

## KOMPOZİSYON TİPİ SINAVLARDA KOMPOZİSYONUN BİÇİMSEL ÖZELLİKLERİNDEN KESTİRİLEN PUANLARIN ANAHTARLA VE GENEL İZLENİMLE ELDE EDİLEN PUANLARLA İLİŞKİSİ

### THE RELATION BETWEEN SCORES PREDICTED FROM STRUCTURED FEATURES OF ESSAY AND SCORES BASED ON SCORING KEY AND OVERALL IMPRESSION IN ESSAY TYPE EXAMINATIONS

Bayram ÇETİN\*, Hülya KELEÇİOĞLU\*\*

**ÖZET:** Bu araştırmada, kompozisyon sınavlarında, kompozisyonun biçimsel özelliklerinin anahtarla puanlamayı ve genel izlenimle puanlamayı ne derece yordadığı ve en iyi yordayıcıların hangi biçimsel özellikler olduğu incelenmiştir. Ayrıca anahtar ile ve genel izlenimle puanlamanın güvenilirlikleri belirlenmeye çalışılmıştır. Bu amaç doğrultusunda Hacettepe Üniversitesi Eğitim Fakültesi İlköğretim Bölümü Sınıf Öğretmenliği Anabilim Dalında 2001-2002 yılında 1. sınıfta öğrenim gören 143 öğrenciye bir kompozisyon sınavı uygulanmıştır. Bu kompozisyonlar iki puanlayıcıya hem anahtarla hem de genel izlenimle puanlatılmıştır. Kompozisyonların içerik analizi yapılarak, her bir kompozisyonda önceden tanımlanan biçimsel özellikler ve anahtar sözcükler belirlenmiştir. Araştırmanın sonunda biçimsel özelliklerin, anahtarla yapılan puanlamanın 0,52'sini, genel izlenimle yapılan puanlamanın 0,60'ını açıkladığı görülmüştür. Anahtarla puanlamanın güvenilirliği 0,45, genel izlenimle puanlamanın güvenilirliği de 0,53 bulunmuştur.

**Anahtar sözcükler:** kompozisyon sınavları, biçimsel özelliklerin puanlanması, anahtarla puanlama, genel izlenimle puanlama, yordama, puanlama güvenilirliği.

**ABSTRACT:** The purpose of this research is to determine the extent to which structural features of essays and keywords predict scoring by key and overall impression in writing test; and which structural features are the best predictors; and the degree of inter-reliability of scoring by key and overall impression. For that purposes, a writing test has been administered to 143 undergraduate freshman students enrolled at Hacettepe University Faculty of Education Department of Elementary Education Program in Primary Education in 2001-2002 school year. The writing of the students have been scored by two Turkish Language and Literature teachers using both scoring key and overall impression scoring methods. Predetermined structural features and keywords were identified by doing content analysis for the compositions written by students. As a result, the explanation level of the structural features to the key scoring was found 0,52 and the explanation level of the structural features to the overall impression was found 0,60. Inter-rater reliability of the key scoring was found 0,45, and overall impression scoring was 0,53.

**Keywords:** composition exams, scoring the structural features, key scoring, overall impression scoring, prediction, inter-rater reliability.

### 1. GİRİŞ

Eğitimde öğrenci davranışlarının ölçülmesinde çeşitli ölçme araçları kullanılmaktadır. Bunlardan başlıcaları yazılı yoklamalar, kısa cevaplı testler, doğru-yanlış tipi testler, çoktan seçmeli testler, sözlü yoklamalardır (Tekin 1993; Tekin 1997).

Kompozisyon sınavları, belirlenmiş bir doğru cevabı bulunmaması yönüyle diğer sınav türlerinden ayrılır. Kompozisyon sınavları öğrencilerin dili kullanması, bilgileri ve düşüncelerini düzenlemesi ve bunları etkili bir anlatımla ifade etmesi gibi özellikleri ölçmeye yöneliktir. Bu nedenle bu sınavlarda ölçülen özelliklerin çoktan seçmeli testlerle geçerli olarak yoklanamaz. Yazılı anlatımda ölçülmek istenilen beceri dilbilgisi yeterliliğinin yanında düşüncüyü de etkili bir şekilde ifade etme yeteneğidir.

