



Okuduğunu Anlama Seviyesinin Ölçümünde Boşluk Tamamlama Testleri *

M. Rahman Kalyoncu ¹, Muhammet Memiş ²

Öz

Bireylerin okuduğunu anlama düzeyini, genel anlamda dil becerisini ve yazılı materyallerin okunabilirliğini ölçmek gibi birçok amaç için kullanılan boşluk tamamlama testlerinin Türkçe için geçerli sonuçlar verip vermediğini nesnel bir düzlemde sorgulamak ve testlerin okuduğunu anlama düzeyinin ölçümünde kullanımına yönelik işe koşulan prosedürleri test etmek amacıyla yürütülen bu çalışma, korelasyonel bir araştırmadır. Bu kapsamda işlevselliği ve nesneliği dolayısıyla okuduğunu anlama düzeyinin ölçümünde oldukça sık kullanılan ve ulusal sınavlarda başat ölçme aracı konumundaki çoktan seçmeli testler ile farklı prosedürler izlenerek oluşturulmuş ve puanlanmış boşluk tamamlama testleri arasındaki tutarlılık incelenmiştir. İnceleme sürecinde 4 farklı metne yönelik 4 çoktan seçmeli okuduğunu anlama testi ve her birinde farklı sıradaki bir sözcüğün düzenli şekilde silindiği 4 boşluk tamamlama testi olmak üzere 8 farklı ölçme aracı, 7. sınıf düzeyindeki öğrencilerden oluşan 90 kişilik çalışma grubuna uygulanmıştır. Boşluk tamamlama testlerinin puanlanmasında, birincisi orijinal sözcüklerin doğru kabul edildiği, ikincisi bağlamı yani cümlenin anlamını koruyan sözcüklerin doğru kabul edildiği iki farklı puanlama yöntemi kullanılmıştır. Bu kapsamda çalışma grubunda yer alan öğrencilere yaklaşık 23.000 soru maddesi sunulmuş, bu maddeler üzerinde yaklaşık 43.000 değerlendirme yapılmıştır. 2023-2024 eğitim öğretim döneminde yüz yüze elde edilen veriler üzerinde Pearson korelasyon analizi gerçekleştirilmiştir. Yapılan analizlere göre; genel olarak boşluk tamamlama testleri ile çoktan seçmeli okuduğunu anlama testleri arasında anlamlı ve pozitif bir ilişki olduğu, orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testlerinin çoktan seçmeli okuduğunu anlama testleri ile daha tutarlı sonuçlar verdiği, bağlamı koruyan sözcüklerin doğru kabul edildiği durumlarda bu korelasyonun genellikle azaldığı ve çoktan seçmeli okuduğunu anlama testleri ile boşluk tamamlama testleri arasında en yüksek düzeyde korelasyonun düzenli olarak 6. sıradaki sözcüğün silindiği durumlarda elde edildiği sonuçlarına ulaşılmıştır. Ulaşılan sonuçlardan hareketle boşluk tamamlama testlerinde bağlamı koruyan sözcüklerden ziyade orijinal sözcüklerin doğru kabul edilmesi, düzenli olarak 6. sözcüğün silinmesi ve testlerin sistematik bir şekilde ölçme değerlendirme faaliyetlerine dâhil edilmesi şeklinde önerilerde bulunulmuştur.

Anahtar Kelimeler

Boşluk tamamlama testi
Çoktan seçmeli test
Okuduğunu anlama
Korelasyon
Türkçe

Makale Hakkında

Gönderim Tarihi: 30.09.2024

Kabul Tarihi: 30.12.2024

Elektronik Yayın Tarihi: 03.03.2025

DOI: 10.15390/EB.2025.14079

* Bu çalışmanın bir bölümü 4-6 Ekim 2024 tarihleri arasında düzenlenen Uluslararası Ölçme, Seçme ve Yerleştirme Sempozyumu'nda sözlü bildiri olarak sunulmuştur.

¹ Sinop Üniversitesi, Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Türkiye, mrkalyoncu@sinop.edu.tr

² Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Türkiye, muhammet.memis@omu.edu.tr

Giriş

Eğitim, insanlık tarihinde önemi tartışılmaz bir olgudur. Bu olgunun asli unsurlarından birisi ise ölçme ve değerlendirmedir. Sistemli bir şekilde bir gerçekleştirilen eğitim ve öğretim faaliyetlerinin belirli bir amaca ulaşma gayesi bulunmaktadır. Belirlenen amaca ulaşmayan eğitim ise niteliksiz ve başarısız bir eğitim olarak değerlendirilmektedir. Bu noktada eğitimlerin kalitesini ve niteliğini ölçmek adına devreye ölçme ve değerlendirme kavramları girmektedir. Ölçme, bir varlığın belirli özelliklere sahip olup olmadığını ya da ne derece sahip olduğunun gözlenip gözlem sonuçlarının sembollerle ifade edilmesi (Tekin, 1982); değerlendirme ise ölçme işleminden elde edilen sonuçları belirli bir ölçütle kıyaslayarak karara ulaşma süreci (Yılmaz, 1998) olarak tanımlanmaktadır.

