

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

PUANLAMA YÖNERGESİ (RUBRİK)
KULLANILMASI DURUMUNDA
PUANLAYICI GÜVENİRLİĞİNİN İNCELENMESİ

Gözde ATMAZ

YÜKSEK LİSANS TEZİ

Mersin, 2009

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

PUANLAMA YÖNERGESİ (RUBRİK)
KULLANILMASI DURUMUNDA
PUANLAYICI GÜVENİRLİĞİNİN İNCELENMESİ

Gözde ATMAZ

Danışman: Prof. Dr. Adnan ERKUŞ

YÜKSEK LİSANS TEZİ

Mersin, 2009

Mersin Üniversitesi, Sosyal Bilimler Enstitüsü Müdürlüğüne,

Gözde ATMAZ tarafından hazırlanan Puanlama Yönergesi (RUBRİK) Kullanılması Durumunda Puanlayıcı Güvenirliğinin İncelenmesi başlıklı bu çalışma, jürimiz tarafından Eğitim Bilimleri / Eğitimde Ölçme ve Değerlendirme Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Başarılı

Başarısız



Başkan

Doç. Dr. Adnan KAN



Üye

Prof. Dr. Adnan ERKUŞ

(Danışman)



Üye

Yrd. Doç. Dr. Devrim ALICI

Onay

Yukarıdaki imzaların, adı geçen öğretim elemanlarına ait olduklarını onaylarım.

... / ... /

Prof. Dr. Mustafa AKSAN

Enstitü Müdürü

ÖNSÖZ

Lisansüstü eğitim ve öğretim sınav yönetmeliğinin Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı için öngördüğü yüksek lisans tezi olarak hazırlanan bu çalışmada, puanlama yönergesi (rubrik) kullanımının puanlamaya ve puanlayıcı güvenilirliğine etkisi incelenmiştir. Bu çalışmanın Eğitimde Ölçme ve Değerlendirme alanına küçük de olsa bir katkı sağlayacağı, çalışma kapsamında ve benzer konularda araştırma yapacak ilgili kişilere yardımcı olacağı umulmaktadır.

Bu çalışmanın tamamlanmasında bir çok kişinin özverili yardım ve desteklerinden faydalanılmıştır.

Çalışmamın başından sonuna kadar önerilerde bulunarak çalışmamı yönlendiren danışmanım Prof. Dr. Adnan ERKUŞ ve Doç. Dr. Adnan KAN'a; çalışmamla ilgili tavsiyelerde bulunan Prof. Dr. Ata TEZBAŞARAN ve Yrd. Doç. Dr. Devrim ÖZDEMİR ALICI'ya; çalışmamın sonuna kadar desteklerini esirgemeyen eşim ve aileme; emeği geçen tüm öğretmenlere teşekkürü bir borç bilirim.

Gözde ATMAZ

Mersin/ Haziran,2009

ÖZET

Bu arařtırmada, grafik yorumlama becerisine iliřkin bir ölçeęi puanlamada puanlama yönergesi (rubrik) kullanımının, ayrıca aynı ve farklı öęretmenler tarafından yapılan puanlamalar arasında geen zamanın güvenilirlik üzerine etkisi incelenmiřtir.

Bu amala, grafik yorumlama becerisine iliřkin 4 aık ulu maddeden oluřan bir ölme aracı ve bu ölme aracını puanlamak iin bir puanlama yönergesi (rubrik) hazırlanmıřtır.

Arařtırma kapsamında hazırlanan sorular Mersin Toros Koleji 8. sınıflarında okuyan 50 öęrenciye uygulanmıřtır. Sorulara verilen yanıtlar ise Tarsus ve Mersin'de alıřmakta olan veya emekli 17 matematik öęretmeni tarafında puanlanmıřtır.

Öęretmenler, öncelikle öęrencilerin yanıtlarını puanlama yönergesi kullanmadan puanlamıřlardır. Daha sonra bellek etkisini önlemek iin öęretmenler 2 hafta beklemiřler ve sonra aynı kaęıtları puanlama yönergesi kullanarak puanlamıřlardır. Zamanın puanlar üzerindeki etkisini ölçmek iin aynı uygulama 10 hafta sonra tekrarlanmıřtır. Bu yaklařımla, aynı öęretmenlerin farklı zamanlarda verdikleri puanlar arasında anlamlı bir iliřki ve farklılık olup olmadıęı belirlenmeye alıřılmıřtır.

Aynı öęretmenler tarafından verilen farklı puanlar arasındaki iliřki Pearson Momentler arpımı Korelasyon Katsayısı ile incelenmiř ve elde edilen korelasyon katsayıları t-testi ile test edilmiřtir. Korelasyon katsayıları arasında anlamlı bir farklılık olup olmadıęını bulmak iin Fisher'in z' dönüşümü yapılmıřtır. Ayrıca aynı öęretmenlerin verdikleri puanların ortalamaları arasında anlamlı bir farklılık olup olmadıęını bulmak iin baęımlı örneklemler iin t testi kullanılmıřtır.

Her uygulama iin, farklı öęretmenlerin puanları ikiřer ikiřer karşılaştırılmıř ve puanlar arasındaki iliřki Pearson Momentler arpımı Korelasyon Katsayısı ile incelenmiř ve

elde edilen korelasyon katsayıları t-testi ile test edilmiştir. Farklı öğretmenlerin verdikleri puanların ortalamaları arasında anlamlı bir fark olup olmadığını bulmak için tek yönlü varyans analizi yöntemi kullanılmıştır. Her puanlama tipi için, puanlayıcılar arası güvenilirlik, Hoyt'un varyans analizine dayalı G katsayısı hesaplanarak belirlenmiştir.

Sonuç olarak, her uygulamada, aynı öğretmenlerin puanlamaları arasında pozitif yönde, oldukça yüksek ve anlamlı bir ilişki bulunmuştur. Fakat puanların ortalamaları arasında anlamlı farklılıklar vardır. Bu noktada, öğretmenler ya da uygulamalar arasında geçen zaman bu farklılığın sebebi olarak düşünülmektedir. Her uygulamada, farklı öğretmenlerin puanları arasında da pozitif yönde, oldukça yüksek ve anlamlı bir ilişki bulunmuştur. Fakat farklı öğretmenlerin puan ortalamaları arasında tutarsızlığa neden olacak kadar manidar farklılıklar gözlenmiştir. Öğretmen özelliklerine dayalı bilişsel değişkenlerin ya da uygulamalar arasında geçen zamanın bu farklılığa neden olduğu düşünülebilir.

Anahtar Kelimeler: puanlama yönergesi (rubrik), grafik yorumlama becerisi

EXAMINING THE INTER-RATER RELIABILITY, WHEN USING SCORING RUBRIC

ABSTRACT

In this research, the effects of using “Scoring Rubric” to score a graph interpretation scale and the time between scorings by same and different teachers, on the reliability were examined.

For this reason, a graph interpretation scale consisting of 4 open-ended questions and a scoring rubric for scoring the scale were prepared.

In the borders of the research, prepared questions were practiced on 50 students attending 8th grade in Mersin Toros College. Answers to the questions were scored by 17 employed and retired Mathematics teachers located in Tarsus and Mersin.

Teachers firstly scored the answers of students without using a scoring rubric. Later on, in order to prevent the memory effect, the teachers waited for 2 weeks and then scored the same papers using a scoring rubric. In order to measure the effect of time on the scores, the above procedure was repeated after 10 weeks. With this approach, it was tried to find out whether a meaningful relationship and difference between the scores by same teachers at different times.

The relationship between the different scores by the same teachers was examined with Pearson Moment- Product Correlation Coefficient Method and the resulting correlation coefficients were tested with t- test. In order to find out whether a meaningful difference between correlation coefficients Fisher’s z’ transform was used. Also, in order to find out whether a meaningful difference between the means of the given scores by same teachers t test for related samples was used.

For each experiment, scores of different teachers were compared in pairs and the relationship between scores were examined with Pearson Moment- Product Correlation Coefficient Method and the resulting correlation coefficients were tested with t- test. In order to find out whether a meaningful difference between the means of the given scores by different teachers, one way variance analysis method was used. The interrater reliability for each scoring type was determined by computing the G coefficient which depends on Hoyt's variance analysis.

In conclusion, between the scores of the same teachers in each experiment, a positively, highly related, meaningful relation is reached. On the other hand, it is observed that there are meaningful differences between the means of scores. At this point, it is considered that teachers or the awaiting time between the experiments are the reason for the difference. Between the scores of different teachers in each experiment, a positive, highly related, meaningful relation is reached. Conversely, it is observed that there were unexpected differences between the scores of different teachers which caused great inconsistency in the observed mean values. Cognitive variables dependent on teacher characteristics or awaiting time between the experiments are considered to be the reason of these differences.

Keywords: scoring rubric, graph interpretation skill

İÇİNDEKİLER

	Sayfa no
ÖNSÖZ.....	i
ÖZET.....	ii
ABSTRACT.....	iv
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ.....	viii
GİRİŞ.....	1
Araştırmanın Amacı ve Önemi.....	11
Problem Cümlesi.....	11
Alt Problemler.....	12
Sayıtlar.....	14
Sınırlılıklar.....	14
BÖLÜM I (YÖNTEM).....	15
I.1.Araştırmanın Türü.....	15
I.2.Çalışma Grubu.....	15
I.3.Verİ Toplama Araçları.....	16
I.3.1.Grafik Yorumlama Becerisi Ölçeği.....	16
I.3.2.Değerlendirme Ölçeği.....	18
I.4.İşlem.....	20
I.5.Verilerin Çözümlemesi.....	23
BÖLÜM II (BULGULAR VE YORUM).....	28

BÖLÜM III (SONUÇ VE ÖNERİLER).....	77
III.1.Sonuçlar.....	77
III.2.Öneriler.....	78
KAYNAKÇA.....	80
EKLER.....	84

TABLOLAR

	Sayfa no
TABLO I.1. Puanlamalara ve Öğretmenlere İlişkin Deney Düzenegi.....	22
TABLO II.1. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Kullanarak Vermiş Oldukları Puanlar Arasındaki Korelasyon Katsayıları.....	29
TABLO II.2. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Kullanarak Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	31
TABLO II.3. Aynı Öğretmenlerin Aradan Belli Bir Süre Geçtikten Sonra Puanlama Yönergesi Kullanmadan ve Kullanarak Vermiş Oldukları Puanlar Arasındaki Korelasyon Katsayıları.....	34
TABLO II.4. Aynı Öğretmenlerin aradan Belli Bir Süre Geçtikten Sonra Puanlama Yönergesi Kullanmadan ve Kullanarak Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	35
TABLO II.5. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Aradan Belli Bir Süre Geçtikten Sonra Tekrar Puanlama Yönergesi Kullanmadan Vermiş Oldukları Puanlar Arasındaki Korelasyon Katsayıları.....	38
TABLO II.6. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Aradan Belli Bir Süre Geçtikten Sonra Tekrar Puanlama Yönergesi Kullanmadan Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	39
TABLO II.7. Aynı Öğretmenlerin Puanlama Yönergesi Kullanarak ve	

Aradan Belli Bir Süre Geçtikten Sonra Tekrar Puanlama Yönergesi Kullanarak Vermiş Oldukları Puanlar Arasındaki Korelasyon Katsayıları.....	42
TABLO II.8. Aynı Öğretmenlerin Puanlama Yönergesi Kullanarak ve Aradan Belli Bir Süre Geçtikten Sonra Tekrar Puanlama Yönergesi Kullanarak Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	43
TABLO II.9. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Belli Bir Süre Sonra Puanlama Yönergesi Kullanarak Vermiş Oldukları Puanlar Arasındaki Korelasyon Katsayıları.....	46
TABLO II.10. Aynı Öğretmenlerin Puanlama Yönergesi Kullanmadan ve Belli Bir Süre Sonra Puanlama Yönergesi Kullanarak Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	47
TABLO II.11. Aynı Öğretmenlerin Puanlama Yönergesi Kullanarak ve Belli Bir Süre Sonra Puanlama Yönergesi Kullanmadan Verdiği Puanlar Arasındaki Korelasyon Katsayıları.....	50
TABLO II.12. Aynı Öğretmenlerin Puanlama Yönergesi Kullanarak ve Belli Bir Süre Sonra Puanlama Yönergesi Kullanmadan Vermiş Oldukları Puanlara İlişkin t-testi Sonuçları.....	52
TABLO II.13. Aynı Öğretmenlerin Farklı Okumalarda Verdikleri Puanlar Arasındaki Pearson Momentler Çarpımı Korelasyon Katsayıları ve Ortalamalar Arasındaki Farkı Gösteren t Puanlarının Yer Aldığı Özet Tablo.....	55
TABLO II.14. Farklı Okumalar Arasındaki Korelasyon Katsayılarının Fisher'in z' Dönüşümü ile Karşılaştırılması.....	56

TABLO II.15. İkili Karşılaştırıldıklarında Farklı Öğretmenlerin	
1. Okuma Sonrasında Verdikleri Puanlar Arasındaki İlişki.....	59
TABLO II.16. Farklı Öğretmenlerin 1. Okuma Sonrasında	
Verdikleri Puanlara İlişkin Varyans Analizi Sonuçları.....	60
TABLO II.17. Farklı Öğretmenlerin 1. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Scheffe Testi Sonuçları.....	61
TABLO II.18. İkili Karşılaştırıldıklarında Farklı Öğretmenlerin	
2. Okuma Sonrasında Verdikleri Puanlar Arasındaki İlişki.....	63
TABLO II.19. Farklı Öğretmenlerin 2. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Varyans Analizi Sonuçları.....	64
TABLO II.20. Farklı Öğretmenlerin 2. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Scheffe Testi Sonuçları.....	65
TABLO II.21. İkili Karşılaştırıldıklarında Farklı Öğretmenlerin	
3. Okuma Sonrasında Verdikleri Puanlar Arasındaki İlişki.....	68
TABLO II.22. Farklı Öğretmenlerin 3. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Varyans Analizi Sonuçları.....	69
TABLO II.23. Farklı Öğretmenlerin 3. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Scheffe Testi Sonuçları.....	70
TABLO II.24. İkili Karşılaştırıldıklarında Farklı Öğretmenlerin	
4. Okuma Sonrasında Verdikleri Puanlar Arasındaki İlişki.....	72
TABLO II.25. Farklı Öğretmenlerin 4. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Varyans Analizi Sonuçları.....	73
TABLO II.26. Farklı Öğretmenlerin 4. Okuma Sonrasında Verdikleri	
Puanlara İlişkin Scheffe Testi Sonuçları.....	74

GİRİŞ

Toplumların teknoloji çağından bilgi çağına doğru ilerleyişleri, toplumun ve iş dünyasının geleceğin yetişkinlerinden beklentilerini de farklılaştırmıştır. Öğrencilere temel bilgi ve becerilerin yanında analitik düşünme, eleştirel düşünme, yaratıcı düşünme, problem çözme gibi üst düzey bilgi ve becerilerin kazandırılmasını zorunlu hale getirmiştir. Bu noktada, öngörülen bu bilgi, beceri, tutum ve davranışların kazanılıp kazanılmadığının ya da ne derece kazanıldığının doğru bir şekilde belirlenmesi gerekmektedir.

Eğitimde ölçme ve değerlendirme, öğrenme-öğretme sürecinde öğrencilerin bilişsel, duyuşsal, devinimsel becerilerini saptamak, eksikliklerini belirlemek, öğretim yöntemlerinin etkililiğini anlamak, programın zayıf ve kuvvetli yanlarını ortaya çıkarmak için yapılabilir.

Yeni ilköğretim programının öğrencilerin öğrenme ihtiyaçlarını merkeze alan, bireyselleştirilmiş bir öğrenme programını temel vizyon olarak edinmesi, ölçme ve değerlendirmede hem yöntem hem de araç olarak çeşitliliğin artmasına sebep olmuştur. Sonuç değerlendirmenin yanında süreç değerlendirme önem kazanmış, ölçme araçları öğrencilerin bireysel farklılıklarına göre değerlendirilebilmesine olanak sağlamıştır.

Öğretmenler, gözlem, performans ödevleri, görüşmeler, öz değerlendirme formları, akran değerlendirme formları, öğrenci ürün dosyaları (portfolyo), projeler, posterler, puanlama yönergeleri (dereceli puanlama anahtarı, rubrik) vb. araç, yöntem ve eğitim çıktılarını kullanarak öğrencilerin bilgiyi nasıl yapılandığı ve üst zihinsel becerilerini ne kadar geliştirdiğini süreç içinde değerlendirir. Ayrıca çoktan seçmeli,

boşluk doldurmalı, eşleştirmeli, kısa cevaplı, açık uçlu vb. maddelerden oluşan testler de kullanarak hem süreci hem de sonucu değerlendirmiş olur (Baki, 2006).

Ölçme işlemi sonucunda kişiler hakkında belli konularda belli kararlar verilir. Verilen kararların doğruluğu, değerlendirmenin dayandığı ölçme işleminin hatasız ya da olabildiğince az hatalı olmasına, tekrarlı ölçümlerde tutarlı sonuçlar vermesine ve verilecek kararlar ilgili olmasına bağlıdır. Başka bir deyişle kullanılacak ölçme araçlarının güvenilirlik ve geçerliklerinin sağlanması gereklidir.

Pratikte bir nesne ya da özellik değişik zamanlarda ölçüldüğünde, ölçmelerin hepsinde aynı ölçümü elde etmek olanaksızdır. Çünkü çeşitli nedenlerle tam anlamıyla hatadan arınık bir ölçme yapılamaz. Bu nedenle bir testten elde edilen bireysel puanlar bir ölçmeden diğerine değişecektir. Ölçme sonuçlarına karışan hatalar, ölçmeciden, ölçme aracı ve ölçme yönteminden, ölçmenin yapıldığı ortamdan, ölçülen bireylerin sözü edilen bu etkenlerle etkileşiminden ileri gelebilir (Tekin, 2004).

Yeni öğretim programında, temel kavram ve becerilerin yanında, akıl yürütme, ilişki kurma, problem çözüme, eleştirel ve yaratıcı düşünme gibi üst düzey zihinsel becerileri kazandırmak hedeflenmiştir. Bu hedeflere ulaşıp ulaşılmadığının değerlendirilmesinde de açık uçlu durumlara ihtiyaç duyulmaktadır. Suzanne Lane'e (1993) göre açık uçlu maddeler, öğrencilerin matematiksel düşünme ve muhakeme yapma becerilerini göstermesini sağlar. Dahası bilgiyi sentezlemesini gerektiren açık uçlu durumlar, çözüm süreçlerini gözden geçirerek cevaplarını değerlendirmelerini sağlayarak üst düzey becerilerini ön plana çıkarır. Fakat açık uçlu maddelere verilen cevapların ya da açık uçlu ürünlerin puanlanması, puanlayıcıların kişisel yargılarından büyük ölçüde etkilendiğinden puanlamada subjektiflik sorunu ortaya çıkar.

Puanlayıcıların öznel yargılarından etkilenmeyen, uzman olsun olmasın herkes tarafından aynı şekilde puanlanan objektif testlerin, eğitimde ölçme ve değerlendirmedeki yeri, objektif puanlamanın öneminin göstergesidir. Kan'a (2001) göre, objektif testlerin kullanılmadığı yerlerde puanların belirli bir ölçek (yanıt anahtarı) ile puanlanması, objektif puanlama için oldukça önemli bir yere sahiptir. Bu amaçla Kan (2001), objektif bir puanlama metodu olarak düşünülen ve hangi soruda hangi davranışa kaç puan verileceğinin ayrıntılı olarak belirtildiği puanlama cetveli ve yanıt anahtarının, puanlama güvenilirliğini artırıp artırmadığını irdelemek üzere, yazılı yoklamalara, puanlama cetveli ve yanıt anahtarı kullanmadan ve kullanarak yapılan puanlamalar arasında bir ilişki olup olmadığını ve bu ilişkinin farklı zamanlarda ve farklı puanlayıcılar söz konusu olduğunda nasıl bir değişme gösterdiğini incelemiştir. Sonuç olarak aynı veya farklı zamanlarda puanlama cetveli ve yanıt anahtarı kullanıldığında verilen puanların tutarlılığının daha yüksek olduğunu gözlemiştir. Fakat öğretmenler arasında puanlama tutarlılığının tam olarak sağlanamadığını ve bunu sağlayabilecek subjektif etkenlerin (öğretmen yargıları, kanaat, kişilik vb) ortadan kaldırılamadığını ortaya koymuştur.