\* Araştırma Görevlisi. Hacettepe Üniversitesi Eğitim Fakültesi Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı - Ankara

\*\* Yrd. Doç. Dr. Hacettepe Üniversitesi Eğitim Fakültesi Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı - Ankara

Yazılı anlatım becerilerini ölçmeye yönelik olarak yapılan kompozisyon sınavlarının güvenilirliğinin ve geçerliğinin istenilen düzeye getirilmesi diğer sınavlara göre daha güçtür. Bu güçlük, yazılı anlatım becerilerinin, öznel öğeleri fazlaca içerdiğinden dolayı güvenilir olarak puanlanamamasından kaynaklanmaktadır. Bu tür sınavlarda öğrencilerin duygu, düşünce ve hayallerini dilin kurallarına uygun bir biçimde etkili bir anlatım kullanılarak düzenlemeleri istenir. Bu nedenle bir doğru cevap bulunmaz ve doğru cevap anahtarı da hazırlanamaz. Bunun yerine içerik, organizasyon, dilbilgisi, anlatım yapısı, sözcük dağarcığı, noktalama ve yazım kuralları gibi ana başlıklar altında uygunluk ve yaratıcılıkları kapsayan bir anahtar hazırlanabilir.

Kompozisyon tipi sınavların puanlanmasında izlenen yollardan biri genel izlenimle puanlamadır. Bu tür bir puanlama, puanlama anahtarı kullanılmadığı için geçerlik ve güvenilirlik konusunda sakıncalar içerebilir.

Enginarlar (1991) kompozisyonların puanlanmasında temel olarak üç yöntemin olduğunu; diğer yöntemlerin bu temel yöntemlerin farklı birer uygulaması olduğunu belirtmektedir. Bu üç yöntem hata sayma, bütünsel (holistic) anahtar kullanarak puanlama ve analitik anahtar ile puanlama yöntemleridir.

Hata sayma; içerik, örgütlenme, dilbilgisi, sözcük dağarcığı, noktalama ve yazım kuralları ana başlıklarına göre yapılan hatalara bakılarak hatanın önemine göre puan kesme şeklinde yapılan bir puanlama metodudur. Bütünsel puanlama metodunda da 5 ya da 6 sınıflama belirlenmektedir. Bunlar, “Çok İyi”, “İyi”, “Yeterli”, “Yeterlik Altı” ve “Zayıf” gibi sınıflamalardan oluşmakta ve bu sınıfların niteliklerini taşıyan kompozisyonların özelliklerini belirten bir anahtar hazırlanmaktadır. Bir kompozisyon hangi sınıfın özelliklerini taşırsa o sınıfın puanını almaktadır. Analitik puanlama metodunda ise ana başlıklar altında ayrıntılı olarak kompozisyonda bulunması gereken özellikleri gösteren bir liste hazırlanmakta ve bu listedeki özelliklerin bulunma derecesine göre bir puanlama yapılmaktadır.

Kompozisyon tipi sınavların puanlanmasında, birden fazla puanlayıcı kullanılmasının güvenilirliği artırıcı rolü olmaktadır. Ancak birden çok puanlayıcı kullanmak her zaman mümkün olmadığından bu sınavların güvenilirliğini artırmak için bazı arayışlara gidilmiştir. Bunlardan biri de toplam puanı, puanların kestiricisi olabilecek bazı biçimsel özelliklerden yola çıkarak kestirmektir. (Burstein ve arkadaşları, 1998b)

Kompozisyon tipi sınavların biçimsel özelliklerinden yola çıkarak toplam puanın kestirilebilmesi için, biçimsel özelliklerin belirlenmesi ve analiz edilmesi gerekmektedir. Burstein ve arkadaşları (1998c) kompozisyonları puanlarken üç genel özellik kullanıldığını belirtmişlerdir. Bunlar cümlelerin yapıları (syntactic structure), anlatım düzeni yapısı (rhetorical structure) ve içerik (topical content) dir. Bu üç ana başlığın kompozisyondaki muhtemel göstergelerinin belirlenmesi ve kompozisyondaki bulunuşluk derecelerinin tespit edilmesi ile uzman puanlarını kestirmek mümkün olabilmektedir.

Cümle yapıları özelliklerinin göstergeleri olarak farklı cümle tür ve yapıları alınmakta; bu tür ve yapıların frekansları cümle ve sözcük yapılarının yordayıcıları olarak düşünülmektedir (Burstein ve Chodorow 1999).

