 ile oldukça yüksektir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Ö X S X P ve (Ö:P) X S desenlerine ait Genellenebilirlik düzeyleri Tablo 31'de verilmiştir.

Tablo 31. ÖXSXP ve (Ö:P) X S Desenlerine Ait Genellenebilirlik Düzeyleri

Ö X S X P deseni			(Ö:P) X S deseni		
Kitapçık 2	G Katsayısı	0,89	Kitapçık 2	G Katsayısı	0,99
	Phi Katsayısı	0,81		Phi Katsayısı	0,98
Kitapçık 8	G Katsayısı	0,71	Kitapçık 8	G Katsayısı	0,95
	Phi Katsayısı	0,68		Phi Katsayısı	0,85

Ö X S X P ve (Ö:P) X S desenlerinden kestirilen G ve Phi katsayıları karşılaştırıldığında, (Ö:P) X S deseninden elde edilen G ve Phi katsayılarının, Ö X S X P desene oranla daha yüksek hesaplandığı görülmüştür. Özellikle, 8. kitaba ilişkin olarak elde edilen çapraz desen sonuçlarında G katsayısının kabul edilebilir düzeyde olmadığı ve Phi katsayısının da bu değer de altında olduğu görülmektedir. Ancak aynı kitapçık için yuvalanmış desenden elde edilen sonuçlarda G ve Phi katsayılarının sırasıyla 0,95 ve 0,85 olduğu, dolayısıyla kabul edilebilir düzeyin üstünde hesaplandığı görülmektedir.

3.4. Ö X S X P ve (Ö : P) X S DESENLERİNDE PUANLAYICI VE ÖĞRENCİ SAYILARININ ARTTIRILIP AZALTILMASIYLA YAPILAN KARAR ÇALIŞMALARINDA ELDE EDİLEN SONUÇLARIN DEĞİŞİMİ

Bu bölümde Ö X S X P ve (Ö:P) X S desenlerine ait puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla gerçekleştirilen K çalışması sonuçları iki alt başlık halinde verilecektir.

3.4.1. Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışmalarında Elde Edilen Bağıl ve Mutlak Hata Varyanslarının Değişimi

Kitapçık 2 için, Ö X S X P ve (Ö:P) X S desenlerine ait puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla gerçekleştirilen K çalışması sonuçları Tablo 32'de verilmiştir.

Tablo 32. Kitapçık 2’de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

Öğrenci		Çapraz desen					Yuvalanmış Desen				
		2 P	3 P	4 P	5 P	6 P	2 P	3 P	4 P	5 P	6 P
100	Bağıl Hata Varyansı	0.02604	0.01736	0.01302	0.01042	0.00868	0.00218	0.00145	0.00109	0.00087	0.00073
	Mutlak Hata Varyansı	0.04911	0.03274	0.02455	0.01964	0.01637	0.00336	0.00261	0.00223	0.00201	0.00186
90	Bağıl Hata Varyansı	0.02737	0.01825	0.01369	0.01095	0.00912	0.00273	0.00182	0.00137	0.00109	0.00091
	Mutlak Hata Varyansı	0.05018	0.03346	0.02509	0.02007	0.01673	0.00398	0.00306	0.00260	0.00233	0.00214
80	Bağıl Hata Varyansı	0.02558	0.01705	0.01279	0.01023	0.00853	0.00308	0.00205	0.00154	0.00123	0.00103
	Mutlak Hata Varyansı	0.04827	0.03218	0.02414	0.01931	0.01609	0.00447	0.00344	0.00293	0.00262	0.00241
70	Bağıl Hata Varyansı	0.02640	0.01760	0.01320	0.01056	0.00880	0.00352	0.00234	0.00176	0.00141	0.00117
	Mutlak Hata Varyansı	0.04663	0.03110	0.02333	0.01867	0.01557	0.00511	0.00393	0.00335	0.00299	0.00276
60	Bağıl Hata Varyansı	0.02612	0.01741	0.01306	0.01045	0.00871	0.00410	0.00273	0.00205	0.00164	0.00137
	Mutlak Hata Varyansı	0.04538	0.03029	0.02275	0.01822	0.01521	0.00596	0.00459	0.00390	0.00349	0.00322
50	Bağıl Hata Varyansı	0.02968	0.01978	0.01484	0.01187	0.00989	0.00492	0.00328	0.00246	0.00197	0.00164
	Mutlak Hata Varyansı	0.04770	0.03184	0.02392	0.01916	0.01599	0.00716	0.00551	0.00468	0.00419	0.00386

Kitapçık 2 için, çapraz desende ve yuvalanmış desende elde edilen bağıl ve mutlak hata varyansları Tablo 32’de verilmiştir. Bu tabloda öğrenci sayısı sabit tutulup, puanlayıcı sayısının azaltılıp artırıldığı durumlarda elde edilen bağıl ve mutlak hata varyansları incelendiğinde, puanlayıcı sayısı arttıkça her iki desen için bağıl ve mutlak hata varyanslarının düştüğü görülmektedir. Çapraz desende 100 öğrencinin 2 puanlayıcı tarafından puanlanmasıyla elde edilen bağıl hata varyansı 0,02604 iken mutlak hata varyansı 0,04911’dir. Ancak 100 öğrenci için 6 puanlayıcı kullanıldığında bağıl hata varyansı 0,00862’ye mutlak hata varyansı 0,01637’e düşmektedir.

100 öğrenci için yuvalanmış desende 2 puanlayıcı tarafından yapılan puanlama sonucunda bağıl hata varyansının 0,00218 mutlak hata varyansının 0,00336 olduğu görülmektedir. 100 öğrenci için 6 puanlayıcının kullanıldığı yuvalanmış desende bağıl hata varyansı 0,00073’e mutlak hata varyansı 0,00186’ya düşmektedir.

Öğrenci sayısı azaltıldığında, bağıl ve mutlak hata varyansları artmaktadır. Ancak çapraz desende bağıl ve mutlak hata varyansları yuvalanmış desene göre daha yüksektir. Çapraz desende 100 öğrencinin 2 puanlayıcı tarafından puanlanmasıyla elde edilen mutlak hata varyansı 0,02604 iken öğrenci sayısının yarıya indirildiği durumda 0,02968’dir. Yuvalanmış desende ise 100 öğrenci 2 puanlayıcı durumunda elde edilen bağıl hata varyansı 0,00218 iken öğrenci sayısı yarıya düşürüldüğünde bağıl hata varyansı 0,00716’dır. Bu değer çapraz desene göre oldukça düşük bir değerdir.

Öğrenci sayısının yarıya indirilip puanlayıcı sayısının 6 olduğu durumda, çapraz desende bağıl hata varyansı 0,00989’a düşmektedir. Bu değer yuvalanmış desende 0,00164’dür. 6 puanlayıcı kullanıldığı durumda bağıl hata varyansının düştüğü görülmektedir. Bağıl hata varyansının azalmasında Puanlayıcı sayısı öğrenci sayısından daha etkilidir. Bu durumda orijinal puanlayıcı sayısı değiştirilirse de öğrenci sayısının yarıya indirilmesi zamandan tasarruf anlamında iyi bir seçenek olabilir.

Kitapçık 8 için, Ö X S X P ve (Ö:P) X S desenlerine ait puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla gerçekleştirilen K çalışması sonuçları Tablo 33'te verilmiştir.

Tablo 33. Kitapçık 8’de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

Öğrenci		Çapraz Desen					Yuvalanmış Desen				
		2 P	3 P	4 P	5 P	6 P	2 P	3 P	4 P	5 P	6 P
100	Bağlı Hata Varyansı	0.07270	0.04847	0.03635	0.02908	0.02423	0.00121	0.00081	0.00061	0.00049	0.00040
	Mutlak Hata Varyansı	0.08473	0.05650	0.04239	0.03392	0.02828	0.00308	0.00230	0.00191	0.00168	0.00153
90	Bağlı Hata Varyansı	0.08095	0.05402	0.04055	0.03247	0.02709	0.00136	0.00091	0.00068	0.00055	0.00045
	Mutlak Hata Varyansı	0.06826	0.04551	0.03413	0.02730	0.02275	0.00343	0.00255	0.00211	0.00184	0.00167
80	Bağlı Hata Varyansı	0.06818	0.04546	0.03409	0.02727	0.02273	0.00150	0.00100	0.00075	0.00060	0.00050
	Mutlak Hata Varyansı	0.08099	0.05399	0.04050	0.03240	0.02700	0.00366	0.00277	0.00233	0.00206	0.00188
70	Bağlı Hata Varyansı	0.06767	0.04512	0.03384	0.02707	0.02256	0.00418	0.00317	0.00266	0.00235	0.00215
	Mutlak Hata Varyansı	0.08204	0.05474	0.04108	0.03289	0.02743	0.00171	0.00114	0.00086	0.00069	0.00057
60	Bağlı Hata Varyansı	0.07198	0.04799	0.03599	0.02879	0.02399	0.00200	0.00133	0.00100	0.00080	0.00067
	Mutlak Hata Varyansı	0.08863	0.05914	0.04440	0.03555	0.02966	0.00488	0.00369	0.00310	0.00275	0.00251
50	Bağlı Hata Varyansı	0.07119	0.04746	0.03559	0.02848	0.02373	0.00240	0.00160	0.00120	0.00096	0.00080
	Mutlak Hata Varyansı	0.08940	0.05972	0.04488	0.03598	0.03005	0.00585	0.00443	0.00372	0.00329	0.00301

Kitapçık 8 için, çapraz desende elde edilen bağıl ve mutlak hata varyanslarına bakıldığında bu değerlerin çok yüksek olduğu görülmektedir. Bu yükseklik Genellenebilirlik ve Phi katsayılarına da yansımış, dolayısıyla G ve Phi katsayıları kabul edilebilir düzeyin altında kalmıştır. Çapraz desende puanlayıcı sayısının artırılması bağıl ve mutlak hata varyanslarında düşüşe sebep olmaktadır. Yuvalanmış desende bağıl ve mutlak hataların çapraz desene oranla oldukça küçük olduğu görülmektedir. Puanlayıcı sayısı artıkça, bağıl ve mutlak hata varyansları daha da küçülmektedir.

100 öğrenci için 2 puanlayıcının kullanıldığı çapraz desende bağıl hata varyansı 0,07270, mutlak hata varyansı 0,08473'dür. Yuvalanmış desene bakıldığında 100 öğrenci 2 puanlayıcı için bağıl hata varyansı 0,00121 ve mutlak hata varyansı 0,00308'dir. Bağıl ve Mutlak hata varyanslarının küçüklüğü G ve Phi katsayılarına da yansımış, bu katsayılar kabul edilebilir değerlerin oldukça üstünde hesaplanmıştır.