Farklı zamanlar ve farklı öğretmenler söz konusu olduğunda puanlama tutarlılığı sağlayabileceği düşünülen bir başka teknik de son zamanlarda oldukça popüler olan puanlama yönergeleridir (dereceli puanlama anahtarı, rubrik).

Puanlama yönergeleri, öğrencinin belli bir alandaki bilgisini göstermesi ya da bir görevi yerine getirmesindeki performansının ya da ortaya koyduğu ürünün analizine rehberlik eden öğretmen veya diğer değerlendiriciler tarafından geliştirilen betimsel puanlama şemalarıdır (Brookhart, 1999; Aktaran: Moskal, 2000).

Moskal'a (2000) göre, puanlama yönergeleri, performansı tanımlayan her bir kriterin ne derece karşılandığını gösteren geniş bir değerlendirme ölçeğidir.

Puanlama yönergeleri, performans kriterleri ve performans düzeyleri olmak üzere iki ana bölüm ve bu bölümlere bağlı olarak belirlenen performans tanımlarından oluşur. Performans kriterleri, öğrencide, performansını ortaya koyarken gözlenmek istenen kritik davranışlardır. Performans düzeyleri, performansı oluşturan kriterlerin ne düzeyde karşılandığını gösteren nitel ya da nicel kategorik ifadelerdir. Performans tanımları ise performans kriterleri ile düzeylerini betimleyen gözlenebilir özelliklerdir (Kan, 2007).

Değerlendirmenin amacına bağlı olarak puanlama yönergeleri iki türdür:

a) Bütüncül puanlama yönergesi (Holistik rubrik): Performansın ya da ortaya çıkan ürünün bileşenlerini yargılamadan ürünü ya da performansı bir bütün olarak değerlendirmeyi gerektirir (Moskal, 2000).

Bütüncül puanlama yönergeleri;

-Ölçülecek performans parçalara ayrıştırılmıyorsa ve performansa ait kriterler arasında binişiklik varsa (Moskal, 2000);

-Ortaya çıkan ürün ya da performansın mükemmel olması durumunda, bu performansı ortaya çıkaran bazı süreç ve işlem basamaklarındaki hataların göz ardı edilebileceği durumlarda;

-Öğrenciden kendi öğrenme stilini taşıyan bir cevap ya da çözüm bekleniyorsa ve belirli ve kesin bir cevap yoksa;

-Öğrenci performansına ilişkin bütüncül bir puan elde edilmek isteniyorsa;

-Öğrenci çalışmaları kısa sürede değerlendirilmek isteniyorsa;

-Düşük eğitim kademelerinde kullanılacaksa tercih edilir (Moskal, 2000; Nitko, 2001; Mertler, 2001; Kan, 2007).

b) Analitik puanlama yönergesi (Analitik rubrik): Performansın ya da ortaya çıkan ürünün bileşenlerinin belli başlı kriterler doğrultusunda ayrı ayrı ve birbirinden bağımsız değerlendirilmesini gerektirir (Moskal, 2000; Nitko, 2001; Mertler, 2001; Kan, 2007).

Analitik puanlama yönergeleri;

-Ölçülecek performans çok boyutlu ve bileşenlerine ayrılabiliriyorsa;

-Performans görevine ilişkin öğretmene ve öğrenciye anlamlı geri bildirimler sağlanması isteniyorsa;

-Performansı değerlendirme için yeterli süre olduğunda;

-Üst eğitim kademelerinde kullanılacaksa tercih edilir (Mertler, 2001; Nitko, 2001; Kan, 2007).

Puanlama yönergesi hazırlanırken izlenmesi gereken işlem basamakları, bu konuda çalışma yapmış birçok eğitim araştırmacısının bulgularına dayanarak şu şekilde sıralanmıştır (Nitko, 2001; Goodrich, 2005; Mertler, 2001; Popham, 1997; Airasian, 1991):

- 1) Rubrik puanlama yönergesi oluşturmak için öğretmen, öncelikle ölçülecek değişkenin ya da performansın doğasını göz önünde tutarak performansı ya da ürünü analitik mi yoksa holistik bir yaklaşımla mı değerlendireceğine karar vermelidir (Airasian ,1991).
- 2) Performans kriterleri ve her bir kritere ilişkin performansın gözlenebilir özellikleri uygun ve açık bir dille tanımlanmalıdır. Airasian (1991) performans kriterlerini belirlemek için aşağıdaki basamakların izlenmesini önermiştir:

- a) Değerlendirilecek performans ya da görev tamamıyla tanımlanır ve öncelikle öğretmen tarafından yerine getirilerek işlem basamakları belirlenir.
 - b) Ortaya çıkan performans ya da ürünün önemli yönleri belirlenir.
 - c) Performans ölçütleri ekonomik bir şekilde sınıflandırılmaya çalışılır.
 - d) Olanaklıysa hedeflere ilişkin gözlenebilir önemli davranışlara karar vermek için bir grup öğretmenle çalışılır.
 - e) Performans ölçütleri, ürünün karakteristik özellikleri veya gözlenebilir öğrenci davranışları şeklinde ifade edilir.
 - f) Performans kriterlerinin anlamını gölgeleyecek belirsiz, açık olmayan ifade veya kelimeler kullanmamaya özen gösterilir.
 - g) Performans kriterleri gözlenebilir olarak düzenlenir.
- 3) Performans düzeyleri ve her bir performans düzeyine ilişkin karakteristik özellikler belirlenir.
 - 4)
 - a) Analitik rubrik hazırlamak için, her bir performans kriterine karşılık gelen performans düzeylerine ilişkin tanımlar yapılır. Bu, her bir performans düzeyi ve her bir performans kriterine ilişkin gözlenebilir özelliklerin eşleştirilmesini gerektirir. Her bir performans kriterine ait en mükemmel performans düzeyinden en düşük performans düzeyine ilişkin tanımların daha önceden listelenen gözlenebilir özelliklere dayalı olarak yapılmasını içerir.
 - b) Holistik rubrik hazırlamak için, performansla ilişkin bütün gözlenebilir özellikleri içinde barındıran çeşitli performans düzeylerinde tanımlar yapılır.
 - 5) Rubrik puanlama yönergesi tekrar gözden geçirilir ve uzman görüşleri alınır.

Puanlama yönergelerini kullanmak öğretmen ve öğrenciye pek çok avantaj sağlayabilir:

*Öğrenciler, değerlendirme sürecine ilişkin öğretmenin beklentilerinin farkına varırlar.

*Öğretmenler, öğrencilerini performansı ortaya çıkaran belirli ayrıntılara odaklayıp, bu ayrıntılar üzerinde dikkati toplayarak öğretimin kalitesini artırabilirler.

*Beklentileri açık hale getirerek değerlendirmeye şeffaflık katar.

*Beklentiler ve ölçütler açık hale geldiği için geri bildirimler daha anlamlı olur ve kendini değerlendirme kolaylaşır (Moskal, 2000; Goodrich, 2005).

Öğrencinin ortaya koyduğu ürünün ya da performansının değerlendirilmesinde objektifliğin sağlanmaması önemli bir sorundur. Geçerlik ve güvenilirliği sağlanmamış bir ölçme aracıyla yapılan ölçümlere dayanarak verilen kararların ne derece doğru olduğu tartışılır. Güvenilir sonuçlar elde etmek adına yapılan önerilerden biri de puanlamada puanlama yönergeleri (rubrik) kullanmaktır.

Johnson, Penny ve Gordon (2000) tarafından yürütülen bir çalışmada, 5. sınıf öğrencilerine ait toplam 120 kompozisyon, her birisi iki puanlayıcı tarafından 6 gelişim düzeyini temsil eden bir holistik rubrik ile puanlanmıştır. Tamsayı puanları rapor edildikten sonra puanlayıcılardan bir de tamsayı puanlarını, eğer kompozisyon ölçütten daha iyi ise (+) ile daha kötü ise (-) ile genişletmeleri istenmiştir (1-, 1, 1+, 2-, 2, 2+,... şeklinde). Genişletme ile birlikte puanlayıcılar arasındaki uyum %59'dan %96'ya, uzmanlar arasındaki uyum ise %63'ten %94'e çıkmıştır. Bütün kağıtlar için puanlayıcılar arasındaki korelasyon katsayısı, genişletmeden önce .72 iken, genişlettikten sonra .75 olmuştur. Uzmanlar arasındaki korelasyon katsayısı, genişletmeden önce .62 iken genişlettikten sonra .74 bulunmuştur. Puanlayıcılar ve uzmanlar arasındaki güvenilirliğin phi

katsayı kestirimleri 4 farklı yöntem için hesaplanmıştır. Tamsayı derecelemelerde ortalama phi katsayısı .62, standart sapması .14'tür. Genişletilmiş puanlarda ortalama phi katsayısı .69, standart sapması .13'tür. Bu iki grubun varyansları birbirine çok yakın çıkmıştır. Her durumda genişletme yapmak daha yüksek bir phi katsayısının oluşmasını sağlamıştır.

Marzano (2002) tarafından yürütülen bir çalışmada 2 farklı deney yapılmıştır. 1. deneyde 6 maddeli bir 8. sınıf Fen testine 10 öğrencinin verdiği cevaplar üzerinde yapılandırılmamış puan yöntemi ile genel rubrik yöntemi karşılaştırılmıştır. Yapılandırılmamış puan yaklaşımında, öğretmen öncelikle her maddeye olası toplam puanlar atamış, daha sonra öğrencilerin cevaplarına göre bu maddeler üzerinde alabilecekleri en yüksek puanları vermiş ve öğrencinin yüzdeler puanını hesaplamıştır. Genel rubrik yaklaşımında ise, puanlayıcılara genel bir rubrik verilmiştir ki, onlar bütün maddeler için öğrenci cevaplarına uygulamışlardır. 2. deneyde yine 1. deneyde kullanılan öğrenciler ve aynı test kullanılmıştır. Bu defa da yapılandırılmış puan yöntemi ile konu odaklı rubrik karşılaştırılmıştır. Yapılandırılmış puan yaklaşımında, puanlayıcılar, puan değerleri önceden belirlenmiş maddelere göre öğrencilerin cevaplarını puanlamışlardır. Konu odaklı rubrik yaklaşımında ise puanlayıcılar değerlendirilen konuya yönelik bir rubrik kullanmışlardır. Bütün koşullarda genellenebilirlik (G) ve alternatif karar (D) çalışmaları ele alınmıştır. Daha yüksek G ve D katsayılarına göre konu odaklı rubrik, yapılandırılmış puan yönteminden, yapılandırılmış puan yöntemi genel rubrikten, genel rubrik yapılandırılmamış puan yönteminden daha güvenilir ve genellenebilir sonuçlar vermiştir.

Parmar ve Signer (2005) tarafından yürütülen bir çalışmada, öğrenme güçlüğü olan ve olmayan 4. ve 5. sınıfta okuyan toplam 91 öğrencinin grafikleri anlamaları ve yorumlama becerileri dört durumda incelenmiştir. Bunlardan ikisi grafik oluşturma ile

ilgili, diğeri ikisi de grafikleri yorumlama ile ilgilidir. Çalışmanın amacı, öğrenme güçlüğü olan ve olmayan 4. ve 5. sınıf öğrencilerinin anlamada ve yorumlamada performanslarında bir değişim olup olmadığını görmektir. Grafik oluşturma ile ilgili durumlar; sütunları oluşturma, etiketlendirmeye vs. dayalı olarak 1-0 şeklinde puanlanmış, grafik yorumlama ile ilgili durumlar ise öğrencilerin grafiklere dayalı oluşturdukları hikayelerin, hazırlanan holistik rubriklerle (5 düzeyli) değerlendirilmesiyle incelenmiştir. $2 \text{ (grup)} \times 2 \text{ (sınıf)} = 4$ durum için varyans analizi yapılarak puanların ortalamaları karşılaştırılmıştır. Sonuçta tüm durumlarda öğrenme güçlüğü olmayan öğrencilerin puanlarının daha fazla olduğu görülmüştür. Ayrıca 4. sınıf öğrencileri, 5. sınıf öğrencilerinden tüm gruplar için daha başarılı çıkmıştır. Bu ise 4. sınıf müfredatında grafiklere daha çok yer ayrılmış olmasına, öğrencilerin konuya daha aşina olmalarına bağlanmıştır.

Öğrencilerin grafiğe dayalı hikaye oluşturmalarını değerlendirmeye yönelik hazırlanan rubrikte yer alan boyutlardan “kavramsal algılama”da öğrenme güçlüğü olan 4. ve 5. sınıf öğrencilerinin yaklaşık % 25’i grafikte ilişkisiz cümleler kurmuşlardır. Diğer bir boyut olan “çizgi aralıklarının doğruluğu”nda, öğrenme güçlüğü olan öğrencilerin % 70’den fazlası, öğrenme güçlüğü olmayan öğrencilerin ise % 45’i grafikteki aralıkları doğru bir şekilde yorumlayamamıştır. “Eksenlere başvurma” boyutunda öğrenme güçlüğü olmayan 4. ve 5. sınıf öğrencilerinin yaklaşık % 30’u x ve y eksenlerindeki noktalara doğru bir şekilde başvurmuştur. Öğrenme güçlü olan öğrencilerin hiçbiri 1 noktadan (genellikle başlama noktası) fazlasına doğru bir şekilde başvuramamıştır. “Yaratıcılık” boyutunda öğrenme güçlüğü olmayan öğrencilerin % 10’u hikayeyi oluştururken yaratıcılıklarını biraz kullanmışlardır.

Akkuş ve Paksu (2006) da çalışmalarında, ilköğretim öğrencilerinin orantısal akıl yürütme becerilerini ölçmeyi amaçlayan bir ölçme aracı ve bu araca uygun bir

puanlama yönergesi geliřtirmiřtir. Ölçme aracında 7 tane verilmeyen değeri bulma, 3 tane niceliksel karşılařtırma, 4 tane niteliksel karşılařtırma, 1 tane ters orantı sorusu bulunmaktadır. Uzman görüşlerinin alınmasından sonra ölçme aracı, iki farklı ildeki üç farklı okuldan yedinci (162) ve sekizinci (142) sınıf olmak üzere toplam 304 öğrenciye uygulanmıřtır.

Oluřturulan ölçme aracının değeriendirilmesinde, verilmeyen değeri bulma ve ters orantı, niceliksel karşılařtırma ve niteliksel karşılařtırma sorularına yönelik olmak üzere 3 ayrı dereceli puanlama anahtarı hazırlanmıřtır. 11 son sınıf ilköğretim matematik öğretmen adayı, 12 saatlik seminer sonucunda 304 öğrenciye ait testleri puanlarken rubrięi kullanmıřlardır. Öğretmen adaylarının maddelere verdikleri puanlar üzerinde %91,1 oranında bir uzlařma saęlanmıřtır. Ölçme aracına ait uygulama sonuçlarından elde edilen iç tutarlılık katsayısı Cronbach α kullanılarak hesaplanmıř ve .86 bulunmuřtur. Madde toplam test korelasyonlarına bakılarak da madde ayırıcılık indekslerinin tümünün .50 ile .71 arasında olduęu gözlenmiřtir. Ayrıca orantısal akıl yürütmeye yönelik hazırlanan ölçme aracındaki maddelerin ortaya koydukları yapıyı belirlemek amacıyla faktör analizi yapılmıř ve iki faktörlü bir yapı olduęu görölmüřtür. İlk faktörde toplanan maddelerin sayı kullanılarak oluřturulan (niceliksel karşılařtırma, ters orantı, verilmeyen değeri bulma) maddeler, ikinci faktörde toplanan maddelerin ise sayı kullanılmadan niteliksel karşılařtırma yapmayı gerektiren maddeler olduęu görölmüřtür. Aynı türden soruların aynı faktörde yer alması, ölçme aracının yapı geçerlilięinin yüksek olduęuna kanıt olarak kabul edilmiřtir.

Sonuç olarak, alandaki çalıřmalar incelendięinde, puanlama yönergesi (rubrik) kullanımının, öğretmen, farklı zamanlarda olsa bile kararlı puanlar vermesini ya da farklı öğretmenler arasında birbirleriyle daha tutarlı puanlamalar yapılmasını, başka bir deyiřle

puanlama güvenilirliğini sağlayıp sağlamadığına ilişkin bir çalışmaya rastlanmamıştır. Bu konuda yapılacak bir çalışmanın güvenilirlik konusunda Eğitimde Ölçme ve Değerlendirme bilim dalına önemli katkılar sağlayacağına inanılmaktadır.

Araştırmanın Amacı ve Önemi

Bu çalışmada puanlama yönergesi kullanmanın puanlayıcı güvenilirliğini artırıp artırmadığı, puanlama yönergesi kullanıldığında ve kullanılmadığında yapılan puanlamalar arasında bir ilişki olup olmadığı, bir ilişki varsa bunun farklı zamanlarda ve farklı puanlayıcılar söz konusu olduğunda nasıl bir değişme gösterdiği araştırılması amaçlanmıştır. Özellikle üst düzey bilişsel becerilerin objektif bir şekilde ölçülmesinde bu çalışmanın bulgularından faydalanılabileceğine inanılmaktadır. Bir öğrenme ürününün, bir seçki dosyasının veya bir ödevin puanlanması aşamasında, bağımsız puanlayıcıların puanlama yönergesi kullanarak aynı puanları elde edip etmediğinin ya da bir puanlayıcının puanlama yönergesi kullanarak zamanla kararlı puanlar verip vermediğinin belirlenmesi güvenilirlik konusunda öğretim sürecine önemli katkılar sağlayabilir.

I.4.Problem Cümlesi

Aynı ve farklı öğretmenlerin, aynı ve farklı zamanlarda puanlama yönergesi (rubrik) kullanarak ve kullanmadan verdikleri puanlar arasında bir ilişki var mıdır; varsa nasıldır? Verilen bu puanların ortalamaları arasında manidar bir fark var mıdır; varsa nasıldır?

I.4.1.Alt Problemler

A) Aynı puanlayıcıların

- a) puanlama yönergesi (rubrik) kullanarak ve kullanmadan
 - i) verdikleri puanlar arasında manidar bir ilişki var mıdır;
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?
- b) aradan belli bir süre geçtikten sonra puanlama yönergesi (rubrik) kullanarak ve kullanmadan
 - i) verdikleri puanlar arasında manidar bir ilişki var mıdır;
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?
- c) puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra tekrar puanlama yönergesi kullanarak
 - i) verdikleri puanlar arasında manidar bir ilişki var mıdır;
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?
- d) puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra tekrar puanlama yönergesi kullanmadan
 - i) verdikleri puanlar arasında manidar bir ilişki var mıdır;
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?
- e) puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra puanlama yönergesi kullanmadan
 - i) verdikleri puanlar arasında manidar bir ilişki var mıdır;

ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

f) puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra puanlama yönergesi kullanarak

i) verdikleri puanlar arasında manidar bir ilişki var mıdır;

ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

B) Farklı puanlayıcıların

a) puanlama yönergesi (rubrik) kullanarak

i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki var mıdır;

ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

b) puanlama yönergesi (rubrik) kullanmadan

i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki var mıdır;

ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

c) belli bir süre sonra puanlama yönergesi kullanarak

i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki var mıdır;

ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

- d) belli bir süre sonra puanlama yönergesi kullanmadan
- i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki var mıdır;
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?

Sayıtlar

- 1) Öğretmenler içtenlikle ve ciddiyetle puanlama yapmışlardır.
- 2) Her bir öğretmenin puanlama yönergesi kullanmadan ve kullanarak verdiği puanlar birbirini etkilememektedir.
- 3) Farklı öğretmenlerin yaptıkları puanlamalar birbirlerinden bağımsızdır.