Eğitim sürecinde ölçülmesi gereken özelliklerden birisi de okuduğunu anlamadır. Çünkü okuma ve okuduğunu anlama, eğitim ve öğretim faaliyetlerinin büyük bir bölümüne kaynaklık etmektedir (Balcı, 2016; Karatay, 2018). Dolayısıyla okuduğunu anlama becerisi yeterince gelişmeyen bireyler, akademik hayatın genelinde de başarısız olmaktadır (Akyol, 2020; Çelenk, 2006; Geske ve Ozola, 2008; Uyanık, 2012). Bu bağlamda okuduğunu anlama becerisinin ölçümünün oldukça önemli ve gerekli olduğu ifade edilebilir. Ancak okuduğunu anlamayı ya da genel olarak anlamayı ölçmek uzun yıllardır tartışma konusudur. Çünkü anlama, bilişsel bir süreçtir ve beyindeki işlemleri doğrudan gözlemek mümkün değildir. Dolayısıyla anlamaya dair yapılan tüm ölçümler dolaylıdır ve mutlak bir kesinlik içermemektedir (Akyol, 2020). Günümüz şartlarında ise okuduğunu anlama düzeyinin ölçülmesi; metinden hareketle üretilen ürünün, metne yönelik oluşturulmuş sorulara verilen cevapların ya da doğrudan sürecin gözlemlenmesiyle gerçekleştirilmektedir.

Eğitim kurum ve kuruluşlarında okuduğunu anlama düzeyinin ölçülmesi, çoktan seçmeli testler, açık uçlu sorular, eşleştirme testleri, doğru-yanlış testleri gibi birçok farklı araç ile gerçekleştirilebilmektedir. Ancak bu ölçme araçlarından en çok başvuru alanları açık uçlu ve çoktan seçmeli sorulardır. Nitekim bu testlerin birbirlerine göre belirli avantajları ve dezavantajları bulunmaktadır. Çoktan seçmeli testler, temelde soru kökü ve seçeneklerden oluşmaktadır ve bu testlerde öğrencilerin soru kökünü okuduktan sonra seçenekler arasından doğru olduğunu düşündükleri şıkkı işaretlemeleri gerekmektedir. Bu testler, kolay ve nesnel bir puanlama sürecine sahip olsa da Türkçe eğitimi açısından belirli sorunlara ve sınırlılıklara sahiptir. Çünkü okuduğunu anlama düzeyinin ölçülmesi için oluşturulan çoktan seçmeli sorularda genelde bir paragraf ve paragrafla ilgili bazı sorular yer almaktadır. Ancak bu sorular, sınırları ve seçenekleri belli olduğu için öğrencilerin yaratıcılıklarını göstermelerine, yorumlama yapmalarına ve soru üzerine detaylıca düşünmelerine izin vermemekte (Özbay, 1997; Temizkan ve Sallabaş, 2011; Üstüner ve Şengül, 2004), doğru cevabı kendi içerisinde hazır olarak sunduğu için okuduğunu anlayamayan hatta okumaya yönelik bir becerisi, metne yönelik bir bilgisi olmayan öğrencilerin dahi doğru cevaba ulaşabilmesine izin vermekte (Başaran, 2013; Katz ve Lautenschlager, 1994; Özbay, 1997), beceri ölçümünde işe koşulamamakta (Aşılıoğlu, 1993) daha çok olgusal bilgiyi ölçmeye yönelik durumda olup üst düzey zihinsel becerileri sınamamakta (Temizkan ve Sallabaş, 2011) dolayısıyla okuduğunu anlamayı tam olarak ölçmemekte (Ömeroğlu, 2018) ve bir beceri dersi olan Türkçeye uygun bir araç görünümü sergileyememektedir. Diğer yandan açık uçlu sorular da bünyesinde belirli sorunları barındırmaktadır. Bu sorunların başında ise öznellik gelmektedir. Çünkü açık uçlu sınavlarda tek bir doğru cevabı bulunmamaktadır ve sorudan elde edilen puan, puanlayıcının inisiyatifi ile şekillenmektedir. Dolayısıyla bu sorular, puanlayıcı özelliklerinin ve görüşlerinin puana en fazla nüfuz edebildiği soru türüdür ve cevapların nesnel olarak değerlendirilmesi oldukça zordur. Keza yeterli bilgi, tecrübe ve uzmanlığa sahip olmayan bir Türkçe öğretmenin açık uçlu sorularla geçerli ve güvenilir bir sınav yapması mümkün değildir (Temizkan ve Sallabaş, 2011). Ancak bu durum çoktan seçmeli testler için de geçerlidir. Çünkü çoktan seçmeli soru hazırlamak, oldukça zorlu ve deneyim gerektiren bir beceridir. Çoktan seçmeli sorularla gerçekleştirilen sınavlarda sık sık hatalara rastlanması da bu durumun bir göstergesidir (Üstüner ve Şengül, 2004). Bu dezavantajlardan yola çıkarak alan yazında okuma becerisinin ölçülmesinde tek başına yeterli olan, beceriyi tüm yönleriyle ölçebilen ve en iyi yöntem olarak değerlendirebilecek bir yöntem olmadığı (Arıcı, 2018) vurgulanmaktadır. Çünkü açıklandığı