3.4.2. Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan Karar Çalışmalarında Elde Edilen G ve Phi Katsayılarının Değişimi

Kitapçık 2 için, Ö X S X P ve (Ö:P) X S desenlerine ait K çalışması sonuçlarından elde edilen G ve Phi Katsayıları Tablo 34'te verilmiştir

Tablo 34. Kitapçık 2'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

Öğrenci Sayısı		Çapraz desen					Yuvalanmış Desen				
		2 P	3 P	4 P	5 P	6 P	2 P	3 P	4 P	5 P	6 P
100	G	0.80	0.86	0.89	0.91	0.92	0.95	0.96	0.97	0.98	0.98
	Phi	0.68	0.76	0.81	0.84	0.87	0.92	0.94	0.95	0.95	0.95
90	G	0.80	0.86	0.89	0.91	0.92	0.95	0.97	0.98	0.98	0.98
	Phi	0.69	0.77	0.81	0.85	0.87	0.93	0.95	0.95	0.96	0.96
80	G	0.82	0.87	0.90	0.92	0.93	0.95	0.96	0.97	0.98	0.98
	Phi	0.71	0.79	0.83	0.86	0.88	0.92	0.94	0.95	0.95	0.96
70	G	0.82	0.87	0.90	0.92	0.93	0.94	0.96	0.97	0.97	0.98
	Phi	0.72	0.80	0.84	0.87	0.89	0.91	0.93	0.94	0.95	0.95
60	G	0.83	0.88	0.91	0.92	0.94	0.93	0.95	0.96	0.97	0.98
	Phi	0.73	0.81	0.85	0.87	0.89	0.90	0.92	0.93	0.94	0.94
50	G	0.81	0.87	0.90	0.91	0.93	0.92	0.94	0.96	0.96	0.97
	Phi	0.73	0.80	0.84	0.87	0.89	0.88	0.91	0.92	0.93	0.93

Tablo 34'de okuma becerileri Kitapçık 2'de yer alan 6 maddenin, çapraz desende 100 öğrencinin 4 puanlayıcı tarafından puanlanmasıyla elde edilen G katsayısı 0,89 ve Phi katsayısı 0,81 olarak hesaplanmıştır. Yine 100 öğrenci için yuvalanmış desende 4 puanlayıcı tarafından yapılan puanlamada G katsayısı 0,97 ve Phi katsayısı 0,95 gibi oldukça yüksek değerler olarak hesaplanmıştır. 100 öğrenci için yapılan puanlamada çapraz desende puanlayıcı sayısını 2'ye indirdiğimizde G katsayısı 0,80 ve Phi Katsayısı 0,68 olarak elde edilmektedir. Yuvalanmış desende elde edilen G katsayısı 0,95 ve Phi katsayısı 0,92'dir.

Hem çapraz desende, hem de yuvalanmış desende öğrenci sayılarını aşamalı olarak 100'den 50'ye kadar indirdiğimizde her iki desende de 4 puanlayıcı için G ve Phi katsayıları kabul edilebilir düzey olan 0,80'in üzerinde yer almaktadır. Ancak öğrenci sayısını 100'den 50'ye indirirken puanlayıcı sayısını da yarıya indirdiğimizde, çapraz desende G katsayıları 0,80'in üzerinde hesaplanırken Phi katsayıları her bir durumda kabul edilebilir düzey olan 0,80'in altında kalmaktadır. Ancak yuvalanmış desende gerek öğrenci sayısındaki azalma,

gerekse puanlayıcı sayısının 2'ye indirilmesi G ve Phi katsayılarını kabul edilebilir düzey olan 0,80'nin üzerinde tutmaya devam etmektedir. Yuvalanmış desende 50 öğrenci için 2 puanlayıcı tarafından yapılan değerlendirme işleminde G katsayısı 0,92 ve Phi katsayısı 0,88 olmaktadır.

Kitapçık 8 için, Ö X S X P ve (Ö:P) X S desenlerine ait K çalışması sonuçlarından elde edilen G ve Phi Katsayıları Tablo 35'te verilmiştir

Tablo 35. Kitapçık 8'de, Ö X S X P ve (Ö:P) X S Desenlerinde Puanlayıcı ve Öğrenci Sayılarının Arttırılıp Azaltılmasıyla Yapılan K Çalışması Sonuçları

Öğrenci Sayısı		Çapraz desen					Yuvalanmış Desen				
		2 P	3 P	4 P	5 P	6 P	2 P	3 P	4 P	5 P	6 P
100	G	0.56	0.65	0.71	0.76	0.79	0.80	0.86	0.89	0.91	0.92
	Phi	0.52	0.62	0.68	0.73	0.76	0.61	0.68	0.72	0.74	0.76
90	G	0.58	0.67	0.73	0.78	0.81	0.81	0.87	0.90	0.91	0.93
	Phi	0.54	0.64	0.70	0.74	0.78	0.64	0.70	0.73	0.76	0.77
80	G	0.60	0.69	0.75	0.79	0.82	0.79	0.85	0.88	0.91	0.92
	Phi	0.56	0.65	0.72	0.76	0.79	0.61	0.67	0.71	0.74	0.75
70	G	0.59	0.69	0.74	0.78	0.81	0.77	0.83	0.87	0.89	0.91
	Phi	0.55	0.64	0.71	0.75	0.78	0.58	0.64	0.68	0.71	0.73
60	G	0.55	0.65	0.71	0.76	0.79	0.74	0.81	0.85	0.88	0.90
	Phi	0.50	0.60	0.67	0.71	0.75	0.54	0.61	0.65	0.68	0.70
50	G	0.52	0.62	0.68	0.73	0.77	0.70	0.78	0.83	0.86	0.88
	Phi	0.46	0.56	0.63	0.68	0.72	0.49	0.56	0.61	0.63	0.66

Tablo 35'te okuma becerileri Kitapçık 8'de yer alan 8 maddenin, çapraz desende 100 öğrencinin 4 puanlayıcı tarafından puanlanmasıyla elde edilen G katsayısı 0,71 ve Phi katsayısı 0,68 olarak hesaplanmıştır. Bu Genellenebilirlik düzeyi kabul edilebilir düzey olan 0,80'nin oldukça altındadır. 100 öğrenci için yuvalanmış desende 4 puanlayıcı tarafından yapılan puanlamada ise G katsayısı 0,89 ve Phi katsayısı 0,72 olarak hesaplanmıştır.

100 öğrenci için yapılan puanlamada, çapraz desende Puanlayıcı sayısını 2'ye

indirdiğimizde G katsayısı 0,56 ve Phi Katsayısı 0,52'ye kadar düşmektedir. Yuvalanmış desende elde edilen G katsayısı ise 0,80 ve Phi katsayısı 0,61'dir. Öğrenci sayılarını aşamalı olarak 100'den 50'ye kadar indirdiğimizde her iki desende de 4 puanlayıcı için G ve Phi katsayıları bir miktar düşüş göstermektedir. Çapraz desende 4 puanlayıcı 50 öğrenciyi puanladığında G katsayısı 0,71'den 0,68'e, Phi katsayısı 0,68'den 0,63'e düşmektedir. Yuvalanmış desende 50 öğrenci 4 puanlayıcı tarafından puanlandığında, G katsayısı 0,89'dan 0,83'e, Phi katsayısı ise 0,72'den 0,61'e düşmektedir.

3.5. HER BİR KİTAPÇIKTA KABUL EDİLEBİLİR DÜZEYDE BİR GENELLENEBİLİRLİK SEVİYESİ ELDE ETMEK İÇİN GEREKLİ MİNİMUM SORU SAYISI

PISA 2009 okuma becerileri testinde, birden fazla puanlayıcı tarafından puanlama yapılan soru sayısı her kitap için birbirinden farklılık göstermektedir. Kitapçık 4 ve 5'te 5'er soru, Kitapçık 2 ve 12'de 6'şar soru, Kitapçık 6 ve 7'de 7'şer soru ve Kitapçık 8'de 8 soru yer almaktadır.

Kitapçıklarda birden fazla puanlayıcı tarafından puanlama yapılan soruların sayısında bir değişikliğe gidildiğinde Genellenebilirlik katsayısının ne oranda değiştiğine ve kabul edilebilir düzeyde bir Genellenebilirlik elde etmek için en az kaç soruya ihtiyaç olduğunu gösteren bilgilere aşağıda yer verilmiştir.

Kitapçık 2'de, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 36'da verilmiştir.

Tablo 36. Kitapçık 2’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4	
	Düzyev	Evren	Düzyev	Evren	Düzyev	Evren	Düzyev	Evren	Düzyev	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	6	INF	2	INF	3	INF	4	INF	5	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF	4	INF
Gözlem	2400		800		1200		1600		2000	
G Katsayısı	0.89		0.75		0.81		0.85		0.87	
Phi Katsayısı	0.81		0.60		0.69		0.75		0.78	
Bağıl Hata V.	0.01302		0.03560		0.02431		0.01866		0.01528	
Ölçm. Std. H.	0.11410		0.18867		0.15591		0.13661		0.12360	
Mutlak H. V.	0.02455		0.07020		0.04738		0.03597		0.02912	
Ölç. Std. H.	0.15670		0.26496		0.21767		0.18965		0.17065	

Kitapçık 2’de birden fazla puanlayıcı tarafından puanlanan 6 soru yer almaktadır. 6 sorunun 4 puanlayıcı tarafından puanlanmasından elde edilen Genellenebilirlik katsayısı 0,89, Phi katsayısı 0,81 iken, soru sayısı 5’e düşürüldüğünde Genellenebilirlik katsayısı 0,87’ye, Phi katsayısı 0,78’e düşmektedir. Soru sayısı 4 iken G katsayısı 0,85, Phi katsayısı 0,75 olmaktadır. Soru sayısı 3’e indirildiğinde, Genellenebilirlik katsayısı 0,81’e, Phi katsayısı 0,69’a düşmektedir. G katsayısını dikkate aldığımızda 4 soru kabul edilebilir en alt sınır olarak değerlendirilebilir.

Kitapçık 4’te, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 37’de verilmiştir.