Anlatım düzeni yapısının bazı ipucu sözcüklerle belirlenebileceği belirtilmektedir (Burstein ve arkadaşları, 1998c). Anlatım düzeni yapısının yordayıcıları olarak başlığın varlığı ve sözcük sayısı, kompozisyonun uzunluğu, paragraf sayısı, ortalama paragraf uzunluğu; örnek verirken kullanılan sözcükler; kişisel kanıları anlatan ifadeler vb. alınmaktadır (Burstein ve arkadaşları 1998a).

Kompozisyonun içerik kısmının yordayıcıları olan anahtar sözcükler, kompozisyon konusunun içeriğiyle ilgili yazılmış makaleler ve yüksek puan almış kompozisyonlarda yer alan sözcükler taranarak belirlenebilmektedir. Belirlenen bu anahtar sözcükler çok fazla olacağından, bu konuda yapılmış çalışmalarda kullanılan anahtar sözcüklerin ağırlıklandırılarak toplamları alınabilmektedir (Burstein ve Marcu 2000). Bu ağırlıklandırma hem sözcüklerin kompozisyon içindeki diğer sözcükler bakımından bulunuşluk derecesine göre hem de bu sözcüklerin diğer kompozisyonlarda bulunma derecelerine göre yapılabilir.

Kompozisyon sınavlarında uzman kanıları, objektif testlere göre daha fazla kullanıldığından puanlamaya karışan öznel yargılar da daha fazla olmaktadır. Uzman puanlayıcıların puanlarına ilişkin objektif olarak ölçülebilen bazı yordayıcıların elde edilmesi, kompozisyonların puanlama güvenilirliğini artırabilir. Kompozisyonun biçimsel özelliklerine dayalı, diğer özelliklerine göre daha objektif puanlanabilen, sayılabilecek, bazı yordayıcılar tanımlanabilir ve bunlar kompozisyon puanının yordayıcıları olarak alınabilir (Clauser 2000). Bu yordayıcılar, daha sonra bir bilgisayar programıyla belirlenerek ikinci bir puanlayıcı olarak kullanılabilir. Bu yöntem, hem zaman hem de maliyet açısından daha ekonomik bir puanlama yöntemi olabilir.

Bu araştırmada kompozisyonların içeriği, anlatım düzeni ve cümle ile ilgili yapıları biçimsel özelliklerine göre analiz edilmiş, bu ana başlıklar etrafında alt yapılar belirlenerek bunların kompozisyonda bulunma dereceleri belirlenmiş ve bu biçimsel özelliklere göre uzman puanları yordayıcıları olarak bir puan elde edilmiştir. Ayrıca alan uzmanı iki öğretmene hem genel izlenimle hem de analitik bir anahtar ile kompozisyon puanlaması yaptırılmıştır. Böylece, kompozisyon tipi sınavlarda uzman puanlarının yordayıcıları olan biçimsel özellikleri belirleme ve bu tür sınavlarda kullanılacak geçerliği ve güvenilirliği yüksek bir puanlama yöntemi geliştirebilme amaçlanmıştır.

Belirlenen yordayıcılarla geçerliği ve güvenilirliği yüksek bir puanlama yöntemi elde edilebilirse hem ekonomik olarak hem de zaman açısından daha uygun bir puanlama yöntemi geliştirilmiş olacaktır. Ayrıca kişisel kanı ortadan kalkacağından, daha objektif bir puanlama yapılabilecektir.

### **1.1. Problem Cümlesi**

Kompozisyon tipi sınavlarda kompozisyonun biçimsel özellikleri, anahtarlar ve genel izlenimle elde edilen puanları ne derece yordamaktadır? Anahtarlar puanlamanın ve genel izlenime göre puanlamanın puanlama güvenilirlikleri nasıldır?

### **1.2. Alt Problemler**

1. Kompozisyon tipi sınavlarda, kompozisyonun biçimsel özellikleri anahtara göre puanlamayı ne derecede yordamaktadır?
2. Kompozisyon tipi sınavlarda, kompozisyonun biçimsel özellikleri genel izlenime göre puanlamayı ne derecede yordamaktadır?
3. Kompozisyon tipi sınavlarda anahtarlar ve genel izlenimle puanlama yöntemlerinin güvenilirlikleri nasıldır?