üzere kullanılan araçların nesnellik, şans faktörü, oluşturulma ve puanlama zorluğu gibi bariz zayıf yönleri bulunmaktadır. Ancak bu zayıf yönleri bünyesinde barındırmayan ve okuduğunu anlama düzeyinin ölçümünde işe koşulabilecek araçlara rastlamak da mümkündür. Bu çalışmanın devamında ise bahsedilen araçlara uygun bir örnek olduğu ve hâlihazırdaki yöntemlerin eksik yönlerini telafi edebilecek özelliklere sahip olduğu düşünülen boşluk tamamlama testleri tanıtılacaktır.

Okuduğunu anlama düzeyinin ölçümde kullanılan araçlardan biri olan boşluk tamamlama tekniği, 1953 yılında Wilson Taylor (1953) tarafından "*cloze procedure*": *a new tool for measuring readability* isimli çalışma ile tanıtılmıştır. Bu teknik, temelde bir Gestalt kavramı olan tamamlama, yani zihnin tamamlanmamış olguları tamamlama yeteneği üzerine kurulmuştur (Ulusoy, 2009). Boşluk tamamlama testleri; yazılı bir materyalin okunabilirliğini, bireyin; belirli bir metin üzerindeki okuma seviyesini, özel bir alan ya da konu üzerindeki sözcük bilgisini, dil yeteneğini, genel anlama seviyesini ve bir sınıfın ya da küçük bir grubun genel okuma profilini tespit etmek için kullanılan (Loewe, 1983; Mariotti ve Homan, 2001) bir araçtır. En basit hâliyle bir metindeki her "n." sıradaki sözcüğün silindiği boşluk tamamlama testleri (Harmer, 2002), literatüre tanıtıldığı günden beri onlarca çalışmaya konu olmuş, birçok farklı amaç için bir araç olarak kullanılmıştır (Oller, Bowen, Dien ve Mason, 1972). Öğrencilerin okuma güçlüklerinin giderilmesi (Dağ, 2010), bireysel farklılıkların ve konuya ilişkin algıların tespit edilmesi (Manis ve Dawes, 1961), dil yeterliliklerinin geliştirilmesi (Booth, 1998; James, 2004), öğrencilerin okuduğunu anlama becerisinin geliştirilmesi (Sukarni, 2021; Wahdaniah, Marbun ve Husin, 2013), farklı dilsel geçmişe sahip okuyucuların belirlenmesi (Craker, 1971) ve okunabilirlik formülü geliştirme sürecinde metnin zorluğunu ölçmek için bir araç olarak kullanılması (Coleman, 1965; Çetinkaya, 2010) bu amaçlar arasında yer almaktadır. Keza uluslararası alan yazında, bazı araştırmacılar (Carvalho ve Souza, 2023) boşluk tamamlama testlerinin günümüzde okuduğunu anlama düzeyinin ölçülmesinde en sık kullanılan araçlar arasında olduğunu belirtmektedir.

Boşluk tamamlama testlerinde okuyucular, metinde yer alan boşlukları; metni okuyup, anlamlandırıp, yorumlayarak doldurmaktadır. Doğru tamamlanan boşlukların yüzdesi, okuyucunun puanını oluşturmakta ve okuyucunun puanı düştükçe metnin okuyucu açısından zorlaştığı kabul edilmektedir (Dubay, 2007). Ancak boşluk tamamlama testleri, öğrenciler açısından bazı zorluklara sahiptir. Bu zorluklara; öğrencilerin kimi zaman testin güçlüğünden dolayı testi cevaplamaktan kaçınması, dolayısıyla her zaman geçerli ve güvenilir sonuçlar elde edilememesi (Karatay, Bolat ve Güngör, 2013; Mariotti ve Homan, 2009), kolay bir metinde usta okurların dahi genellikle %65'ten fazla başarı gösterememesi (Bormuth, 1966; Froese, 1971) örnek olarak gösterilebilir. Bu bağlamda boşluk tamamlama testleriyle ilk kez karşılaşılacak öğrencilerden geçerli puanlar elde edilmesi beklenmemeli ve asıl uygulamalardan önce öğrencilerin boşluk tamamlama testlerine aşina olmaları sağlanmalıdır (Ruddell, 2005; Vacca ve Vacca, 2005). Keza tek bir boşluk tamamlama testiyle öğrencilerin dil yeterliliklerin ölçülmesi beklenmemelidir (Brown ve Grüter, 2020). Diğer yandan boşluk tamamlama testlerinin oldukça nesnel olması, kolayca hazırlanabilmesi, kullanımının ve çözümlenmesinin oldukça basit olması ve belki de en önemlisi metnin doğrudan test aracı olarak kullanılabilmesi (Çetinkaya, 2010); bu testleri işlevsel ve sıkça işe koşulacak araçlar hâline getirerek yaygın bir şekilde kullanılmasını sağlamaktadır. Ancak bu testlerin hazırlanışında, uygulanışında ve puanlanışında bir standarttan bahsetmek oldukça güçtür.