Sınırlılıklar

- 1) Araştırma, 2008-2009 öğretim yılı Mersin ilinde görev yapan 17 Matematik öğretmeni ile sınırlıdır.
- 2) Araştırma, grafik yorumlama becerisine yönelik hazırlanan 4 açık uçlu maddeden oluşan bir ölçme aracı ile sınırlıdır.
- 3) Araştırma, Özel Toros Koleji 8. sınıfında okuyan 50 öğrenci ile sınırlıdır.

BÖLÜM I

YÖNTEM

I.1.Araştırmanın Türü

Bu araştırma, grafik yorumlama becerisini ölçerken puanlayıcıların, puanlama yönergesi (rubrik) kullanarak ve kullanmadan yaptıkları puanlamalar arasında bir ilişki olup olmadığını ve bu ilişkinin puanlamaların, farklı zamanlarda ve farklı puanlayıcılar tarafından yapılması halinde nasıl bir değişiklik gösterdiğini saptamak amacıyla yapılmıştır. Puanlamada rubrik kullanımının objektif bir puanlama sağlayıp sağlamadığı ve puanlayıcı güvenilirliğini değiştirip değiştirmediği irdelenmiştir. Fakat araştırma sonucunda elde edilen bulgular, araştırma kapsamına girebilecek evrene genellenemeyeceğinden ve var olan bilgilere yalnızca yenilerini katmak amacıyla yapılan bir çalışma olduğundan bu araştırma “betimsel araştırmalar” grubuna girmektedir.

I.2.Çalışma Grubu

Bu araştırma betimsel bir araştırma olduğu için, bir evren ve bu evreni temsil eden bir örneklem üzerinde çalışılmamıştır. Yapılan araştırma, 2008-2009 öğretim yılında, Mersin Özel Toros Koleji 8. sınıflarında okuyan 50 öğrenci ile Tarsus ve Mersin’de çeşitli okul ve dersanelerde çalışmakta olan veya emekli 17 Matematik öğretmeni üzerinde yürütülmüştür. Araştırma başlangıçta 21 öğretmenin katılımıyla başlamış ancak 4 öğretmen çeşitli sebeplerle (katılımı gönüllü olarak devam ettirmeme, kağıtları zamanında okumama, istenen formata uygun değerlendirme yapmama vb.) araştırma kapsamı dışında bırakılmıştır.

I.3. Veri Toplama Araçları

I.3.1. Grafik Yorumlama Becerisi Ölçeği

Araştırma kapsamında, grafik yorumlama becerisine ilişkin 4 açık uçlu maddeden oluşan bir ölçme aracı geliştirilmiştir. Grafik yorumlama becerisine ilişkin ölçme aracının oluşturulması aşamasında geniş bir literatür taraması yapılmış, söz konusu becerinin tanımı, alt becerileri ve nasıl ölçüldüğü irdelenmiştir.

Pek çok araştırmacı (Bertin, 1967/1983; Carswell, 1992; Curcio, 1981a, 1981b, 1987; McKnight, 1990; Wainer, 1992) grafiği kavramanın üç düzeyi ile ilgilenmişlerdir: Grafikteki verilerin genişletilmesine odaklanan bir temel düzey (yerleştirme, çevirme), grafikte gösterilen verilerdeki ilişkileri ve ara değerleri bulmak için bir orta düzey (bütünleştirme, yorumlama), verilerden genelleme yapmayı ve grafikte örtük olan ilişkileri analiz etmeyi gerektiren bir ileri düzey (genelleme, yordama) (Friel, Bright, Curcio, 2001).

Curcio'nun (1987) terminolojisine göre, grafik kavramanın üç düzeyi şöyledir: “Verileri okuma”, “Veriler arasını okuma” ve “Verilerin ötesini okuma”.

“Verileri okuma” alt düzeyde bir bilişsel durumdur ve grafikte yer alan bilgileri göstermeyi gerektirir.

“Veriler arasını okuma”, nicelikleri karşılaştırmayı, matematiksel ilişkileri açıklamayı gerektirir.

“Verilerin ötesini okuma”, grafikte açıkça veya örtük olarak verilen bilgilerden yordama ve çıkarımlarda bulunmayı gerektirir.

Söz konusu beceriyi ölçmeye yönelik yapılmış çalışmalar incelendiğinde, her üç madde tipi hakkında ilginç bir fikir birliği ile karşılaşılmıştır. Öğrenciler, “verileri okuma” konusunda çok az sıkıntı yaşamaktadırlar, fakat “veriler arasını okuma” konusunda oldukça fazla hatalar yapılmaktadır (Pereira, Mendoza ve Mellor, 1991;

Wainer, 1980; Curcio, Bright ve Friel, 2001; Curcio, 1987). Bu hatalar daha çok matematik bilgisi, okuma /dil hataları, ölçek hataları veya eksenleri okuma hataları olmaktadır. “Verilerin ötesini okuma” maddeleri ise daha belirsiz görülmektedir. Öğrenciler, verileri yorumlamak, veri gruplarını karşılaştırmak, bilinmeyen bir durumla ilgili yordama yapmak, genelleme yapmak veya bir eğilimi belirlemek için, gösterimden sonuçlar çıkarmak zorundadırlar.

Bright, Friel ve Curcio'nun (2001) oluşturduğu taksonomiye göre başlangıç düzeyindeki maddelerde, verilerden bilgi çıkarmak amaçlanmıştır. Grafikte açıkça cevabı duran maddeleri cevaplamak için grafikten veriler taşınır. Orta düzeydeki maddeleri cevaplamak için, veriler arasındaki ilişkileri bulmak amaçlanmıştır. İleri düzeydeki maddeleri cevaplamak için verilerin ötesine geçmek amaçlanmıştır. Verilen gösterimden sonuçlar çıkarılır, yordamalarda bulunulur.

Grafik yorumlama becerisini ölçmek için ikisi çizgi ve ikisi sütun grafiği olmak üzere 4 açık uçlu maddeden oluşan bir ölçme aracı hazırlanmıştır (bkz. Ek 1). Çizgi grafiğine yönelik hazırlanmış olan maddelerden biri tek değişkenli (zamana bağlı olarak bisikletlinin aldığı yol), diğeri çift değişkenlidir (aylara göre ev ve cep telefonu faturaları). Aynı şekilde sütun grafiğine yönelik hazırlanmış olan maddelerden biri tek değişkenli (zamana bağlı olarak havuzda biriken su miktarı), diğeri çift değişkenlidir (saatlere göre trenlerden inen ve trenlere binen insan sayıları). Maddelerin ortak yönleri ise, öğrencilerden grafiklere dayalı olarak hikaye oluşturmalarının istenmesidir. Gözlenecek olan beceri ve beklenen davranışlar hepsinde aynıdır. Hikayeyi oluştururken öğrencilerin kurmuş oldukları cümlelerden, yapmış oldukları yorum ve çıkarımlardan bu beceriyi kazanma düzeylerine ilişkin değerlendirmeler yapılmıştır.

İlk ölçme aracı, araştırmacı tarafından hazırlandıktan sonra, matematik öğretmenlerinin ve ölçme-değerlendirme uzmanlarının değerlendirmelerine sunulmuştur. 21 matematik öğretmeni maddeleri, amaca uygunluk ve ifadelerin açıklığı açısından incelemişler, 2 ölçme-değerlendirme uzmanı ise teknik açıdan incelemiş ve önerileri doğrultusunda ölçme aracı yeniden düzenlenmiştir.

Hazırlanan ölçme aracı ile öncelikle Tarsus Özel Toros Koleji 8. sınıfından 10 öğrenci üzerinde pilot uygulama yapılmış, öğrencilerin de görüşleri alınıp aksayan yönler olup olmadığı tespit edilmiştir. Öğrencilerin görüşleri (uygulama süresinin yetersiz olması, cevapların yazılmasında yeterince boşluk bırakılmamış olması vb.) üzerinde yapılan nitel analizler sonucunda ölçme aracına son şekli verilip asıl uygulama Mersin Özel Toros Koleji 8. sınıflarında okuyan 50 öğrenci üzerinde yapılmıştır. Ölçme aracının son hali EK 1’de sunulmuştur.

I.3.2.Değerlendirme Ölçeği

Oluşturulan ölçme aracına verilen yanıtların değerlendirilmesinde, öğrencilerin grafik yorumlama becerilerini belirlemek ve bu sürecin her adımını inceleyebilmek için bir puanlama yönergesi (rubrik) hazırlanması amaçlanmıştır.

Grafik yorumlama becerisine ilişkin puanlama yönergesi hazırlanırken çalışmanın “Giriş” kısmında belirtilen işlem basamakları takip edilmiştir. Performansı oluşturan bileşenlerin, ayrı ayrı belirlenen kriterler doğrultusunda birbirinden bağımsız değerlendirilmesi söz konusu olduğundan analitik rubrik tercih edilmiştir. Daha sonra performans kriterleri, başka bir deyişle performansı ortaya koyarken öğrencide gözlenmek istenen kritik davranışlar geniş literatür taraması sonucu belirlenmiştir. Bu davranışlar şöyledir:

*Grafikte yer alan eksenlerin temsil ettikleri anlamları fark eder.

*Grafığın bileşenlerini fark eder.

*Grafığın bileşenleri arasındaki ilişkileri fark eder.

*Grafikteki bileşenlerin, bilgilerin gösterimi üzerindeki etkilerini fark eder.

*Grafikte yer alan değerler ve aralıklar arasında ayrıntılı karşılaştırmalar yapar.

*Grafikte yer alan artış-azalış, azlık-çokluk, büyüklük-küçüklük ilişkilerini fark eder.

*Grafikteki yönelimden hareketle grafiğin bütününde nasıl bir eğilim olduğuna karar verir (Curcio, 1987; Wood, 1968; Joliffe, 1991; NCTM, 2000).

Performans kriterleri gözlenebilir bir şekilde belirlendikten sonra performans alt alanları oluşturulmuştur. Eksenlerin Anlamı, Kesişme Noktalarının Anlamı, Karşılaştırmalar ve Bütünleştirme olmak üzere 4 performans alt alanı belirlenmiştir.

Performans kriterleri belirlendikten sonra bu kriterlerin ne derece karşılandığını gösteren performans düzeylerinin belirlenmesi aşamasına geçilmiştir. Rubriğin performans düzeyini gösteren kategoriler, rakamlarla veya betimleyici ifadelerle ya da her ikisi birlikte kullanılarak belirtilebilir. Kategori sayısına ilişkin belirli bir sınırlama yoktur. Fakat ölçülen davranışın gözlenmesini güç hale getirecek kadar çok, tam anlamıyla kapsamayacak kadar az olmamalıdır. Genellikle 3-5 arasında düzey sayısı önerilir (Kan, 2007). Bu çalışmada performans düzeyleri 0-1-2-3-4 olarak belirlenmiştir (bkz. Ek 2).

Bir sonraki aşamada ise performans kriterlerine ve düzeylerine ilişkin gözlenebilir özellikleri içeren performans tanımları oluşturulmuştur. Performans tanımları yapılmadan önce performans görevine ilişkin gözlenebilir özellikler ve bu özelliklere ait

göstergeler listelenmiştir. Özelliklere ait göstergeler, her bir kriterin karşılanması için öğrenciden ne beklendiğine ilişkin detaylı bilgi sağlar. Göstergeler, performans kriterlerine en üst düzeyde sahip öğrenci davranışından en düşük düzeyde sahip öğrenci davranışına doğru sıralanmış ve puanlama yönergesine (rubrik) düzeylere uygun olacak şekilde yerleştirilmiştir.

Geliştirilen puanlama yönergesi, uygulanmadan önce son şeklinin verilmesi için 21 matematik öğretmeni, 2 ölçme ve değerlendirme uzmanı ve 2 matematik eğitimi alan uzmanının değerlendirmelerine sunulmuştur. Matematik öğretmenleri ve matematik alan uzmanları, geliştirilen puanlama yönergesinin grafik yorumlama becerisini ölçme amaçlı hazırlanan ölçme aracına uygunluğunu performans, kriter ve tanımları açısından incelemişlerdir. Ölçme ve Değerlendirme uzmanları ise puanlama yönergesini teknik ve dil açısından incelemişlerdir. Uzmanların öneri ve eleştirileri doğrultusunda puanlama yönergesine son şekli verilmiş ve öğretmenler tarafından değerlendirmelerde kullanılacak duruma getirilmiştir. Puanlama yönergesinin son hali Ek 2’de sunulmuştur.

I.4. İşlem

Tablo I.1.’de verilen deney düzeneğinde de görüldüğü gibi öğretmenler, grafik yorumlama becerisini ölçmeye yönelik 50 öğrenciye uygulanmış 4 adet açık uçlu maddeden oluşan testleri önce kendi kriterleri doğrultusunda puanlamışlardır (1. okuma). İlk uygulamadan sonra öğretmenlere puanlama yönergeleri konusunda kısa bir eğitim verilmiştir. Bellek etkisini önlemek için belirlenen 2 haftalık süre sonunda puanlama yönergesini kullanarak aynı öğretmenler, aynı kağıtları tekrar puanlamışlardır (2. okuma). Çalışmada zamanın, puanlayıcı güvenilirliği üzerindeki etkisini saptamak açısından belirlenen 10 haftalık süre geçtikten sonra, aynı kağıtlar aynı öğretmenlere puanlama

yönergesini kullanmadan tekrar puanlatılmıştır (3. okuma). Tekrar bellek etkisini önlemek için belirlenen 2 haftalık süre geçtikten sonra, aynı öğretmenlere aynı kağıtlar puanlama yönergelerini kullanarak tekrar puanlatılıp (4. okuma) sonuçlar rapor edilmiştir. Puanlamada objektifliği sağlamak için de öğrenci isimleri yerine kod numaraları kullanılmıştır.

TABLO I.1.

PUANLAMALARA VE ÖĞRETMENLERE İLİŞKİN DENEY DÜZENEGİ

	1.öğretmen	2.öğretmen	3.öğretmen	...	n.öğretmen			
Okuma1	1.öğrenci			...		Ortalama	Önce	
Puanlama	2.öğrenci					Okuma1		
yönergesiz	...							
	m.öğrenci							
	ortA(Öğretmen1)							
2 haftalık süre								
Okuma2	1.öğrenci			...		Ortalama	Önce	
Puanlama	2.öğrenci					Okuma2		
yönergeli	...							
	m.öğrenci							
	ortB(Öğretmen1)							
10 haftalık süre								
Okuma3	1.öğrenci			...		Ortalama	Sonra	
Puanlama	2.öğrenci					Okuma3		
yönergesiz	...							
	m.öğrenci							
	OrtC(Öğretmen1)							
2 haftalık süre								
Okuma4	1.öğrenci			...		Ortalama	Sonra	
Puanlama	2.öğrenci					Okuma4		
yönergeli	...							
	m.öğrenci							
	ortD(Öğretmen1)							

I.5.Verilerin Çözümlemesi

A) Aynı puanlayıcıların

a) Puanlama yönergesi (rubrik) kullanarak ve kullanmadan

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı

Pearson Momentler Çarpımı Korelasyon Katsayısı ile

hesaplanmış, sonuçların manidarlığı t testi ile;

ii) verdikleri puanların ortalamaları arasında manidar bir fark

olup olmadığı bağımlı örneklem için t testi ile

incelenmiştir.

b) Aradan belli bir süre geçtikten sonra puanlama yönergesi (rubrik)

kullanarak ve kullanmadan

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı

Pearson Momentler Çarpımı Korelasyon Katsayısı ile

hesaplanmış, sonuçların manidarlığı t testi ile;

ii) verdikleri puanların ortalamaları arasında manidar bir fark

olup olmadığı bağımlı örneklem için t testi ile

incelenmiştir.

c) Puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra tekrar

puanlama yönergesi kullanarak

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı

Pearson Momentler Çarpımı Korelasyon Katsayısı ile

hesaplanmış, sonuçların manidarlığı t testi ile;.

ii) verdikleri puanların ortalamaları arasında manidar bir fark

olup olmadığı bağımlı örneklem için t testi ile

incelenmiştir.

d) Puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra tekrar puanlama yönergesi kullanmadan

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmış, sonuçların manidarlığı t testi ile;

ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı bağımlı örneklem için t testi ile incelenmiştir.

e) Puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra puanlama yönergesi kullanmadan

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmış, sonuçların manidarlığı t testi ile;

ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı bağımlı örneklem için t testi ile incelenmiştir.

f) Puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra puanlama yönergesi kullanarak

i) verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmış, sonuçların manidarlığı t testi ile;

ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı bağımlı örneklem için t testi ile

incelenmiştir.

- iii) puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un (1941) varyans analizi üzerine temellenen güvenilirlik kestirimleri (G katsayısı) hesaplanmıştır.

B) Farklı puanlayıcıların

a) Puanlama yönergesi (rubrik) kullanarak

- i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmıştır. İlişkinin manidarlığı ise t testi ile test edilmiştir.

- ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı tek yönlü varyans analizi ile test edilmiştir. Varsa farklılığın hangi gruplardan kaynaklandığını bulmak amacıyla post hoc testlerinden biri olan Scheffe testine başvurulmuştur.

- iii) puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un (1941) varyans analizi üzerine temellenen güvenilirlik kestirimleri (G katsayısı) hesaplanmıştır.

b) Puanlama yönergesi (rubrik) kullanmadan

- i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmıştır. İlişkinin manidarlığı ise t testi ile test edilmiştir.

- ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı tek yönlü varyans analizi ile test edilmiştir. Varsa farklılığın hangi gruplardan kaynaklandığını bulmak amacıyla post hoc testlerinden biri olan Scheffe testine başvurulmuştur.
 - iii) puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un (1941) varyans analizi üzerine temellenen güvenilirlik kestirimleri (G katsayısı) hesaplanmıştır.
- c) Belli bir süre sonra puanlama yönergesi kullanarak
- i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmıştır. İlişkinin manidarlığı ise t testi ile test edilmiştir.
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı tek yönlü varyans analizi ile test edilmiştir. Varsa farklılığın hangi gruplardan kaynaklandığını bulmak amacıyla post hoc testlerinden biri olan Scheffe testine başvurulmuştur.
 - iii) puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un (1941) varyans analizi üzerine temellenen güvenilirlik kestirimleri (G katsayısı) hesaplanmıştır.

- d) Belli bir süre sonra puanlama yönergesi kullanmadan
- i) ikili karşılaştırıldıklarında verdikleri puanlar arasında manidar bir ilişki olup olmadığı Pearson Momentler Çarpımı Korelasyon Katsayısı ile hesaplanmıştır. İlişkinin manidarlığı ise t testi ile test edilmiştir.
 - ii) verdikleri puanların ortalamaları arasında manidar bir fark olup olmadığı tek yönlü varyans analizi ile test edilmiştir. Varsa farklılığın hangi gruplardan kaynaklandığını bulmak amacıyla post hoc testlerinden biri olan Scheffe testine başvurulmuştur.
 - iii) puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un (1941) varyans analizi üzerine temellenen güvenilirlik kestirimleri (G katsayısı) hesaplanmıştır.

BÖLÜM II

BULGULAR VE YORUM

Bu bölümde araştırmada elde edilen bulgular ve bu bulgulara dayalı yorumlar her bir alt problem için ayrı ayrı verilmiştir.

Alt problemlere ilişkin analizler yapılmadan önce, testin faktör yapısı belirlenmiştir (bkz. Ek 3). Temel bileşenler faktör analizi sonucunda, 4 farklı okumada da testin tek faktörlü bir yapıya sahip olduğu görülmüştür. 1. okumada öz değeri 1'den büyük olan tek bir faktör vardır ve bu faktör toplam varyansın % 81.098'ini açıklamaktadır. 2. okumada öz değeri 1'den büyük olan yine tek bir faktör vardır ve bu faktör toplam varyansın % 88.098'ini açıklamaktadır. 3. okumada da öz değeri 1'den büyük olan bir faktör vardır ve bu faktör toplam varyansın % 87.311'ini açıklamaktadır. 4. okumada da öz değeri 1'den büyük olan bir faktör vardır ve bu faktör toplam varyansın % 90.108'ini açıklamaktadır. Test, tek faktörlü olduğundan, başka bir deyişle tüm maddeleri aynı yapıyı ölçtüğünden toplam puanlar üzerinden istatistiksel analizler yapılmıştır.