Tablo 37. Kitapçık 4'te, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF
Soru	5	INF	2	INF	3	INF	4	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF
Gözlem		2000		800		1200		1600
G Katsayısı		0.90		0.78		0.84		0.88
Phi Katsayısı		0.88		0.75		0.82		0.86
Bağıl Hata V.		0.01050		0.02625		0.01750		0.01312
Ölçm. Std. H.		0.10247		0.16201		0.13228		0.11456
Mutlak H. V.		0.01239		0.03097		0.02065		0.01548
Ölç. Std. H.		0.11130		0.17598		0.14369		0.12444

Kitapçık 4'te birden fazla puanlayıcı tarafından puanlama yapılan 5 soru bulunmaktadır. 5 sorunun 4 puanlayıcı tarafından puanlanmasından elde edilen Genellenebilirlik katsayısı 0.90, Phi katsayısı 0,88'dir. Soru sayısı 4'e düşürüldüğünde, G katsayısı 0,88'e, Phi katsayısı 0,86'ya düşmektedir. Soru sayısını 3'e indirdiğimizde, G katsayısı 0,84, Phi katsayısı 0,82 olmaktadır. 3 soru için elde edilen Genellenebilirlik ve Phi katsayıları, kabul edilebilir bir düzeydedir.

Kitapçık 5'te, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 38'de verilmiştir.

Tablo 38. Kitapçık 5'te, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF
Soru	5	INF	2	INF	3	INF	4	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF
Gözlem		2000		800		1200		1600
G Katsayısı		0.91		0.80		0.86		0.89
Phi Katsayısı		0.85		0.69		0.77		0.82
Bağıl Hata V.		0.00718		0.01794		0.01196		0.00897
Ölçm. Std. H.		0.08471		0.13393		0.10935		0.09470
Mutlak H. V.		0.01300		0.03251		0.02167		0.01626
Ölç. Std. H.		0.11404		0.18031		0.14722		0.12750

Kitapçık 5'te birden fazla puanlayıcı tarafından puanlama yapılan 5 soru vardır. 5 sorunun 4 puanlayıcıyla puanlanmasından elde edilen G katsayısı 0,91, Phi katsayısı 0,85'tir. Soru sayısını 4'e düşürdüğümüzde, G katsayısı 0,89'a, Phi katsayısı 0,82'ye, soru sayısını 3'e düşürdüğümüzde, G katsayısı 0,86'ya, Phi katsayısı 0,77'ye düşmektedir. Soru sayısı 2'ye indirildiğinde Genellenebilirlik katsayısı kabul edilebilir bir düzey olan 0,80'de kalmakta, Phi katsayısı 0,69'a düşmektedir. Kabul edilebilir G katsayısı için en az 2 soru gerekmektedir. Phi katsayısı dikkate alındığında en az 4 soruya ihtiyaç vardır.

Kitapçık 6'da, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 39'da verilmiştir.

Tablo 39. Kitapçık 6'da, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düz.	Evr	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.
Oğrenci	100	INF	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	7	INF	2	INF	3	INF	4	INF	5	INF	6	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
Gözlem	2800		800		1200		1600		2000		2400	
G Katsayısı	0.89		0.70		0.77		0.82		0.85		0.87	
Phi Katsayısı	0.84		0.59		0.69		0.74		0.78		0.81	
Bağıl Hata V.	0.00643		0.02251		0.01501		0.01125		0.00900		0.00750	
Ölçm. Std. H.	0.08019		0.15003		0.12250		0.10609		0.09489		0.08662	
Mutlak H. V.	0.01017		0.03558		0.02372		0.01779		0.01423		0.01186	
Ölç. Std. H.	0.10083		0.18863		0.15401		0.13338		0.11930		0.10890	

Kitapçık 6'da, birden fazla puanlayıcı tarafından puanlama yapılan 7 soru bulunmaktadır. 7 sorunun 4 puanlayıcı tarafından puanlanmasından elde edilen Genellenebilirlik katsayısı 0,89, Phi katsayısı 0,84'tür. Soru sayısı 6'ya düşürüldüğünde, G katsayısı 0,87'ye, Phi katsayısı 0,81'e düşmektedir. Soru sayısı 5'e düşürüldüğünde Genellenebilirlik katsayısı 0,85, Phi katsayısı 0,78 olmaktadır. Soru sayısı 4 olduğunda, Genellenebilirlik katsayısı 0,82, Phi katsayısı 0,74 olmaktadır. Soru sayısı 3'e düşürüldüğünde Genellenebilirlik katsayısı kabul edilebilir düzeyin altına düşmektedir. Dolayısıyla 4 soru kabul edilebilir düzeyde Genellenebilirlik katsayısı için yeterlidir. Ancak Phi katsayısı da dikkate alındığında en az 5 soruya ihtiyaç vardır.

Kitapçık 7'de, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 40'ta verilmiştir.

Tablo 40. Kitapçık 7’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	7	INF	2	INF	3	INF	4	INF	5	INF	6	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
Gözlem	2800		800		1200		1600		2000		2400	
G Katsayısı	0.92		0.77		0.84		0.87		0.90		0.91	
Phi Katsayısı	0.88		0.68		0.76		0.81		0.84		0.86	
Bağıl Hata V.	0.00962		0.03367		0.02244		0.01683		0.01347		0.01122	
Ölçm. Std. H.	0.09808		0.18348		0.14981		0.12974		0.11605		0.10593	
Mutlak H. V.	0.01593		0.05574		0.03716		0.02787		0.02230		0.01858	
Ölç. Std. H.	0.12620		0.23609		0.19277		0.16694		0.14932		0.13631	

Kitapçık 7’de birden fazla puanlayıcı tarafından puanlama yapılan 7 soru yer almaktadır. 7 sorunun 4 puanlayıcı tarafından puanlanmasıyla elde edilen Genellenebilirlik katsayısı 0,92, Phi katsayısı 0,88’dir. Soru sayısının 6’ya düşürülmesiyle, Genellenebilirlik katsayısı 0,91’e, Phi katsayısı 0,86’ya düşmektedir. Soru sayısının 5 olması durumunda, G katsayısı 0,90, Phi katsayısı 0,84 olmaktadır. Soru sayısının 4’e düşürülmesiyle G katsayısı 0,87, Phi katsayısı 0,81 olmaktadır. Soru sayısının 3’e düşürülmesiyle Genellenebilirlik katsayısı 0,84, Phi katsayısı 0,76 olmaktadır. 3 soru ile Genellenebilirlik kabul edilebilir bir düzeyde kalmaktadır.

Kitapçık 8’de, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum soru sayısı Tablo 41’de verilmiştir.

Tablo 41. Kitapçık 8’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4		Seçenek 5	
	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.	Düz.	Evr.
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	8	INF	3	INF	4	INF	5	INF	6	INF	7	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF	4	INF	4	INF
Gözlem	3200		1200		1600		2000		2400		2800	
G Katsayısı	0.71		0.66		0.68		0.69		0.70		0.71	
Phi Katsayısı	0.68		0.63		0.65		0.66		0.67		0.68	
Bağıl Hata V.	0.03635		0.04726		0.04290		0.04028		0.03853		0.03729	
Ölçm. Std. H.	0.19066		0.21739		0.20711		0.20070		0.19630		0.19310	
Mutlak H. V.	0.04239		0.05427		0.04952		0.04667		0.04477		0.04341	
Ölç. Std. H.	0.20589		0.23296		0.22253		0.21603		0.21158		0.20835	

Kitapçık 8’de birden fazla puanlayıcı tarafından puanlama yapılan 8 soru bulunmaktadır. 8 sorunun 4 puanlayıcı tarafından puanlanmasından elde edilen Genellenebilirlik katsayısı 0,71, Phi katsayısı 0,68’dir ve bu durumda G ve Phi katsayıları kabul edilebilir düzeyin altında yer almaktadır. Madde sayısı azaldıkça Genellenebilirlik ve Phi katsayıları da düşmektedir. Dolayısıyla 8. Kitapçıkta madde sayısında bir değişiklik yapmak söz konusu değildir.

Kitapçık 12’de, kabul edilebilir düzeyde bir Genellenebilirlik Katsayısı için gerekli minimum madde sayısı Tablo 42’de verilmiştir.

Tablo 42. Kitapçık 12’de, Kabul Edilebilir Düzeyde Bir Genellenebilirlik Katsayısı İçin Gerekli Minimum Soru Sayısı

	G-Çalışması		Seçenek 1		Seçenek 2		Seçenek 3		Seçenek 4	
	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren	Düzy	Evren
Öğrenci	100	INF	100	INF	100	INF	100	INF	100	INF
Soru	6	INF	2	INF	3	INF	4	INF	5	INF
Puanlayıcı	4	INF	4	INF	4	INF	4	INF	4	INF
Gözlem	2400		800		1200		1600		2000	
G Katsayısı	0.79		0.67		0.73		0.76		0.78	
Phi Katsayısı	0.74		0.62		0.68		0.71		0.73	
Bağıl Hata V.	0.01342		0.02602		0.01972		0.01657		0.01468	
Ölçm. Std. H.	0.11585		0.16132		0.14044		0.12873		0.12117	
Mutlak H. V.	0.01792		0.03171		0.02482		0.02137		0.01930	
Ölç. Std. H.	0.13387		0.17808		0.15753		0.14618		0.13893	

Kitapçık 12’de birden fazla puanlayıcı tarafından puanlama yapılan 6 soru bulunmaktadır. 6 sorunun 4 puanlayıcı tarafından puanlanmasıyla elde edilen Genellenebilirlik katsayısı 0,79, Phi katsayısı 0,74’tür. Bu Genellenebilirlik düzeyi, kabul edilebilir düzeyin biraz altında yer almaktadır, dolayısıyla bunu kabul edilebilir alt limit olarak alabiliriz. Ancak soru sayısı azaldıkça Genellenebilirlik ve Phi katsayılarının değeri de düşmektedir, dolayısıyla bu kitapçıkta kabul edilebilir Genellenebilirlik düzeyi için en az 6 sorunun kullanılması gerekmektedir.

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu araştırmada, Genellenebilirlik Kuramına göre PISA 2009 açık uçlu soruların puanlanmasında öğrencilerin birden fazla puanlayıcı tarafından, birlikte ve dönüşümlü olarak puanlamalarıyla oluşturulan farklı desenlerin Genellenebilirlik (G) ve Karar (K) çalışmaları sonuçları ele alınmıştır.

Araştırmadan elde edilen bulgular doğrultusunda sonuç ve önerilere aşağıda yer verilmiştir.

