



KORKUT ATA TÜRKİYAT ARAŞTIRMALARI DERGİSİ

Uluslararası Dil, Edebiyat, Kültür, Tarih, Sanat ve Eğitim Araştırmaları Dergisi

The Journal of International Language, Literature, Culture, History, Art and Education Research

Sayı/Issue 14 (Şubat/February 2024), s. 993-1007.

Geliş Tarihi-Received: 11.12.2023

Kabul Tarihi-Accepted: 22.01.2024

Araştırma Makalesi-Research Article

ISSN: 2687-5675

DOI: 10.51531/korkutataturkiyat.1403091

Yazılı Üretimlerin Analitik Değerlendirilmesi: Ölçüt Ayırımının Çok Yüzeyle Rasch Analiziyle İncelenmesi

Analytical Evaluation of Essays: Examining Criterion Discrimination Using Multi-Facet Rasch Analysis

Ayfer SAYIN*

Öz

Bu çalışmada ortaöğretim öğrencilerinin yazılı üretimlerinin analitik bir şekilde değerlendirilmesinin çok yönlü bir şekilde incelenmesi amaçlanmıştır. Nicel araştırma yöntemlerinden tarama araştırması deseninde yürütülen çalışmaya 225 ortaöğretim 9. sınıf öğrencisi katılmıştır. Öğrencilere bir yazma görevi verilmiş ve öğrencilerin oluşturduğu yazılı üretimler iki farklı puanlayıcı tarafından puanlanmıştır. Puanlama işlemi üç aşamada gerçekleştirilmiştir: Birinci aşamada dört puanlayıcı, rastgele seçilen 10 üretimi puanlamış, ardından odak grup görüşmesi sonucunda puanlama anahtarı üzerinde gerekli değişiklikler yapılmıştır. İkinci aşamada puanlayıcılar yeniden farklı 10 yazılı üretimi puanlamış ve odak grup görüşmesi sonucunda puanlama anahtarına son şekli verilmiştir. Puanlayıcılar arası fark çok düşük olduğu için üçüncü aşamada seçilen iki puanlayıcı 225 öğrenci tarafından ortaya konulan yazılı üretimleri, analitik dereceli puanlama anahtarı ile puanlamıştır. Analitik dereceli puanlama anahtarında yönergeye uygunluk, içerik (fikirler), anlatım (metin organizasyonu) ve yazım-noktalama olmak üzere dört ölçüt yer almıştır. Araştırma kapsamında tüm değişkenlerin aynı anda analiz edilerek değerlendirilebilmesi amacıyla "yazılı üretimler (225) x puanlayıcılar (2) x ölçütler (4)" olmak üzere üç yüzeyle çaprazlanmış desen oluşturulmuş ve çok yüzeyle Rasch analizi yapılmıştır. Analiz sonrasında öğrenci puanlarının anlamlı bir şekilde ayrıştığı; yazılı üretimleri düşük, orta ve yüksek olarak değerlendirilen öğrenciler olduğu belirlenmiştir. Analiz sonrasında aynı puanlama anahtarını kullanmasına karşın puanlayıcılar arasında anlamlı bir farklılık olduğu belirlenmiş, birinci puanlayıcının daha cimri/katı puanlama yaptığı tespit edilmiştir. Araştırma sonucunda ayrıca puanlama anahtarındaki ölçütlerden içerik ve anlatım ölçütlerinin aynı logit düzeyinde yer aldığı, başka bir anlatımla puanlayıcılar tarafından ayrıştırılmadığı belirlenmiştir. Araştırmadan elde edilen bulgular doğrultusunda içerik ve anlatım ölçütlerinin neden ayrıştırılmadığı tartışılarak bu konuda önerilere yer verilmiştir.

Anahtar Kelimeler: Türkçe eğitimi, analitik dereceli puanlama anahtarı, yazma becerisi, puanlayıcı güvenilirliği, çok yüzeyle Rasch analizi.

Abstract

This study aimed to examine the analytical evaluation of essays by high school students in a multidimensional manner. A total of 225 high school students participated in the study. The

* Doç. Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, e-posta: ayfersayin@gazi.edu.tr, ORCID: 0000-0003-1357-5674.

current research employed a survey design, which is quantitative research. The students were assigned a writing task, and two raters assessed their essays. The scoring process took place in three stages: In the first stage, the raters scored ten randomly selected essays. Subsequently, necessary revisions were made to the rubric based on the outcomes of the focus group interview. In the second stage, the raters scored ten different essays again, and the rubric was finalized following the focus group interview. In the third stage, both raters evaluated the essays of 225 students using the analytic rubric. The rubric included four criteria: instructions, content (ideas), expression (text organization), and spelling-punctuation. To analyze and evaluate all variables simultaneously, a three-facet crossed design was created as "essays (225) x raters (2) x criteria (4)," and a multi-faceted Rasch analysis was conducted. After the examination, it was determined that student scores differed significantly, with some students having low, medium, and high-rated essays. Additionally, a noteworthy difference between the raters' scores despite using the same rubric. The first rater was identified as being more stingy/strict in scoring. As a result of the study, it was also noted that the content and expression criteria were at the same logit scale, indicating that the raters could not differentiate them. Based on the findings from the research, the reasons behind the inability to distinguish between content and expression criteria were discussed, and suggestions were provided.

Keywords: Turkish education, analytical rubric, essays, rater reliability, writing skill, many-facet Rasch analysis.

Giriş

Wittgenstein (1996) "Dilimin sınırları, dünyamın sınırlarıdır." sözüyle dil gelişiminin önemine dikkat çekmektedir. Wittgenstein kişilerin kullandıkları dilin özelliklerinin; kendilerini ifade ettikleri ve konumlandıkları yerin temel belirleyicisi olarak görmektedir. Örneğin yabancı bir ülkede o ülkenin ana diline hâkim olmayan bir kişi; çevresindekileri anlamakta güçlük çekecek, kendini yeterince ifade edemeyecek ve dünyası bu anlamda daralacaktır. Bu nedenle tüm dünyada küçük yaşlardan itibaren dil becerilerinin gelişimi önem kazanır. Bu gelişim sürecinde dört temel dil becerisi olan dinleme/izleme, okuma, konuşma ve yazma alan becerilerine odaklanılır. Bu araştırmada öğrencilerin yazma becerileri ele alınmıştır.

Yazma; bireylerin bilgilerini ve temel becerilerini ifade etmenin bir yolu olarak tanımlanır (Walker vd., 2005). Bireyler yazma becerilerinin gelişmesiyle birlikte bilgiyi aktarabilir, düşüncelerini etkili bir şekilde ifade edebilir, bilgilerini organize edebilir ve iletişim yeteneklerini geliştirebilir (Ungan, 2007). Yazma, bireylerin entelektüel kapasitelerini genişletirken aynı zamanda eleştirel düşünme becerilerini de güçlendirir. Bu nedenle örgün eğitimlerde öğrencilerin yazma becerilerinin geliştirilmesi önemli görülür ve bu süreçte yazma becerisinin bileşenlerine odaklanılır. Yapılan araştırmalar; yazma becerisinin geliştirilmesi sürecinin; bilgi, beceri, bireysel farklılık, hafıza, öğretim yöntem ve tekniği, öğretmen tutumu, yazılı ürünlerin değerlendirilmesi ve geri bildirim gibi değişkenlerden etkilendiğini ortaya koymaktadır (Kellogg ve Raulerson, 2007; Özbay, 2015; Tangpermpoon, 2008; Yoleri ve Tanış, 2014). Bu araştırmada yazma becerilerinin geliştirilme sürecine etki eden yazılı ürünlerin değerlendirilmesi üzerine odaklanılmıştır.