Araştırmada elde edilen verilerin dağılımlarını incelemek üzere Kolmogorov-Smirnov normallik testi yapılmış, sonuçlar Ek 4'te sunulmuştur. Her öğretmenin, her okumada vermiş olduğu puan dağılımları incelendiğinde, çoğunun normal dağıldığı görülmektedir. Değişkenlerin her biri normal dağılsa veya normale dönüştürülse bile, bu değişkenlerin tüm lineer kombinasyonlarının normal dağılacığının bir garantisi yoktur (Tabachnick ve Fidell, 1996). Bu durum göz önünde bulundurularak, normal dağılmamış olan verilerin çarpıklık ve basıklık katsayıları incelenmiş, normal sayılabilecek değerlerde oldukları görülmüştür. Böylece tüm dağılımların normal oldukları varsayılarak istatistiksel analizler yapılmıştır.

II.1. “Aynı öğretmenlerin puanlama yönergesi kullanmadan ve kullanarak verdiği puanlar arasında manidar bir ilişki var mıdır? Verilen bu puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemine yanıt bulabilmek için, aynı öğretmenlerin puanlama yönergesi kullanmadan (1.okumalar) ve kullanarak (2. okumalar), 50 öğrencinin cevapladığı 4 açık uçlu maddeden oluşan teste vermiş oldukları puanlar arasındaki korelasyon katsayıları hesaplanmış, korelasyon katsayılarının manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 1. okumalarında hem de 2. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı “bağımlı örneklem için t testi” ile test edilmiştir.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve kullanarak verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.1’de verilmiştir.

TABLO II.1.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve kullanarak vermiş olduğu puanlar arasındaki korelasyon hesaplamaları

Öğretmen	1. ve 2. okuma arasındaki korelasyon	p
1	0,790	0,00
2	0,802	0,00
3	0,746	0,00
4	0,837	0,00
5	0,753	0,00
6	0,841	0,00
7	0,751	0,00

8	0,805	0,00
9	0,925	0,00
10	0,730	0,00
11	0,796	0,00
12	0,851	0,00
13	0,586	0,00
14	0,669	0,00
15	0,702	0,00
16	0,741	0,00
17	0,836	0,00

($p < 0.01$ düzeyinde anlamlı)

Tablodaki bulgular, öğretmenlerin puanlama yönergesi (rubrik) kullanmadan ve kullanarak verdiği puanlar arasındaki korelasyonların yüksek ve manidar olduğuna dair ipuçları vermektedir. Öğretmenlerin farklı türde yapmış olduğu puanlamalar arasında pozitif yönde bir ilişki bulunmaktadır. Puanlama yönergesi kullanarak ve kullanmadan yapılan puanlamalar arasında en yüksek korelasyon 0,925 ile 9. öğretmene, en düşük korelasyon ise 0,586 ile 13. öğretmene aittir. Bu durumda 13. öğretmenin 1. ve 2. okumalarında verdiği puanların birbiriyle çok ilişkili olmadığı, 9. öğretmenin ise birbiriyle ilişkili puanlar verdiği söylenebilir. Bu durum, çoğu öğretmen için puanlamada rubrik kullanıp kullanmamının fark etmediği şeklinde yorumlanabilir. Ancak, öğretmenlerin 1. ve 2. okumalara ilişkin korelasyon değerlerinin arasındaki farklılıklar ($s = 0,078$), puanlamada, puanlayıcıların bir değişim kaynağı olabileceğini de göstermektedir.

TABLO II.2.

**Aynı öğretmenlerin puanlama yönergesi kullanmadan ve kullanarak vermiş olduğu
puanlara ilişkin t-testi sonuçları**

Öğretmen	Okumalar	ort	Std.sapma	N	t	p
1	1.okuma	51,52	23,28	50	-7,519	0,00
	2.okuma	66,90	16,00			
2	1.okuma	55,90	20,49	50	-7,372	0,00
	2.okuma	68,90	18,83			
3	1.okuma	50,36	17,57	50	0,394	0,695
	2.okuma	49,65	17,80			
4	1.okuma	59,16	25,38	50	-8,781	0,00
	2.okuma	76,75	18,53			
5	1.okuma	53,84	24,37	50	3,069	0,003
	2.okuma	46,87	18,56			
6	1.okuma	59,86	18,04	50	8,059	0,00
	2.okuma	48,72	15,70			
7	1.okuma	44,20	22,87	50	-8,691	0,00
	2.okuma	63,28	13,55			
8	1.okuma	51,74	20,40	50	-10,600	0,00
	2.okuma	70,03	17,92			
9	1.okuma	74,20	16,64	50	-4,633	0,00
	2.okuma	78,43	14,03			
10	1.okuma	50,22	19,49	50	-0,230	0,819
	2.okuma	50,65	15,85			
11	1.okuma	48,64	22,39	50	-6,407	0,00

	2.okuma	61,15	15,21			
12	1.okuma	67,64	16,21	50	8,687	0,00
	2.okuma	56,96	12,02			
13	1.okuma	47,06	14,29	50	-7,919	0,00
	2.okuma	61,40	13,84			
14	1.okuma	46,22	19,92	50	-4,084	0,00
	2.okuma	54,78	14,15			
15	1.okuma	40,70	20,22	50	-11,022	0,00
	2.okuma	63,65	17,23			
16	1.okuma	50,04	19,40	50	-2,127	0,039
	2.okuma	53,96	13,41			
17	1.okuma	43,88	20,65	50	-13,016	0,00
	2.okuma	65,87	13,49			

Aynı öğretmenlerin rubrik kullanmadan ve kullanarak verdiği puanlar arasındaki farkın manidarlığını test etmek için öğretmenlerin rubrik kullanmadan ve kullanarak vermiş olduğu puanların ortalamaları arasında fark olup olmadığı bağımlı örneklem için t testi ile test edilmiştir.

Tabloda görüldüğü gibi, sadece 3. ve 10. öğretmenlerin rubrik kullanmadan ve kullanarak verdikleri puanlar arasında manidar farklılıklar yok iken, diğer öğretmenlerin verdikleri puanlar arasında manidar farklılıklar bulunmaktadır ($p < 0,01$). 5., 6. ve 12. öğretmenlerde rubrik kullanımının, verdikleri puanların ortalamalarında bir artışa sebep olduğu, diğerlerinde ise düşürdüğü söylenebilir. Artış ve azalış şeklinde puanlarının değişmesi, puanlayıcıların önemli bir değişim kaynağı olduğu şeklinde yorumlanabilir. Bu

ise, puanlamada, hangi bilişsel süreçlerin cimrilik ya da cömertliğe yol açtığıнын incelenmesinin gerekliliğine işaret etmektedir. Bernardin ve Orban'a (1990) göre dereceleme biçimi öğretmenlerin cömert ya da cimri davranmalarını etkilemektedir.

II.2. “Aynı öğretmenlerin aradan belli bir süre geçtikten sonra puanlama yönergesi (rubrik) kullanmadan ve kullanarak verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemine yanıt bulabilmek için, aynı öğretmenlerin aradan belli bir süre geçtikten sonra puanlama yönergesi kullanmadan (3.okumalar) ve kullanarak (4. okumalar) vermiş oldukları puanlar arasındaki korelasyon katsayıları hesaplanmış, korelasyon katsayılarının manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 3. okumalarında hem de 4. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı “bağımlı örneklem için t testi” ile test edilmiştir.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve kullanarak verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.3'te verilmiştir.

TABLO II.3.

Aynı öğretmenlerin aradan belli bir süre geçtikten sonra puanlama yönergesi kullanmadan ve kullanarak vermiş olduğu puanlar arasındaki korelasyon hesaplamaları

Öğretmen	3. ve 4.okuma arasındaki korelasyon	p
1	0,852	0,00
2	0,868	0,00
3	0,838	0,00
4	0,811	0,00
5	0,749	0,00
6	0,846	0,00
7	0,755	0,00
8	0,667	0,00
9	0,894	0,00
10	0,674	0,00
11	0,807	0,00
12	0,711	0,00
13	0,814	0,00
14	0,749	0,00
15	0,923	0,00
16	0,747	0,00
17	0,819	0,00

($p < 0.01$ düzeyinde anlamlı)

Tablodaki bulgular, öğretmenlerin aradan belli bir süre geçtikten sonra puanlama yönergesi (rubrik) kullanmadan ve kullanarak 3. ve 4. okumalarda verdikleri puanlar arasındaki korelasyonların yüksek ve manidar olduğuna dair ipuçları vermektedir. Öğretmenlerin rubrik kullanarak ve kullanmadan verdiği puanlar arasındaki en yüksek korelasyon 0,923 ile 15. öğretmene, en düşük korelasyon ise 0,667 ile 8. öğretmene aittir. Bu durumda en yüksek tutarlılık 15. öğretmene, en düşük tutarlılık ise 8. öğretmene aittir. Bu durum, çoğu öğretmen için puanlamada rubrik kullanıp kullanmamanın fark etmediği şeklinde yorumlanabilir. Ancak, öğretmenlerin 3. ve 4. okumalara ilişkin korelasyon değerlerinin arasındaki farklılıklar ($s = 0,073$), puanlamada, puanlayıcıların bir değişim kaynağı olabileceğini de göstermektedir.

TABLO II.4.

Aynı öğretmenlerin aradan belli bir süre geçtikten sonra puanlama yönergesi kullanmadan ve kullanarak vermiş olduğu puanlara ilişkin t-testi sonuçları

Öğretmen	Okumalar	ort	Std.sapma	N	t	p
1	3.okuma	45,24	21,10	50	-10,693	0,00
	4.okuma	62,09	16,49			
2	3.okuma	63,20	21,80	50	-3,644	0,001
	4.okuma	68,81	17,78			
3	3.okuma	45,06	16,88	50	-3,876	0,00
	4.okuma	50,56	18,16			
4	3.okuma	64,70	22,84	50	-10,598	0,00
	4.okuma	85,75	14,18			
5	3.okuma	58,70	20,89	50	5,071	0,00
	4.okuma	48,31	19,87			

6	3.okuma	56,80	15,90	50	14,924	0,00
	4.okuma	38,87	13,70			
7	3.okuma	51,52	20,62	50	-8,230	0,00
	4.okuma	67,40	13,79			
8	3.okuma	56,44	23,49	50	-2,389	0,021
	4.okuma	62,37	14,29			
9	3.okuma	75,40	16,74	50	-3,849	0,00
	4.okuma	79,50	14,41			
10	3.okuma	55,26	20,99	50	4,921	0,00
	4.okuma	44,47	13,91			
11	3.okuma	61,16	20,70	50	-1,133	0,263
	4.okuma	63,15	14,34			
12	3.okuma	57,76	14,68	50	-4,436	0,00
	4.okuma	64,65	14,19			
13	3.okuma	40,24	14,22	50	-16,353	0,00
	4.okuma	61,28	15,55			
14	3.okuma	51,56	23,44	50	0,224	0,824
	4.okuma	51,06	15,48			
15	3.okuma	50,10	23,91	50	-8,592	0,00
	4.okuma	62,84	17,05			
16	3.okuma	36,34	14,51	50	-15,516	0,00
	4.okuma	58,78	14,25			
17	3.okuma	56,46	15,85	50	-7,352	0,00
	4.okuma	65,93	13,43			

Aynı öğretmenlerin aradan belli bir süre geçtikten sonra aynı kağıtlara rubrik kullanmadan ve kullanarak verdikleri puanların ortalamaları arasında fark olup olmadığı bağımlı örneklem için t testi ile test edilmiştir.

Tabloda görüldüğü gibi, 8., 11. ve 14. öğretmenlerin rubrik kullanmadan ve kullanarak vermiş olduğu puanların ortalamaları arasında manidar farklılıklar bulunmazken, diğerlerinde manidar farklılıklar gözlenmiştir. Artış ve azalış şeklinde puanlarının değişmesi puanlayıcıların, önemli bir değişim kaynağı olduğu şeklinde yorumlanabilir. Öğretmenlerin, puanlama yönergesi kullandıklarında, genel olarak daha cömert davrandıkları görülmektedir.

II. 3.“Aynı öğretmenlerin puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra tekrar puanlama yönergesi kullanmadan verdiği puanlar arasında manidar bir ilişki var mıdır? Verilen puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemi için, aynı öğretmenlerin aynı kağıtlara puanlama yönergesi kullanmadan (1. okuma) verdikleri puanlar ile aradan belli bir süre geçtikten sonra aynı öğretmenlerin yine puanlama yönergesi kullanmadan (3. okuma) verdikleri puanlar arasındaki korelasyonlar hesaplanmış ve manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 1. okumalarında hem de 3. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı “bağımlı örneklem için t testi” ile test edilmiştir.

Aynı öğretmenlerin farklı zamanlarda puanlama yönergesi kullanmadan verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.5’te verilmiştir.

TABLO II.5.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve aradan belli bir süre geçtikten sonra tekrar puanlama yönergesi kullanmadan vermiş olduğu puanlar arasındaki korelasyon hesaplamaları

Öğretmen	1. ve 3. okuma arasındaki korelasyon	p
1	0,920	0,00
2	0,834	0,00
3	0,852	0,00
4	0,926	0,00
5	0,741	0,00
6	0,818	0,00
7	0,874	0,00
8	0,828	0,00
9	0,854	0,00
10	0,703	0,00
11	0,843	0,00
12	0,804	0,00
13	0,666	0,00
14	0,522	0,00
15	0,771	0,00
16	0,625	0,00
17	0,769	0,00

($p < 0,01$)

Tabloda görüldüğü gibi, aynı öğretmenlerin farklı zamanlarda puanlama yönergesi (rubrik) kullanmadan verdiği puanlar arasındaki korelasyonlar yüksek ve

manidardır. Aynı öğretmenlerin farklı zamanlarda rubrik kullanmadan verdikleri puanlar arasındaki en yüksek korelasyon, 0,926 ile 4. öğretmene, en düşük korelasyon ise 0,522 ile 14. öğretmene aittir. En yüksek tutarlılık 4. öğretmene, en düşük tutarlılık ise 14. öğretmene aittir. Öğretmenlerin 1. ve 3. okumalara ilişkin korelasyon değerlerinin arasındaki farklılıklar ($s= 0,106$), puanlamada, puanlayıcıların bir değişim kaynağı olabileceğini de göstermektedir.

TABLO II.6.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve aradan belli bir süre geçtikten sonra tekrar puanlama yönergesi kullanmadan vermiş oldukları puanlara ilişkin t-testi sonuçları

Öğretmen	Okumalar	ort	Std.sapma	N	t	p
1	1.okuma	51,52	23,28	50	4,856	0,00
	3.okuma	45,24	21,10			
2	1.okuma	55,90	20,49	50	-4,212	0,00
	3.okuma	63,20	21,80			
3	1.okuma	50,36	17,57	50	3,990	000
	3.okuma	45,06	16,88			
4	1.okuma	59,16	25,38	50	-4,072	0,00
	3.okuma	64,70	22,84			
5	1.okuma	53,84	24,37	50	-2,068	0,044
	3.okuma	58,70	20,89			
6	1.okuma	59,86	18,04	50	2,071	0,044
	3.okuma	56,80	15,90			

7	1.okuma	44,20	22,87	50	-4,646	0,00
	3.okuma	51,52	20,62			
8	1.okuma	51,74	20,40	50	-2,517	0,015
	3.okuma	56,44	23,49			
9	1.okuma	74,20	16,64	50	-0,942	0,351
	3.okuma	75,40	16,74			
10	1.okuma	50,22	19,49	50	-2,276	0,027
	3.okuma	55,26	20,99			
11	1.okuma	48,64	22,39	50	-7,275	0,00
	3.okuma	61,16	20,70			
12	1.okuma	67,64	16,21	50	7,135	0,00
	3.okuma	57,76	14,68			
13	1.okuma	47,06	14,29	50	4,137	0,00
	3.okuma	40,24	14,22			
14	1.okuma	46,22	19,92	50	-1,764	0,084
	3.okuma	51,56	23,44			
15	1.okuma	40,70	20,22	50	-4,333	0,00
	3.okuma	50,10	23,91			
16	1.okuma	50,04	19,40	50	6,316	0,00
	3.okuma	36,34	14,51			
17	1.okuma	43,88	20,65	50	-6,735	0,00
	3.okuma	56,46	15,85			

Yukarıdaki tablo incelendiğinde 5., 6., 8., 9., 10. ve 14. öğretmenlerin farklı zamanlarda rubrik kullanmadan verdikleri puanların ortalamaları arasında manidar

farklılıklar olmadığı, diğerlerinde manidar farklılıklar ($p<0,01$) olduğu söylenebilir. 1., 3., 5., 12., 13. ve 16. öğretmenlerde araya zaman girmesinin, verdikleri puanların ortalamalarında bir düşüşe sebep olduğu, diğerlerinde ise artırdığı yorumu yapılabilir. Artış ve azalış şeklinde puanlarının değişmesi puanlayıcıların ya da aradan geçen zamanın, önemli birer değişim kaynağı olabileceği şeklinde yorumlanabilir.

II.4. “Aynı öğretmenlerin puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra tekrar puanlama yönergesi kullanarak verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemi için aynı öğretmenlerin aynı kağıtlara puanlama yönergesi kullanarak (2. okuma) verdikleri puanlar ile aradan belli bir süre geçtikten sonra aynı öğretmenlerin yine puanlama yönergesi kullanarak (4. okuma) verdikleri puanlar arasındaki korelasyonlar hesaplanmış ve manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 2. okumalarında hem de 4. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı bağımlı örneklemeler için t testi ile test edilmiştir.

Aynı öğretmenlerin farklı zamanlarda puanlama yönergesi kullanarak verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.7’de verilmiştir.

TABLO II.7.

Aynı öğretmenlerin puanlama yönergesi kullanarak ve aradan belli bir süre geçtikten sonra tekrar puanlama yönergesi kullanarak vermiş olduğu puanlar arasındaki korelasyon hesaplamaları

Öğretmen	2. ve 4. okuma arasındaki korelasyon	p
1	0,943	0,00
2	0,866	0,00
3	0,751	0,00
4	0,877	0,00
5	0,868	0,00
6	0,832	0,00
7	0,850	0,00
8	0,694	0,00
9	0,957	0,00
10	0,761	0,00
11	0,778	0,00
12	0,744	0,00
13	0,799	0,00
14	0,727	0,00
15	0,900	0,00
16	0,846	0,00
17	0,926	0,00

($p < 0,01$ düzeyinde anlamlı)

Tabloda görüldüğü gibi, aynı öğretmenlerin farklı zamanlarda puanlama yönergesi (rubrik) kullanarak verdiği puanlar arasındaki korelasyonlar manidar diğer okumalara kıyasla oldukça yüksek görünmektedir. Öğretmenlerin farklı zamanlarda aynı türde yapmış olduğu puanlamalar arasında pozitif yönde yüksek bir ilişki bulunmuştur.

Farklı zamanlarda rubrik kullanarak verilen puanlar arasındaki en yüksek korelasyon 0,957 ile 9. öğretmene, en düşük korelasyon ise 0,694 ile 8. öğretmene aittir. Bu durumda aynı öğretmenlerin farklı zamanlarda rubrik kullanarak yaptığı puanlamalar arasındaki tutarlılık en yüksek düzeyde 9. öğretmen tarafından sağlanırken, en düşük düzeyde 1. öğretmen tarafından sağlanmaktadır. Öğretmenlerin 2. ve 4. okumalara ilişkin korelasyon değerlerinin arasındaki farklılıklar ($s = 0,078$), puanlamada, puanlayıcıların ya da aradan geçen zamanın birer değişim kaynağı olabileceğini de göstermektedir.

TABLO II.8.