4.1. SONUÇLAR

PISA 2009 uygulamasında, okuma becerilerinin yer aldığı kitapçıklar 2, 4, 5, 6, 7, 8 ve 12 nolu kitapçıklardır. Bu kitapçıklarda Ö X S X P çapraz deseninin Genellenebilirlik çalışması sonuçları:

1. Öğrencilerin bütün kitapçıklarda, okuma becerileri seviyesi bakımından farklılaştığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranlarının, Kitapçık 2'de % 16, Kitapçık 4'te %27,7, Kitapçık 5'te %21,9, Kitapçık 6'da %15,4, Kitapçık 7'de %20,6, kitapçık 8'de %19,8, kitapçık 12'de %19,9 olduğu görülmektedir.
2. Bütün kitapçıklarda, soru değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama düzeyi %0'dır. Bu durumda kitapçık bazında bakıldığında soruların zorluk kolaylık derecesinin birbirinden farklı olmadığı görülmektedir.
3. Puanlayıcı değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama düzeyi, 2, 4, 5, 6 ve 7 nolu kitapçıklar için %0 iken, kitapçık 8'de %4,8, kitapçık 12'de %6'dır. PISA uygulamasında 2, 8 ve 12 nolu

kitapçıklarda puanlama sistemi tam puan 2, kısmi puan 1, yanlış/ilgisiz yanıt 0 olarak belirlenmiştir. Dolayısıyla bu kitapçıklar diğer kitapçıklarda yer alan tam puan 1, yanlış/ilgisiz yanıt 0'dan daha aşamalı bir puanlama rehberine sahiptir. Bu durumun sonucu etkilediği düşünülmektedir.

4. Öğrenci ve soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama düzeyinin bütün kitapçıklarda %0 olduğu, dolayısıyla öğrencilerin performanslarının sorudan soruya değişmediği görülmüştür.
5. Öğrenci ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama düzeyi 2, 4, 5, 6 ve 7 nolu kitapçık için %0 iken, kitapçık 8'de %26, kitapçık 12'de %10,9 olduğu görülmektedir. Daha önce de ifade edildiği gibi 8. ve 12. kitapçıklarda puanlama sistemi tam puan 2, kısmi puan 1, yanlış/ilgisiz yanıt 0 olarak belirlenmiştir. Bu durumun sonucu etkilediği düşünülmektedir.
6. Soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama düzeylerine bakıldığında, neredeyse bütün kitapçıklarda yüksek bir yüzdeyle karşılaşılmaktadır. Kitapçık 2'de bu oran %41,9, kitapçık 4'te %11, kitapçık 5'te %35, kitapçık 6'da %31, kitapçık 7'de %31,4, kitapçık 8'de %3,7 ve kitapçık 12'de %5,4'dür. Puanlayıcıların soruları puanlamada farklılaştığı görülmüştür.
7. Öğrenci, soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdelerinin oldukça yüksek olduğu görülmektedir. Kitapçık 2'de %41, kitapçık 4'te %61,3, kitapçık 5'te %43,1, kitapçık 6'da %53,5, kitapçık 7'de %47,9, kitapçık 8'de %45,7 ve kitapçık 12'de %57,8'dir. Bu oranların yüksek olması, puanlayıcı ve soru ortak etkisinin yüksek olması ya da tesadüfi hata kaynaklarının yüksek olması anlamına gelmektedir ya da her ikisi de söz konusudur.

Ö X S X P Çapraz Desende Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışması Sonuçları:

1. Kitapçık 2, 4, 5, 6 ve 7'de, öncelikle öğrenci sayıları sabit tutulup, puanlayıcı sayıları artırılıp azaltıldığında G katsayısının değişimine bakılmıştır. Puanlayıcı sayısının artırılmasının G katsayısında bir artışa sebep olduğu, ancak bu artışın küçük bir miktar olduğu, dolayısıyla puanlayıcı sayısı 4'ten 2'ye düşürülse de G katsayısının istenilen düzeylerde (0,80 ve üzeri) kaldığı görülmektedir.
2. Kitapçık 2, 4, 5, 6 ve 7'de, öğrenci sayılarının 100'den 40'a aşamalı olarak düşürülmesinde G katsayısının çok etkilenmediği, istenilen düzeylerde kaldığı görülmektedir.
3. Kitapçık 8 ve 12'de durum, kitapçık 2, 4, 5, 6 ve 7'den farklılık göstermektedir. 8. Kitapçıkta istenilen G katsayısı düzeyi 4 puanlayıcıyla sağlanamamaktadır, en az 6 puanlayıcıya ihtiyaç duyulmaktadır. Öğrenci sayısının 100'den 40'a aşamalı olarak düşürülmesi G katsayısında büyük değişikliklere sebep olmamakta asıl farkı puanlayıcı sayısının artırılması ya da azaltılması yaratmaktadır.

Kitapçık 12'de, istenilen düzeyde G katsayısı elde edebilmek için puanlayıcı sayısının en az 4 olması gerekmektedir. Puanlayıcı sayısının artırılması G katsayısını artırmaktadır, puanlayıcı sayısının 4'ten daha az olması durumunda G katsayısı istenilen düzeyin altına düşmektedir. Öğrenci sayısının 100'den aşamalı olarak 40'a indirilmesi G katsayısında önemli bir değişikliğe sebep olmamaktadır.

(Ö:P) X S Desenin Genellenabilirlik Sonuçları:

1. Öğrencilerin, okuma becerileri seviyesi bakımından farklılaştığı görülmektedir. Öğrencilere ait varyans bileşeninin toplam varyansı açıklama oranları, Kitapçık 2'de % 16,8, kitapçık 8'de %15,5'tir.
2. Soru değişkeni için kestirilen varyans bileşeninin toplam varyansı açıklama düzeyi 2. kitapçıkta %8,2'dir. Bu durum soruların zorluk düzeyi açısından birbirinden farklılık gösterdiğinin bir göstergesidir. Kitapçık 8 için bu değer %1,1'dir. Soruların zorluk ve kolaylık düzeyleri 8. Kitapçıkta %1 oranında farklılık göstermektedir.
3. Öğrenci değişkeniyle, puanlayıcı değişkeninin yuvalandığı (Ö:P) deseninde kestirilen varyans bileşeninin toplam varyansı açıklama düzeyi, kitapçık 2 için %0,3 iken, kitapçık 8 için %36,5'tir. Bu değer büyük olması, öğrenci puanlayıcı etkileşiminin farklılaştığı, puanlayıcıların yaptığı puanlamanın bir öğrenciden diğerine farklılık gösterdiği şeklinde yorumlanabilir.
4. Öğrenci ve soru ortak etkisine ait varyans bileşeninin toplam varyansı açıklama düzeyinin her iki kitapçıkta %0 olduğu, dolayısıyla öğrencilerin performanslarının sorudan soruya değişmediği görülmüştür.
5. Öğrenci, soru ve puanlayıcı ortak etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdelerinin oldukça yüksek olduğu görülmektedir. Kitapçık 2'de %74,7, kitapçık 8'de %46,9'dur. Bu oranların yüksek olması, puanlayıcı ve soru ortak etkisinin yüksek olması ya da tesadüfi hata kaynaklarının yüksek olması anlamına gelmektedir ya da her ikisi de söz konusudur.

(Ö:P) X S Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışması Sonuçları:

1. (Ö:P) X S desenine göre, Kitapçık 2 ve 8'de, öncelikle öğrenci sayıları sabit tutulup, puanlayıcı sayıları artırılıp azaltıldığında G katsayısının değişimine bakılmıştır. Her iki kitapçıkta da, puanlayıcı sayısının artırılmasının G katsayısında bir artışa sebep olduğu, ancak bu artışın küçük bir miktar olduğu, puanlayıcı sayısı 4'ten 2'ye düşürüldüğünde de G katsayısındaki düşüşün çok küçük miktarda olduğu ve puanlayıcı sayısı 4'ten 2'ye düşürülse de G katsayısının istenilen düzeylerde (0,80 ve üzeri) kaldığı görülmektedir.
2. (Ö:P) X S desenine göre Kitapçık 2'de, öğrenci sayılarının 186'dan 100'e aşamalı olarak düşürülmesiyle elde edilen G katsayısının öğrenci sayısının düşürülmesinden etkilenmediği ve puanlayıcı sayısı da 4'den 2'ye düşürüldüğünde de istenilen düzeylerde kaldığı görülmektedir.
3. (Ö:P) X S desenine göre Kitapçık 8'de, öğrenci sayılarının 186'dan 100'e aşamalı olarak düşürülmesiyle elde edilen G katsayısının öğrenci sayısının düşürülmesinden etkilendiği ve 186 öğrencinin 4 puanlayıcıyla puanlandığı durumda G katsayısının 0,95'den, 100 öğrencinin 4 puanlayıcı tarafından puanlandığı durumda G katsayısının 0,89'a düştüğü görülmektedir. Ancak asıl düşüş, öğrenci sayısı 100'e ve puanlayıcı sayısı 2'ye indirildiğinde yaşanmaktadır. Bu durumda G katsayısı 186 öğrenci 4 puanlayıcı için 0,95'ten 0,80'e ve Phi katsayısı 0,85'ten 0,61'e düşmektedir.

Ö X S X P ve (Ö:P) X S Desenlerinden Elde Edilen G Çalışması Sonuçları:

1. Kitapçık 2 için Ö X S X P ve (Ö:P) X S desenlerinden elde edilen G çalışması sonuçları incelendiğinde, her iki desende de öğrencilerin

okuma becerileri bakımından farklılaştığı görülmektedir. Ö X S X P deseninde sorular kolaylık zorluk açısından farklılaşmazken, (Ö:P) X S deseninde soru değişkenine ait varyansın toplam varyansın %6,5'ini açıkladığı, dolayısıyla farklılaştığı görülmektedir. Her iki desende de puanlayıcı etkisine ait varyans bileşeninin oldukça küçük olduğu dolayısıyla puanlayıcıların her iki desende de öğrencileri tutarlı puanladığı görülmüştür. Her iki desende de artık varyans oldukça yüksek çıkmıştır.

2. Kitapçık 8 için Ö X S X P ve (Ö:P) X S desenlerinden elde edilen G çalışması sonuçları incelendiğinde, her iki desende de öğrencilerin okuma becerileri bakımından farklılaştığı görülmektedir. Her iki desende de sorular kolaylık zorluk açısından farklılaşmamaktadır. Her iki desende de öğrenci puanlayıcı ortak etkisinin oldukça yüksek olduğu, Ö X S X P deseninde ÖXP ortak etkisinin toplam varyansı açıklama yüzdesi % 26, (Ö:P) ortak etkisinin toplam varyansı açıklama yüzdesi %40,9'dur. Bu durumda, öğrenci puanlayıcı etkileşiminin farklılaştığı, puanlayıcıların yaptığı puanlamanın bir öğrenciden diğerine farklılık gösterdiği söylenebilir.