Boşluk tamamlama testlerinin hazırlanma ve uygulanma süreçlerine geçilmeden önce, bu konularda birbirinden farklı birçok prosedür, puanlama yöntemi, hazırlanış süreci, görüş ve tespit olduğu ifade edilmelidir. İlk olarak boşluk tamamlama testine dönüştürülecek metnin uzunluğu ele alınmalıdır. Bu uzunluk, esasında test içerisinde kaç boşluk olacağına ve düzenli olarak kaçınıcı sıradaki sözcüğün silineceğine bağlı olarak şekillenmektedir. Bu konuda uluslararası alan yazındaki en yaygın kabul, 300-400 kelimelik bir metnin ilk ve son cümlesini orijinal hâlinde bırakılarak 50 boşluk elde edilinceye kadar düzenli olarak 5. sıradaki sözcüğün silinmesidir. Ancak burada belirtilen tüm ifadeler ve standartlar ayrı ayrı tartışma konusudur. Örnek olarak Nation (2009), boşluk tamamlama testlerinde ideal boşluk sayısının 40-50 arasında olması gerektiğini belirtirken Shahnazari-Dorcheh, Roshan ve Hesabi (2012), boşluk sayısının okuyucunun düzeyine göre değişmesi gerektiğini savunmakta ve

başlangıç seviyeleri için 25-30 boşluğun ideal olduğunu ifade etmektedir. Bu durumun yanı sıra sözcüklerin silinme prosedürü de ayrıca tartışılmaktadır. Çünkü boşluk tamamlama testleri, sabit oranlı silme veya düzensiz olarak silme formatında olmak üzere iki farklı şekilde hazırlanabilmektedir (Koda, 2005). Sabit oranlı silme işlemlerinde belirli bir sıradaki sözcük düzenli olarak silinmektedir. Düzensiz silme işlemi ise belirli bir sözcük türünün ya da bağlam içerisinden seçilen sözcüklerin silinmesi söz konusudur. Sabit oranlı silme işlemine yönelik Bormuth (1964), ana dil için hazırlanan bir boşluk tamamlama testinde "n." sıradaki sözcüğün sağındaki ya da solundaki sözcüğün silinmesinin anlamlı bir fark yaratmayacağını, Oller ve diğerleri (1972), silme işleminin düzensiz olsa da geçerli sonuçlar verebileceğini ancak kullanıma uygun olmadığını, Bachman (1982) ise silme işleminin sabit oranlı bir şekilde gerçekleştirilmesi gerektiğini savunmaktadır. Diğer yandan sabit oranlı silme işlemiyle hazırlanan testlerin genel anlamdan ziyade yüzeysel ve cümle kapsamındaki yerel anlamlara odaklandığını (Carlisle ve Rice, 2004), dilsel yeterliliği tam anlamıyla ölçemeyeceğini ve anlama becerisinden çok bellek performansını yansıttığını, dolayısıyla düzensiz silme işleminin daha geçerli olduğunu ifade eden araştırmacılar (Alderson, 2000; Cain ve Oakhill, 2006) da bulunmaktadır. Ancak her dilin eşsiz özellikleri olduğu ve bir dil için ortaya koyulan standartların, kendisi için oluşturulduğu dil dışında güvenilir sonuçlar vermeyebileceği (Kalyoncu ve Memiş, 2024; Kurudayıoğlu ve Karadağ, 2005) dikkate alınmalıdır. Bu bağlamda açıklanan görüşlerin ve standartların Türkçe için ayrıca test edilmesi, bugüne kadar gerçekleştirilmediği için bilimsel açıdan bir gerekliliktir.

Alan yazında sürmekte olan tartışmalardan bağımsız olarak boşluk tamamlama testleri ile ilgili genel bir süreç açıklaması yapılması gerekirse boşluk tamamlama testlerinin; öğrencilerin aşına olmadığı, 300-400 kelimelik bir metin seçilmesi, seçilen metnin ilk ve son cümlesine işlem yapılmadan 50 boşluk elde edilinceye kadar belirlenen sıradaki sözcüklerin silinmesi ve bu sözcüklerin yerine boşluk bırakılmasıyla oluşturulduğunu söylemek mümkündür. Bu boşluk uzunlukları ise tüm metin genelinde standart (Stubbs ve Tucker, 1974) ve Türkçe için 12 yazı birim uzunluğunda (Çetinkaya, 2010) olması gerektiği ifade edilmiştir. Boşluğa dönüştürme işlemi sırasında eğer boşluğa dönüştürülecek kelime bir özel isme denk gelmişse kelime atlanarak bir sonraki kelime boşluğa dönüştürülmektedir. Açıklanan prosedürler dikkate alındığında ise ideal bir boşluk tamamlama testi elde edilmektedir. Bu doğrultuda ölçme değerlendirme uygulamaları ve eğitim konularında yeterli tecrübeye sahip olmayan bir kişinin bile kolayca bir ideal bir boşluk tamamlama testi hazırlayabileceği ifade edilebilir.