Aynı öğretmenlerin puanlama yönergesi kullanarak ve aradan belli bir süre geçtikten sonra tekrar puanlama yönergesi kullanarak vermiş oldukları puanlara ilişkin t-testi

sonuçları

Öğretmen	Okumalar	ort	Std.sapma	N	t	p
1	2.okuma	66,90	16,00	50	6,163	0,00
	4.okuma	62,09	16,49			
2	2.okuma	68,90	18,83	50	0,07	0,945
	4.okuma	68,81	17,78			
3	2.okuma	49,65	17,,80	50	-0,505	0,616
	4.okuma	50,56	18,16			
4	2.okuma	76,75	18,53	50	-6,970	0,00
	4.okuma	85,75	14,18			

5	2.okuma	46,87	18,56	50	-1,022	0,312
	4.okuma	48,31	19,87			
6	2.okuma	48,72	15,70	50	7,971	0,00
	4.okuma	38,87	13,70			
7	2.okuma	63,28	13,55	50	-3,899	0,00
	4.okuma	67,40	13,79			
8	2.okuma	70,03	17,92	50	4,150	0,00
	4.okuma	62,37	14,29			
9	2.okuma	78,43	14,03	50	-1,802	0,078
	4.okuma	79,50	14,41			
10	2.okuma	50,65	15,85	50	4,183	0,00
	4.okuma	44,47	13,91			
11	2.okuma	61,15	15,21	50	-1,431	0,159
	4.okuma	63,15	14,34			
12	2.okuma	56,96	12,02	50	-5,667	0,00
	4.okuma	64,65	14,19			
13	2.okuma	61,40	13,84	50	0,093	0,926
	4.okuma	61,28	15,55			
14	2.okuma	54,78	14,15	50	2,386	0,021
	4.okuma	51,06	15,48			
15	2.okuma	63,65	17,23	50	0,747	0,458
	4.okuma	62,84	17,05			
16	2.okuma	53,96	13,41	50	-4,409	0,00
	4.okuma	58,78	14,25			
17	2.okuma	65,87	13,49	50	-0,086	0,932
	4.okuma	65,93	13,43			

Aynı öğretmenlerin, farklı zamanlarda puanlama yönergesi (rubrik) kullanarak yapmış olduğu puanlamaların ortalamaları arasında manidar bir farklılık olup olmadığı “bağımlı örneklem için t-testi” ile test edilmiştir.

Tablodaki bulgular incelendiğinde 2., 3., 5., 9., 11., 13., 14., 15. ve 17. öğretmenlerin farklı zamanlarda puanlama yönergesi kullanarak verdikleri puanların ortalamaları arasında manidar farklılıklar bulunmazken, diğer 8 öğretmenin verdikleri puanların ortalamaları arasında manidar farklılıklar ($p < 0,01$) gözlenmiştir. 4., 7., 12. ve 16. öğretmenlerde, araya zaman girmesinin, rubrik kullanarak verdikleri puanların ortalamalarında bir yükselmeye neden olduğu, diğer 4 öğretmende ise düşürdüğü söylenebilir. Bu durumda aynı öğretmenlerin aynı kağıtları aradan belli bir süre geçtikten sonra tekrar puanlamaları durumunda rubrik kullanarak verdikleri puanların ortalamaları arasındaki farklılığın diğer okumalara kıyasla daha az olduğu yorumu yapılabilir. Artış ve azalış şeklinde puanlarının değişmesi puanlayıcıların ya da aradan geçen zamanın, önemli birer değişim kaynağı olabileceği şeklinde yorumlanabilir.

II.5. “Aynı öğretmenlerin puanlama yönergesi (rubrik) kullanmadan ve belli bir süre sonra puanlama yönergesi kullanarak verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemi için, aynı öğretmenlerin aynı kağıtlara puanlama yönergesi kullanmadan (1. okuma) verdikleri puanlar ile aradan belli bir süre geçtikten sonra aynı öğretmenlerin puanlama yönergesi kullanarak (4. okuma) verdikleri puanlar arasındaki korelasyonlar hesaplanmış ve manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 1.

okumalarında hem de 4. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı “bağımlı örneklem için t testi” ile test edilmiştir.

Aynı öğretmenlerin puanlama yönergesi kullanmadan verdiği puanlar ile belli bir süre sonra puanlama yönergesi kullanarak verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.9’da verilmiştir.

TABLO II.9.

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve belli bir süre sonra puanlama yönergesi kullanarak vermiş oldukları puanlar arasındaki korelasyon hesaplamaları

Öğretmen	1. ve 4. okuma arasındaki korelasyon	p
1	0,848	0,00
2	0,872	0,00
3	0,774	0,00
4	0,752	0,00
5	0,791	0,00
6	0,785	0,00
7	0,750	0,00
8	0,678	0,00
9	0,916	0,00
10	0,643	0,00
11	0,695	0,00
12	0,749	0,00
13	0,651	0,00

14	0,555	0,00
15	0,794	0,00
16	0,733	0,00
17	0,766	0,00

(p<0,01)

Tablodaki bulgular, aynı öğretmenlerin farklı zamanlarda puanlama yönergesi kullanmadan (1. okuma) ve kullanarak (4.okuma) verdiği puanlar arasındaki korelasyonların çok yüksek olmadığına dair ipuçları vermektedir. Farklı zamanlarda ve farklı teknikler kullanılarak yapılan puanlamalar arasındaki korelasyonların düşük çıkması zaten beklenen bir sonuçtur. En yüksek korelasyon 0,916 ile 9. öğretmene, en düşük korelasyon ise 0,555 ile 11. öğretmene aittir. Fakat korelasyonlar daha çok 0,60 ile 0,80 arasında (13 öğretmen) yığılmıştır. Öğretmenlerin 1. ve 4. okumalara ilişkin korelasyon değerlerinin arasındaki farklılıklar ($s= 0,088$), puanlamada, puanlayıcıların aradan geçen zamanın ya da kullanılan tekniğin birer değişim kaynağı olabileceğini de göstermektedir.

TABLO II.10

Aynı öğretmenlerin puanlama yönergesi kullanmadan ve belli bir süre sonra puanlama yönergesi kullanarak vermiş oldukları puanlara ilişkin t-testi sonuçları

Öğretmen	Okumalar	ort	Std.sapma	N	t	P
1	1.okuma	51,52	23,28	50	-5,853	0,00
	4.okuma	62,09	16,49			
2	1.okuma	55,90	20,49	50	-9,093	0,00
	4.okuma	68,81	17,78			
3	1.okuma	50,36	17,57	50	-0,119	0,906
	4.okuma	50,56	18,16			

4	1.okuma	59,16	25,38	50	-10,785	0,00
	4.okuma	85,75	14,18			
5	1.okuma	53,84	24,37	50	2,616	0,012
	4.okuma	48,31	19,87			
6	1.okuma	59,86	18,04	50	13,269	0,00
	4.okuma	38,87	13,70			
7	1.okuma	44,20	22,87	50	-10,587	0,00
	4.okuma	67,40	13,79			
8	1.okuma	51,74	20,40	50	-5,011	0,00
	4.okuma	62,37	14,29			
9	1.okuma	74,20	16,64	50	-5,583	0,00
	4.okuma	79,50	14,41			
10	1.okuma	50,22	19,49	50	2,713	0,009
	4.okuma	44,47	13,91			
11	1.okuma	48,64	22,39	50	-6,358	0,00
	4.okuma	63,15	14,34			
12	1.okuma	67,64	16,21	50	1,929	0,06
	4.okuma	64,65	14,19			
13	1.okuma	47,06	14,29	50	-8,028	0,00
	4.okuma	61,28	15,55			
14	1.okuma	46,22	19,92	50	-1,998	0,051
	4.okuma	51,06	15,48			
15	1.okuma	40,70	20,22	50	-12,682	0,00
	4.okuma	62,84	17,05			
16	1.okuma	50,04	19,40	50	-4,682	0,00
	4.okuma	58,78	14,25			

17	1.okuma	43,88	20,65	50	-11,554	0,00
	4.okuma	65,93	13,43			

Aynı öğretmenlerin farklı zamanlarda rubrik kullanmadan ve kullanarak verdiği puanlar arasındaki farkın manidarlığını test etmek ve iki farklı tür puanlama arasında geçen zamanın manidar bir farklılığa neden olup olmadığını belirlemek amacıyla, aynı öğretmenlerin farklı zamanlarda rubrik kullanmadan ve kullanarak yapmış oldukları puanların (1. ve 4. okumalar) ortalamaları arasındaki farklılık bağımlı örneklem için t-testi ile test edilmiştir.

Tablodaki bulgulara göre 3., 5., 12. ve 14. öğretmenler hariç, diğer tüm öğretmenlerin rubrik kullanmadan (1.okuma) ve kullanarak (4.okuma) farklı zamanlarda verdikleri puanların ortalamaları arasında manidar farklılıklar ($p<0,01$) bulunmuştur. 6. ve 10. öğretmenlerde, aradan geçen zamanın ve değişen yöntemin verdikleri puanların ortalamalarında bir düşüşe sebep olduğu, diğerlerinde ise artırdığı şeklinde bir yorum yapılabilir. Aradan geçen zamanın, aynı öğretmenin yapmış olduğu iki farklı tür puanlama için tutarlılığın bozulmasında etkili olduğu yorumu yapılabilir. Puanlamaların çoğunlukla artarak değişmesi öğretmenlerin rubrikli puanlamalarda daha cömert davrandıklarını göstermektedir.

Buna göre, aynı öğretmenlerin aynı kağıtları farklı zamanlarda rubrik kullanarak ve kullanmadan tutarlı puanlar veremediği, ayrıca aradan geçen zamanın bu durum üzerinde manidar bir farklılık yarattığı yani tutarlılığı bozduğu söylenebilir.

II.6. “Aynı öğretmenlerin puanlama yönergesi (rubrik) kullanarak ve belli bir süre sonra puanlama yönergesi kullanmadan verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemi için, aynı öğretmenlerin aynı kağıtlara puanlama yönergesi kullanarak (2. okuma) verdikleri puanlar ile aradan belli bir süre geçtikten sonra aynı öğretmenlerin puanlama yönergesi kullanmadan (3. okuma) verdikleri puanlar arasındaki korelasyonlar hesaplanmış ve manidarlığı t testi ile test edilmiştir. Verilen bu puanlar arasındaki farkın manidarlığını test etmek amacıyla, aynı öğretmenlerin kağıtları hem 2. okumalarında hem de 3. okumalarında vermiş oldukları puanların ortalamaları hesaplanmış ve ortalamalar arasında fark olup olmadığı “bağımlı örneklem için t testi” ile test edilmiştir.

Aynı öğretmenlerin puanlama yönergesi kullanarak verdiği puanlar ile aradan belli bir süre geçtikten sonra puanlama yönergesi kullanmadan verdiği puanlar arasında hesaplanan Pearson Momentler Çarpımı korelasyon katsayıları Tablo II.11’de verilmiştir.

TABLO II.11.

Aynı öğretmenlerin puanlama yönergesi kullanarak ve belli bir süre sonra puanlama yönergesi kullanmadan verdiği puanlar arasındaki korelasyon hesaplamaları

Öğretmen	2. ve 3. okuma arasındaki korelasyon	p
1	0,819	0,00
2	0,867	0,00
3	0,841	0,00
4	0,869	0,00
5	0,704	0,00
6	0,792	0,00

7	0,753	0,00
8	0,905	0,00
9	0,860	0,00
10	0,805	0,00
11	0,853	0,00
12	0,854	0,00
13	0,691	0,00
14	0,804	0,00
15	0,837	0,00
16	0,679	0,00
17	0,905	0,00

($p < 0,01$)

Tablodaki bulgular, aynı öğretmenlerin farklı zamanlarda puanlama yönergesi (rubrik) kullanarak (2. okuma) ve kullanmadan (3.okuma) verdikleri puanlar arasındaki korelasyonların yüksek ve manidar olduğuna dair ipuçları vermektedir. İlişkinin düzeyi en yüksek 0,905 ile 8. öğretmene, en düşük ise 0,679 ile 16. öğretmene aittir. Buna göre aynı öğretmenlerin, farklı zamanlarda rubrik kullanarak ve kullanmadan yapmış oldukları puanlamalar arasındaki tutarlılığın en yüksek düzeyde 8. öğretmen tarafından, en düşük düzeyde 16. öğretmen tarafından sağlandığı söylenebilir. Öğretmenlerin 2. ve 3. okumalardaki korelasyon değerlerinin arasındaki farklılıklar ($s=0,070$) puanlamada, puanlayıcıların, aradan geçen zamanın ya da kullanılan yöntemin birer değişim kaynağı olabileceğini göstermektedir.

TABLO II.12.

Aynı öğretmenlerin puanlama yönergesi kullanarak ve belli bir süre sonra puanlama yönergesi kullanmadan vermiş oldukları puanlara ilişkin t-testi sonuçları

Öğretmen	Okumalar	ort	Std.sapma	N	t	p
1	2.okuma	66,90	16,00	50	12,585	0,00
	3.okuma	45,24	21,10			
2	2.okuma	68,90	18,83	50	3,713	0,001
	3.okuma	63,20	21,80			
3	2.okuma	49,65	17,80	50	3,307	0,002
	3.okuma	45,06	16,88			
4	2.okuma	76,75	18,53	50	7,479	0,00
	3.okuma	64,70	22,84			
5	2.okuma	46,87	18,56	50	-5,455	0,00
	3.okuma	58,70	20,89			
6	2.okuma	48,72	15,70	50	-5,603	0,00
	3.okuma	56,80	15,90			
7	2.okuma	63,28	13,55	50	6,069	0,00
	3.okuma	51,52	20,62			
8	2.okuma	70,03	17,92	50	9,117	0,00
	3.okuma	56,44	23,49			
9	2.okuma	78,43	14,03	50	2,516	0,015
	3.okuma	75,40	19,74			
10	2.okuma	50,65	15,85	50	-2,602	0012
	3.okuma	55,26	20,99			

11	2.okuma	61,16	15,21	50	-0,001	0,999
	3.okuma	61,16	20,70			
12	2.okuma	56,96	12,02	50	-0,73	0,469
	3.okuma	57,76	14,68			
13	2.okuma	61,40	13,84	50	13,571	0,00
	3.okuma	40,24	14,22			
14	2.okuma	54,78	14,15	50	1,550	0,128
	3.okuma	51,56	23,44			
15	2.okuma	63,65	17,23	50	7,167	0,00
	3.okuma	50,10	23,91			
16	2.okuma	53,96	13,41	50	11,093	0,00
	3.okuma	36,34	14,51			
17	2.okuma	65,87	13,49	50	9,778	0,00
	3.okuma	56,46	15,85			

Aynı öğretmenlerin farklı zamanlarda rubrik kullanarak ve kullanmadan yapmış oldukları puanların (2. ve 3. okumalar) ortalamaları arasında manidar bir farklılık olup olmadığı bağımlı örneklem için t-testi ile test edilmiştir.

Tablodaki bulgulara göre, 9., 10., 11., 12. ve 14. öğretmenlerin dışındaki tüm öğretmenlerin aynı kağıtları farklı zamanlarda rubrikli ve rubriksiz okumaları sonucunda vermiş oldukları puanların ortalamaları arasında manidar farklılıklar bulunmuştur. 5. ve 6. öğretmenlerde, aradan geçen zamanın ve değişen yöntemin, puanların ortalamalarında bir artışa sebep olduğu, diğerlerinde ise düşürdüğü şeklinde bir sonuç çıkarılabilir. Buna göre, aynı öğretmenlerin aynı kağıtlara farklı zamanlarda rubrik kullanarak ve kullanmadan tutarlı puanlar veremediği, ayrıca aradan geçen zamanın bu durum üzerinde manidar bir

farklılık yarattığı yani tutarlılığı bozduğu söylenebilir. Puanlamaların çoğunlukla azalarak değişmesi öğretmenlerin rubriksiz puanlamalarda daha cimri davrandıklarını göstermektedir.

TABLO II.13.
Aynı öğretmenlerin farklı okumalarda verdikleri puanlar arasındaki pearson momentler çarpımı korelasyon katsayıları ve ortalamalar arasındaki farkı gösteren t puanlarının yer aldığı özet tablo

Ögrt.	r_{12}	t_{12}	r_{34}	t_{34}	r_{13}	t_{13}	$r_{Öğretmenler}$	t_{24}	r_{14}	t_{14}	r_{23}	t_{23}
1	0,790*	-7,519*	0,852*	-10,693*	0,920*	4,856*	0,943*	6,163*	0,848*	-5,853*	0,819*	12,585*
2	0,802*	-7,372*	0,868*	-3,644*	0,834*	-4,212*	0,866*	0,07	0,872*	-9,093*	0,867*	3,713*
3	0,746*	0,394	0,838*	-3,876*	0,852*	3,990*	0,751*	-0,505	0,774*	-0,119	0,841*	3,307*
4	0,837*	-8,781*	0,811*	-10,598*	0,926*	-4,072*	0,877*	-6,970*	0,752*	-10,785*	0,869*	7,479*
5	0,753*	3,069*	0,749*	5,071*	0,741*	-2,068	0,868*	-1,022	0,791*	2,616	0,704*	-5,455*
6	0,841*	8,059*	0,846*	14,924*	0,818*	2,071	0,832*	7,971*	0,785*	13,269*	0,792*	-5,603*
7	0,751*	-8,691*	0,755*	-8,230*	0,874*	-4,646*	0,850*	-3,899*	0,750*	-10,587*	0,753*	6,069*
8	0,805*	-10,600*	0,667*	-2,389	0,828*	-2,517	0,694*	4,150*	0,678*	-5,011*	0,905*	9,117*
9	0,925*	-4,633*	0,894*	-3,849*	0,854*	-0,942	0,957*	-1,802	0,916*	-5,583*	0,860*	2,516
10	0,730*	-0,230	0,674*	4,921*	0,703*	-2,276	0,761*	4,183*	0,643*	2,713*	0,805*	-2,602
11	0,796*	-6,407*	0,807*	-1,133	0,843*	-7,275*	0,778*	-1,431	0,695*	-6,358*	0,853*	-0,001
12	0,851*	8,687*	0,711*	-4,436*	0,804*	7,135*	0,744*	-5,667*	0,749*	1,929	0,854*	-0,73
13	0,586*	-7,919*	0,814*	-16,353*	0,666*	4,137*	0,799*	0,093	0,651*	-8,028*	0,691*	13,571*
14	0,669*	-4,084*	0,749*	0,224	0,522*	-1,764	0,727*	2,386	0,555*	-1,998	0,804*	1,550
15	0,702*	-11,022*	0,923*	-8,592*	0,771*	-4,333*	0,900*	0,747	0,794*	-12,682*	0,837*	7,167*
16	0,741*	-2,127*	0,747*	-15,516*	0,625*	6,316*	0,846*	-4,409*	0,733*	-4,682*	0,679*	11,093*
17	0,836*	-13,016*	0,819*	-7,352*	0,769*	-6,735*	0,926*	-0,086	0,766*	-11,554*	0,905*	9,778*

(*p<0,01)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Korelasyon Katsayıları Arasındaki Fark	$r_{12} - r_{34}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-2,36	p>.05	p<.05 z'=-3,65	p>.05	p>.05	
	$r_{12} - r_{13}$	p<.05 z'=-2,59	p>.05	p>.05	p<.05 z'=-2,065	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{12} - r_{24}$	p<.05 z'=-3,395	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-2,1	p>.05	p<.05 z'=-3,01	p>.05	p<.05 z'=-2,095	
	$r_{12} - r_{14}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{12} - r_{23}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{34} - r_{13}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=1,965	p<.05 z'=2,875	p>.05	p>.05
	$r_{34} - r_{24}$	p<.05 z'=-2,45	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-2,265	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{34} - r_{14}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=2,60	p>.05	
	$r_{34} - r_{23}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-3,47	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{13} - r_{24}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-3,15	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-2,235	p<.05 z'=-2,525	p<.05 z'=-3,015
	$r_{13} - r_{14}$	p>.05	p>.05	p>.05	p<.05 z'=3,215	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	
	$r_{13} - r_{23}$	p<.05 z'=2,195	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-2,665	p>.05	p>.05	p<.05 z'=-2,395
	$r_{24} - r_{14}$	p<.05 z'=2,55	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=1,96	p>.05	p<.05 z'=3,065
	$r_{24} - r_{23}$	p<.05 z'=3,00	p>.05	p>.05	p>.05	p<.05 z'=2,225	p>.05	p>.05	p>.05	p<.05 z'=-3,245	p<.05 z'=3,035	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=2,065	p>.05
$r_{14} - r_{23}$	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p>.05	p<.05 z'=-3,37	p>.05	p>.05	p<.05 z'=-2,01	p>.05	p>.05	p<.05 z'=-2,435	p>.05	p>.05	p<.05 z'=-2,445	

TABLO II.14.