Yazma becerilerinin değerlendirilmesi, öğrencilerin bu alandaki gelişimini anlamak ve öğretim stratejilerini en verimli şekilde kullanmak için önemlidir. Değerlendirme süreci, öğrencilerin yazılı ifadelerini analiz ederek güçlü yönleri belirlemeye ve geliştirilmesi gereken alanları tanımlamaya yardımcı olur. Bu sayede öğretmenler öğrencilerini daha etkili bir şekilde yönlendirerek becerinin gelişimini etkili ve hızlı hâle getirir. Çünkü değerlendirme süreci sonrasında sağlanacak geri bildirim, yazma becerilerindeki ilerlemeleri izleme fırsatı sunar (Demir Bilican ve Yıldırım, 2019; Göçer, 2010; Kellogg ve Raulerson, 2007; Uyar, 2016; Walker ve Riu, 2008). Bu nedenle de yazılı üretimlerin objektif bir şekilde değerlendirilmesi ve dolayısıyla kullanılan puanlama anahtarlarının geçerli ve güvenilirlik kanıtları öne çıkar. Yazılı üretimlerin

değerlendirilmesinde mutlaka bir puanlama anahtarı kullanılması gerekir. Çünkü puanlama anahtarıyla öğrenci performansı belirli kriterlere göre değerlendirildiği için nesnel sonuçlara ulaşılır. Başka bir anlatımla puanlama anahtarları ile gerçekleştirilen yazılı değerlendirmeler daha güvenilir sonuçlar üretir (Demir Bilican ve Yıldırım, 2019; Jonsson ve Svingby, 2007; Popham, 1997; Silvestri ve Oescher, 2006). Ayrıca puanlama anahtarları, öğretmenlere öğrenci ürünlerini standart bir çerçeve üzerinden analiz etme ve değerlendirme imkânı sunar. Bu sayede puanlama anahtarları, öğretmenlerin öğrenci başarılarını daha tutarlı bir şekilde ölçmelerine ve öğrencilere daha etkili geri bildirimler sunmalarına olanak sağlar. Puanlama anahtarları aynı zamanda yazma becerilerindeki belirli güçlü ve zayıf yönleri belirleme konusunda rehberlik eder (Aslanoglu, 2022; Attali, 2007; Durmuş, 2022; Göçer, 2010). Bu noktada öğrencilerin yazılı üretimlerinde kullanılan puanlama anahtarı sonuçlarının incelenmesi ön plana çıkar.

Yazma becerilerini değerlendirmek için kullanılan puanlama anahtarları, belirli ölçütlere dayalı olarak öğrenci performansını değerlendirme amacı taşır. Öğrencilerin yazılı ürünlerinin puanlanmasında temel özelliklere dayalı puanlama ile analitik ve bütüncül puanlama anahtarları kullanılabilir. Temel özelliklere dayalı yapılan puanlamada yazılı ürünlerin değerlendirilmesi için belirlenen temel özellikleri içeren bir puanlama anahtarı kullanılır. Buradaki temel felsefe öğrencilerin belirli bir anlatım biçimini ne düzeyde iyi yazabildiklerine odaklanmaktır, bunun için de öğrenci ürünlerindeki ortak ve öğrenciye özgü özelliklere odaklanılır (Spandel, 2005; Zorbaz, 2013). Analitik dereceli puanlama anahtarları, ölçülecek özelliğin boyutlarını gösteren bir çeşit puanlama ölçüsüdür ve öğrencilerin farklı özelliğe sahip performanslarının ayrı ayrı değerlendirilmesine olanak tanır (Beaglehole, 2014; Popham, 1997). Puanlama sürecinde performansı bölümlere ayırdığı için daha fazla bilgi sağlamakta (Moskal, 2003); puanlanan ve puanlayıcı için ortak bir algı oluşturmaktadır (Arter ve McTighe, 2001; Beaglehole, 2014; Moskal, 2009). Bir diğer puanlama anahtarı ise bütüncül puanlama anahtarıdır. Analitik dereceli puanlama anahtarının aksine bütüncül puanlama anahtarlarında genel bir değerlendirme sunulur. Bütüncül puanlama anahtarında puanlayıcı değerlendirmeye esas olan bileşenleri tek bir ölçüt hâline getirir ve özellikleri bir bütün olarak ele alarak değerlendirme yapar (Carr, 2000; Popham, 1997). Bütüncül puanlama anahtarının temel bakış açısı; yazma ürününün bütünü, onun parçaları toplamından daha değerli olmasına dayanır (Espin vd., 2017). TOEFL gibi uluslararası geniş ölçekli uygulamalardaki yazılı üretimlerin bütüncül bir şekilde değerlendirildiği, ölçülmek istenen özelliğin bir bütün teşkil ettiği ve daha kullanışlı olduğu ifade edilmektedir (Papageorgiou ve diğerleri, 2021). Açıklamalardan da anlaşılacağı üzere öğrencilerin yazılı üretimlerinin değerlendirilmesinde değerlendirme amacı ve kullanılan ölçütlerin özellikleri, kullanılacak puanlama anahtarının belirleyicisidir. Bu noktada alan yazında en iyi sonuç üreten puanlama anahtarı türünü bulmaya yönelik çalışmalar mevcuttur ve bu çalışmalarda genellikle analitik ve bütüncül puanlama anahtarına dayalı puanlamaların karşılaştırıldığı görülür.

Analitik ve bütüncül puanlama anahtarlarından elde edilen sonuçların karşılaştırıldığı çalışmalarda, analitik dereceli puanlama anahtarının bütüncül puanlama anahtarından daha güvenilir sonuçlar verdiğini gösteren araştırmalar vardır (Ghalib ve Al-Hattami, 2015; Johnson ve Svingby, 2007). Bununla birlikte bazı çalışmalarda bütüncül puanlamanın analitik dereceli puanlamadan daha güvenilir sonuç ürettiği tespit edilmiştir (Çetin, 2011; Çetin ve Kelecioğlu, 2004). Ancak bu konudaki araştırmalar; genel olarak bütüncül puanlamalar ile analitik dereceli puanlamalar arasında benzer sonuçlar elde edildiğini göstermektedir (Anıl ve Şahhüseyinoğlu, 2008; Carr, 2000; Parlak ve Doğan, 2014; Putri ve Melani, 2022). Bu doğrultuda sonuçları göz önüne alınarak puanlama anahtarına karar verilebilir. Örneğin ölçütlerin birbirinden net bir şekilde

ayrışabildiği durumlarda analitik dereceli puanlama anahtarı kullanmak, ürünler hakkında daha detaylı bilgi sağlar. Analitik dereceli puanlama anahtarında bir özellik, sadece bir ölçüt kapsamında puanlandığı için öğrencilerin yazma becerilerindeki güçlü ve zayıf yönleri belirleyerek onlara etkili geri bildirim vermek de daha kolaydır (Attali, 2007). Ancak bazen yazılı üretimdeki özellikler birbirinden net bir şekilde farklılaşamayabilir. Bu durumda puanlayıcılar bir özelliği birden fazla ölçütte puanlar ise ölçütler arasında analitik bir ayrışma gerçekleşemez. Bu da sonuçların güvenilirliğini ve geçerliğini olumsuz yönde etkiler. Bu nedenle de analitik dereceli puanlama anahtarında kullanılacak ölçütlerin belirlenmesi, tanımlanması ve derecelendirilmesi önem arz eder.