## **2. YÖNTEM**

Bu araştırma, öğrencilerin yazılı anlatım becerilerini ölçmede kullanılan kompozisyon türü sınavları puanlamada, kompozisyonun biçimsel özelliklerinden uzman puanlarının yordayıcılarını belirlemeyi ve varolan bilgilere yenilerini katmayı amaçladığından temel araştırma niteliği taşımaktadır. Araştırmada kompozisyonların, biçimsel özellikleri ile uzmanlar tarafından anahtarlar ve genel izlenime göre puanlanması arasındaki ilişkiler ve puanlama çeşitlerinin puanlama güvenilirlikleri incelendiğinden araştırmanın modeli ilişkisel taramadır.

### **2.1. Çalışma Grubu**

Araştırma, 2001-2002 öğretim yılında Hacettepe Üniversitesi Eğitim Fakültesi İlköğretim Bölümü Sınıf Öğretmenliği Anabilim Dalı birinci sınıfında öğrenim gören 143 öğrenci üzerinde yürütülmüştür.

## 2.2. Veri Toplama Araçları

Araştırmada öğrencilere “Bazı İnsanlar Reklamların İnsanları Gerçekte İhtiyaç Olmayan Ürünleri Almaya Teşvik Ettiğini, Bazıları ise Reklamların İnsanlara Yaşamı Kolaylaştıracak Yeni Ürünleri Tanıtıldığını Düşünür. Hangi Görüşle Aynı Fikirdeyiniz? Açık Sebepler ve Örneklerle Fikrinizi Destekleyecek Bir Kompozisyon Yazınız” başlıklı bir kompozisyon sınavı uygulanmış ve kompozisyonlar elde edilmiştir. Bu kompozisyonlar alan uzmanı iki tane Türk Dili ve Edebiyatı öğretmenine hem bir puanlama anahtarına göre hem de genel izlenime göre puanlatılmıştır.

## 2.3. Verilerin Çözümlemesi

143 öğrencinin yazdığı kompozisyonlar Türk Dili ve Edebiyatı alanında uzman iki öğretmene hem genel izlenimle, hem de puanlama anahtarı kullanılarak puanlatılmıştır.

Kompozisyonun biçimsel özellikleri belirlenirken içerik analizi tekniğinden yararlanılmıştır. İçerik analizinde biçimsel özelliklerin bir kısmı elle sayma şeklinde yapılmış, bir kısmı da bilgisayarda Microsoft Word programı ile belirlenmiştir. Bu biçimsel özelliklerden cümle yapısı ve anlatım yapısı değişkenleri Tablo 2’de verilmiştir.

Kompozisyonun uzunluğu, cümle sayısı, sözcük sayısı, paragraf sayısı, başlığın varlığı, sözcük sayısı ve uzunluğu; kompozisyonun anlatım yapısının yordayıcıları olarak belirlenen, örnek verirken kullanılan ipucu sözcükleri, konunun devam ettiğini gösteren ipucu sözcükleri, kişisel kanaatleri belirtirken kullanılan ipucu sözcükleri ve sonuç bölümünde kullanılan ipucu sözcükleri ile kompozisyondaki yazım hataları sayısı Microsoft Word programıyla belirlenmiştir.

Anahtar sözcükler ise iki grupta düşünülmüştür. Birinci grupta, puanlamadan önce, konu ile ilgili makaleler taranarak anahtar sözcükler elde edilmiştir. İkinci grupta ise, kompozisyonların puanlanmasından sonra yüksek puan almış kompozisyonlardan çıkarılan anahtar sözcükler yer almıştır.

Anahtar sözcüklerin bulunmuşluk dereceleri Microsoft Word programının “Bul” ve “Değiştir” komutlarıyla belirlenmiş ve sayısallaştırılmıştır. Birinci grupta yer alan, dergilerde yayınlanmış makalelerden çıkarılan anahtar sözcükler her kompozisyon için Burstein ve Marcu (2000)’dan alınan yönteme göre, aşağıdaki formül kullanılarak yapılmıştır.

$$\text{ağırlık}_{wj} = \frac{\text{freq}_{wj}}{\max \text{freq}_j} \log \left( \frac{n}{\text{komp}_w} \right)$$

**ağırlık<sub>wj</sub>** : w sözcüğünün j kompozisyonundaki ağırlığı

**freq<sub>wj</sub>** : w sözcüğünün j kompozisyonundaki frekansı

**maxfreq<sub>j</sub>** : j kompozisyonundaki en çok tekrarlanan sözcüğün frekansı

**n** : regresyon analizinde kullanılan toplam kompozisyon sayısı

**komp<sub>w</sub>** : w sözcüğünün bulunduğu toplam kompozisyon sayısı

Yukarıdaki formüle göre ağırlıklandırma yapıldıktan sonra her kompozisyondaki sözcükler için hesaplanan ağırlıkların toplamları alınarak tüm anahtar sözcükler için her bir öğrencinin bir tek ağırlıklı anahtar sözcük puanı elde edilmiştir.