Farklı Okumalar Arasındaki Korelasyon Katsayılarının Fisher'in z' Dönüşümü ile Karşılaştırılması

Aynı öğretmenlerin farklı okumalarda, kararlı puanlar verip veremediklerini test etmek üzere hesaplanan Pearson Momentler Çarpımı Korelasyon Katsayıları ve öğretmenlerin verdikleri puanların ortalamaları arasında farklılıklar olup olmadığı, varsa artış ve azalışları test etmek üzere hesaplanan t puanları Tablo II.13.'te birlikte verilmiştir.

Tablodaki t değerleri incelendiğinde + ve - 'lerin bulunması değerlendiricilerin de birer değişim kaynağı olduklarını göstermektedir. Tabloda hangi öğretmenlerin, hangi puanlamalarda cömert, hangi puanlamalarda cimri davrandığı görülmektedir. Tablo, genel olarak incelendiğinde çoğu öğretmenin puanlama yönergesi (rubrik) kullanıldığında, daha yüksek puan verme eğiliminde oldukları ifade edilebilir. Bu bulgu, cömertlik etkisinin, dereceleme biçiminin bir fonksiyonu olduğunu savunan Bernardin ve Orban'ı (1990) desteklemektedir. Bu durum, puanlama yönergelerinin ayrıntılı bir puanlama biçimi önermesi, daha açık bir ifade ile hangi davranışa kaç puan verileceğinin belirlenmiş olmasına bağlanabilir. Farklı zamanlarda rubrikle yapılan puanlamalar (2. ve 4. okumalar) incelendiğinde, ortalamalar arasındaki manidar farklılıkların diğer okumalara kıyasla daha az olduğu görülmektedir. Başka bir deyişle, öğretmenler rubrikle puanlama yaptıklarında farklı zamanlarda birbirine çok yakın puanlar vermişlerdir.

Tablodaki korelasyon katsayıları incelendiğinde artış ve azalışlar olduğu gözlenirse bile, net bir karşılaştırma yapabilmek için korelasyon katsayıları üzerinde Fisher'in z' dönüşümü hesaplanmıştır. Hesaplamaların sonucu Tablo II.14'te gösterilmiştir. Tabloda her öğretmen için, hangi korelasyon katsayıları arasında $p=0,05$ düzeyinde manidar farklılıklar bulunmadığı, hangilerinde bulunduğu ve farklılıkların hangi yönde olduğu görülmektedir. Bu durumda bazı puanlamalardaki korelasyon değerleri arasında $p=0,05$ düzeyinde istatistiksel olarak anlamlı farklılıklar bulunmuştur. Örneğin, farklı zamanlarda rubrik kullanmadan verilen puanlar arasındaki korelasyon katsayıları (r_{13})

yüksektir. Farklı zamanlarda rubrik kullanarak verilen puanlamalar arasındaki korelasyon katsayıları (r_{24}) da yüksektir. Fakat iki korelasyon katsayısı Fisher'in z' dönüşümü yapılarak karşılaştırıldığında 4 öğretmende (9., 15., 16., 17.) istatistiksel olarak manidar ($p < 0,05$) farklılıklar gözlenmiştir. Bu 4 öğretmen, farklı zamanlarda rubrik kullanarak daha kararlı puanlamalar yapmışlardır.

Aynı ya da farklı zamanlarda, farklı yöntemler arasındaki korelasyon katsayıları karşılaştırıldığında ($r_{12} - r_{34}$, $r_{12} - r_{14}$, $r_{12} - r_{23}$, $r_{34} - r_{14}$, $r_{34} - r_{23}$) genel olarak manidar farklılıklar çıkmamıştır. Bu durum, araya zaman faktörü girse de girmese de farklı teknikler arasında benzer korelasyon katsayılarının bulunduğu göstergesidir. Başka bir deyişle, farklı teknikler arasında gözlenen uyumun derecesi, zamana göre istatistiksel olarak farklılaşmamaktadır.

Tablo, öğretmenler açısından incelendiğinde, 10 yılın üzerinde tecrübeye sahip öğretmenlerin (2., 3., 4., 5., 9., 11., 12. öğretmenler) genel olarak daha kararlı ve tutarlı puanlamalar yaptıkları gözlenmektedir. Bu bulgu, deneyimli öğretmenlerin daha doğru yargılamalarda bulunduğunu savunan Jorgenson'u (1975) desteklemektedir.

II.7. “Farklı öğretmenlerin ikili karşılaştırıldıklarında puanlama yönergesi (rubrik) kullanmadan verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemine yanıt bulabilmek için, öncelikle farklı öğretmenlerin puanlama yönergesi kullanmadan yapmış oldukları puanlamalar (1.okuma) arasındaki ilişkiyi belirlemek üzere Pearson Momentler Çarpımı korelasyon Katsayısı hesaplanmış ve sonuçlar Tablo II.15'te sunulmuştur:

TABLO II.15.

İkili karşılaştırıldıklarında farklı öğretmenlerin 1.okuma sonrasında verdikleri puanlar arasındaki ilişkinin Pearson momentler çarpımı korelasyon katsayısı ile hesaplanması

Öğrt.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1.	1,00																
2.	,798	1,00															
3.	,365	,619	1,00														
4.	,819	,777	,416	1,00													
5.	,698	,766	,446	,714	1,00												
6.	,821	,853	,649	,796	,751	1,00											
7.	,818	,778	,393	,813	,765	,791	1,00										
8.	,785	,812	,573	,770	,719	,777	,765	1,00									
9.	,738	,829	,597	,694	,699	,827	,742	,778	1,00								
10.	,462	,734	,675	,620	,506	,601	,509	,681	,688	1,00							
11.	,778	,842	,564	,834	,825	,823	,825	,856	,805	,639	1,00						
12.	,718	,881	,679	,719	,702	,846	,700	,810	,893	,739	,846	1,00					
13.	,494	,574	,673	,439	,460	,618	,412	,697	,589	,646	,574	,652	1,00				
14.	,609	,695	,467	,587	,696	,637	,621	,738	,725	,569	,729	,719	,607	1,00			
15.	,645	,777	,766	,632	,581	,801	,617	,702	,713	,744	,622	,764	,719	,574	1,00		
16.	,658	,742	,645	,697	,744	,802	,675	,748	,692	,620	,726	,713	,608	,546	,720	1,00	
17.	,465	,692	,621	,491	,573	,577	,464	,713	,707	,761	,651	,737	,701	,695	,731	,593	1,00

Farklı öğretmenler ikişer ikişer karşılaştırıldıklarında rubrik kullanmadan verdikleri puanlar arasındaki tüm korelasyonlar $p=0,01$ düzeyinde anlamlı çıkmıştır. En yüksek korelasyon 0,893 ile 12. ve 9. öğretmenler arasında, en düşük korelasyon ise 0,365 ile 1. ve 3. öğretmenler arasında bulunmuştur.

Tabloya göre 3., 10. ve 13. öğretmenlerin genel olarak diğer öğretmenlerle düşük korelasyonlara sahip oldukları, başka bir deyişle daha az uyum gösterdikleri söylenebilir.

TABLO II.16.

Farklı öğretmenlerin 1. okuma sonrasında verdikleri puanlara ilişkin varyans analizi sonuçları

	Kareler toplamı	df	Kareler ortalaması	F	Sig.
Gruplar arası	59765,075	16	3735,317	9,059	,000
Gruplar içi	343484,300	833	412,346		
Toplam	403249,375	849			

Farklı öğretmenlerin puanlama yönergesi kullanmadan vermiş oldukları puanların ortalamalarının karşılaştırılması tek yönlü varyans analizi ile yapılmış ve öğretmenlerin 1. okumada vermiş oldukları puanların ortalamaları arasında istatistiksel olarak manidar bir farklılık bulunmuştur. Bu aşamada, değişkenlik kaynağı olarak puanlayıcıları ve puanlayıcılardan gelen potansiyel hataları hesaba katarak, puanlayıcılar arası güvenilirliği belirlemek için Hoyt'un varyans analizi yönteminden faydalanılmıştır. Hoyt (1941), varyans analizi üzerine temellendirdiği ve güvenilirlik katsayısı olarak da yorumlanabilen bir G katsayısı elde etmiştir (Erkuş,2003). G katsayısı, şu şekilde hesaplanır:

$$G = \frac{\sigma_{\text{gruplararası}}}{\sigma_{\text{gruplararası}} + \frac{\sigma_{\text{grupiçi}}}{n}}$$

$\sigma_{\text{gruplararası}}$ = Bireyler arasındaki varyans

$\sigma_{\text{grupiçi}}$ = Hata varyansı

Farklı öğretmenlerin puanlama yönergesi kullanmadan vermiş oldukları puanlara ilişkin hesaplanan G katsayısı, bir öğretmen için 0,67 olarak bulunurken, 17 öğretmen için 0,971 çıkmıştır. Bu durum, öğretmen sayısının artmasının, rubrik

kullanmadan yapılan puanlamalarda hatayı azalttığını, dolayısıyla güvenilirliği yükselttiğini göstermektedir.

1. okumalarda farklılığın hangi gruplar arasında olduğunu belirlemede Scheffe testine başvurulmuş ve sonuçlar Tablo II.17’de sunulmuştur.

TABLO II.17.

Farklı Öğretmenlerin 1. Okuma Sonrasında Verdikleri Puanlara İlişkin

Scheffe Testi Sonuçları

Puanlayıcılar	Ortalama farklar	p
1-9	-22,68	0,014
3-9	-23,84	0,005
7-9	-30,00	0,000
7-12	-23,44	0,008
8-9	-22,46	0,017
9-10	23,98	0,005
9-11	25,56	0,001
9-13	27,14	0,000
9-14	27,98	0,000
9-15	33,50	0,000
9-16	24,16	0,004
9-17	30,32	0,000
12-14	21,42	0,035
12-15	26,94	0,000
12-17	23,76	0,006

(p<0,05)

Scheffe testinin sonuçlarına göre yukarıdaki öğretmen çiftlerinin puanlama yönergesi (rubrik) kullanmadan verdikleri puanlar arasında tutarlılığı bozacak düzeyde anlamlı farklılıklar çıkmıştır. Farklılığı en çok yaratan öğretmen ise 9. öğretmen olarak görülmektedir. 9. öğretmen diğer öğretmenlere göre daha yüksek puanlar verme eğilimindedir.

Sonuç olarak farklı öğretmenlerin aynı kağıtları, aynı yöntemlerle okusalar dahi tutarlı puanlamalar yapamadıkları, bu aşamada işin içine öğretmen yargıları, tecrübeleri, bilişsel süreçleri vb. farklı değişkenlerin girdiği düşünülebilir.

II.8. “Farklı öğretmenlerin ikili karşılaştırıldıklarında puanlama yönergesi (rubrik) kullanarak verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?”alt problemine yanıt bulabilmek için, öncelikle farklı öğretmenlerin puanlama yönergesi kullanarak yapmış oldukları puanlamalar (2.okuma) arasındaki ilişkiyi belirlemek üzere Pearson Momentler Çarpımı korelasyon Katsayısı hesaplanmış ve sonuçlar Tablo II.18’de sunulmuştur:

TABLO II.18.

İkili karşılaştırıldıklarında farklı öğretmenlerin 2. okuma sonrasında verdikleri puanlar arasındaki ilişkinin Pearson momentler çarpımı korelasyon katsayısı ile hesaplanması

Öğrt.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00																
2	,823	1,00															
3	,713	,783	1,00														
4	,847	,776	,574	1,00													
5	,713	,602	,520	,740	1,00												
6	,740	,789	,713	,759	,595	1,00											
7	,847	,746	,601	,792	,714	,638	1,00										
8	,841	,808	,619	,907	,698	,808	,820	1,00									
9	,910	,817	,671	,874	,693	,770	,862	,857	1,00								
10	,826	,764	,590	,783	,669	,644	,761	,754	,792	1,00							
11	,855	,786	,644	,854	,645	,807	,798	,843	,906	,751	1,00						
12	,813	,816	,740	,841	,614	,891	,715	,853	,850	,671	,814	1,00					
13	,751	,759	,687	,771	,610	,762	,689	,757	,771	,638	,711	,830	1,00				
14	,841	,806	,719	,842	,751	,840	,767	,837	,834	,728	,825	,893	,788	1,00			
15	,864	,795	,661	,863	,692	,827	,748	,878	,839	,792	,862	,848	,699	,855	1,00		
16	,843	,836	,732	,823	,642	,730	,808	,801	,819	,690	,776	,827	,779	,804	,769	1,00	
17	,779	,810	,734	,585	,448	,647	,720	,630	,735	,693	,706	,631	,640	,628	,694	,680	1,00

Farklı öğretmenlerin rubrik kullanarak verdikleri puanlar arasındaki korelasyonlar $p=0,01$ düzeyinde anlamlıdır. En yüksek korelasyon 0,910 ile 1. ve 9. öğretmen arasında, en düşük korelasyon ise 0,520 ile 3. ve 5. öğretmenler arasında bulunmuştur.

1. okumalarda 0,80'in üzerinde korelasyona sahip 22 öğretmen çifti varken, 2. okumalarda 0,80'in üzerinde korelasyona sahip 52 öğretmen çifti bulunmaktadır.

1. okumalarda en düşük korelasyonlara sahip 3. öğretmen 2. okumalarda da en düşük korelasyonlara sahiptir fakat ilişkinin düzeyinin tek tek karşılaştırıldığında artmış olduğu görülmektedir. Örneğin; 1. okumada 1. ve 3. öğretmenler arasındaki korelasyon katsayısı 0,365 iken 2. okumalarda 0,713 çıkmıştır. Katsayılara Fisher'in z' dönüşümü yapıldığında korelasyon katsayıları arasında $p=0,05$ düzeyinde istatistiksel olarak anlamlı bir farklılık olduğu görülür ($z'=-2,53$). Benzer şekilde 1. ve 2. okumalara ait korelasyon katsayıları karşılaştırıldığında puanlama yönergesi kullanılarak yapılan puanlamalarda (2. okuma) öğretmen çiftleri arasındaki korelasyon katsayılarının istatistiksel olarak daha yüksek olduğu görülür.

Sonuç olarak farklı öğretmenlerin rubrik kullanarak vermiş oldukları puanlar arasındaki korelasyonların, rubrik kullanmadan vermiş oldukları puanlar arasındaki korelasyonlardan daha yüksek olduğu yorumu yapılabilir. Buna göre, puanlama yönergesi kullanımının farklı öğretmenlerin birbiriyle daha tutarlı puanlama yapmalarında katkısının olabileceği söylenebilir.

TABLO II.19

Farklı Öğretmenlerin 2. okuma sonrasında verdikleri puanlara ilişkin varyans analizi

sonuçları

	Kareler toplamı	df	Kareler ortalaması	F	Sig.
Gruplar arası	72881,970	16	4555,123	18,250	,000
Gruplar içi	207911,961	833	249,594		
Toplam	280793,932	849			

Farklı öğretmenlerin puanlama yönergesi (rubrik) kullanarak vermiş oldukları puanların ortalamalarının karşılaştırılması tek yönlü varyans analizi ile yapılmış ve öğretmenlerin 2. okumada vermiş oldukları puanların ortalamaları arasında istatistiksel olarak manidar bir farklılık bulunmuştur. Bu aşamada yine G katsayısı hesaplanmış ve bir öğretmen için 0,74 olarak bulunurken, 17 öğretmen için 0,979 çıkmıştır. Bu durum, öğretmen sayısının artmasının, rubrik kullanarak yapılan puanlamalarda da hatayı azalttığını ve güvenilirliği yükselttiğini, ayrıca öğretmenlerin rubrikle daha güvenilir puanlamalar yaptıklarını göstermektedir.

2. okumalarda farklılığın hangi gruplar arasında olduğunu belirlemede Scheffe testine başvurulmuş ve sonuçlar aşağıdaki tabloda verilmiştir.

TABLO II.20.

Farklı Öğretmenlerin 2. Okuma Sonrasında Verdikleri Puanlara İlişkin

Scheffe Testi Sonuçları

Puanlayıcılar	Ortalama farklar	p
1-3	17,24	0,021
1-5	20,03	0,001
1-6	18,18	0,008
2-3	19,25	0,002
2-5	22,03	0,000
2-6	20,18	0,001
2-10	18,25	0,007
3-4	-27,09	0,000
3-8	-20,37	0,001
3-9	-28,78	0,000

4-5	29,87	0,000
4-6	28,03	0,000
4-10	26,09	0,000
4-12	19,78	0,001
4-14	21,97	0,000
4-16	22,78	0,000
5-7	-16,40	0,044
5-8	-23,15	0,000
5-9	-31,56	0,000
5-15	-16,78	0,032
5-17	-18,99	0,003
6-8	-21,31	0,000
6-9	-29,71	0,000
6-17	-17,15	0,023
7-10	19,37	0,002
9-10	27,78	0,000
9-11	17,28	0,020
9-12	21,46	0,000
9-13	17,03	0,025
9-14	23,65	0,000
9-16	24,46	0,000

($p < 0,05$)

Scheffe testinin sonuçlarına göre Tablo II.20'deki öğretmen çiftlerinin puanlama yönergesi (rubrik) kullanarak verdikleri puanlar arasında tutarlılığı bozacak

düzeyde anlamlı farklılıklar çıkmıştır. Bu durumda farklı öğretmenlerin bazılarının, aynı kağıtları rubrikle okusalar bile tutarlı puanlamalar yapamadıkları, işin içine öğretmen yargıları, tecrübeleri, bilişsel süreçleri vb. farklı değişkenlerin girdiği düşünülebilir.

Rubrik kullanarak ve kullanmadan yapılan puanlamalarda, bazı öğretmenlerin vermiş oldukları puanların ortalamaları arasında anlamlı farklılıklar bulunmuştur. Bu durum rubrik kullanımının farklı öğretmenlerin puanlamaları arasındaki farklılığı gidermede tek başına yeterli olmadığının göstergesi olarak kabul edilebilir.

II.9. “Farklı öğretmenlerin ikili karşılaştırıldıklarında belli bir süre sonra puanlama yönergesi (rubrik) kullanmadan verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemine yanıt bulabilmek için, öncelikle farklı öğretmenlerin puanlama yönergesi kullanmadan yapmış oldukları puanlamalar (3.okuma) arasındaki ilişkiyi belirlemek üzere Pearson Momentler Çarpımı korelasyon Katsayısı hesaplanmış ve sonuçlar Tablo II.21.’de sunulmuştur:

TABLO II.21.

İkili karşılaştırıldıklarında farklı öğretmenlerin 3. okuma sonrasında verdikleri puanlar arasındaki ilişkinin Pearson momentler çarpımı korelasyon katsayısı ile hesaplanması

Öğrt.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00																
2	,628	1,00															
3	,566	,863	1,00														
4	,886	,666	,614	1,00													
5	,770	,740	,740	,869	1,00												
6	,587	,749	,774	,658	,782	1,00											
7	,818	,650	,572	,805	,692	,575	1,00										
8	,847	,714	,700	,878	,880	,729	,772	1,00									
9	,713	,776	,806	,794	,872	,765	,717	,820	1,00								
10	,693	,813	,789	,773	,821	,732	,695	,794	,833	1,00							
11	,713	,866	,796	,780	,823	,812	,748	,882	,832	,835	1,00						
12	,603	,755	,804	,710	,876	,771	,612	,826	,830	,793	,822	1,00					
13	,386	,666	,855	,456	,619	,757	,349	,542	,624	,690	,642	,689	1,00				
14	,554	,761	,838	,636	,819	,802	,520	,753	,748	,762	,801	,890	,808	1,00			
15	,764	,818	,770	,756	,836	,808	,723	,849	,786	,807	,870	,801	,606	,801	1,00		
16	,777	,621	,661	,749	,670	,617	,762	,756	,711	,689	,723	,574	,521	,606	,694	1,00	
17	,524	,822	,747	,553	,636	,819	,548	,625	,674	,725	,737	,674	,611	,724	,724	,585	1,00

Farklı öğretmenlerin, aradan belli bir süre geçtikten sonra, puanlama yönergesi (rubrik) kullanmadan tekrar yapmış oldukları puanlamalar (3.okuma) arasındaki ilişkiyi belirlemek üzere Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmış ve

sonular Tablo II.21.'de verilmiřtir. Farklı ğretmenlerin vermiř oldukları puanlar arasında hesaplanan korelasyon katsayıları $p=0,01$ düzeyinde anlamlıdır.