Ö X S X P ve (Ö:P) X S Desenlerinden Elde Edilen K Çalışması Sonuçları:

1. Ö X S X P ve (Ö:P) X S desenlerinden Kitapçık 2 için elde edilen K çalışması sonuçlarına bakıldığında, Puanlayıcı sayısı 4 ile sabit tutulup öğrenci sayısı kademeli olarak 100'den 50'ye indirildiğinde, G katsayısında bir değişiklik olmadığı, neredeyse sabit kaldığı görülmektedir. Öğrenci sayısı sabit tutulup puanlayıcı sayısı artırıldığında G katsayısında bir artış, puanlayıcı sayısı azaltıldığında da bir azalma söz konusudur. Ancak değişim miktarı oldukça küçüktür ve G katsayısı her durumda 0,80 seviyesinin üzerinde kalmaktadır.

2. Ö X S X P ve (Ö:P) X S desenlerinden Kitapçık 8 için elde edilen K çalışması sonuçlarına bakıldığında, Ö X S X P deseninde 100 öğrenci ve 6 puanlayıcı kullanıldığında istenilen G katsayısı seviyesine ulaşılabildiği, öğrenci sayısı ve puanlayıcı sayısı azaldıkça G ve Phi katsayılarının oldukça düştüğü ve kabul edilebilir sınırların altında kaldığı görülmektedir. (Ö:P) X S deseninde 100 öğrencinin 2, 3, 4, 5 ve 6 puanlayıcı tarafından puanlandığı durumlarda G katsayısı 0,80'in üzerinde yer alırken Phi katsayısı her bir durum için 0,80'nin altında kalmaktadır. Öğrenci sayısı aşamalı olarak 100'den 50'ye düşürüldüğünde 3, 4, 5 ve 6 puanlayıcı için Genellenebilirlik düzeyi 0,80'in üzerinde yer alırken 2 puanlayıcı için kabul edilebilir düzeyin altına inmektedir. Ancak (Ö:P) X S deseninde bütün durumlar için Phi katsayısı kabul edilebilir düzeyin çok altındadır. Çünkü Ö:P etkileşimi çok yüksektir, bu durum da ölçmeye hata varyansı olarak yansımaktadır.

Her Bir Kitapçıkta İstenen Genellenebilirlik Seviyesini Elde Etmek İçin Gerekli Minimum Soru Sayısı:

1. Kitapçık 2, 4, 5, 6 ve 7'de, madde sayısını neredeyse yarıya indirdiğimizde bile istenilen Genellenebilirlik katsayısı elde edilebilmektedir.
- Kitapçık 2'de, madde sayısı 6'dan 3'e düşürüldüğünde G katsayısı 0,81,
 - Kitapçık 4'te, madde sayısı 5'ten 3'e düşürüldüğünde G katsayısı 0,84,
 - Kitapçık 5'te, madde sayısı 5'ten 2'ye düşürüldüğünde G katsayısı 0,80,
 - Kitapçık 6'da, madde sayısı 7'den 4'e düşürüldüğünde G katsayısı 0,82,
 - Kitapçık 7'de, madde sayısı 7'den 3'e düşürüldüğünde G katsayısı 0,82 olarak hesaplanmaktadır.

2. Kitapçık 8 ve 12'de madde sayılarının değiştirilmesi durumunda G katsayısı istenilen düzeyin altına düşmesi nedeniyle, kitapçık 8'de 8 sorunun, kitapçık 12'de 6 sorunun sabit kalmasının uygun olacağı düşünülmektedir.

Elde edilen sonuçları kısaca özetlersek;

Öncelikle kitapçıkları 1-0 olarak puanlanan ve 2-1-0 olarak kısmi puanlama yapılan kitapçıklar diye iki gruba ayırmanın yararlı olacağı düşünülmektedir. Puanlamanın 1-0 olarak yapıldığı 4, 5, 6, 7 nolu kitapçıkların tamamında öğrencilerin Okuma Becerileri açısından farklılaştığı, soruların kolaylık ve zorluk açısından farklılaşmadığı, öğrenci performanslarının sorudan soruya farklılaşmadığı, puanlayıcıların birbirleriyle tutarlı puanlama yaptığı ve öğrenci puanlayıcı etkileşiminin çok düşük düzeyde olduğu, puanlayıcıların puanlama yaparken öğrenciden öğrenciye farklı puanlama yapmadığı görülmüştür. Ancak, soru puanlayıcı ortak etkisi oldukça yüksektir. Bu durumda puanlayıcıların sorudan soruya farklı puanlama yaptığı söylenebilir.

Puanlama işleminin kısmi puanın dahil edilerek 2-1-0 olarak yapıldığı kitapçıklar ise 2, 8 ve 12 nolu kitapçıklardır. Kitapçık 2'de, öğrencilerin Okuma Becerileri açısından farklılaştığı, soruların kolaylık ve zorluk açısından farklılaşmadığı, öğrenci performanslarının sorudan soruya farklılaşmadığı, puanlayıcıların birbirleriyle tutarlı puanlama yaptığı ve öğrenci puanlayıcı etkileşiminin çok düşük düzeyde olduğu, puanlayıcıların puanlama yaparken öğrenciden öğrenciye farklı puanlama yapmadığı görülmüştür. Ancak, soru puanlayıcı ortak etkisi oldukça yüksektir. Bu durumda puanlayıcıların sorudan soruya farklı puanlama yaptığı söylenebilir.

Kitapçık 8 ve Kitapçık 12'de puanlayıcıların birbirinden farklılaştığı, öğrenci puanlayıcı ortak etkisinin toplam varyansı açıklama oranının 8. Kitapçık için %26 ve 12. Kitapçık için %10,9 olduğu görülmektedir. Dolayısıyla puanlayıcılar öğrenciden öğrenciye farklı puanlama yapmışlardır. Bu durum 8. Kitapçıkta

Genellenebilirlik katsayısında kendini göstermiş ve Genellenebilirlik düzeyi 0,71'de kalmıştır. 12 kitapçıkta Genellenebilirlik düzeyi 0,79 ile kabul edilebilir düzeydedir denilebilir. Kitapçık 2'de Genellenebilirlik düzeyi 0,89'dur.

Ancak, bütün kitapçıklarda Öğrenci Soru Puanlayıcı ortak etkisine ait varyans bileşenin toplam varyansı açıklama yüzdesi oldukça yüksektir. Üç değişkenlik kaynağının etkileşimden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın yüksek olması öğrenci, soru, puanlayıcı etkileşiminin ve/veya tesadüfi hata kaynaklarının büyük olabileceğinin bir göstergesi olabilir.

Bu sonuçlara bakıldığında kısmi puanlama yapılan kitapçıklardan 8 ve 12 nolu kitapçıkların Genellenebilirlik katsayılarının diğer bütün kitapçıklardan daha düşük düzeyde olduğu söylenebilir. Puan kategorisinin 2'den 3'e çıkmış olması bu durumda etkili olabileceği gibi, puanlama rehberinde puanlayıcılar tarafından net anlaşılmayan ve kendi yorumlamalarını ilave etmelerine neden olan durumlar olabilir. Ayrıca bu kitapçıkta yer alan sekiz sorudan bazıları aynı metinle ilgili sorular olduğu için puanlayıcılar cevapları farkında olmadan kendi içinde gruplandırmış olabilirler. Diğer olası sebeplerin araştırılmasında fayda olduğu düşünülmektedir.

Ö X S X P ve (Ö:P) X S desenleri karşılaştırıldığında, (Ö:P) X S deseni ile kestirilen bağıl ve mutlak hata varyanslarının Ö X S X P desenine göre daha küçük olduğu, dolayısıyla G ve Phi katsayılarının daha büyük değerler aldığı görülmektedir. Bu iki desende yapılan Karar çalışmaları incelendiğinde, her iki desende de puanlayıcı sayısını artırmanın G ve Phi katsayılarında artış sağladığı, ancak puanlayıcı sayısını artırarak harcanacak işgücü ve ekonomiklik bakımından G katsayısındaki artışın bir avantaj sağlayacak büyüklükte olmadığı görülmüştür. Kitapçık 8 ve 12 hariç, kitapçıklarda puanlayıcı sayısını yarıya indirerek ya da madde sayısını yarıya indirerek, çoğu zaman her ikisinin sayısını da yarıya indirerek kabul edilebilir düzeylerde G katsayısına ulaşmanın mümkün olduğu görülmektedir. Madde sayılarının 8 ve 12. Kitapçıklar hariç diğer bütün kitapçıklarda yarıya indirilmesi istenilen düzeyde bir Genellenebilirlik

katsayısı elde etmeyi mümkün kılmaktadır. Bu nedenle, puanlamanın 1-0 şeklinde yapıldığı kitapçıklarda ve 2. kitapçıkta, soru sayısını ve puanlayıcı sayısını yarıya indirip, kısmi puanlama yapılan kitapçıklardan 8 ve 12 nolu kitapçıklarda madde sayısını sabit tutup puanlayıcı sayısının artırılmasında yarar olduğu sonucuna ulaşılmıştır.