Literatür incelendiğinde boşluk tamamlama testlerinin standartlaştırılmış okuduğunu anlama testleri ile korelasyonunu inceleyen ve bu testler ile boşluk tamamlama testleri arasında yüksek korelasyonlar tespit eden birçok çalışmaya (Abraham ve Chapelle, 1992; Darnell, 1970; Hinofotis, 1980; Irvine, Atai ve Oller, 1974; Williams, Ari ve Santamaria, 2011) rastlamak mümkündür. Bu yüksek korelasyonlara ek olarak boşluk tamamlama testlerinin bir diğer güçlü yönü de Fransızca (Tremblay ve Garrison, 2010), İspanyolca (Vari-Cartier, 1980), Portekizce (Suehiro ve Santos, 2015), Endonezce (Sukarni, 2021), Korece (Taylor ve Lee, 1954), Tayca (Oller vd., 1972), Arapça (Abanami, 1982), Japonca (Shiba, 1957) gibi oldukça farklı özelliklere sahip birçok farklı dilde geçerli sonuçlar vermesi ve dolayısıyla evrensel özellikler taşımasıdır. Diğer yandan boşluk tamamlama testlerinin oluşturma, uygulama ve yönetme süreçlerindeki kolaylık, materyal çeşitliliği ve nesnellik ele alındığında eğitim faaliyetleri için oldukça işlevsel olduğu sonucuna ulaşılmaktadır. Ancak daha önce bahsedildiği üzere ne kadar geçerli olursa olsun farklı bir dil için oluşturulmuş standardın veya prosedürün Türkçede de doğru sonuçlar vereceği varsayılmamalı ve ilgili standartların ya da prosedürlerin geçerliliği Türkçe için de test edilmelidir. Bu gerekliliğin başlıca sebepleri Türkçenin ortalama cümle uzunluğunun ve söz diziminin birçok dilden farklılaşmasıdır. Örnek olarak R. Flesch'e (1948) göre İngilizcenin ortalama cümle uzunluğu 14-15 sözcükken Ateşman'a (1997) göre Türkçenin ortalama cümle uzunluğu 9-10 sözcüktür. Bu farklılık bir cümle içerisinde sıralanan sözcüklerin niteliğinin değişmesi sebep olabileceken diğer yandan söz diziminin farklılaşması da doğal olarak cümle içerisindeki öğelerin konumunu değiştirecektir. Bu doğrultuda Türkçe metinlerde silinmesi gereken sözcüklerin sırasının da değişmesi olasıdır.

Boşluk tamamlama testleri ile ilgili ulusal alan yazında gerçekleştirilen çalışmalar incelendiğinde Türkçe için standart bir boşluk tamamlama prosedürü üzerinde uzlaşmadığı ve testlerin standartlaştırılmış nesnel araçlarla tutarlılığının yeterince sorgulanmadığı görülmektedir. Gerçekleştirilen çalışmalar, daha çok yeterince nesnel olmayan ölçme araçlarıyla ya da görüşlerle boşluk tamamlama testlerinin tutarlılığını, hangi sıradaki sözcüğün silinmesinin daha tutarlı sonuçlar vereceğini ve eş anlamlı kelimelerin doğru kabul edilip edilmemesini araştırmaya yoğunlaşmış durumdadır. Örneğin Ulusoy (2009), çalışmasında öğrencilerin boşluk tamamlama testlerinden aldıkları puanlarla öğretmenlerin öğrencilere yönelik görüşlerini karşılaştırmış ve her 5. sözcüğün silindiği testlerin güvenilir sonuç vermediği, her 6. sözcüğün silindiği testlerin daha güvenilir olduğu sonucuna ulaşmıştır. Diğer yandan Kaplan ve Çiftçi (2021), boşluk tamamlama testlerinde eş anlamlı sözcüklerin doğru kabul edilmemesinin dezavantaja neden olduğunu belirtmiş, her 4. kelimenin silindiği ve bağlamı koruyan cevapların doğru kabul edildiği "ikame kelime" uygulamasının daha tutarlı sonuçlar verdiğini tespit etmiştir. Tunçer ve Erden (2015) ise her 10. sözcüğün silindiği 2 boşluk tamamlama testi oluşturmuş, bir testte ilk harfleri ipucu olarak bırakmışlardır. Boşluk tamamlama puanlarının açık uçlu sorularla korelasyonunu inceleyen Tunçer ve Erden (2015), ilk harfin ipucu olarak verildiği testte daha tutarlı sonuçlar alındığını belirtmişlerdir. Benzer şekilde boşluk tamamlama testleri ile 6 açık uçlu okuduğunu anlama sorusuna sahip testler arasındaki korelasyonu inceleyen Uyanık (2011), 7 ve 8. sıradaki sözcükler arasındaki korelasyonun 6. sıradaki sözcüğün silindiği boşluk tamamlama testleriyle elde edilen korelasyondan yüksek olduğu sonucuna ulaşmıştır.