Farklı ğretmenlerin rubrik kullanmadan verdikleri puanlar arasındaki korelasyonlar yksek ve manidar bulunmuřtur. En yksek korelasyon 0,886 ile 1. ve 4. ğretmen arasında, en dřk korelasyon ise 0,386 ile 1. ve 13. ğretmen arasında bulunmuřtur.

3. okumalarda farklı ğretmenlerin verdikleri puanlar arasındaki korelasyonlar Fisher'in z' dnřm yapılarak karřılařtırıldıėında 1. okumalara gre daha yksektir. Bařka bir deyiřle farklı ğretmenlerin rubrik kullanmaksızın verdikleri puanlar arasında geen srenin, ğretmenlerin puanlamaları arasındaki tutarlılıėı arttırdıėı sylenebilir. ğretmenlerin sorulara alıřmıř olmaları da buna etken olarak dřnlebilir.

TABLO II.22.

Farklı ğretmenlerin 3. Okuma sonrasında Verdikleri Puanlara İliřkin Varyans

Analizi Sonuları

	Kareler toplamı	df	Kareler ortalaması	F	Sig.
Gruplar arası	72374,798	16	4523,425	11,747	,000
Gruplar ii	320762,780	833	385,069		
Toplam	393137,578	849			

Farklı ğretmenlerin puanlama ynergesi (rubrik) kullanmaksızın vermiř oldukları puanların ortalamalarının karřılařtırılması tek ynl varyans analizi ile yapılmıř ve ğretmenlerin 3. okumalarında vermiř oldukları puanların ortalamaları arasında istatistiksel olarak manidar bir farklılık bulunmuřtur. Bu ařamada G katsayısı, bir ğretmen iin 0,71 olarak bulunurken, 17 ğretmen iin 0,976 çıkmıřtır. Yine ğretmen

sayısının artmasının, rubrik kullanmadan yapılan puanlamalarda güvenilirliği yükselttiği görülmektedir.

3. okumalarda farklılığın hangi gruplar arasında olduğunun belirlenmesinde Scheffe testine başvurulmuş ve sonuçlar aşağıdaki tabloda verilmiştir.

TABLO II.23.

Farklı Öğretmenlerin 3. Okuma Sonrasında Verdikleri Puanlara İlişkin

Scheffe Testi Sonuçları

Puanlayıcılar	Ortalama farklar	p
1-9	-30,16	0,000
2-13	22,96	0,006
2-16	26,86	0,000
3-9	-30,34	0,000
4-13	24,46	0,001
4-16	28,36	0,000
5-16	22,36	0,010
6-16	20,46	0,042
7-9	-23,88	0,002
9-13	35,16	0,000
9-14	23,84	0,0003
9-15	25,30	0,001
9-16	39,06	0,000
11-13	20,92	0,030
11-16	24,82	0,001
12-16	21,42	0,021

($p < 0,05$)

Scheffe testinin sonuçlarına göre belirli bir sürenin sonunda yukarıdaki öğretmen çiftlerinin puanlama yönergesi (rubrik) kullanmadan verdikleri puanlar arasında tutarlılığı bozacak düzeyde anlamlı farklılıklar çıkmıştır. Bazı öğretmenlerin, aynı kağıtları rubrik kullanmadan okuduklarında tutarlı puanlamalar yapamadıkları ifade edilebilir. Bu aşamada hangi bilişsel süreçlerin tutarsızlığı yarattığının incelenmesi gerekmektedir.

II.10. “Farklı öğretmenlerin ikili karşılaştırıldıklarında belli bir süre sonra puanlama yönergesi (rubrik) kullanarak verdikleri puanlar arasında manidar bir ilişki var mıdır? Verdikleri puanların ortalamaları arasında manidar bir fark var mıdır?” alt problemine yanıt bulabilmek için, öncelikle farklı öğretmenlerin puanlama yönergesi kullanarak yapmış oldukları puanlamalar (4.okuma) arasındaki ilişkiyi belirlemek üzere Pearson Momentler Çarpımı korelasyon Katsayısı hesaplanmış ve sonuçlar Tablo II.24’te sunulmuştur.

TABLO II.24

İkili karşılaştırıldıklarında farklı öğretmenlerin 4. okuma sonrasında verdikleri puanlar arasındaki ilişkinin Pearson momentler çarpımı korelasyon katsayısı ile hesaplanması

Öğrt.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00																
2	,804	1,00															
3	,766	,807	1,00														
4	,786	,835	,653	1,00													
5	,680	,610	,463	,630	1,00												
6	,683	,657	,756	,660	,571	1,00											
7	,763	,803	,659	,720	,674	,579	1,00										
8	,748	,717	,562	,730	,613	,476	,751	1,00									
9	,889	,826	,733	,887	,645	,692	,800	,727	1,00								
10	,721	,744	,687	,704	,565	,532	,638	,613	,719	1,00							
11	,749	,814	,760	,757	,505	,631	,671	,678	,776	,650	1,00						
12	,666	,701	,659	,760	,607	,700	,539	,476	,686	,594	,664	1,00					
13	,667	,812	,775	,693	,553	,739	,651	,542	,679	,619	,714	,667	1,00				
14	,678	,782	,700	,650	,688	,705	,697	,604	,672	,616	,680	,707	,742	1,00			
15	,889	,880	,812	,832	,671	,772	,780	,690	,863	,766	,784	,673	,777	,737	1,00		
16	,804	,830	,751	,866	,652	,744	,720	,619	,865	,794	,754	,766	,721	,694	,893	1,00	
17	,688	,771	,744	,640	,387	,581	,754	,614	,758	,688	,704	,506	,634	,645	,700	,645	1,00

Farklı öğretmenlerin puanlamaları arasındaki korelasyon katsayıları $p=0,01$ düzeyinde anlamlı çıkmıştır. Farklı öğretmenlerin rubrik kullanarak yaptığı puanlamalar arasında en yüksek korelasyon 0,893 ile 15. ve 16. öğretmenler arasında, en düşük korelasyon ise 0,387 ile 5. ve 17 öğretmenler arasında bulunmuştur.

4. okumalarda bazı öğretmen çiftlerinin verdikleri puanlar arasındaki korelasyonlar 2. okumalara göre daha düşük gibi gözükse de Fisher'in z' dönüşümü yapılarak karşılaştırıldığında istatistiksel olarak anlamlı farklar bulunamamıştır. Bu durum puanlama yönergesi kullanılarak yapılan puanlamalarda farklı zamanlarda olsa bile öğretmenler arasında uyum olduğunu göstermektedir.

TABLO II.25.

Farklı Öğretmenlerin 4. Okuma Sonrasında Verdikleri Puanlara İlişkin Varyans

Analizi Sonuçları

	Kareler toplamı	df	Kareler ortalaması	F	Sig.
Gruplar arası	112107,266	16	7006,704	29,323	,000
Gruplar içi	199042,659	833	238,947		
Toplam	311149,925	849			

Farklı öğretmenlerin, aradan belli bir süre geçtikten sonra, puanlama yönergesi kullanarak verdikleri puanların ortalamalarının karşılaştırılması tek yönlü varyans analizi ile yapılmış ve öğretmenlerin, 4. okumalarında vermiş oldukları puanların ortalamaları arasında istatistiksel olarak anlamlı bir farklılık bulunmuştur. Bu aşamada hesaplanan G katsayısı bir öğretmen için 0,69 olarak bulunurken, 17 öğretmen için 0,974 çıkmıştır. Bu durum, öğretmen sayısının artmasının, rubrik kullanarak yapılan puanlamalarda hatayı azalttığını, dolayısıyla güvenilirliği yükselttiğini göstermektedir.

Tüm okumalar karşılaştırıldığında, 17 puanlayıcı ile puanlama yapıldığında güvenilirlik açısından yöntemler arasındaki farklılıklar azalmaktadır. Fakat tek öğretmen açısından bakıldığında rubrikle daha güvenilir sonuçlar elde edilmektedir.

4. okumalardaki farklılığın hangi gruplar arasında olduğunun belirlenmesinde Scheffe testine başvurulmuş ve sonuçlar Tablo II.26'da sunulmuştur.

TABLO II.26.

Farklı Öğretmenlerin 4. Okuma Sonrasında Verdikleri Puanlara İlişkin

Scheffe Testi Sonuçları

Puanlayıcılar	Ortalama farklar	p
1-4	-23,65	0,000
1-6	23,21	0,000
1-9	-17,40	0,012
1-10	17,62	0,00
2-3	18,25	0,005
2-4	-16,93	0,019
2-5	20,50	0,000
2-6	29,93	0,000
2-10	24,34	0,000
2-14	17,74	0,008
3-4	-35,18	0,000
3-7	-16,84	0,021
3-9	-28,93	0,000
4-5	37,43	0,000
4-6	46,87	0,000
4-7	18,34	0,004
4-8	23,37	0,000
4-10	41,28	0,000
4-11	22,59	0,000
4-12	21,09	0,000

4-13	24,46	0,000
4-14	34,68	0,000
4-15	22,90	0,000
4-16	26,96	0,000
4-17	19,81	0,001
5-7	-19,09	0,002
5-9	-31,18	0,000
5-12	-16,34	0,034
5-17	-17,62	0,010
6-7	-28,53	0,000
6-8	-23,49	0,000
6-9	-40,62	0,000
6-11	-24,28	0,000
6-12	-25,78	0,000
6-13	-22,40	0,000
6-15	-23,96	0,000
6-16	-19,90	0,001
6-17	-27,06	0,001
7-10	22,93	0,000
7-14	16,34	0,034
8-9	-17,12	0,016
8-10	17,90	0,007
9-10	35,03	0,000
9-11	16,34	0,034
9-13	18,21	0,005
9-14	28,43	0,000
9-15	16,65	0,026
9-16	20,71	0,000
10-11	-18,68	0,003
10-12	-20,18	0,000
10-13	-16,81	0,022
10-15	-18,37	0,004
10-17	-21,46	0,000

Scheffe testinin sonuçlarına göre belirli bir sürenin sonunda Tablo II.26'da belirtilen öğretmen çiftlerinin puanlama yönergesi (rubrik) kullanarak verdikleri puanlar arasında tutarlılığı bozacak düzeyde anlamlı farklılıklar çıkmıştır. Bu durum son okumaların artık öğretmenlerde sıkılma etkisi yaratmış olabileceği şeklinde yorumlanabilir. Bu nedenle öğretmenlerin birbirleriyle tutarlı puanlar vermelerinde, tek başına rubrik kullanımının olumlu etkisinin olduğu söylenemez.

BÖLÜM III

SONUÇ VE ÖNERİLER

Bu araştırma, daha önce de belirtildiği gibi, grafik yorumlama becerisine ilişkin 4 açık uçlu maddeden oluşan bir ölçme aracını, puanlayıcıların, puanlama yönergesi (rubrik) kullanarak ve kullanmadan yaptıkları puanlamalar arasında bir ilişki olup olmadığını ve varsa bu ilişkinin puanlamaların, farklı zamanlarda ve farklı puanlayıcılar tarafından yapılması halinde nasıl bir değişiklik gösterdiğini saptamak amacıyla yapılmıştır.

Bu bölümde, araştırmanın belirtilen amaç doğrultusunda elde edilen bulgulara dayalı olarak sonuçları özetlenmekte ve bu sonuçlara dayalı olarak bazı öneriler sunulmaktadır.

III.1. Sonuçlar

Araştırmanın başlangıcında puanlama yönergesinin öğretmenlerin farklı zamanlarda yapmış oldukları puanlamalar arasındaki kararlılığı artıracağı ve puanlamalar arasındaki farklılığı ortadan kaldıracağı beklenmekteydi. Araştırma sonuçları gösteriyor ki; puanlayıcılar, puanlama yönergesi kullanarak farklı zamanlarda kararlı puanlamalar yapmışlardır fakat, puanlamalar arasındaki farklılık, diğer okumalara kıyasla azalsa da ortadan kalkmamıştır. Bu noktada, puanlayıcılar, değişkenlik kaynağı olarak düşünülmektedir.

Araştırmada ayrıca, farklı öğretmenlerin puanlama yönergesi kullanarak birbirleriyle daha uyumlu puanlar vermeleri ve ortalamaları arasındaki farkların da en düşük düzeyde olması beklenmekteydi. Araştırma sonuçları gösteriyor ki; puanlama yönergesi kullanarak puanlama yapan öğretmenler arasında, puanlama yönergesi

kullanmadan puanlama yapan öğretmenlere kıyasla daha fazla uyum bulunmaktadır. Fakat puanlama yönergesi kullanılması, farklı öğretmenlerin puanlamaları arasındaki farklılıkları ortadan kaldıramamıştır. Yine de hesaplanan G katsayısına bakılarak öğretmenler arasındaki tutarlılığın puanlama yönergeleri kullanıldığında, daha fazla olduğu söylenebilir.

III.2. Öneriler

Bu çalışmada, aynı öğretmenlerin farklı zamanlarda puanlama yönergesi kullanarak verdikleri puanların kararlılığının, puanlama yönergesi kullanmadan verdikleri puanlara göre daha yüksek olduğu, fakat öğretmenler arasında puanlama tutarlılığının tam olarak sağlanamadığı ve bunu sağlayabilecek subjektif etkenlerin ortadan kaldırılamadığı düşünülmektedir. Grafik yorumlama becerisine ilişkin ölçme aracı ve değerlendirme ölçeği araştırmacı tarafından hazırlanmış ve öğretmenlerin değerlendirmelerine sunulmuş son şekli verilmiştir. Öğretmenler arasındaki tutarsızlığın giderilmesi için bütün öğretmenlerin bir araya gelip bir komisyon oluşturarak ortak fikir birliği içinde grafik yorumlama testini ve puanlama yönergesini hazırlamaları bu konuda daha yararlı olabilir.

Bu çalışma daha önceden belirtilen sayıltı ve sınırlılıklar kapsamında gerçekleştirilmiştir. Dolayısı ile elde edilen bulgular çalışma kapsamı ile sınırlıdır. Başka faktörler de araştırma kapsamına alınabilir. Farklı sınıf düzeyleri, farklı dersler, farklı beceriler, cinsiyet, öğrencilerin başarı durumları, öğretmenlerin tecrübeleri gibi çeşitli faktörler de araştırma kapsamında düşünülebilir.

Feinberg ve Shapiro'ya (2009) göre, öğrencilerin akademik düzeyleri ne kadar yüksek ise öğretmenlerin, öğrenci performanslarını yargılamaları o kadar doğru olmaktadır. Öğretmen yargılamalarındaki doğruluğun, öğrencilerin başarı düzeylerinin bir

fonksiyonu olarak düşünülmesini başka arařtırmalar da dođrulamaktadır (Coladarci, 1986; Bates ve Nettlebeck, 2001; Salvensen ve Undheim, 1994; Aktaran: Feinberg ve Shapiro, 2009). Bu açıdan bakıldığında, öğrenciler düzeylere ayrılıp çalışma tekrarlandığında daha farklı bulgulara rastlanabilir.

Deneyimli öğretmenler, daha az deneyimli öğretmenlere göre, daha dođru yargılamalarda bulunmaktadır (Jorgenson, 1975). Bu durum göz önünde bulundurulduğunda, öğretmenler tecrübelerine göre sınıflandırılıp araştırma kapsamında incelenebilir.

Grafik yorumlama becerisi, her sınıf düzeyine yönelik bir beceri olduğundan araştırma kapsamında hem ilköğretim hem lise öğretmenleri ile çalışılmıştır. Jorgenson'a (1975) göre, öğretmenlerin farklı sınıf düzeylerindeki deneyimleri de yargılamalarındaki dođruluđu etkilemektedir. Bu açıdan bakıldığında, belli sınıf düzeylerinde deneyim kazanmış öğretmenlerle çalışılarak araştırmanın tekrarlanması, öğretmenler arasındaki uyum üzerinde etkili olabilir.

Kan (2001), çalışmasında, puanlama cetveli ve yanıt anahtarı kullanımının puanlamada güvenilirliđi artırıp artırmadığını irdelemiştir. Puanlama cetveli ve yanıt anahtarı kullanıldığı takdirde öğretmenlerin farklı zamanlarda daha kararlı puanlamalar yaptıklarını, fakat öğretmenler arasında puanlama tutarlılığının tam olarak sağlanamadığını, buna sebep olan subjektif etkenlerin ortadan kaldırılamadığını görmüştür. Puanlama cetveli ve yanıt anahtarı ile puanlama yönergesi (rubrik) kullanımının karşılaştırılıp, puanlayıcılar arası tutarlılığı sağlamada hangi tekniğin daha etkili olacağının irdelendiđi bir çalışma da, alana önemli katkılar sağlayabilir.

KAYNAKÇA

- Airasian, P. W. (1991). *Classroom assessment*. New York: McGraw-Hill.
- Akkuş, O. ve Duatepe-Paksu, A. (2006). Orantısal akıl yürütme becerisi testi ve teste yönelik dereceli puanlama anahtarı geliştirilmesi. *Eğitim Araştırmaları*, 6, 25, 1-10.
- Baki, A. (2006). *Öğretmenler ve öğrenenler için ek açıklamalarla "Yeni ilköğretim programları" (1-5. sınıflar)*. Ed: Kasım Kıroğlu. Ankara: PegemA Yayıncılık.
- Bernardin, H. J. ve Orban, J. A. (1990). Leniency effect as a function of rating format, purpose for appraisal and rater individual differences. *Journal of Business and Psychology*, 5 (2), 197-211.
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382-393.
- Curcio, F., Friel, S. ve Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal of Research in Mathematics Education*, 32 (2), 124-158.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Feinberg, A. B. ve Shapiro, E. S. (2009). Teacher accuracy: An examination of

teacher based judgments of students' reading with differing achievement levels. *The Journal of Education Research*, 102 (6), 453-461.

Goodrich, A. H. (2005). Teaching with rubrics: The good, the bad and the ugly. *College Teaching*, 53.

Jolliffe, F. R. (1991). Assessment of the understanding statistical concepts. *Proceedings of the third international conference on teaching statistics*, 1, 461-466.

Jorgenson, G. W. (1975). An analysis of teacher judgments of reading level. *American Educational Research Journal*, 12 (1), 67-75.

Kan, A. (2001). *Yazılı yoklamaların puanlanmasında puanlama cetveli ve yanıt anahtarının kullanımının puanlamaya ve puanlayıcı güvenilirliğine etkisi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü .Yayınlanmamış Yüksek Lisans Tezi.

Kan, A. (2007). Performans değerlendirme sürecine katkıları açısından yeni program anlayışı içerisinde kullanılabilir bir değerlendirme yaklaşımı: Rubrik puanlama yönergeleri. *Kuram ve Uygulamada Eğitim Bilimleri*, 7 (1), 129-152.

Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement: Issues and Practice*, 12 (2), 16-23.

- Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessment. *Applied Measurement in Education*, 15(3), 249-267.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7 (25).
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment Research & Evaluation*, 7 (3).
- NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM Publications.
- Nitko, A. J. (2001). *Educational assessment of student*. Upper Saddle River, NJ: Merrill.
- Parmar, R. ve Signer, B. (2005). Sources of error in constructing and interpreting graphs: A study of fourth- and fifth- grade students with LD. *Journal of Learning Disabilities*, 38 (3), 250-261.
- Penny, J., Johnson, R. L. ve Gordon, B. (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing*, 7, 143-169.

Pereira-Mendoza, L. ve Mellor, J. (1991). Students' concepts of bar graphs- Some preliminary findings. *Proceedings of the third international conference on teaching statistics*, 1, 150-157.