Türk dilinde üretilen yazılı üretimlerin değerlendirilmesine yönelik alan yazında farklı puanlama anahtarları bulunmaktadır. Bu puanlama anahtarlarında kullanılan ölçütler de farklılık göstermektedir. Örneğin Göçer (2005) kompozisyonların değerlendirilmesinde plan fikri, olaylara bakış açısı, anlatımda duruluk ve doğallık, konuyu anlama ve anlam bütünlüğü, imla (yazım) ve noktalama ölçütlerinin belli ağırlıklarla kullanılabileceğini ifade etmiştir. Özkara (2007) ise yazılı anlatımların fikirler, organizasyon, üslup, kelime seçimi, cümle akıcılığı, imla ölçütleri bağlamında değerlendirilmesine yönelik bir puanlama anahtarı geliştirmiştir. Özdemir (2014) analitik yazma süreçlerinin değerlendirilmesinde fikirler, düzenleme, ifade biçimi, kelime seçimi, cümle akıcılığı, imla ve noktalama kuralları ve paylaşım ölçütleri bağlamında değerlendirilebileceğini belirtmiştir. Demir Bilican ve Yıldırım (2019) içerik, söylem yönetimi, anlaşılabilirlik, bağlama uygun dil kullanımı, yazım kuralları ve dil bilgisi ölçütlerini kullanmıştır. Demir (2020) öğrencilerin akademik yazmalarını değerlendirmek için ölçüt ve alt ölçütler tanımlamış; görev tanımı, içerik, kaynak kullanımı, anlatım ve biçimsel özellikler ölçütler bağlamında ürünleri değerlendirmiştir. Temizkan ve Yıldırım (2023) yazma eğitimine yönelik lisansüstü çalışmalarında kullanılan ölçme araçlarını incelemiştir. Öğrencilerin yazma beceri düzeylerini belirlemek üzere gerçekleştirilen doktora tez çalışmalarında 21 farklı puanlama anahtarı yer aldığını belirlemişlerdir. Bunlar Yazılı Anlatımı Değerlendirme İçin Derecelendirilmiş Puanlama Anahtarı, Yazılı Anlatım Değerlendirme Rubriği, Bilgilendirici Metin Puanlama Anahtarı, Hikâye Unsurları Analitik dereceli puanlama anahtarı gibi farklı isimde olduğu gibi farklı sayıda ölçüt barındırmaktadır (Temizkan ve Yıldırım, 2023). Bu konuda yapılan çalışmalar yazma becerisinin değerlendirmesinde; ürünün amacına (akademik, sınıf içi vb.), ortaya konulan metin türüne (bilgilendirici, öyküleyici gibi), öğrencilerin gelişim düzeyine (ilkokul, ortaokul vb.) göre kullanılan ölçütlerin farklılaştığını göstermektedir. Ancak yapılan araştırmalarda ölçütlerin birbirinden ne düzeyde ayrışabildiğine; ölçütler arasındaki ilişkilere yönelik bilgiler yer almamaktadır. Bu durum; puanlama anahtarında yer alan ölçütlerin işlevliliğinin incelenmesi, farklı sayıda ve içerikteki ölçütlerden hangilerinin ne düzeyde sonuç ürettiğinin görülmesi bakımından önemlidir. Çünkü ölçütlerin doğru bir şekilde yapılandırılması, puanlama güvenilirliğini doğrudan etkileyecektir. Aksi hâlde puanlayıcılar bir özelliği hangi ölçütte puanlayacağını bilemeyecek ve puanlar puanlayıcılara göre değişkenlik gösterebilecektir. Bu konuda yapılan bazı çalışmalar, puanlayıcıların puanlama ölçütlerini yorumlama biçimlerinde önemli farklılıklar olduğunu ve ölçütlerin birbirinden ayrışmadığını göstermektedir (Engelhard, 1994; Eckes, 2008; Lumley ve McNamara, 1995). Bu araştırma kapsamında ortaöğretim öğrencilerinin yazılı üretimleri değerlendirilirken kullanılan ölçütlerin puanlayıcılar tarafından nasıl anlaşıldığı ve ölçütlerin birbirinden ne düzeyde ayrıştığına odaklanılmaktadır. Buradan hareketle araştırmada cevap aranan alt problemler şu şekildedir:

1. Ortaöğretim 9. sınıf öğrencilerinin yazma becerileri, öğrenciler arasında anlamlı bir farklılık göstermekte midir?

2. Ortaöğretim 9. sınıf öğrencilerinin yazma becerilerinin değerlendirilmesinde puanlayıcılar arasında anlamlı bir fark bulunmakta mıdır?

3. Ortaöğretim 9. sınıf öğrencilerinin yazma becerilerinin değerlendirilmesinde puanlama anahtarındaki ölçütler anlamlı bir şekilde ayrılmakta mıdır?

Yöntem

Bu bölümde araştırmanın modeli, çalışma grubu, verilerin toplanması ve verilerin analizine ilişkin bilgilere yer verilmiştir.

Araştırmanın Modeli

Bu araştırmada ortaöğretim öğrencilerinin yazılı ürünlerinin analitik değerlendirilmesinde kullanılan ölçütlerin puanlayıcılar tarafından ne düzeyde ayrıştığının belirlenmesi amaçlanmıştır. Bu doğrultuda öğrencilerin yazılı ürünleri x puanlayıcılar x ölçütler olmak üzere üç yüzeyle bir model test edilmiştir. Bu doğrultuda çalışma, tarama araştırmasında yürütülmüştür. Tarama araştırmaları, araştırma konusunun var olan durumu, özellikleri, yaygınlığı, nedenleri ve sonuçları hakkında bilgi edinmek amacıyla gerçekleştirilen çalışmalardır (Creswell ve Poth, 2016). Bu araştırma, Gazi Üniversitesi Etik Komisyonunun 07.11.2023 tarihli 19 sayılı toplantı kararı doğrultusunda etik ilkelere uygun bulunarak yürütülmüştür.

Çalışma Grubu

Bu araştırmanın temel odağı, ortaöğretim öğrencilerinin yazma becerilerini değerlendirmek değil, puanlayıcıların ölçütleri ne düzeyde ayrıştırabildiğini belirlemektir. Bu nedenle evren ve örneklem seçimine gidilmemiş; çalışma grubu olarak belirlenen 9. sınıf öğrencileri ile araştırma yürütülmüştür. Araştırmada kullanılan veri toplama aracı 9. sınıf öğrenci düzeyinde geliştirildiği için çalışma bu sınıf düzeyi ile gerçekleştirilmiştir. Çalışmanın verileri, bir özel eğitim kurumunun 9. sınıfına devam eden 225 öğrenciden toplanmıştır. Araştırmaya katılan öğrencilerin %44'ü (n=98) İstanbul'da, %30'u (n=68) Ankara'da, %26'sı (n=59) Antalya'da öğrenim görmektedir. Öğrencilerin %59'unu (n=133) kadın, %92'sini (n=41) erkek öğrenciler oluşturmaktadır. Araştırma kapsamında ayrıca öğrencilerin yazılı ürünleri iki uzman tarafından puanlanmıştır. Puanlayıcıların Türkçe eğitimi alanında doktora eğitimleri bulunmaktadır ve yazma becerileri üzerine çalışmaları vardır.

Verilerin Toplanması

Araştırma kapsamında verilerin toplanması için öğrencilere öncelikle bir yazma görevi verilmiştir. Ardından da puanlayıcıların öğrencilerin ürünlerini değerlendirmeleri için analitik dereceli puanlama anahtarı hazırlanmıştır. Veriler öğrencilerden dijital ortamda toplanmıştır.

Yazma Görevi: Ortaöğretim 9. sınıf öğrencilerinin yazılı bir ürün ortaya koyabilmeleri için yazma görevi, KAGGLE platformundan seçilmiştir. KAGGLE; veri bilimcileri ve makine öğrenimi uzmanları için veri kümeleri, eğitim kaynakları gibi özellikler sunan çevrim içi bir platformdur. Bu platformda daha önce uygulamaları yapılmış, geçerlik ve güvenilirlik kanıtları elde edilmiş görevler ve puanlar yer almaktadır (Bojer ve Meldgaard, 2021; Iglovikov vd., 2017; Tolachev vd., 2020). Bu araştırma için KAGGLE platformundan öğrencilerin ikna edici yazı yazmaları için tasarlanmış bir görev seçilmiştir. Görev öncelikle Türkçeye çevrilmiştir. Ardından ifadelerin öğrenci düzeyine ve Türk kültürüne uygunluğunu belirlemek için üç Türkçe eğitim uzmanı ile (iki uzman

öğretim üyesi, bir uzman doktora öğrencisi olmak üzere) odak grup görüşmesi gerçekleştirilmiştir. Uzmanların görüşleri doğrultusunda göreve son şekli verilmiştir. Görev üç bölümden oluşmaktadır: Birinci bölümde konu başlığı altında uzaktan eğitimden söz edilmiş ve öğrencilerin seçim yapabilmeleri için “uzaktan eğitimin devam etmesini isteyenler” ile “yüz yüze eğitim yapılmasını isteyenler” olmak üzere iki seçenek sunulmuştur. Seçeneklere yönelik de kısa açıklamalarda bulunulmuştur. Görevin ikinci bölümünde yönerge bulunmaktadır. Yönergede kelime sayısı, süre ile metin türü hakkında öğrencilere bilgi verilmiştir. Görevin üçüncü bölümünü de analitik dereceli puanlama anahtarı oluşturmuştur.