Sonuç olarak ilgili ulusal alan yazında boşluk tamamlama testlerine yönelik Türkçe özelinde net bir prosedür ile sunulmadığı, işe koşulan yöntemlerin geçerliliğinin nesnel araçlarla test edilmediği görülmektedir. Açıklanan gerekçelerden yola çıkarak bu çalışmada farklı prosedürlerle oluşturulmuş boşluk tamamlama testleri ile geçerliliği ve güvenilirliği test edilmiş çoktan seçmeli testler arasındaki tutarlılığı inceleyerek boşluk tamamlama testlerinin Türkçe özelinde okuduğunu anlama düzeyinin ölçülmesinde geçerli sonuçlar verip vermediğini belirlemek ve testlerin hazırlanma, uygulanma ve puanlanma süreçleri ile ilgili net bir prosedür ortaya koymak amaçlanmıştır. Bu amaç doğrultusunda araştırma boyunca cevap aranan sorular şu şekildedir:

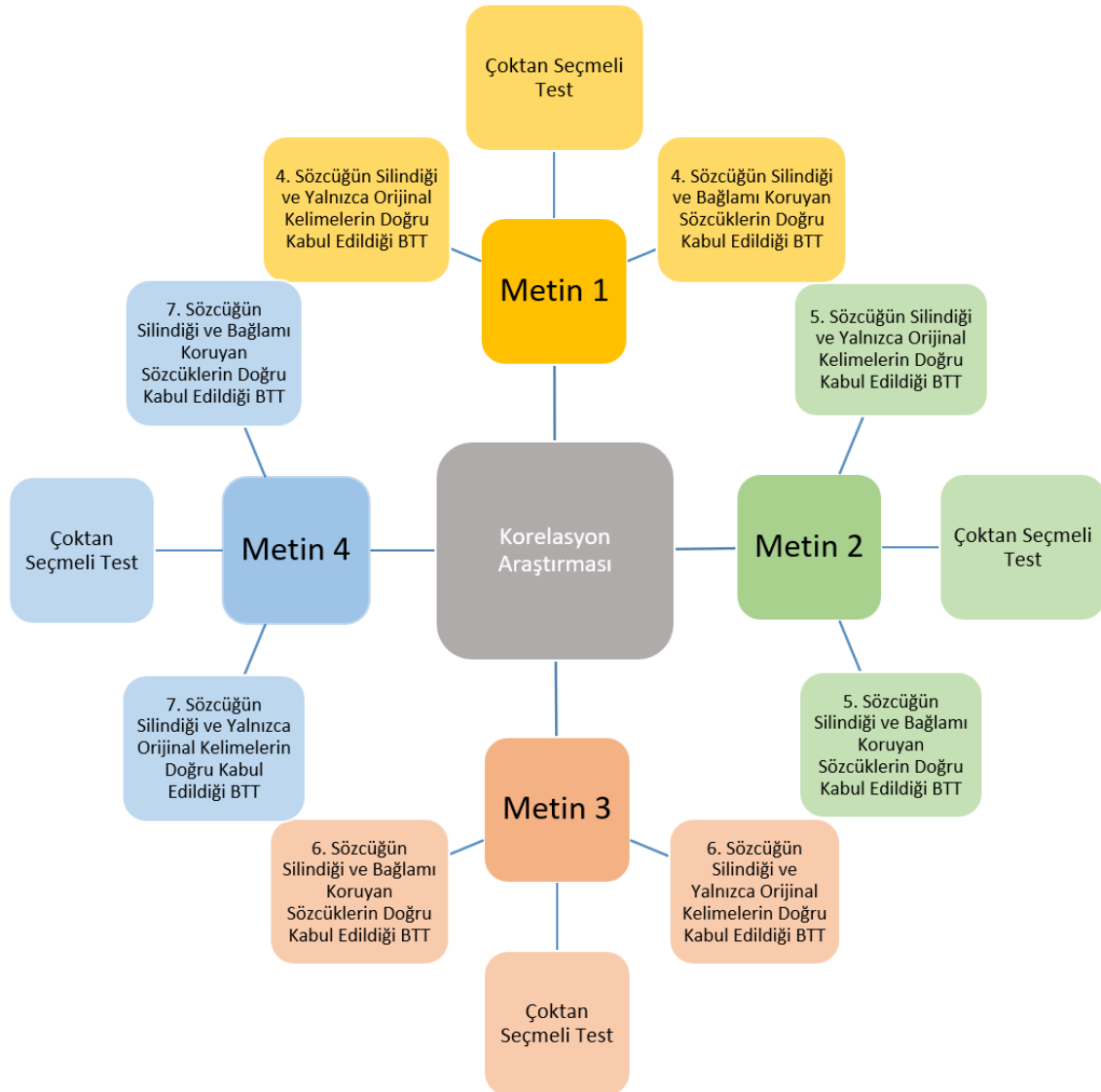
1. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 4. kelimenin silindiği ve orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
2. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 4. kelimenin silindiği ve bağlamı koruyan sözcüklerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
3. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 5. kelimenin silindiği ve orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
4. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 5. kelimenin silindiği ve bağlamı koruyan sözcüklerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
5. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 6. kelimenin silindiği ve orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
6. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 6. kelimenin silindiği ve bağlamı koruyan sözcüklerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
7. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 7. kelimenin silindiği ve orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?
8. Çalışma grubunda yer alan öğrencilerin, aynı metne yönelik her 7. kelimenin silindiği ve bağlamı koruyan sözcüklerin doğru kabul edildiği boşluk tamamlama testi puanları ile çoktan seçmeli okuduğunu anlama testi puanları arasında anlamlı bir ilişki var mıdır?