4.2. ÖNERİLER

1. Öğrenci sayısının çok olduğu ve birden fazla puanlayıcı tarafından puanlama yapılması gereken performans belirleme sınavlarında, puanlayıcılar arası tutarlık sağlandığında, her öğrencinin puanlama yapan bütün puanlayıcılar tarafından puanlandığı desenler yerine, puanlayıcıların bir kısmının, öğrencilerin bir kısmını dönüşümlü olarak puanladığı puanlama desenlerinin kullanılması, hem zaman hem iş gücü açısından daha ekonomik olacaktır.
2. 1-0 şeklinde puanlama yapılan kitapçıklarda, birden fazla puanlayıcı tarafından puanlanan öğrenci sayısı yarıya indirilerek ya da madde sayısı yarıya indirilerek ya da puanlayıcı sayısı yarıya indirilerek ya da bunların bir kombinasyonundan elde edilecek desenlerin kullanılması ve diğer soruların tek puanlayıcı desenine göre puanlanması zaman ve iş gücü açısından daha ekonomiktir.
3. 1-0 şeklinde puanlanan kitapçıklarda puanlayıcı tutarlılığı açısından bir sorun olmadığı ve Genellenebilirlik düzeylerinin oldukça yüksek olduğu görülmüştür. 2-1-0 şeklinde kısmi puanlama yapılan kitapçıklarda puanlayıcı tutarlılıklarında ve Genellenebilirlik katsayılarında bir miktar düşüş olduğu görülmüştür. Dolayısıyla, kısmi puanlama yapılan kitapçıklar için puanlayıcı eğitimi verilirken daha çok örnek çalışma yapılması, yerel örnek sayısının artırılması ve sorular puanlanırken her kitapçıkta belirli bir sorunun puanlanması bitmeden bir sonraki soruya geçilmediğinden emin olunması gerekmektedir. Bu, puanlama işleminin

aynı ünitenin içinde yer alan bütün maddelerin ayrı ayrı kodlanması ve bütün kitaplarda bu ünitelerin tamamlandıktan sonra yeni bir üniteye geçilmesi anlamına gelmektedir

4. Uluslararası çalışmalarda puanlayıcı olarak görev yapacak bir grubun çalışmalar başlamadan önce tanımlanması önem taşımaktadır. Ders dönemi içinde özellikle de yıl sonuna denk gelen bir dönemde, okullardan öğretmen görevlendirmesi yapmak motivasyonu düşüren bir durumdur. Öğretmenler yıl sonu itibarıyla okuldaki görevlerini de aksatmamak adına, okul çalışmaları ve puanlama çalışmaları arasında, zaman kazanma derdine düşmeden puanlama çalışmalarına ağırlık verebilmelidirler.
5. TIMSS, PIRLS, PISA, ICILS gibi uluslararası düzeyde katıldığımız performans değerlendirme sınavlarında farklı değişkenlik kaynakları (kitapçıklar, modüller vb..) ve farklı desenlerin karşılaştırıldığı çalışmalar yapılabilir.
6. TIMSS, PISA gibi uluslararası performans değerlendirme sınavlarında, matematik okuryazarlığı ve fen okuryazarlığı gibi alt alanların tümünü ele alan, puanlama desenleri açısından aralarındaki ilişkileri inceleyen çalışmalar yapılabilir.
7. Bu çalışmada EduG programı kullanılmıştır. Genellenebilirlik kuramında kullanılan diğer programların (Genova, SPSS..) aynı sonuçları verip vermediğine ilişkin karşılaştırma çalışmaları yapılabilir.

KAYNAKÇA

- Alharby, E. R. (2006). *A Comparison Between Two Scoring Methods, Holistic vs. Analytic Using Two Measurement Models, The Generalizability Theory and The Many Facet Rasch Measurement Within The Context of Performance Assessment*. Yayınlanmamış Doktora Tezi, The Pennsylvania State University.
- Atılgan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi: Ankara.
- Atılgan, H. ve Tezbaşaran, A. A. (2005). Genellenebilirlik kuramı alternatif karar çalışmaları ile senaryolar ve gerçek durumlar için elde edilen G ve Phi katsayılarının tutarlılığının incelenmesi. *Eğitim Araştırmaları Dergisi*, 18, 28-40.
- Atılgan, H. (2005). Genellenebilirlik Kuramı ve Puanlayıcılar Arası Güvenirlik için Örnek Bir Uygulama. *Eğitim Bilimleri ve Uygulama*, 4 (7).
- Atılgan, H. (2008) Using Generalizability Theory to assess the Score reliability of the Special Ability Selection Examinations for Music Education Programmes in Higher Education. *International Journal of Research and Method Education*. 31(1) 63-76.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM.
- Bernardin, H. J., Villanova, P. (2005) Research Streams in Rater Self-efficacy. *Group and Organizational Management*, 30, 61-88

- Biemer, L. (1993). Authentic Assessment. *Educational Leadership*, 50 (8), 81-82
- Brennan, R. L. (1983). *Elements of generalizability theory*. Iowa City, IA. American College Testing.
- Brennan, R. L. (1992). *Elements of generalizability theory (rev. ed)*. Iowa City, IA. American College Testing.
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlog.
- Brennan, R. L. (2011). *Using Generalizability Theory to Address Reliability Issues for PARCC Assessments: A white Paper*
- Büyüköztürk, Ş., Çakmak, E., Akgün, Ö., Karadeniz, Ş., Demirel, F. (2010). *Bilimsel Araştırma Yöntemleri*. Ankara, PEGEM Akademi
- Calkins, D.S., Erlich, O., Marston, P.T., & Malitz, D. (1978). An empirical investigation of the distributions of generalizability coefficients and various estimates for an application of generalizability theory. *Paper presented at the Annual Meeting of the American Educational Research Association, Toronto.*
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Harcourt Brace Javanovich College Publishers, USA.
- Cronbach, L. J., Linn. R.L., Brennan, R.L. ve Haertel, E.H. (1995). *Generalizability analysis for educational assessment*. Evaluation Comment, Summer. Report by UCLA's Center for the Study of Evaluation and the National Center for Research on Evaluation, Standards and Student Testing.

- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Cronbach, L. J., Rajaratnam, N., Gleser, G. C. (1963). Theory of Generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.
- Darst, P.W., Zakrajsek, D.B., & Mancini, V.H. (Eds.) (1989). *Analyzing physical education and sport instruction* (2nd edition), Champaign, IL: Human Kinetics.
- EARGED (2005). *PISA 2003 Projesi, Ulusal Nihai Raporu*
- EARGED (2010). *PISA 2009 Projesi, Ulusal Ön Raporu.*, <http://earged.meb.gov.tr/pdf/pisa2009rapor.pdf>
- Erlish, O., Shavelson, R. (1976). The Application of Generalizability Theory to the Study of Teaching, Technical Report. *Beginning Teacher Evaluation Study*. 76-9-1, San Francisco, CA: Far West Laboratory for Educational Research and Development.
- Goodwin, L. D. (2001). Interrater Agreement and Reliability. *Measurement in Psychical Education and Exercises Science*, 5 (1), 13-34.
- Güler, N. (2008). *Klasik Test Kuramı, Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- Güler, N. ve Çetin, B. (2010). Genellenabilirlik Kuramı ve Klasik Test Kuramı Açısından Gözetmen Sayısı Etkisinin İncelenmesi. *Sözel Bildiri. II. Ulusal*

Eđitimde ve Psikolojide Ölçme ve Deđerlendirme Kongresi, Mersin: Türkiye.

Güler, N. (2011). Rasgele Veriler Üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramına Göre Güvenirliđin Karşılaştırılması. *Eđitim ve Bilim*, Cilt 36, Sayı 162

Hill, B.M. (1970). *Some contrasts between Bayesian and classical influence in the analysis of variance and in the testing of models*. In D.L. Meyer & R.O. Collier, Jr. (Eds.) *Bayesian statistics*. Itasca, IL: F.E. Peacock.

Hovardaođlu, S. (2007). *Davranış Bilimleri İçin Araştırma Teknikleri*. Ankara: Seçkin Yayıncılık, 2. Baskı

Kan, A. (2005). Yazılı Yoklamaların Puanlanmasında Puanlama Cetveli ve Yanıt Anahtarı Kullanımının Puanlayıcı Güvenirliđine Etkisi, *Eurasian Journal of Educational Research*. Sayı:19

Karasar, N. (1998). *Araştırmalarda Rapor Hazırlama Yöntemi*. Ankara: Pars Matbaacılık

Kieffer, K. M. (1998). Why Generalizability Theory is Essential and Classical Test Theory Often Inadaquate? *Paper presented at the annual meeting of the Southwestern Psychological Association*. New Orleans, LA. USA.

Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., Comfort, K., & Othman, A. R. (1998). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education*, 11, 121-138.

Kutlu, Ö. (2006). Üst düzey zihinsel süreçleri belirleme yolları: Yeni durum belirleme yaklaşımları, *Çađdaş Eđitim* Sayı 335

- Kutlu, Ö., Doğan, D., Karakaya, İ.(2009). *Öğrenci Başarılarının Belirlenmesi, Performansa ve Portfoyaya Dayalı Durum Belirleme*, Pegem Akademi
- Lee, Y., W., (2005) *Dependability of Scores for a New ESL Speaking Test: Evaluating Prototype Tasks*, ETS, Princeton, NJ
- Lee, G., Frisbie, D. A., (1999), Estimating Reliability Under a Generalizability Theory Model for Test Scores Composed of testlets. *Applied Measurement in Education*, 12(3), 237-255
- Mcbee, M., Barnes, L. (1998), The Generalizability of a Performance Assessment Measuring Achivement in Eighth-Grade Mathematics. *Applied Measurement in Education*, 11(2), 179-194, Lawrence Erlbaum Associates, Inc.
- Musquash, C. , O'connor, B. P. (2006). SPSS and SAS Programs for Generalizability Theory Analysis. *Behavior Research Methods*. 38 (3), 542-545
- Nalbantoğlu, F. (2009). *Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- OECD (2009) *Assessment Framework: Key Competencies in reading, matematics and science*. OECD Publications.
- Özçelik, D. A., (1992), *Ölçme ve Değerlendirme*, Ankara, ÖSYM Yayınları
- Sanders, N. M.(1996) *Classroom questions: What kinds?* Newyork, Harper &Row

- Schoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing* 2005 22 (1) 1-30
- Scullen SE, Mount MK, Goff M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85, 956–970.
- Sharma, F. ; Weathers, D. (2003), Assessing generalizability of scales used in cross-national research. *International Journal of Research in Marketing* 20 (2003) 287-295
- Smith, Teresa A. (1997). *Paper presented at the Annual Meeting of the American Educational Research Association* (Chicago, IL, March 24-28, 1997)
- Shavelson, J. R., Webb, N. M., Rowley, G. (1989). Generalizability Theory *American Psychologist* Vol. 44, No:6, 922-932
- Shavelson, J. R., Webb, N. M. (1991). *Generalizability Theory: A Primer*. Sage Publications, USA
- Tekindal, S., (1996) *Klasik Yazılı Sınavla ve Çok Sorulu Testle Elde Edilen Ölçümlerin Güvenirlik ve Geçerliği*; Ankara, Pegem
- Tobar, D. ; Stegner, A. ; Kane, M. (1999), The Use of Generalizability Theory in Examining the Dependability of Scores on the Profile of Mood States, *Measurement in Physical Education and Exercise Science*, 3(3), 141-156
- Turgut, F. M. (1992) *Eğitimde Ölçme ve Değerlendirme Metotları*. 9 baskı. Ankara: Saydam Matbaacılık.