Kompozisyonlar puanlandıktan sonra yüksek puan almış kompozisyonlardan daha önceki anahtar sözcüklerde olmayan yeni anahtar sözcükler belirlenmiştir. Bu anahtar sözcüklerin sayıları toplanarak ağırlıklandırmaya gidilmeden ayrı bir anahtar sözcük puanı elde edilmiştir.

Kompozisyonların olası yordayıcıları uzmanların anahtara göre ve genel izlenime göre verdikleri puanlarla çoklu regresyon analizine tabi tutularak yordayıcılık güçleri belirlenmiştir.

Genel izlenim ve anahtarla puanlama yöntemlerinin güvenilirliği, iki puanlayıcının aynı yöntemle verdikleri puanlar arasındaki korelasyon hesaplanarak elde edilmiş ve iki puanlama yönteminin güvenilirlikleri arasında fark, z testi ile test edilmiştir (Spiegel 1995).

### 3. BULGULAR VE YORUMLAR

*Birinci alt probleme* yanıt bulabilmek için kompozisyonlardan çıkarılan biçimsel özelliklere ait veriler ile kompozisyonun anahtarla puanlanması sonucu elde edilen puanlar arasında çoklu regresyon analizi yapılmıştır. Regresyon denklemiyle ilgili ANOVA tablosu Tablo 1’de verilmiştir.

Tablo 1’de görüldüğü gibi oluşturulan regresyon denklemi 0,05 manidarlık düzeyinde önemlidir. Elde edilen bulgu, kestirilen kısmi regresyon katsayılarıyla oluşturulacak denklemi kullanarak puanlayıcının anahtarla yaptığı puanlamanın kestirebileceği şeklinde yorumlanabilir.

**Tablo 1:** Kompozisyonun Biçimsel Özellikleri ile Anahtarla Puanlamanın Regresyon Analizine İlişkin ANOVA Tablosu

Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Regresyon	8334,93	28	297,68	4,32	0,00
Hata	7847,28	114	68,84		
Toplam	16182,21	142			

**Tablo 2:** Kompozisyonun Biçimsel Özelliklerinden Anahtarla Puanlama ve Genel İzlenimle Puanlamanın Yordama Bağlılıkları