Yöntem

Araştırma modeli

Boşluk tamamlama testlerinin Türkçe metinler özelinde okuduğunu anlama düzeyini doğru ve isabetli ölçüp ölçmediğini sorgulamak ve boşluk tamamlama testlerinin hazırlanış, uygulanış ve puanlanış sürecine yönelik öne sürülen prosedürleri test etmek amacıyla yürütülen bu çalışmada nicel araştırma yöntemlerinden biri olan korelasyonel araştırmadan yararlanılmıştır. İki ya da daha fazla değişken arasındaki ilişkileri belirlemek, betimlemek ve neden-sonuç ile ilgili ipuçları elde etmek amacıyla yürütülen korelasyonel araştırmalarda (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2023; Karakaya, 2014; Özdemir ve Doğruöz, 2020), ilişkilerin daha iyi anlaşılabilmesi için kısmen aynı zamanda aynı bireylerden doğal ortamlar içerisinde elde edilen farklı değişkenler arasındaki ilişkiler incelenmektedir (Mertens, 2015; Tuncer, 2020).

Bu bağlamda çalışmanın amaçları ve süreci doğrultusunda en uygun yöntemin korelasyonel araştırma olduğu ifade edilebilir. Çünkü araştırma sürecinde birbirinden farklı 4 metnin her birine yönelik oluşturulmuş 2 farklı ölçme aracı kullanılmış, bu araçlardan biri iki farklı şekilde puanlanmış ve bu araçlar arasındaki korelasyonlar incelenmiştir. Araştırma modeline yönelik oluşturulan Şekil 1 aşağıda yer almaktadır:



BTT: Boşluk Tamamlama Testi

Şekil 1. Araştırmanın Modeli

Şekil 1 incelendiğinde araştırmanın esasında 4 farklı metin üzerinden yürütüldüğü görülmektedir. Bu metinlerin her birine yönelik hazırlanmış olan biri çoktan seçmeli okuduğunu anlama testi, biri boşluk tamamlama testi olmak üzere iki farklı ölçme aracı bulunmaktadır. Bu ölçme araçlarından boşluk tamamlama testleri ise birincisi yalnızca orijinal sözcüklerin doğru kabul edildiği, diğeri bağlamı koruyan sözcüklerin doğru kabul edildiği iki farklı puanlama yöntemiyle değerlendirilmiştir. Bu doğrultuda çalışma içerisinde 4 farklı metne yönelik 8 farklı ölçme aracı ve 12 farklı değerlendirme süreci kullanılmıştır.

Çalışma Grubu

Bu araştırmanın çalışma grubu, Samsun ve Balıkesir illerinde öğrenim gören 7. sınıf düzeyindeki 90 öğrenciden oluşmaktadır. Çalışma grubunda yer alan öğrenciler, kolay ulaşılabilir durumdaki okullardan gönüllük esası ile belirlenmiştir. Ancak araştırmanın yapısından kaynaklı, çalışma grubunda yer alan bir öğrencinin aynı metne yönelik hazırlanmış testlerin tamamına katılım sağlaması gerekliliği, aynı metne yönelik hazırlanmış testlerin tamamına katılım sağlayamayan öğrencilerin değerlendirme sürecinden çıkartılmasını gerektirmiştir. Bu nedenle yalnızca, bir metne yönelik tüm testlere katılım sağlayan öğrencilerin puanları arasındaki korelasyon değerlendirilmiştir. Keza korelasyon karşılaştırmalarının sağlıklı bir şekilde gerçekleştirilebilmesi için de çalışma grubu, katılımın en düşük olduğu metne yönelik testlerdeki katılımcı sayısına (60) düşürülmüştür. Tam katılım sağladığı hâlde grupları eşitlemek adına çalışma grubundan çıkartılan öğrencilerin çıkartılma işlemi, puanlarına göre en yüksek ve en düşük öğrenciler arasından eşit olacak şekilde gerçekleştirilmiştir. Bu işlemler sonucunda araştırmanın çalışma grubu, 7. sınıf düzeyinde öğrenim gören 60 öğrenci olarak belirlenmiş ve bu çalışma grubu üzerinden elde edilen veriler değerlendirilmiştir. Aşağıda yer alan Tablo 1 içerisinde çalışma grubunun özelliklerine dair detaylar yer almaktadır:

Tablo 1. Çalışma Grubunun Genel Özellikleri

Özellik	f	Yüzde
Sınıf Düzeyi (7)	60	%100
Cinsiyet (Erkek)	22	%36.6
Cinsiyet (Kadın)	38	%63.4
Şehir (Samsun)	36	%60
Şehir (Balıkesir)	24	%40

Tablo 1 incelendiğinde çalışma grubunda yer alan 22 öğrencinin (%36,6) erkek, 38 öğrencinin (%63,4) kadın olduğu anlaşılmaktadır. Diğer yandan 36 öğrenci (%60) Samsun ilinde öğrenim görmekteyken 24 öğrenci (%40) Balıkesir ilinde eğitim almaktadır.