Analitik dereceli puanlama anahtarı: Öğrencilerin yazılı üretimlerinin puanlanması için analitik dereceli puanlama anahtarı hazırlanmıştır. KAGGLE’de yer alan yazılı üretimin puanlanmasında kullanılan puanlama anahtarı Türkçeye uyarlanmıştır. Bu süreçte puanlama anahtarı öncelikle Türkçeye çevrilmiştir. Ardından puanlama için dört ölçüt (yönergeye uygunluk, içerik (fikirler), anlatım (ifade biçimi) ve yazım noktalama) tanımlanmıştır. Her bir ölçüt 1-3 arasında derecelendirilerek her bir derecenin karşılığı açıklanmıştır. Puanlama anahtarı için çeviriler tamamlandıktan sonra üç Türkçe eğitim uzmanı ile (iki öğretim üyesi, bir doktora öğrencisi) odak grup görüşmesi ile ifadelerle son şekli verilmiştir. Puanlama anahtarının geçerlik ve güvenilirliğine ilişki kanıt toplanması amacıyla puanlama için ön uygulama çalışmaları yapılmış, puanlamalar üç aşamada gerçekleştirilmiştir. Bu aşamalardan ilk ikisi hem eğitim sürecini hem de geçerlik ve güvenilirlik incelemesini kapsamaktadır. Puanlamanın birinci aşamasında rastgele 10 yazılı ürün seçilmiş ve dört uzmandan (Türkçe eğitimi alanında doktora derecesine sahip) 10 ürünü puanlama anahtarı doğrultusunda puanlaması istenmiştir. Birinci puanlama aşaması sonrasında ortaya çıkan farklı puanlamaların nedeni odak grup görüşmesi ile tartışılmış ve puanlama anahtarında gerekli düzenlemeler yapılmıştır. Daha sonra yine rastgele 10 yazılı ürün seçilerek dört uzman tarafından puanlanmıştır. İkinci puanlama aşaması sonrasında puanlama anahtarına son şekli verilmiş ve ek açıklamalar yapılmıştır. Lumley (2005) puanlama anahtarındaki ek açıklamaların puanlayıcı güvenilirliğine olumlu katkısı olduğunu ifade etmektedir. Puanlama anahtarına son şekli verildikten sonra üçüncü puanlama aşamasında iki puanlayıcıdan 225 yazılı ürünü bağımsız bir şekilde puanlaması istenmiştir. Puanlama anahtarı Ek 1’de gösterilmiştir.

Verilerin Analizi

Bu çalışmada öğrencilerin yazılı üretimlerinin değerlendirilmesinde kullanılan ölçütlerin birbirinden ne düzeyde ayrıştığı incelenmiştir. Bunun için öğrenci yazılı üretimi x puanlayıcı x ölçüt olmak üzere üç yüzeyle çaprazlanmış Rasch analizi (Myford ve Wolfe, 2004) yapılmıştır. Çok yüzeyle Rasch analizi, bir ölçme aracının farklı yönlerini aynı anda analiz etmeyi sağlayan bir ölçme modelidir. Bu, araştırmacının ölçme aracının güvenilirliğini ve geçerliliğini değerlendirmesini sağlar (Eckes, 2009). Rasch analizinin gerçekleştirilebilmesi için karşılanması gereken varsayımlardan tek boyutluluk, yerel bağımsızlık ve model-veri uyumu kontrol edilmiştir (DeMars, 2021). Öncelikle her bir puanlayıcının 225 yazılı üretimdeki ölçüt puanlarına ilişkin açımlayıcı faktör analizi yapılmıştır. Açımlayıcı faktör analizi, çok sayıda değişkeni (maddesi) daha az sayıda anlamlı yapıya indirgemeyi amaçlayan bir istatistiksel yöntemdir. Analiz öncesinde veri setinin analize uygunluğunun tespit edilmesi için KMO testi ve Barlett anlamlılık testi kullanılır. KMO testi 0,90’ın üzerinde ise veri seti, analize çok uygundur. 0,80’in üzerinde ise iyi, 0,70’in üzerinde ise orta, 0,60’ın üzerinde ise zayıftır. Barlett anlamlılık testi, değişkenler arasında anlamlı korelasyonlar olup olmadığını ölçer. Testin değeri anlamlı ise değişkenler arasında anlamlı korelasyonlar vardır ve veri seti analize uygundur

(Tabachnik, Fidell ve Ullman, 2013). Bu çalışmada yapılan analizde ölçütlere verilen cevapların KMO değeri 0,823 olarak hesaplanmıştır. Bu değer, verilerin analize iyi düzeyde uygun olduğunu göstermektedir. Ayrıca, Barlett testinin değeri anlamlı bulunmuştur ($p < 0,05$). Bu sonuç, değişkenler arasında anlamlı korelasyonlar olduğunu göstermektedir. Veri setinin analize uygun olduğunun tespit edilmesinin ardından faktör yapısının incelenmesi için öz değer incelemesi gerçekleştirilmiştir. Öz değeri 1'den büyük olan tek bir faktör olduğu (2.454) belirlenmiş, bu nedenle de scree plot düşüşü ve paralel analiz yöntemlerine başvurulmamıştır. Dört ölçütün varyansın %61.347'sine açıklık getirdiği belirlenmiştir. Ölçütlerin faktör yük değerlerinin 0.555 (yazım ve noktalama), 0.794 (yönergeye uygunluk), 0.860 (içerik) ve 0.881 (anlatım) olarak hesaplandığı belirlenmiştir. Dört ölçüte verilen puanların güvenilirliğini belirlemek için hesaplanan Cronbach alfa iç tutarlılık katsayısı 0.782 olarak hesaplanmıştır. Araştırma kapsamında öğrenci x puanlayıcı x ölçüt etkileşiminin tek bir analizde incelenmesi istendiğinden çok yüzeyli Rasch yapılmıştır. Analiz öncesinde yüzeydeki hücre sayısının yeterli olduğu (1800) olduğu tespit edilmiştir. Tüm puanların standartlaştırılmış artık değerlerinin ± 2 arasında yer aldığı belirlenmiştir. Varsayımların karşılanmasının ardından çok yüzeyli Rasch analizleri hesaplanmış, bunun için Minifac (Facets) paket programı kullanılmıştır.

Bulgular

Araştırma kapsamında sırasıyla öğrencilerin yazma ürünlerinin, puanlayıcıların ve puanlama anahtarında yer alan ölçütlerin ayrışması çok yüzeyli Rasch analizi ile incelenmiştir. Bu doğrultuda öncelikle öğrenci ölçüm raporu için değerler hesaplanmış ve sonuçlar Tablo 1'de gösterilmiştir.