Bağımsız Değişkenler	Anahtarla Puanlama				Genel İzlenimle			
	Kat sayıları <sup>1</sup>	Std. Hata	t	p	Kat sayıları <sup>1</sup>	Std. Hata	t	p
<i>(Sabit)</i>	44,365	37,499	1,183	,239	28,228	29,689	0,951	,344
Başlığın Varlığı	9,141	3,562	2,566	,012	10,545	2,820	3,740	,000
Başlığın Sözcük Sayısı	0,956	1,063	0,900	,370	1,121	0,842	1,332	,186
Başlığın Uzunluğu	-0,077	0,152	-0,511	,611	-0,127	0,120	-1,060	,292
Kompozisyonun Uzunluğu	0,004	0,024	0,146	,884	-0,011	0,019	-0,570	,570
Sözcük Sayısı	-0,063	0,172	-0,366	,715	0,086	0,136	0,627	,532
Ortalama Sözcük Uzunluğu	-4,034	5,470	-0,737	,462	-0,348	4,331	-0,080	,936
Ortalama Cümle Uzunluğu	0,220	0,077	2,859	,005	0,117	0,061	1,917	,058
Paragraf Sayısı	-0,081	1,026	-0,079	,937	0,581	0,812	0,715	,476
Ortalama Paragraf Uzunluğu	-0,009	0,011	-0,854	,395	0,006	0,009	0,710	,479
Yazım Hatası Sayısı	-1,340	0,404	-3,316	,001	-0,382	0,320	-1,194	,235
Basit Cümle Sayısı	0,998	0,394	2,535	,013	0,317	0,312	1,016	,312
Bileşik Cümle Sayısı	2,951	1,779	1,659	,100	0,565	1,409	0,401	,689
Sıralı Cümle Sayısı	0,516	0,637	0,810	,419	-0,153	0,505	-0,303	,763
Yardımcı Cümlecik Sayısı	-0,820	1,401	-0,586	,559	-0,587	1,109	-0,529	,598
Master Cümlecği Sayısı	0,035	0,280	0,124	,901	-0,063	0,221	-0,284	,777
Sıfat Cümlecği Sayısı	0,266	0,330	0,806	,422	0,222	0,262	0,848	,398
Olasılık ve Olabilirlik anlatan Cümle Sayısı	0,034	0,366	0,092	,927	0,346	0,290	1,195	,235
Gereklilik Cümlecği Sayısı	0,010	0,477	0,021	,983	0,264	0,378	0,698	,487
Dilek-Şart Cümlecği Sayısı	-0,509	0,783	-0,650	,517	0,553	0,620	0,892	,374
Edat Sayısı	0,013	0,169	0,079	,937	-0,100	0,134	-0,749	,455
Soru Cümlesi Sayısı	0,967	0,561	1,724	,087	1,234	0,444	2,779	,006
Ünlem Cümlesi Sayısı	0,071	1,189	0,060	,952	-1,310	0,942	-1,391	,167
Kişisel Kanıları Belirten Sözcükler	1,275	0,742	1,718	,088	0,201	0,588	0,342	,733
Örnek Vermede Kullanılan Sözcükler	-1,245	1,615	-0,771	,442	-3,747	1,279	-2,930	,004
Bu, Bunlar ve türevleri	0,346	0,274	1,264	,209	0,048	0,217	0,217	,828
Sonuç Kısmının Belirteci Sözcükler	-0,989	1,611	-0,614	,540	-2,208	1,276	-1,731	,086
Ağırlıklandırılmış Anahtar Sözcükler	1,312	0,492	2,668	,009	2,669	0,924	2,887	,005
Puanlamadan Sonra Çıkarılan Anahtar Sözcükler	1,780	1,168	1,524	,130	0,902	0,389	2,318	,022

<sup>1</sup>Tabloda italik karakterle yazılmış sayılar *a* katsayısını ve buna ait değerleri, diğerleri *b* katsayıları ve bunlara ait değerleri göstermektedir.

Bağımsız değişkenlerle bağımlı değişken arasındaki çoklu korelasyon katsayısı 0,72, belirtme katsayısı 0,52, kestirmenin standart hatası da 8,30 olarak bulunmuştur. Bu bulgular, kompozisyonun biçimsel özelliklerinin anahtarla puanlamadaki değişkenliğinin 0,52'sini açıklayabildiğini göstermektedir.

Regresyon analizi sonucunda elde edilen kısmi regresyon katsayılarının anlamlılığı t testi ile test edilerek, kompozisyonun anahtarla puanlanmasının en iyi yordayıcıları belirlenmiştir. Ayrıca tüm bağımsız değişkenler yordayıcılık güçlerine göre sıralanmıştır. Bu analiz sonucunda elde edilen kısmi regresyon katsayıları, bunlara ait standart sapmalar, t değerleri ve manidarlık düzeyi Tablo 2'de verilmiştir.

Tablo 2'te görüldüğü gibi uzman puanlayıcıların puanlama anahtarı kullanarak verdiği puanların en iyi yordayıcıları olarak "Yazım Hatası Sayısı", "Ortalama Cümle Uzunluğu", "Yayınlanmış Makalelerden Çıkarılmış ve Ağırlıklandırılmış Anahtar Sözcükler", "Başlığın Var Olup Olmaması" ve "Basit Cümle Sayısı" gibi kompozisyonların biçimsel özelliklerine ilişkin değişkenler belirlenmiştir.

İkinci alt probleme yanıt bulabilmek için kompozisyonlardan çıkarılan biçimsel özelliklere ait veriler ile kompozisyonun genel izlenimle puanlanmasından elde edilen puanlar arasında yine çoklu regresyon analizi yapılmıştır. Bu regresyon analizi sonucunda elde edilen regresyon denkleminin önemliliği F testi ile test edilmiştir. Regresyon denklemiyle ilgili ANOVA tablosu Tablo 3'te verilmiştir.