Popham, W. J. (1997). What's stil wrong- and what's stil right- with rubric. *Educational Leadership*, 55, 72-75.

Tabachnick, B. G. ve Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins College Publishers.

Tekin, H. (2004). *Eđitimde ölçme ve deęerlendirme*. Ankara: Yargı Yayınevi.

Wainer, H. (1980). A test of graphicacy in children. *Applied Psychological Measurement*, 4, 331-340.

Wood, R. (1968). Objectives in the teaching of mathematics, *Educational Research*. 10, 83-98.

EKLER

Ek 1. Grafik Yorumlama Becerisine İlişkin Ölçme Aracı

Ek 2. Grafik Yorumlama Becerisine İlişkin Puanlama Yönergesi
(Rubrik)

Ek 3. Dört Farklı Okumaya İlişkin Faktör Analizi Sonuçları

Ek 4. Normallik Testi Sonuçları

EK 1: GRAFİK YORUMLAMA BECERİSİNE İLİŞKİN ÖLÇME ARACI

Sevgili öğrenciler,

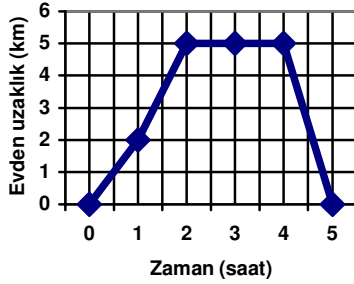
Bu ölçme aracı, sizlerin grafik yorumlama becerinizi ölçmek amacıyla hazırlanmış 4 sorudan oluşmaktadır. Verilen herbir grafiği dikkatlice inceledikten sonra cevaplarınızı belirtilen boşluklara yazılı olarak ifade etmeniz beklenmektedir. Sonuçlar bilimsel bir çalışmada kullanılacağından soruları içtenlikle cevaplamanızı rica eder, katılımınız için şimdiden teşekkür ederim.

Matematik Öğretmeni
Gözde ATMAZ

Öğrencinin Adı ve Soyadı :
Sınıfı :

1) Aşağıda bir öğrencinin bisikletle gezintisinin zamana ve aldığı yola bağlı seyrini gösteren bir grafik verilmiştir.

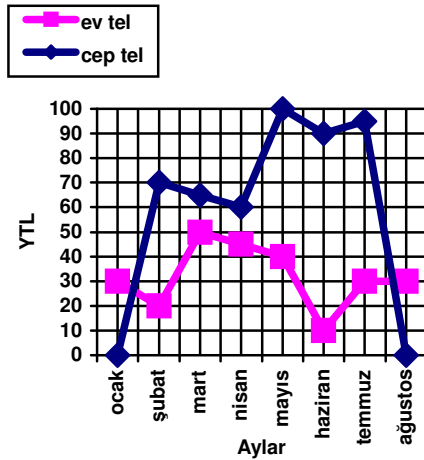
Grafikte yer alan elemanları ve bu elemanlar arası bağlantıları grafiğin bütünü ile ilişkilendirerek öğrencinin bisikletle gezintisi esnasında gerçekleştiğini düşündüğünüz olgu ve olaylar hakkında bir hikaye yazınız.



2) Aşağıdaki grafik yılın ilk 8 ayında bir kişiye ait cep telefonu ve ev telefonu faturalarındaki değişimi göstermektedir.

Grafikte yer alan elemanları ve bu elemanlar arası bağlantıları grafiğin bütünü ile ilişkilendirerek bu kişinin telefon faturalarındaki değişim hakkında bir hikaye yazınız.

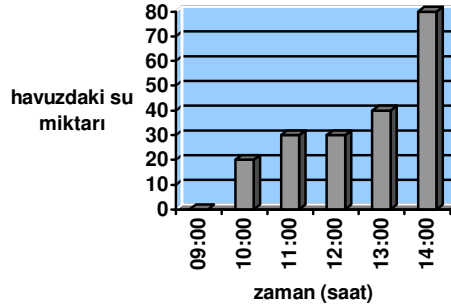
(Not: Hiç konuşulmadığı takdirde ev telefonu hattının sabit faturası 10 YTL, cep telefonu hattının sabit faturası 20 YTL'dir.)



3) Aşağıda 80 m³lük hacme sahip bir havuza, bir musluktan akıtılan suya bağlı olarak havuzda biriken su miktarının zamana bağlı değişim grafiği verilmiştir.

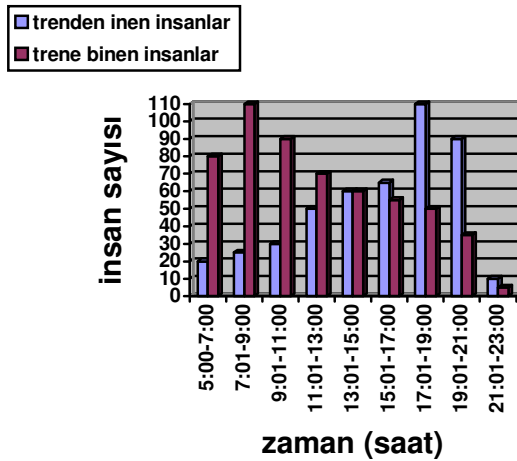
Grafikte yer alan elemanları ve bu elemanlar arası bağlantıları grafiğin bütünü ile ilişkilendirerek, söz konusu havuzun dolma sürecinde, gerçekleştiğini düşündüğünüz olgu ve olaylar hakkında bir hikaye yazınız.

(Not: Havuzu boşaltan herhangi bir deliğin olmadığı varsayılacaktır)



4) Aşağıdaki grafik, bir pazartesi günü X istasyonunda, aynı güzergahta gidip gelen trenlere binen ve bu trenlerden inen yolcuların belli zaman aralıklarındaki sayılarını göstermektedir.

Grafikte yer alan elemanları ve bu elemanlar arası bağlantıları, grafiğin bütünü ile ilişkilendirerek istasyonun konumuna, gelen ve bekleyen insanlara ve trenlere ilişkin düşüncelerinizi bir hikaye oluşturarak yazınız.



EK 2: GRAFİK YORUMLAMA BECERİSİNE İLİŞKİN PUANLAMA YÖNERGESİ (RUBRİK)

Değerli meslektaşım,

Aşağıda grafik yorumlama becerisini değerlendirmeye yönelik bir puanlama yönergesi (rubrik) sunulmaktadır. Rubrikte, öğrencilerin ölçülen beceriye ilişkin gösterebilecekleri performanslar 0'dan 4'e kadar verilmiş performans düzeylerine ve performans alt alanlarına göre ayrıntılı bir şekilde tanımlanmıştır. Sizlerden beklenen, bu beceriyi ölçmeye yönelik, öğrencilere uygulanmış olan ölçme araçlarını, verilen rubriğe dayanarak puanlamanızdır. Bu amaçla öncelikle, ölçme aracında yer alan her bir grafiğe ilişkin öğrencilerin vermiş oldukları cevapları inceleyiniz. Gösterdikleri performansa uygun tanımlamalara karşılık gelen düzeyi puan kabul ediniz. Öğrencinin cevabını her bir performans alt alanı açısından bu şekilde puanlayınız ve bir toplam puan oluşturunuz. Puanlarınızı soruların üzerine, öğrencinin almış olduğu toplam puanı da soru kağıdının üst kısmına yazarak belirtiniz.

Katkılarınız için şimdiden teşekkürler.

Gözde ATMAZ

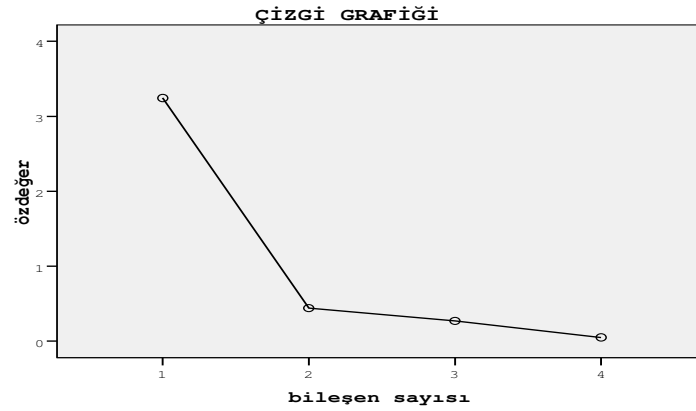
GRAFİK YORUMLAMA BECERİSİNE İLİŞKİN PUANLAMA YÖNERGESİ

PERFORMANSIN ALT ALANLARI	PERFORMANS DÜZEYLERİ				
	0	1	2	3	4
EKSENLERİN ANLAMI	Cevap yok.	Eksenlerin neyi temsil ettiklerini fark edememiştir ya da eksenlerin belirttikleri anlamları birbirine karıştırmıştır.	Eksenlerden sadece birinin neyi temsil ettiğini fark etmiştir. Diğerini yanlış anlamış ya da hiç anlamamıştır.	Eksenlerin neyi temsil ettiklerini fark etmiş fakat bazen karıştırmıştır.	Eksenlerin neyi temsil ettiklerini tamamiyle fark etmiştir.
KESİŞME NOKTALARININ ANLAMI	Cevap yok.	Satır ve sütun elemanlarının kesişme noktalarından çıkarmış olduğu anlam, grafikte aktarılmak istenen gerçek anlama uymamaktadır.	Satır ve sütun elemanlarından sadece birkaçının kesişme noktasından çıkarmış olduğu anlam, grafikte aktarılmak istenen gerçek anlama uymamaktadır.	Satır ve sütun elemanlarından çoğunun kesişme noktasından çıkarmış olduğu anlam, grafikte aktarılmak istenen gerçek anlama uymamaktadır.	Tüm satır ve sütun elemanlarının kesişme noktalarından çıkarmış olduğu anlam, grafikte aktarılmak istenen gerçek anlama uymamaktadır.
KARŞILAŞTIRMALAR	Cevap yok.	Grafikte yer alan değerler ve aralıklar arasında hiçbir karşılaştırma yapamamış ya da yaptığı karşılaştırmalar ve kurduğu azlık-çokluk, büyüklük-küçüklük ilişkileri grafiğe uymamaktadır.	Grafikte yer alan değerler ve aralıklar arasında ayrıntılı bir karşılaştırma yapmamış ve yaptığı karşılaştırmalar ölçüsünde dahi kurduğu azlık-çokluk, büyüklük-küçüklük ilişkilerininin ancak bir kısmı grafiğe uymamaktadır.	Grafikte yer alan değerler ve aralıklar arasında yapmış olduğu ayrıntılı karşılaştırmaların büyük bir kısmı grafiğe uymamaktadır. Artış ve azalışların, azlık-çokluk, büyüklük-küçüklük ilişkilerinin çoğunu doğru bir şekilde fark etmiştir.	Grafikte yer alan değerler ve aralıklar arasında yapmış olduğu ayrıntılı karşılaştırmalar tamamen grafiğe uymamaktadır. Artış ve azalışları, azlık-çokluk, büyüklük-küçüklük ilişkilerini doğru bir şekilde fark etmiştir.
BÜTÜNLEŞTİRME	Cevap yok.	Grafikte aktarılmak istenen olgu ya da olay ile grafik arasında bir bağlantı kuramamıştır. Tamamen grafik dışı olgu ve olaylardan bahsetmiştir.	Sadece grafikteki verileri aktarmış, grafiğin bütününde nasıl bir eğilim olduğuna dair ipuçları vermemiştir. Grafikte aktarılmak istenenleri bir bütün haline getirememiştir.	Grafikte gerçekleştiğini düşündüğü olgu ve olaylarla ilgili olarak kurmuş olduğu cümleler birbirinden kopuktur. Düşüncelerini bir bütün halinde sunamamıştır.	Grafikteki yönelimden hareketle grafiğin bütününde nasıl bir eğilim olduğuna karar vermemiştir. Grafikte gerçekleştiğini düşündüğü olgu ve olayları bir bütün haline getirip hikayeletmiştir.

EK 3: DÖRT FARKLI OKUMAYA İLİŞKİN FAKTÖR ANALİZİ SONUÇLARI

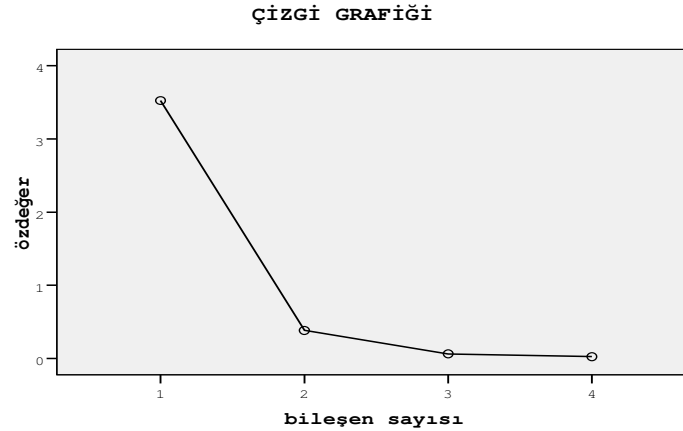
1.OKUMA

BİLEŞENLER	ÖZDEĞER	VARYANSI AÇIKLAMA ORANI (%)
1	3,244	81,098
2	,440	10,997
3	,267	6,683
4	,049	1,222



2.OKUMA

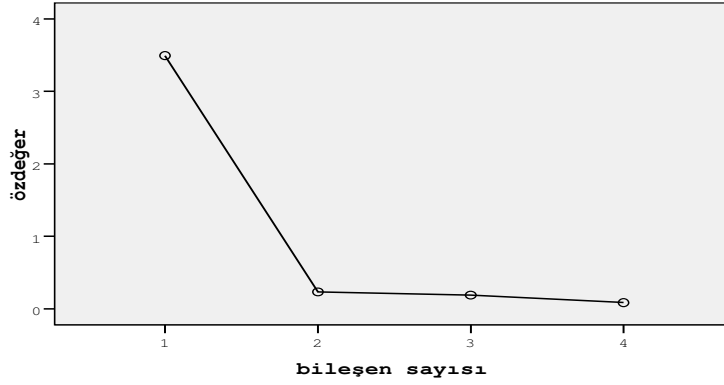
BİLEŞENLER	ÖZDEĞER	VARYANSI AÇIKLAMA ORANI (%)
1	3,524	88,098
2	,384	9,611
3	,065	1,632
4	,026	,659



3.OKUMA

BİLEŞENLER	ÖZDEĞER	VARYANSI AÇIKLAMA ORANI (%)
1	3,492	87,311
2	,234	5,839
3	,189	4,720
4	,085	2,130

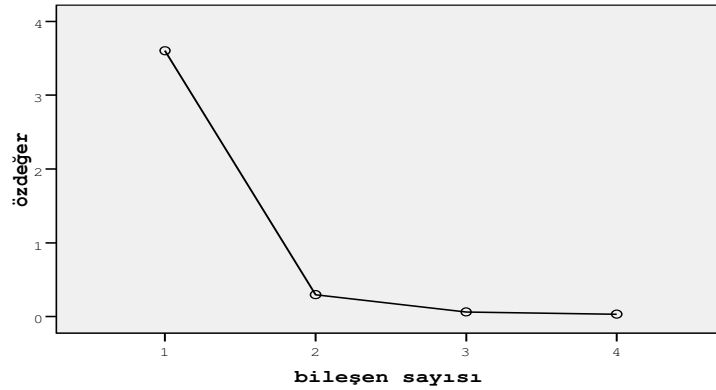
ÇİZGİ GRAFİĞİ



4.OKUMA

BİLEŞENLER	ÖZDEĞER	VARYANSI AÇIKLAMA ORANI (%)
1	3,604	90,108
2	,298	7,442
3	,064	1,597
4	,034	,853

ÇİZGİ GRAFİĞİ



EK 4: NORMALLİK TESTİ SONUÇLARI

PUANLAYICILAR	OKUMALAR	Kolmogorov-Smirnov			Çarpıklık katsayısı	Basıklık katsayısı
		istatistik	s.d.	p		
1.puanlayıcı	1.okuma	,065	50	,200	,145	-,761
	2.okuma	,073	50	,200	-,343	-,383
	3.okuma	,063	50	,200	,358	-,541
	4.okuma	,087	50	,200	-,116	-,555
2.puanlayıcı	1.okuma	,102	50	,200	-,212	-,816
	2.okuma	,110	50	,184	-,421	-,762
	3.okuma	,146	50	,010	-,332	-1,144
	4.okuma	,136	50	,022	-,609	-,526
3.puanlayıcı	1.okuma	,140	50	,016	,337	-1,013
	2.okuma	,082	50	,200	,205	-,156
	3.okuma	,101	50	,200	,178	-,637
	4.okuma	,083	50	,200	-,184	-,671
4.puanlayıcı	1.okuma	,095	50	,200	-,220	-,843
	2.okuma	,113	50	,144	-,827	,044
	3.okuma	,145	50	,010	-,483	-,430
	4.okuma	,158	50	,003	-1,021	,354
5.puanlayıcı	1.okuma	,080	50	,200	,086	-,780
	2.okuma	,110	50	,185	,781	,476
	3.okuma	,145	50	,010	-,552	-,604
	4.okuma	,151	50	,006	,836	-,121
6.puanlayıcı	1.okuma	,151	50	,006	-,713	-,109
	2.okuma	,090	50	,200	-,250	-,659
	3.okuma	,095	50	,200	-,353	-,526
	4.okuma	,104	50	,200	,148	-,715
7.puanlayıcı	1.okuma	,088	50	,200	,184	-,282
	2.okuma	,153	50	,005	-,449	-,462
	3.okuma	,104	50	,200	,209	-,801
	4.okuma	,089	50	,200	-,276	-,529
8.puanlayıcı	1.okuma	,117	50	,082	-,124	-,997
	2.okuma	,117	50	,085	-,697	-,364
	3.okuma	,138	50	,018	-,262	-,995
	4.okuma	,091	50	,200	-,502	,387
9.puanlayıcı	1.okuma	,156	50	,004	-,844	,369
	2.okuma	,156	50	,004	-,823	-,206
	3.okuma	,150	50	,006	-,779	,267
	4.okuma	,112	50	,154	-,669	-,313
10.puanlayıcı	1.okuma	,103	50	,200	-,056	-,705
	2.okuma	,063	50	,200	-,263	-,101
	3.okuma	,116	50	,090	-,530	-,666
	4.okuma	,109	50	,194	,079	,629
11.puanlayıcı	1.okuma	,112	50	,161	-,194	-,881
	2.okuma	,116	50	,092	-,479	-,130
	3.okuma	,129	50	,036	-,494	-,744
	4.okuma	,111	50	,173	-,386	,023

12.puanlayıcı	1.okuma	,118	50	,080	-,595	-,282
	2.okuma	,105	50	,200	-,396	-,296
	3.okuma	,137	50	,020	-,256	-1,063
	4.okuma	,129	50	,038	,177	-,290
13.puanlayıcı	1.okuma	,089	50	,200	,492	-,216
	2.okuma	,120	50	,069	-,276	-1,034
	3.okuma	,117	50	,084	,729	,193
	4.okuma	,111	50	,171	-,236	-,405
14.puanlayıcı	1.okuma	,095	50	,200	,061	-,395
	2.okuma	,105	50	,200	-,400	-,314
	3.okuma	,103	50	,200	,062	-1,064
	4.okuma	,070	50	,200	,537	,097
15.puanlayıcı	1.okuma	,084	50	,200	,102	-,752
	2.okuma	,086	50	,200	-,298	-,752
	3.okuma	,173	50	,001	-,300	-1,199
	4.okuma	,142	50	,013	-,213	-,924
16.puanlayıcı	1.okuma	,100	50	,200	-,207	-,809
	2.okuma	,107	50	,200	-,304	-,652
	3.okuma	,106	50	,200	-,001	-,666
	4.okuma	,146	50	,010	-1,024	,629
17.puanlayıcı	1.okuma	,078	50	,200	,249	-,793
	2.okuma	,099	50	,200	-,039	-,724
	3.okuma	,098	50	,200	,151	-,719
	4.okuma	,107	50	,200	-,104	-,501

Çarpıklığın standart hatası=0,337

Basıklığın standart hatası=0,662