Wexley, K. N. and Youtz, M. A. (1985). Rater Beliefs About Others: Their Effect on Rating Errors and Rater Accuracy. *Journal of Occupational Psychology*, 58, 265-275.

Webb, N.M, Rowley, G.L., & Shavelson, R.J. (1988). Using generalizability theory in counseling and development. *Measurement and Evaluation in Counseling and Development*, 21, 81-90.

Yelboğa, A., (2012) Genellenabilirlik Kuramına Göre İş performans Ölçeklerinde Güvenirlik, *Eğitim ve Bilim*, Cilt 37, sayı 163, 157-164

EK: 1 PISA OKUMA BECERİLERİ ÖRNEK SORULAR

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. Uluslararası 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	Şunları Yapmayınız
Aramaları kısa tutunuz.	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.
Cep telefonunuzu bekleme konumundayken vücudunuzdan uzakta tutunuz.	"SAR" değeri ¹ yüksek olan cep telefonu satın almayınız. SAR değerinin yüksek olması, telefonun daha fazla radyasyon yaydığı anlamına gelir.
Konuşma süresi uzun olan cep telefonları satın alın. Bu daha verimli olur ve daha az radyasyon yayar.	Bağımsız bir kuruluş tarafından test edilmeyen sözde koruyucu cihazları satın almayın.

¹ SAR (belirli emilim oranı) Bir cep telefonu kullanırken vücut dokuları tarafından emilen elektromanyetik dalgaların miktarını belirten bir ölçümdür.

Önceki iki sayfada yer alan metin bir web sitesinden alınmıştır. Aşağıdaki soruları yanıtlamak için bu metinden yararlanınız.

Soru 2: CEP TELEFONU GÜVENLİĞİ

R414Q02

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

- A Cep telefonu kullanmanın tehlikelerini tanımlamak
- B Cep telefonu güvenliğiyle ilgili tartışmaların halen devam ettiğini ileri sürmek
- C Cep telefonu kullananların almaları gereken önlemleri tanımlamak
- D Cep telefonları kaynaklı bilinen bir sağlık problemi olmadığını ileri sürmek

Kullanım amacı: Kamusal

Metnin biçimi: Bağımsız

Metnin türü: Açıklama

Metne yaklaşım: *Bilgileri bir araya getirme ve yorumlama:* Genel bir anlayış oluşturma

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 576 (Düzey 4)

CEP TELEFONU GÜVENLİĞİ 2

SORUNUN AMACI:

Bilgileri birleştirme ve yorumlama: Genel bir anlayış oluşturma

Metinde yer alan tablonun bir kısmının amacını ayırt etme.

Tam Puan

- B. Cep telefonu güvenliğiyle ilgili tartışmaların halen devam ettiğini ileri sürmek

Soru 11: CEP TELEFONU GÜVENLİĞİ

R414Q11

“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?

- A Bu bilgi, evet kısmında yer alan iddiayı destekler fakat bunu kanıtlamaz.
- B Bu bilgi, evet kısmında yer alan iddiayı kanıtlar.
- C Bu bilgi, hayır kısmında yer alan iddiayı destekler fakat bunu kanıtlamaz.
- D Bu bilgi, hayır kısmında yer alan iddianın yanlış olduğunu gösterir.

Kullanım amacı: Kamusal

Metnin biçimi: Bağımsız

Metnin türü: Açıklama

Metne yaklaşım: *Kendi düşüncelerini yansıtma ve metni değerlendirme:* Bir metnin içeriğini değerlendirme ve kendi düşüncelerini yansıtma

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 625 (Düzey 4)

CEP TELEFONU GÜVENLİĞİ PUANLAMA 11

SORUNUN AMACI:

Düşünme ve değerlendirme: Bir metnin içeriği üzerinde derinlemesine düşünme ve değerlendirme.

Metnin dışında yer alan genel ifadeler ve tabloda yer alan bir çift ifade arasındaki ilişkiyi ayırt etme.

Tam Puan

C. Bu bilgi, hayır kısmında yer alan iddiayı destekler fakat bunu kanıtlamaz.

Soru 6: CEP TELEFONU GÜVENLİĞİ

R414Q06 – 0 1 9

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

Kullanım amacı: Kamusal

Metnin biçimi: Bağımsız

Metnin türü: Açıklama

Metne yaklaşım: Kendi düşüncelerini yansıtma ve metni değerlendirme: Bir metnin içeriğini değerlendirme ve kendi düşüncelerini yansıtma

Soru biçimi: Açık uçlu

Zorluk derecesi: 536 (Düzey 3)

CEP TELEFONU GÜVENLİĞİ PUANLAMA 6

SORUNUN AMACI:

Düşünme ve değerlendirme: Bir metnin içeriği üzerinde derinlemesine düşünme ve değerlendirme

Metinde yer alan bilgiye uygulamak için önceki bilgilerinizden yararlanma.

Tam Puan

Kod 1: Modern yaşamda, yorgunluk, baş ağrısı ya da konsantrasyon kaybıyla ilgili olabilecek bir faktörü ifade eder. Verilen cevap yeteri kadar açıktır veya bir açıklama ile verilebilir.

- Yeterince uyuyamamak. Eğer uyuyamazsanız, yorgun olursunuz.
- Çok meşgul olmak. Bu sizi yorar.
- Çok fazla ev ödevi. Bu sizi yorar VE başınızın ağrmasına sebep olur.
- Gürültü– başınızın ağrmasına sebep olur.
- Stres.
- Geç saatlere kadar çalışmak.
- Sınavlar.
- Dünya çok gürültülü.
- İnsanlar rahatlamak için kendilerine eskisi kadar zaman ayırmıyorlar.
- İnsanlar konuları bir önem sırasına koymazlar. Bu nedenle hırçın ve hasta olurlar.
- Bilgisayarlar.
- Hava kirliliği.
- Çok fazla TV seyretmek.
- Uyuşturucular.
- Mikrodalgalar.
- Çok fazla elektronik posta göndermek.

Soru 9: CEP TELEFONU GÜVENLİĞİ

R414Q09

Eğer bir cep telefonu kullanıyorsanız... başlıklı tabloya bakınız.

Tablo aşağıdaki fikirlerden hangisine dayanmaktadır?

- A Cep telefonu kullanımının herhangi bir tehlikesi yoktur.
- B Cep telefonu kullanımının tehlikeli olduğuna dair kanıt vardır.
- C Cep telefonu kullanımı tehlikeli olabilir ya da olmayabilir. Fakat önlem almakta fayda vardır.
- D Cep telefonu kullanımı tehlikeli olabilir ya da olmayabilir, fakat biz emin olana kadar cep telefonları kullanılmamalıdır.
- E **Şunları yapınız** yönergeleri, tehdidi ciddiye alanlar içindir ve **Şunları yapmayınız** yönergeleri diğer insanlar içindir.

Kullanım amacı: Kamusal

Metnin biçimi: Bağımsız

Metnin türü: Açıklama

Metne yaklaşım: *Bilgileri bir araya getirme ve yorumlama:* Bir yorum geliştirme

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 494 (Düzey 3)

CEP TELEFONU GÜVENLİĞİ PUANLAMA 9

SORUNUN AMACI:

Bilgileri birleştirme ve yorumlama: Bir yorum geliştirme

Açıklayıcı metnin bir kısmında yer alan varsayımı ayırt etme.

Tam Puan

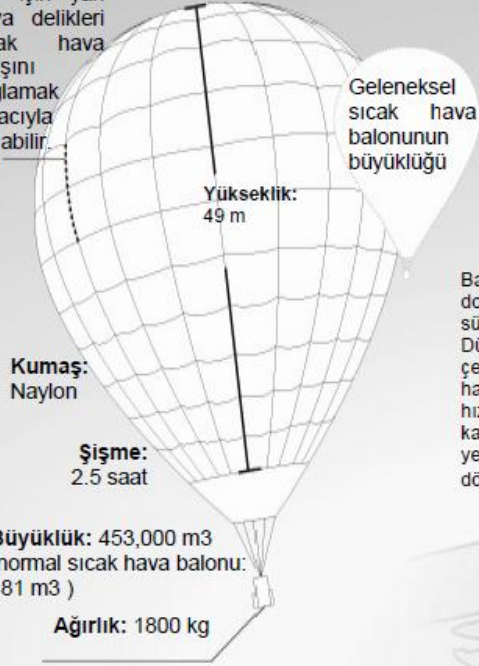
C. Cep telefonu kullanımı tehlikeli olabilir ya da olmayabilir. Fakat önlem almakta fayda vardır.

BALON

Sıcak hava balonları yükseklik rekoru

Hintli pilot Vijaypat Singhania, 26 Kasım 2005'te sıcak hava balonları yükseklik rekorunu kırdı. O, balonla deniz seviyesinden 21.000 metre yüksekliğe çıkan ilk kişiydi.

İniş için yan hava delikleri sıcak hava çıkışını sağlamak amacıyla açılabilir.



Aşağıda verilen soruları yanıtlamak için “Balon” adlı metinden yararlanınız.

Soru 8: BALON

R417Q08

Bu metnin ana fikri nedir?

- A Singhania balon yolculuğu süresince tehlikede idi.
- B Singhania yeni bir dünya rekoru kırdı.
- C Singhania hem deniz hem de kara üzerinde yolculuk yaptı.
- D Singhania'nın balonu çok büyüktü.

Kullanım amacı: Eğitimsel

Metnin biçimi: Bağımsız

Metnin türü: Tanımlama

Metne yaklaşım: *Bilgileri bir araya getirme ve yorumlama: Genel bir anlayış oluşturma*

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 369 (Düzey 1a)

BALON PUANLAMA 8

SORUNUN AMACI:

- Bilgileri birleştirme ve yorumlama: Genel bir anlayış oluşturma
- Resimli bir metinde ana fikri bulma

Tam Puan

Kod 1: B. Singhania yeni bir dünya rekoru kırdı.

Soru 3: BALON

R417Q03 – 0 1 2 9

Vijaypat Singhania, diğer iki ulaşım türünde bulunan bazı teknolojileri kullandı. Bunlar hangi tür ulaşımlardır?

1.

2.