Tablo 3'te görüldüğü gibi oluşturulan regresyon denklemi 0,05 manidarlık düzeyinde önemlidir. Elde edilen bulgu, bu regresyon denkleminin kompozisyon puanlarını kestirmek için kullanılabileceği şeklinde yorumlanabilir.

**Tablo 3:** Kompozisyonun Biçimsel Özellikleri ile Genel İzlenimle Puanlamanın Regresyon Analizine İlişkin ANOVA Tablosu

Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Regresyon	7310,22	28	261,08	6,05	0,00
Hata	4919,03	114	43,15		
Toplam	12229,25	142			

Regresyon analizinde kullanılan bağımsız değişkenler ile bağımlı değişken arasındaki çoklu korelasyon katsayısı 0,77, belirtme katsayısı 0,60, kestirmenin standart hatası ise 6,57 bulunmuştur. Bu bulgular, kompozisyonun biçimsel özelliklerinin, kompozisyonun genel izlenimle puanlanmasındaki değişkenliğinin 0,60'ını açıklayabildiğini göstermektedir. Kısmi regresyon katsayılarının anlamlılığı t testi ile test edilerek, genel izlenimle puanlanmasının en iyi kestiricileri belirlenmeye çalışılmıştır. Regresyon analizi sonucunda elde edilen kısmi regresyon katsayıları, bunların standart sapmaları, t değerleri ve manidarlık düzeyleri Tablo 2'de verilmiştir. Tablo 2'de görüldüğü gibi uzman puanlayıcıların genel izlenimle verdiği puanların en iyi yordayıcıları olarak "Başlığın Var Olup Olmaması", "Örnek Verirken Kullanılan Sözcüklerin Varlığı", "Kompozisyon Puanlandıktan Sonra Çıkarılan Anahtar Sözcükler", "Soru Cümlesi Sayısı" ve "Yayınlanmış Makalelerden Çıkarılmış ve Ağırlıklandırılmış Anahtar Sözcükler" gibi biçimsel özellik değişkenleri belirlenmiştir. Bu yordayıcılar, 0,05 manidarlık düzeyinde anlamlı bulunan yordayıcılardır.

Üçüncü alt problemdeki kompozisyon puanlama yöntemlerinin güvenilirliklerini bulmak için, puanlayıcılardan aynı yöntemle elde edilen puanlar arasında Pearson Momentler Çarpımı korelasyon katsayısı hesaplanmıştır. Sonuçlar Tablo 4'de gösterilmiştir.

**Tablo 4:** Anahtarla Puanlama ve Genel İzlenimle Puanlama Yöntemlerinin Güvenirlikleri

Puanlama Yöntemi	r	z	p
Anahtar	0,45	0,88	0,189
Genel İzlenim	0,53		

Tablo 6'da, anahtarla puanlama yönteminin güvenilirliğini gösteren korelasyon katsayısının 0,45; genel izlenimle puanlama yönteminin güvenilirliğini gös-



teren korelasyon katsayısının 0,53 olduğu görülmektedir. Puanlayıcılar arası tutarlılık anlamındaki bu güvenirlilik katsayıları yeterince yüksek değildir. Bu durum, anahtarla puanlamada daha fazla olmak üzere, puanlayıcılar arasında tutarlılığın düşük olduğunu göstermektedir.

Puanlama anahtarı kullanarak yapılan bir puanlamada tutarlılığın düşük olması puanlamaya esas olan özelliklerin yeterince iyi tanımlanamamasından da kaynaklanıyor olabilir.

Genel izlenimle yapılan puanlamalar arasındaki tutarlılığın anahtarla yapılan puanlamalar arasındaki tutarlılıktan daha yüksek çıkması bu iki korelasyon katsayısı arasında fark olup olmadığını test etmeye yönelmiştir. Korelasyon katsayılarının test edilmesi sonucunda, aralarında 0,05 düzeyinde manidar bir fark bulunamamıştır. İki puanlama türünün puanlayıcılar arası tutarlılık anlamındaki güvenirliliğinin manidar düzeyde farklı olmadığı görülmektedir.

#### 4. SONUÇ VE ÖNERİLER

Araştırmada elde edilen bulgulara bağlantılı olarak sonuçlar ve bunlara dayalı olarak geliştirilen öneriler aşağıda verilmiştir.