Tablo 1. Öğrenci Ölçüm Raporu

Öğrenci	Logit	Std.error	Infit		Outfit	
			MnSq	Zst	MnSq	Zst
162	2.69	0.83	1.33	0.7	1.6	0.8
160	2.09	0.73	2.38	2.1	2.14	1.5
161	2.09	0.73	2.38	2.1	2.14	1.5
158	1.6	0.68	2.03	1.8	1.76	1.3
159	1.6	0.68	2.06	1.8	1.79	1.3
---	---	---	---	---	---	---
185	-3.62	1.11	0.57	-0.3	0.26	-0.3
186	-3.62	1.11	0.57	-0.3	0.26	-0.3
187	-3.62	1.11	0.57	-0.3	0.26	-0.3
188	-3.62	1.11	0.57	-0.3	0.26	-0.3
189	-3.62	1.11	0.57	-0.3	0.26	-0.3
Mean	-1.52	0.75	0.94	0	0.93	0.1
SD	1.34	0.15	0.44	0.8	0.5	0.8

Model, Sample: RMSE = .76 Standard deviation = 1.10
 Discrimination ratio=1.65 Discrimination index = 2.76
 Discrimination index of reliability= 0.78
 Model, Fixed (all same) chi square=676.3 df =224 p= .00
 Model, Random (normal) chi square =189.7 df = 223 p= .95

Araştırma kapsamında 225 öğrencinin yazılı üretimi değerlendirilmiş ve üretimlere ilişkin puanların logit değerlerini göstermek amacıyla en yüksek ve en düşük logit değerlerine sahip olan 5'er öğrenci Tablo 1'de gösterilmiştir. Tablo 1'de yer alan öğrenci ölçüm raporu incelendiğinde öğrenci puanlarının logit değerlerinin -3.62 ile 2.69 arasında değiştiği görülmektedir. Rasch analizinden elde edilen güvenilirlik indeksi 0.78 değeriyle kabul edilebilir düzeydedir. Ayrıca, 2.26'lık ayırma indeksi, öğrencilerin yazılı

üretim puanlarının zorluk açısından önemli ölçüde farklı olduğunu göstermektedir. Yazılı üretim puanlarına ilişkin iç ve dış güvenilirlik kriterinin tüm öğrenciler için ± 2 değerinden aşırı bir sapma göstermediği tespit edilmiştir.

Tablo 2. Puanlayıcı Ölçüm Raporu

Rater	Logit	Std.error	Infit		Outfit	
			MnSq	Zst	MnSq	Zst
1	0.39	0.07	0.95	-0.9	0.84	-2.1
2	-0.39	0.06	1.06	1.2	1.03	0.5
Mean	0	0.07	1.01	0.2	0.93	-0.8
SD	0.39	0	0.05	1.1	0.1	1.3

Model, Sample: RMSE = .07 Standard deviation = 0.39
 Discrimination ratio=5.73 Discrimination index = 7.97
 Discrimination index of reliability= 0.97
 Model, Fixed (all same) chi square=67.7 df =1 p= .00

Tablo 2’de puanlayıcılara ilişkin ölçüm raporu yer almaktadır. Çok yüzeyle Rasch analizi sonucunda birinci puanlayıcıya ilişkin logit değerinin 0.39; ikinci puanlayıcıya ilişkin logit değerinin ise -0.39 olarak hesaplandığı görülmektedir. Hesaplanan değerler, ikinci puanlayıcının birinci puanlayıcıya göre daha cömert/bonkör puanlama yaptığını göstermektedir ($r=0.97$). Hesaplanan ayırma indeksi, puanlayıcılar arasındaki farklılığın puanlarda ve puanlara kaynaklık eden alt bileşenlerde de farklılaştığını göstermektedir. Çünkü puanlayıcılar 4 ölçüt bağlamında * 3 dereceli olmak üzere 12 kategorili bir puanlama yapmışlardır. Hesaplanan değerler, puanlayıcıların söz konusu kategorilerin sekizinde ayrıştığını ortaya koymaktadır.

Araştırmada son olarak puanlama anahtarındaki ölçütlerin ayrışmasına yönelik incelemeler gerçekleştirilmiş ve ölçüt ölçüm raporu Tablo 3’te gösterilmiştir.

Tablo 3. Ölçüt Ölçüm Raporu

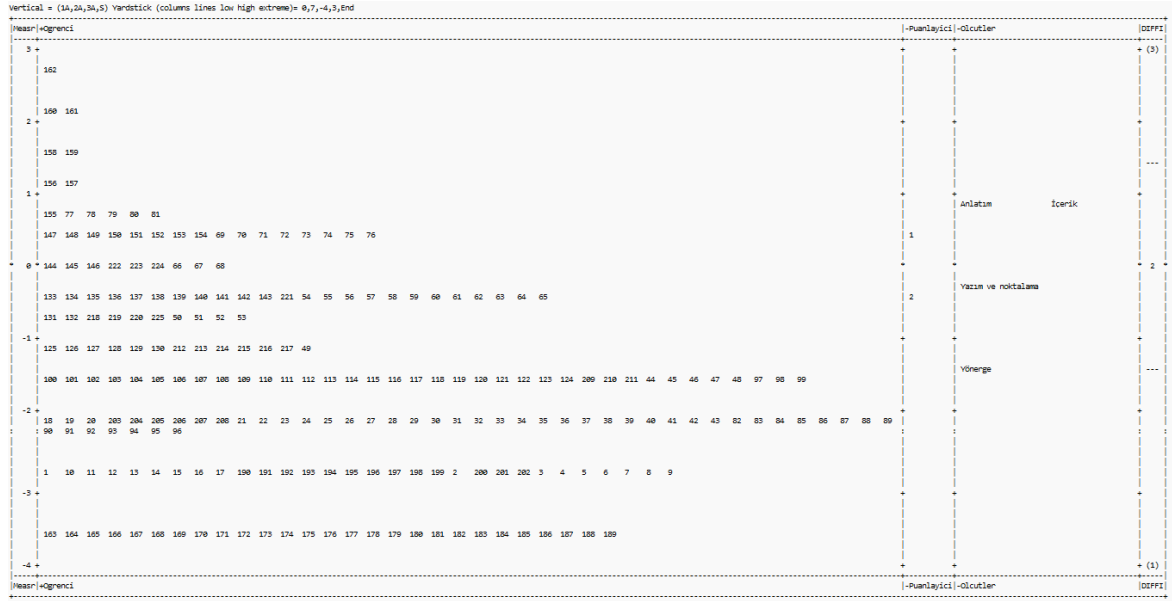
Ölçütler	Logit	Std.error	Infit		Outfit	
			MnSq	Zst	MnSq	Zst
İçerik	0.9	0.11	1.42	4.9	1.2	1.5
Anlatım	0.81	0.1	1.23	2.8	1	0
Yazım ve Noktalama	-0.32	0.09	1.09	1.3	0.99	-0.1
Yönergeye uygunluk	-1.39	0.08	0.54	-9	0.55	-8.6
Mean	0	0.1	1.07	0	0.93	-1.8
SD	0.94	0.01	0.33	5.4	0.24	4

Model, Sample: RMSE = .10 Standard deviation = 0.93
 Discrimination ratio=7.17 Discrimination index = 9.63
 Discrimination index of reliability= 0.99
 Model, Fixed (all same) chi square=3.0 df =2 p= .22

Tablo 3’te analitik dereceli puanlama anahtarında yer alan ölçütlere ilişkin ölçüm raporu görülmektedir. Ölçütlere ilişkin logit değerlerinin -1.39 ile 0.9 arasında değiştiği görülmektedir. Rasch analizinden elde edilen güvenilirlik indeksi 0.93 değeriyle oldukça yüksektir. Hesaplanan ayırma indeksi, ölçütler arasındaki farklılığın ölçütlerde ve ölçütlere kaynaklık eden alt bileşenlerde de farklılaştığını göstermektedir. Daha önce söz edildiği gibi her bir ölçüt (dört ölçüt) * 3 dereceli olmak üzere 12 kategorili bir puanlama yapılmıştır ve dokuzunun anlamlı bir şekilde farklılaştığı belirlenmiştir. Araştırmada en fazla puanlanan (en yüksek puanlanan) ölçütün yönergeye uygunluk, en düşük puanlanan ölçütün ise içerik olduğu belirlenmiştir.

Araştırmada 225 öğrencinin ortaya koyduğu yazılı üretimler x dört ölçütlü bir analitik dereceli puanlama anahtarı ile x iki puanlayıcı tarafından çaprazlanmış bir

şekilde puanlanmıştır. Üç yüzeyli çaprazlanmış desen ile gerçekleştirilen çok yüzeyli Rasch analizi sonucunda tüm facetlerin ölçekleme sonuçları Şekil 1’de gösterilmiştir.