Kullanım amacı: Eğitimsel

Metnin biçimi: Bağımsız

Metnin türü: Tanımlama

Metne yaklaşım: *Bilgiye ulaşma ve bilgiyi hatırlama: Bilgiyi hatırlama*

Soru biçimi: Kısa yanıt

Zorluk derecesi: **Tam puan** 623 (Düzey 4); **Kısmi puan** 458 (Düzey 2)

BALON PUANLAMA 3

SORUNUN AMACI:

- Bilgiye ulaşma ve hatırlama: Bilgiyi hatırlama
- Resimli bir metinde açıkça belirtilen iki parça bilgiyi bulma

Tam Puan

Kod 2: HEM uçak HEM DE uzay gemisine (her hangi bir sırayla) değinir. [iki yanıt da aynı satıra yazılabilir]

- 1. Uçak
- 2. Uzay gemisi
- 1. Uçaklar
- 2. Uzay gemileri
- 1. Hava yolculuğu
- 2. Uzay yolculuğu
- 1. Jumbolar
- 2. Uzay roketleri
- 1. jetler
- 2. roketler

Kısmi Puan

Kod 1: SADECE uçak YA DA uzay gemisine değinir.

- Uzay gemisi
- Uzay yolculuğu
- Uzay roketleri
- Roketler
- Uzay gemisi
- Uçaklar
- Hava yolculuğu
- Jumbolar
- Jetler

Soru 4: BALON

R417Q04 – 0 1 9

Bu metinde bir jumbo jet resmi bulunmasının nedeni nedir?

.....

.....

Kullanım amacı: Eğitimsel

Metnin biçimi: Bağımsız

Metnin türü: Tanımlama

Metne yaklaşım: Kendi düşüncelerini yansıtmaya ve metni değerlendirme: Bir metnin içeriğini değerlendirme ve kendi düşüncelerini yansıtmaya

Soru biçimi: Açık uçlu

Zorluk derecesi: 526 (Düzey 3)

BALON PUANLAMA 4

SORUNUN AMACI:

Düşünme ve değerlendirme: Bir metnin içeriği üzerinde derinlemesine düşünme ve değerlendirme

Resimli bir metinde yer alan bir resmin amacını belirleme

Tam Puan

Kod 1: Yükseklığe değinir. Jumbo jet ve balon arasında karşılaştırma yapabilir.

- Balonun ne kadar yükseğe çıktığını göstermek için.
- Balonun sahiden yükseğe çıktığı gerçeğini vurgulamak için.
- Onun rekorunun gerçekten ne kadar etkileyici olduğunu göstermek için – O jumbo jetlerden daha yükseğe çıktı!
- Yükseklikle ilgili karşılaştırma noktası olarak.

Soru 6: BALON

R417Q06



Neden iki balon çizimi bulunmaktadır?

- A Singhania'nın balonunun şişirilmeden önce ve sonraki büyüklüklerini karşılaştırmak için.
- B Singhania'nın balonu ile diğer sıcak hava balonlarının büyüklüklerini karşılaştırmak için.
- C Singhania'nın balonunun yerden daha küçük görüldüğünü göstermek için.
- D Singhania'nın balonunun diğer balonla neredeyse çarpışacağını göstermek için.

Kullanım amacı: Eğitimsel

Metin biçimi: Bağımsız

Metin türü: Tanımlama

Metne yaklaşım: Kendi düşüncelerini yansıtma ve metni değerlendirme: Bir metnin içeriğini değerlendirme ve kendi düşüncelerini yansıtma

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 414(Düzyey 2)

BALON PUANLAMA 6

SORUNUN AMACI:

Düşünme ve değerlendirme: Bir metnin içeriği üzerinde derinlemesine düşünme ve değerlendirme

Resimli bir metinde bağlantılı şekillerin amacını tanıma

Tam Puan

- B. Singhania'nın balonu ile diğer sıcak hava balonlarının büyüklüklerini karşılaştırmak için.

KAN ARANIYOR



Kan bağışı gereklidir.

İnsan kanının yerini tam olarak tutabilecek başka bir madde yoktur. Kan bağışı, 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 bağışı:

Kan bağışı en çok bilinen bağış 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 bağışı yapanlar 18 ile 65 yaş arasında olabilir.

Her bağış arasında geçmesi gereken zorunlu süre 8 haftadır.

Önceki sayfada yer alan “Kan Aranıyor” adlı metin bir Fransız internet sitesinden alınmıştır. Aşağıdaki sorulara yanıt vermek için bu metinden yararlanınız.

Soru 8: KAN ARANIYOR

R429Q08 – 0 1 9

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?

.....

Kullanım amacı: Kamusal

Metnin biçimi: Akıcı

Metnin türü: Tartışma

Metne yaklaşım: Bilgileri bir araya getirme ve yorumlama: Bir yorum geliştirme

Soru biçimi: Açık uçlu

Zorluk derecesi: 446 (Düzey 2)

KAN ARANIYOR PUANLAMA 8

SORUNUN AMACI:

Bilgileri birleştirme ve yorumlama: Bir yorum geliştirme

Kısa bir metinde bağlantılar kurarak bir sonuca ulaşma.

Tam Puan

Son kan bağışından bu yana yeterli zaman geçmesi gerektiğini vurgular.

- Onun son kan vermesinden bu yana 8 hafta geçip geçmediğine bağlı.
- Yeterince süre geçtiyse kan verebilir aksi takdirde veremez.

Soru 9: KAN ARANIYOR

R429Q09

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

Bu metinde bu bilgi neden yer almaktadır?

- A Kan bağışının güvenli olduğundan emin olmanız için.
- B Kan bağışının gerekli olduğunu vurgulamak için.
- C Alınan kanın kullanım şeklini açıklamak için.
- D Testler ve kontrollerle ilgili ayrıntıları vermek için.

Kullanım amacı: Kamusal

Metnin biçimi: Akıcı

Metnin türü: Tartışma

Metne yaklaşım: Kendi düşüncelerini yansıtmaya ve metni değerlendirme: Bir metnin içeriğini değerlendirme ve kendi düşüncelerini yansıtmaya

Soru biçimi: Çoktan seçmeli

Zorluk derecesi: 365 (Düzey 1a)

KAN ARANIYOR PUANLAMA 9**SORUNUN AMACI:**

Düşünme ve değerlendirme: Bir metnin içeriği üzerinde derinlemesine düşünme ve değerlendirme

Bir ilanda yer alan bir cümlenin ikna edici amacını belirler.

Tam Puan

A. Kan bağışının güvenli olduğundan emin olmanız için.

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."

Aşağıdaki soruları yanıtlamak için bir önceki sayfada yer alan "Cimri ve Altını" adlı masaldan yararlanınız.

Soru 1: CİMRİ

R433Q01 – 0 1 9

Aşağıdaki cümleleri okuyunuz ve metinde yer alan olayların sırasına göre numaralandırınız.

Cimri, bütün parasını altına çevirmeye karar vermiş.

Bir adam, cimrinin altınını çalmış.

Cimri bir çukur kazmış ve altınını onun içine saklamış.

Cimrinin komşusu ona, altının yerine bir taş koymasını söylemiş.

Kullanım amacı: Kişisel

Metnin biçimi: Akıcı

Metnin türü: Anlatım

Metne yaklaşım: Bilgileri bir araya getirme ve yorumlama: Bir yorum geliştirme

Soru biçimi: Kapalı uçlu

Zorluk derecesi: 372 (Düzey 1a)

CİMRİ PUANLAMA 1

SORUNUN AMACI:

Bilgileri birleştirme ve yorumlama: Bir yorum geliştirme

Bir hikâyedeki olayları sıraya koyma.

Tam Puan

Dört seçeneğin tamamı doğru yanıtlanmış: Sırasıyla 1,3,2,4

Soru 7: CİMRİ

R433Q07 – 0 1 9

Cimri bir külçe altını nasıl aldı?

Kullanım amacı: Kişisel

Metnin biçimi: Akıcı

Metnin türü: Anlatım

Metne yaklaşım: Bilgiye ulaşma ve bilgiyi hatırlama: Bilgiyi hatırlama

Soru biçimi: Kısa yanıt

Zorluk derecesi: 301 (Düzey 1b)

CİMRİ PUANLAMA 7**SORUNUN AMACI:**

Bilgiye ulaşma ve hatırlama: Bilgiyi hatırlama

Kısa bir metnin başında açık bir şekilde belirtilen bilgiyi bulur

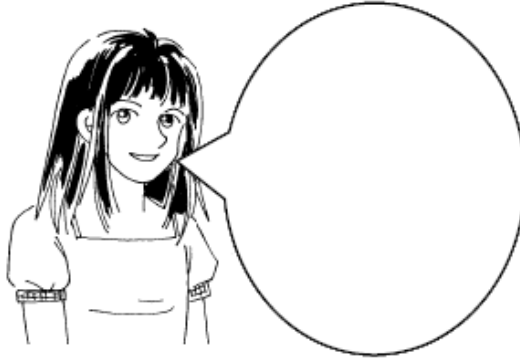
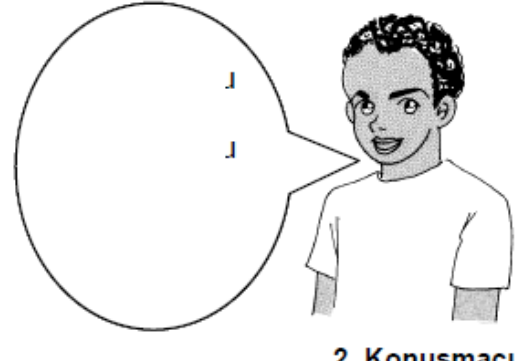
Tam PuanKod 1: Sahip olduğu her şeyi sattığını belirtir. Kendi sözcükleri ya da metinden alıntı olabilir.

- O her şeyini sattı.
- O bütün mallarını sattı.
- O altını satın aldı. [*sahip olduğu her şeyi sattığına dolaylı atıfta bulunma*]

Soru 5: CİMRİ

R433Q05 – 0 1 9

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

**1. Konuşmacı****2. Konuşmacı**

2. Konuşmacı kendi görüşünü desteklemek için ne söyleyebilirdi?

.....

.....

Kullanım amacı: Kişisel**Metnin biçimi:** Akıcı**Metnin türü:** Anlatım**Metne yaklaşım:** Bilgileri bir araya getirme ve yorumlama: Bir yorum geliştirme**Soru biçimi:** Açık uçlu**Zorluk derecesi:** 569(Düzey 4)

CİMRİ PUANLAMA 5**SORUNUN AMACI:**

Bilgileri birleştirme ve yorumlama: Bir yorum geliştirme
Bir masaldaki ayrıntıyı ana fikirle ilişkilendirme

Tam Puan

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 birşey 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 anlatmak istediği buydu.