##### 4.1. Sonuçlar

Araştırmanın bulguları, kompozisyonların puanlanmasında, kompozisyonun bazı biçimsel özelliklerinin hem anahtarla puanlamayı hem de genel izlenimle puanlamayı yordayabileceğini göstermektedir. Kompozisyonun biçimsel özelliklerine ilişkin değişkenler, anahtarla puanlamadaki değişkenliğin % 52'sini, genel izlenimle puanlamadaki değişkenliğin %60'ını açıklayabildiği ortaya çıkmıştır. Her iki yöntemde de beşer yordayıcı anlamlı bulunmuş; bunlardan "*Yayınlanmış Makalelerden Çıkarılmış ve Ağırlıklandırılmış Anahtar Sözcükler*" ile "*Başlığın Var Olup Olmaması*"nın ortak oldukları görülmüştür.

Anahtarla puanlamada puanlayıcı güvenirliliği, genel izlenimle puanlamanın puanlayıcı güvenirliliğinden daha düşük bulunmuştur. Bu durum, sadece iki puanlayıcı kullanılmasından ve puanlayıcıların anahtarla puanlama yeterliliklerinin az olmasından kaynaklanmış olabilir. Puanlayıcı güvenirliliklerinin tam olması, birinci ve ikinci alt problemde pek çok değişkenin yordayıcılık gücünün düşük çıkmasına yol açmış olabilir.

##### 4.2. Öneriler

- 1) Regresyon analizine katılan biçimsel özelliklere ait farklı değişkenler kullanılabilir. Farklı cümlecik türleri; cümlelerin farklı öğeleri de belirlenerek analize katılabilir.
- 2) Anahtarla puanlamanın güvenirliliğini artırmak için, uzmanlara anahtar puanlama kullanma konusunda önceden bir eğitim verilebilir.
- 3) Bu araştırma, bütünsel puanlama ve hata sayma yöntemleri de kullanarak tekrarlanabilir.

#### KAYNAKLAR

- Burstein, J. ve D. MARCU. "Toward using text summarization for essay-based feedback." 20 Eylül 2001 <<http://www.ets.org/research/talsum.doc>>.
- Burstein, J., K. Kukich, S. Wolff, C. Lu ve M. Chodorow (1998b). "Enriching automated essay scoring using discourse marking." 20 Eylül 2001 <<http://www.ets.org/research/dscrfinl.pdf>>.
- Burstein, J., K. Kukich, S. Wolff, C. LU, M. Chodorow, L. B. Harder, ve M. D. Harris (1998c). "Automated scoring using a hybrid feature identification technique." 20 Eylül 2001 <<http://www.ets.org/research/aclfinal.pdf>>.
- Burstein, Jill; M. Chodorow (1999). "Automated essay scoring for nonnative english speakers." 20 Eylül 2001 <<http://www.ets.org/research/ac199rev.pdf>>.

- Burstein, J., K. KUKICH, S. WOLFF, C. LU, ve M. CHODOROW (1998a) "Computer analysis of essays." 20 Eylül 2001 <<http://www.ets.org/research/ncmefinal.pdf>>.
- Clauser, B. E. "Recurrent issues and recent advances in scoring performance assessments." *Applied Psychological Measurement*, Vol. 24 No:4, 2000:310-324.
- Çetin, B. (2002). Kompozisyon tipi sınavlarda kompozisyonun biçimsel özelliklerinden kestirilen puanların anahtarla ve genel izlenimle puanlanmasından elde edilen puanlarla ilişkisi. *Yayımlanmamış Yüksek Lisans Tezi*, Hacettepe Üniversitesi, Ankara.
- Enginarlar, H. "A quantitative and qualitative comparison of three techniques of grading ESL/EFL Essays." *Journal of Human Sciences*, X. sayı 1, 1991: 23-45.
- Spiegel, M. R. *İstatistik 870 çözümlü problem*. (Çevirenler: A. Ayaydın, M. Turanlı, İ. H. Armutlulu, Ş. Bülbül.) Bilim Teknik Yayınevi. İstanbul, 1995.
- Tekin, H. *Eğitimde ölçme ve değerlendirme*. Yargı yayınları. Ankara, 1993.
- Turgut, M. F. *Eğitimde ölçme ve değerlendirme*. Onuncu baskı. Yargıcı Matbaası, Ankara, 1997.