Şekil 1. Tüm yüzeyler için ölçüm raporu

Şekil 1 incelendiğinde öğrencilerin yazılı üretim puanlarının sola çarpık bir dağılım gösterdiği (ortalamanın üzerinde olduğu) görülmektedir. Öğrenci puanları arasında anlamlı bir farklılık olduğu ölçüm raporunda belirlenmiştir ($r>0.70$). Birinci puanlayıcının ikinci puanlayıcıya göre daha katı/cimri puanlama yaptığı; başka bir ifadeyle ikinci puanlayıcının daha cömert olduğu görülmektedir. Puanlayıcılar arasında anlamlı bir farklılık olduğu ölçüm raporunda ifade edilmiştir ($r>0.70$). Araştırmada öğrencilerin yazılı üretimlerinin puanlanmasında kullanılan analitik dereceli puanlama anahtarı ölçütleri arasında anlamlı bir farklılık olduğu belirlenmiştir ($p>0.70$). Şekil 1 incelendiğinde bu farklılığın anlatım ve içerik ölçütlerinde gözlemlenmediği görülmektedir. Yazılı ürünlerin puanlanmasında kullanılan ölçütlerden anlatım ve içerik ölçütünün aynı logit ölçek düzeyinde olduğu ve yazım noktalama ile yönerge ölçütünden daha düşük puanlandığı görülmektedir.

Tartışma ve Sonuç

Bu araştırmada ortaöğretim 9. sınıf öğrencilerinin yazılı üretimlerinin değerlendirilmesinde analitik dereceli puanlama anahtarı sonuçlarının çok yönlü bir şekilde değerlendirilmesi amaçlanmıştır. Bu doğrultuda uluslararası bir veri platformu olan KAGGLE’de ortaöğretim 9. sınıf öğrencileri düzeyinde geliştirilmiş bir yazma görevi ve puanlama anahtarı Türkçeye uyarlanmıştır. Uyarlama sürecinde bire bir çeviriler gerçekleştirildikten sonra üç Türkçe eğitim uzmanıyla yapılan odak grup görüşmesiyle araçlara son şekli verilmiştir. Ardından yazma görevi doğrultusunda 225 öğrenci, yazılı üretimlerini ortaya koymuştur. Yazılı üretimlerin puanlanması üç aşamada gerçekleştirilmiştir. Birinci aşamada rastgele seçilen 10 yazılı ürün puanlama anahtarı ile puanlanmıştır. Birinci aşamada ortaya çıkan farklı puanlamaların nedeni odak grup görüşmesi ile tartışılmış ve puanlama anahtarı düzenlenmiştir. İkinci aşamada farklı 10 yazılı ürün seçilerek yeniden puanlama işlemi gerçekleştirilmiştir. İkinci eğitim aşaması sonrasında puanlama anahtarına son şekli verilmiştir. Üçüncü aşamada son olarak iki uzmandan 225 yazılı ürünü bağımsız bir şekilde puanlaması istenmiştir. Öğrenci ürünleri

(225) x puanlayıcı (2) x puanlama anahtarındaki ölçütler (4) arasındaki ilişkilerin aynı anda incelenbilmesi amacıyla da çok yüzeyle Rasch analizi gerçekleştirilmiştir.

Araştırmanın birinci alt problemde öğrencilerin yazılı üretim becerilerinin anlamlı bir şekilde farklılık gösterip göstermediğinin incelenmiştir. Araştırma sonucunda öğrencilerin yazma ürünlerine verilen puanların anlamlı bir şekilde üç grupta toplandığı belirlenmiştir. Bu durum; yazılı üretimleri yüksek puanlanan öğrenciler olduğu gibi orta ve düşük puan öğrenciler olduğunu da göstermektedir. Araştırmanın ikinci alt problemde puanlayıcıların puanları arasında anlamlı bir farklılık olup olmadığı incelenmiştir. Araştırmada puanlama öncesinde iki aşamalı eğitim puanlaması yapılmasına, puanlama anahtarı ve puanlama şekli odak grup görüşmesi ile belirlenmesine karşın 225 öğrencinin yazılı üretimlerinin puanlanmasında puanlayıcılar arasında anlamlı bir farklılık olduğu belirlenmiştir. İkinci puanlayıcının birinci puanlayıcıya göre birçok ölçüt ve ölçüt alt bileşeninde daha cömert davrandığı belirlenmiştir. Purti ve Melani (2022) de yapmış oldukları çalışmada iki puanlayıcı tarafından verilen puanların ortalamasını benzer ancak puanlayıcılar arasındaki korelasyonu orta düzeyde ($r=0.42$) bulmuştur. Sayın ve Şata (2022) da puanlayıcı özelliklerinin puanlamalara etki ettiğini ve bazı puanlayıcıların daha cimri/katı puanlama gerçekleştirdiklerini ortaya koymuştur. Alan yazında puanlayıcılar arasında yüksek uyumun olduğu çalışmalar bulunmakla birlikte (Hamzadayı ve Dölek, 2022) genel olarak araştırmalar, aynı puanlama anahtarını kullanmalarına ve puanlama eğitimi almalarına karşın puanlayıcıların aynı üretime farklı puanlar verebileceğini göstermektedir (Johnson vd., 2001; Moskal ve Leydens, 2019). Puanlayıcılar arasındaki bu farklılığın nedenlerinden biri de puanlama anahtarının özelliği olduğundan puanlama anahtarında yer alan ölçütlerin ayrışması incelenmiştir çünkü puanlama anahtarındaki ölçütler iyi belirlenmemişse puanlayıcı deneyimi puanlama sürecine karışabilmektedir (Anadol ve Doğan, 2018).

Araştırmanın üçüncü alt problemde ölçütler arasında anlamlı bir farklılık olup olmadığı incelenmiştir. Araştırma sonucunda analitik dereceli puanlama anahtarının ölçütleri arasında anlamlı bir ayrışma olduğu ve bu ayrışmanın ölçütlerin alt bileşenlerinde de yer aldığı tespit edilmiştir. Bununla birlikte içerik ve anlatım ölçütlerinden alınan puanların aynı logit düzeyinde yer aldığı ve anlamlı bir farklılık göstermediği belirlenmiştir. Bu durum, analitik dereceli puanlama anahtarının kullanılma mantığı ile örtüşmemektedir. Çünkü analitik dereceli puanlama anahtarı ölçülen bir yetenek boyutunun öğelere ayrıştırılabildiğinde ve daha ayrıntılı puanlama yapmak istendiğinde kullanılmaktadır (Haladyna 1997). Performansın farklı düzeylerinin ortaya çıkarılması için belirlenmiş ölçütler arasında bir ayrışma bulunmadığı durumlarda bütünsel puanlama anahtarının kullanılması önerilmektedir (Brookhart 1999). Araştırma sonucunda ortaya çıkan bu durum da puanlayıcıların kendi içinde ayrışmasına karşın içerik ve anlatım ölçütlerini benzer bir bakış açısıyla değerlendirmesinden kaynaklı olabilir. Yazılı üretimlerde içerik değerlendirmesi, öğrencinin yazısının temel konu veya argümanını, kullanılan destekleyici bilgileri ve düşünce yapısını değerlendirmeye dayanırken (Jones vd., 2022) anlatımın değerlendirilmesinde öğrencinin yazılı ifadesi, anlaşılabilirliği ve etkileyciliği, cümle yapısı ve kelime seçimi gibi unsurlar değerlendirilir (Lipnevich ve Smith, 2008). Ek 1’de yer alan puanlama anahtarındaki ek açıklamalarda da yer aldığı üzere içerik (fikirler) metnin ne söylediğine odaklanırken anlatım (metin organizasyonu) nasıl söylediğine odaklanmaktadır. Ancak yine de bu iki ölçütün birbirinden ayrılması güç de olabilir. Çünkü içerik her ne kadar metnin mesajını, ana fikrini ve içerdiği bilgileri kapsasa da güzel bir anlatıma sahip olmadığında metnin iyi bir fikre sahip olduğunu söylemek oldukça zordur. Metnin iyi bir fikri ele aldığını söylemek de zordur. Bu nedenle analitik dereceli puanlama anahtarlarında aynı özellik hem içerikte

hem de anlatımda puanlanıyor ve bu nedenle söz konusu ölçütler ayrışmıyor olabilir. Benzer şekilde yazılı üretimlerin analitik ve bütüncül puanlama sonuçları bu nedenle benzerlik de gösteriyor olabilir (Anıl ve Şahhüseyinoğlu, 2008; Carr, 2000; Parlak ve Doğan, 2014). Saxton vd. (2012) puanlayıcıların dereceli puanlama anahtarlarında bazı ölçütlerde tutarlı puanlama yaparken bazı ölçütlerdeki tutarlılıklarının düşük olabileceğine işaret etmektedir. Johnson vd (2000) puanlayıcıların içerik ve biçim ölçüt puanlarındaki uyuşmalarının %69 olduğunu belirtmektedir. Beyreli ve Arı (2009) da gerçekleştirmiş oldukları çalışmada puanlayıcı güvenirliliğinin ölçütler bağlamında farklılık gösterdiğini, en düşük uyumun anlatım ölçütünde olduğunu belirlemiştir.

Öneriler

Analitik dereceli puanlama anahtarının kullanıldığı çalışmalarda ölçütlerin birbirinden çok iyi ayrıştığından emin olunması gerekir. Örneğin anlam ve içerik ölçütünün bir arada puanlandığı bir çalışmada puanlayıcılar arasındaki uyumunun 0.80'in üzerinde hesaplandığı belirlenmiştir (Taşkın vd., 2023). Bu doğrultuda elde edilen sonuçlar bağlamında ölçüt puanları arasındaki korelasyonların incelenmesi, kullanılacak ölçütlerin değerlendirilecek ürün bağlamında ele alınması, ayırma analizleri, Rasch analizi gibi ölçütlerin ölçek düzeylerinin belirlenmesine yönelik işlemlerin gerçekleştirilmesi önerilmektedir. Ayrıca ölçütlerle ilgili puanlayıcılarla birlikte odak grup görüşmeleri gerçekleştirilebilir böylece tüm puanlayıcıların ölçütlere aynı anlam yüklemesi sağlanabilir. Eğer ölçütler birbiri ile aynı anlamı taşıyorsa analitik dereceli puanlama anahtarı yerine bütüncül puanlama anahtarının kullanılması önerilmektedir.

Kaynakça

- Anadol, H. Ö., ve Doğan, C. D. (2018). Dereceli Puanlama Anahtarlarının Güvenirliliğinin Farklı Deneyim Yıllarına Sahip Puanlayıcıların Kullanıldığı Durumlarda İncelenmesi. *Elementary Education Online*, 17(2), 1066-1076.
- Anıl, D., ve Şahhüseyinoğlu, D. (2008). Performansa Dayalı Değerlendirmede Puanlama Tutarlılığı: Holistik ve Analitik Analitik Puanlama Anahtarının Karşılaştırılması. 1. *Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi*, 15-20 Mayıs 2008, Ankara.
- Arter, J., ve McTighe, J. (2001). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Corwin Press.
- Aslanoglu, A. E. (2022). Akran ve Öz Değerlendirme Uygulamalarının Yazma Becerilerine Etkisinin İncelenmesi. *International Journal of Assessment Tools in Education*, 9, 179-196.
- Attali, Y. (2007). Construct Validity of E-rater® in Scoring TOEFL® Essays. *ETS Research Report Series*, 2007(1), 1-22.
- Beaglehole, V. (2014). Assessing Writing Using Rubrics. *Practically Primary*, 19(2), 13-15.
- Beyreli, L., ve Arı, G. (2009). The Use of Analytic Rubric in the Assessment of Writing Performance--Inter-Rater Concordance Study. *Educational Sciences: Theory and Practice*, 9(1), 105-125.
- Bojer, C. S., ve Meldgaard, J. P. (2021). Kaggle Forecasting Competitions: An Overlooked Learning Opportunity. *International Journal of Forecasting*, 37(2), 587-603.
- Carr, N. T. (2000). A Comparison of the Effects of Analytic and Holistic Rating Scale Types in the Context of Composition Tests. *Issues in Applied Linguistics*, 11(2), 207-241.

- Çetin, B., ve Kelecioğlu, H. (2004). Kompozisyon Tipi Sınavlarda Kompozisyonun Biçimsel Özelliklerinden, Kestirilen Puanların Anahtar Puanlarla ve Genel İzlenimle Elde Edilen Puanlarla İlişkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 26(26), 308-320.
- Çetin, Y. (2011). Reliability of Raters for Writing Assessment: Analytic-Holistic, Analytic-Analytic, Holistic-Holistic. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(16), 471-486.
- Creswell, J. W., ve Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications.
- DeMars, C. E. (2021). Violation of Conditional Independence in the Many-Facets Rasch Model. *Applied Measurement in Education*, 34(2), 122-138.
- Demir, B. P. (2020). Akademik Yazı Yazma Becerilerinin Değerlendirilmesi için Analitik Analitik Puanlama Anahtarının Geliştirilmesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 318-327.
- Demir Bilican, S., ve Yıldırım, Ö. (2019). Yazılı Anlatım Becerilerinin Değerlendirilmesi için Analitik Puanlama Anahtarı Geliştirme Çalışması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 47, 457-473.
- Durmuş, A. O. (2022). *Dereceli Puanlama Anahtarına Dayalı Geribildirim Türkenin Yabancı Dil Olarak Öğretiminde Yazma Becerisinin Geliştirilmesine Etkisi*. Doktora Tezi. Ankara: Ankara Üniversitesi.
- Eckes, T. (2008). Rater Types in Writing Performance Assessments: A Classification Approach to Rater Variability. *Language Testing*, 25(2), 155-185.
- Eckes, T. (2009). *Many-facet Rasch measurement*. Routledge: New York, USA.
- Engelhard, G., Jr. (1994). Examining Rater Errors in the Assessment of Written Composition with a Many-Faceted Rasch Model. *Journal of Educational Measurement*, 31, 93-112.
- Espin, C. A., Weissenburger, J. W., ve Benson, B. J. (2017). *Assessing the Writing Performance of Students in Special Education*. Routledge.
- Ghalib, T. K., ve Al-Hattami, A. A. (2015). Holistic Versus Analytic Evaluation of EFL Writing: A Case Study. *English Language Teaching*, 8(7), 225-236.
- Göçer, A. (2005). *İlköğretim İkinci Kademe Türkçe Öğretiminde Ölçme ve Değerlendirme*. Doktora Tezi. Erzurum: Atatürk Üniversitesi.
- Göçer, A. (2010). Türkçe Öğretiminde Yazma Eğitimi. *Uluslararası Sosyal Araştırmalar Dergisi*, 3(12), 178-195.
- Hamzadayı, E., ve Dölek, O. (2022). Bilgilendirici Metin Yazma Başarımını Değerlendirmeye Yönelik Bir Ölçek Önerisi. *Gaziantep Üniversitesi Eğitim Bilimleri Dergisi*, 6(1), 115-130.
- Iglovikov, V., Mushinskiy, S., ve Osin, V. (2017). Satellite Imagery Feature Detection Using Deep Convolutional Neural Network: A Kaggle Competition. *arXiv preprint arXiv:1706.06169*.
- Johnson, R. L., Penny, J., ve Gordon, B. (2000). The Relation Between Score Resolution Methods and Interrater Reliability: An Empirical Study of an Analytic Scoring Rubric. *Applied Measurement in Education*, 13(2), 121-138.

- Johnson, R. L., Penny, J., ve Gordon, B. (2001). Score Resolution and the Interrater Reliability of Holistic Scores in Rating Essays. *Written Communication*, 18(2), 229-249.
- Johnson, A., ve Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review*, 2(2), 130-144.
- Jones, D. M., Cheng, L., ve Tweedie, M. G. (2022). Automated Scoring of Speaking and Writing: Starting to Hit Its Stride. *Canadian Journal of Learning and Technology*, 48(3), 1-22.
- Kellogg, R. T., ve Raulerson, B. A. (2007). Improving the Writing Skills of College Students. *Psychonomic Bulletin and Review*, 14, 237-242.
- Lipnevich, A. A., ve Smith, J. K. (2008). Response to Assessment Feedback: The Effects of Grades, Praise, and Source of Information. *ETS Research Report Series*, 2008(1), i-57.
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective* (Vol. 3). P. Lang.
- Lumley, T. ve McNamara, T.F. (1995). Rater Characteristics and Rater Bias: Implications for Training. *Language Testing*, 12, 54-71.
- Myford, C. M., ve Wolfe, E. W. (2004). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- Özbay, M. (2015). *Yazma Eğitimi*. Ankara: Pegem Akademi Yayıncılık.
- Özdemir, B. (2014). *Analitik Yazma ve Değerlendirme Modelinin Türkçe Öğretmeni Adaylarının Yazma Becerilerine ve Yazma Tutumlarına Etkisi*. Doktora Tezi. Ankara: Gazi Üniversitesi Eğitim Bilimleri Enstitüsü.
- Özkara, Y. (2007). *6+ 1 Analitik Yazma ve Değerlendirme Modelinin 5. Sınıf Öğrencilerinin Hikâye Edici Metin Yazma Becerilerini Geliştirmeye Etkisi*. Doktora Tezi. Ankara: Gazi Üniversitesi.
- Parlak, B., ve Doğan, N. (2014). Analitik Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Uyum Düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 29(29-2), 189-197.
- Popham, W. J. (1997). What's Wrong--and What's Right--with Rubrics. *Educational leadership*, 55(2), 72-75.
- Putri, E. S., ve Melani, M. (2022). Holistic vs. Analytic Evaluation in Writing Test of Eighth Grade Students. *Journal of English Language Studies*, 7(2), 157-175.
- Sayın, A., & Mehmet, Ş. (2022). Using Rasch Analysis to Examine Raters' Expertise Turkish Teacher Candidates' Competency Levels in Writing Different Types of Test Items. *International Journal of Assessment Tools in Education*, 9(4), 998-1012.
- Saxton, E., Belanger, S., ve Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating Intra-Rater and Inter-Rater Reliability of a Scoring Mechanism for Critical Thinking Performance Assessments. *Assessing Writing*, 17(4), 251-270.
- Silvestri, L., ve Oescher, J. (2006). Using Rubrics to Increase the Reliability of Assessment in Health Classes. *International Electronic Journal of Health Education*, 9, 25-30.
- Spandel, V. (2005). *Creating Writers Through 6-Trait Writing Assessment and Instruction* Boston. MA: Pearson.

- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using Multivariate Statistics* (Vol. 6, pp. 497-516). Boston, MA: Pearson.
- Tangpermpoon, T. (2008). Integrated Approaches to Improve Students Writing Skills for English Major Students. *ABAC Journal*, 28(2), 1-9.
- Taşkın, Y., Yorgancı, O. K., ve Harmankaya, M. Ö. (2023). Ortaokul Öğrencileri için Bilgilendirici Metin Analitik Puanlama Anahtarı Geliştirme Çalışması. *Turkish Studies-Language and Literature*, 18(3), 1971-1989.
- Temizkan, M., ve Yıldırım, Ö. (2023). Yazma Eğitimine Yönelik Lisansüstü Çalışmalarda Kullanılan Ölçme Araçları Üzerine Bir İnceleme. *Ana Dili Eğitim Dergisi*, 11(2), 283-306.
- Tolkachev, A., Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., ve Ibragimov, B. (2020). Deep Learning for Diagnosis and Segmentation of Pneumothorax: The Results on the Kaggle Competition and Validation Against Radiologists. *IEEE Journal of Biomedical and Health Informatics*, 25(5), 1660-1672.
- Tomas, C., Whitt, E., Lavelle-Hill, R., ve Severn, K. (2019). Modeling Holistic Marks with Analytic Rubrics. *Frontiers in Education*, 4, 1-19.
- Ungan, S. (2007). Yazma Becerisinin Geliştirilmesi ve Önemi. *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 23(2), 461-472.
- Uyar, Y. (2016). Yazma Becerisini Geliştirmeye Yönelik Araştırmalar: Son Çeyrek Asrın Değerlendirilmesi. *Electronic Turkish Studies*, 11(3), 2273-2294.
- Walker, B., Shippen, M. E., Alberto, P., Houchins, D. E., ve Cihak, D. F. (2005). Using the Expressive Writing Program to Improve the Writing Skills of High School Students with Learning Disabilities. *Learning Disabilities Research ve Practice*, 20(3), 175-183.
- Walker, R., ve Ríu, C. P. (2008). Coherence in the Assessment of Writing Skills. *ELT Journal*, 62(1), 18-28.
- Yoleri, S., ve Tanış, H. M. (2014). İlkokul Birinci Sınıf Öğrencilerinin Okula Uyum Düzeylerini Etkileyen Değişkenlerin İncelenmesi. *Karabük Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 4(2), 130-141.
- Zorbaz, K. Z. (2013). Yazılı Anlatımın Puanlanması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 178-192.

EK 1. Analitik Dereceli Puanlama Anahtarı

Ölçütler	3	2	1	Ek Açıklama
Yönergeye uygunluk	Kelime sayısı 200-400 arasındadır. Metin, bir okul gazetesinde yayınlanacak şekilde düzenlenmiştir.	Kelime sayısı 200-400 arasında değildir ya da metin, okul gazetesi için düzenlenmemiştir.	Kelime sayısı 200-400 arasında değildir ve metin, okul gazetesi için düzenlenmemiştir.	Bağlaçlar da (ve, ya da gibi) kelime sayısının içinde yer alacaktır.
İçerik (Fikirler)	Metinde seçilen görüşün sınırları belirlenmiş,	Metinde seçilen görüş sınırlandırılmamış, ana fikri sınırlı	Metinde seçilen görüş sınırlandırılmamış, ana fikri	Metnin "ne söylendiği"ne odaklanılacaktır.

	uygun örnekler ve detaylarla ana fikir desteklenerek açık bir şekilde ifade edilmiştir. Örnekler mantıksal bütünlük içinde verilmiştir.	sayıda örnek verilerek ifade edilmiştir.	destekleyen örnekler verilememiş ya da açıklamalar yetersiz kalmıştır.	
İfade Biçimi (Üslup)	Metin, geniş bir kelime hazinesiyle ilgi çekici, akıcı cümlelerle okuru ikna etmeye çalışmıştır. Cümleler arasında mantıksal bir akış vardır.	Anlatım akıcılığı yer yer kopmuş, kelime hazinesi bazen yetersiz kalmıştır. Metnin ikna ediciliği bazen zayıf kalmıştır. Cümleler arasında mantıksal akış bazen kopmuştur.	Metin çok az kelime hazinesine sahiptir, metnin akıcılığı ve ikna ediciliği oldukça zayıftır. Cümleler arasındaki mantıksal akış oldukça zayıf kalmıştır.	Akıcılık hem cümle bazında hem de metin bazında değerlendirilecektir. Metnin istenilen mesajı "nasıl söylediği" ne odaklanılacaktır.
İmla ve Noktalama Kuralları	Metinde imla kurallarının, noktalama işaretlerinin kullanımını uygundur. Yanlış yok denecek kadar azdır ve bu yanlışlar anlam tutarlılığını bozmamaktadır.	Metnin imlası ve noktalama işaretinin kullanımında bazı hatalar mevcuttur. Bu yazım ve noktalama hataları okumayı zorlaştırmış ve bazen anlam tutarlılığını bozmuştur.	Metnin imlası ve noktalama işaretinin kullanımında çok fazla hata mevcuttur. Bu hatalar, okumayı güçleştirmiştir. Yanlışlar, anlam tutarlılığını bozmuştur.	Harf hataları, metnin puntosu, yazı karakteri dikkate alınmayacaktır. Yazım ve noktalama yanlışlığında; öğretim programında yer alan geçiş ve bağlantı ifadelerinin yazımı, bazı birleşik kelimelerin yazımı ve bunların doğru anlamda kullanılması dikkate alınacaktır. Metindeki hata sayısının azlığı ya da çokluğu, metnin uzunluğu bağlamında değerlendirilecektir.