Veri Toplama Araçları

Bu çalışmada 4 farklı metinle alakalı 8 farklı veri toplama aracı kullanılmıştır. Kullanılan araçların temel kaynağını Temizkan (2007) tarafından hazırlanan “*Bilgilendirici Metinler Okuduğunu Anlama Ölçeği*” oluşturmaktadır. Ölçek, Temizkan (2007) tarafından doktora tezi kapsamında okuduğunu anlama düzeyini test etmek amacıyla geliştirilmiştir. Bu amaç doğrultusunda oluşturulan ölçek içerisinde “Eğer Gençlik Bilseydi”, “Okumak”, “Roman Okumak”, “Okul ve Meslek Seçimi” olmak üzere 4 metin ve bu metinlerden her birine yönelik 14 adet, toplam 56 çoktan seçmeli okuduğunu anlama sorusu bulunmaktadır. Bahsedilen ölçeğin geçerliliği ve güvenilirliği, ilgili çalışmada test edilmiş ve kanıtlanmıştır. Aynı çalışmada hesaplanan testteki soruların madde ayırt edicilik düzeyleri (rx) Tablo 2 içerisinde gösterilmiştir.

Tablo 2. Bilgilendirici Metinler Okuduğunu Anlama Ölçeği Madde Ayırt Edicilik İndeksleri

Sorular	Eğer Gençlik Bilseydi	Okumak	Roman Okumak	Okul ve Meslek Seçimi
	rjx	rjx	rjx	rjx
1	.33	.83	.33	.33
2	.25	.33	.50	.50
3	.25	.25	.83	.83
4	.50	.25	.83	.83
5	.50	.50	.41	.41
6	.83	.33	.33	.33
7	.83	.41	.83	.83
8	.33	.41	.25	.25
9	.25	.33	.41	.41
10	.33	.25	.33	.33
11	.33	.33	.83	.83
12	.83	.41	.83	.83
13	.41	.50	.33	.33
14	.50	.25	.83	.83

Bir maddenin rjx yani madde ayırt edicilik değerinin “0.20” ile “0.30” değerleri arasında olması, maddenin orta düzeyde ayırcılığının olduğunu; “0.30” değerinden yüksek olması ise maddenin iyi düzeyde ayırcı olduğu anlamına gelmektedir (Güler, 2019). Bu bağlamda testler içerisindeki 56 madde içerisinde 9 tanesinin orta düzeyde ayırt edici, geri kalan 47 maddenin ise iyi düzeyde ayırt edici olduğu görülmektedir.

Temizkan (2007) tarafından hesaplanan madde ayırt edicilik indekslerine ek olarak aynı ölçeğin geçerlik ve güvenilirliği, Özyılmaz (2010) tarafından tekrar ele alınmış ve “.85” KR-20 değerine sahip olduğu tespit edilmiştir. Güvenirlilik katsayısının 0.70 ya da daha yüksek olmasının testin güvenilirliği için yeterli olduğu (Büyüköztürk, 2006) dikkate alındığında ise ölçeğin güvenilir olduğu ifade edilebilir. Özyılmaz (2010) aynı zamanda testin seviyesine yönelik uzman görüşlerine başvurmuş ve testin 7. sınıf öğrencilerine uygun olduğu sonucuna ulaşmıştır. Bu bağlamda ölçeğin, okuduğunu anlama düzeyinin ölçülmesinde bir araç olarak 7. sınıf öğrencileriyle uygulanabileceğini söylemek mümkündür.

Araştırma sürecinde açıklanan çoktan seçmeli testlere ek olarak boşluk tamamlama testlerine de ihtiyaç duyulmuştur. İhtiyaç duyulan bu boşluk tamamlama testleri ise testlerin aynı metne yönelik olması gerekliliği ile “Bilgilendirici metinler okuduğunu anlama ölçeği” içerisinde yer alan metinlerden oluşturulmuştur. Söz konusu ölçek içerisinde; 300 kelimelik “Eğer Gençlik Bilseydi”, 313 kelimelik “Okumak”, 368 kelimelik “Okul ve Meslek Seçimi” ve 370 kelimelik “Roman Okumak” metni olmak üzere 4 metin yer almaktadır. Bu metinlerden “Eğer Gençlik Bilseydi” metni düzenli olarak her 4., “Okumak” metni 5., “Roman Okumak” metni 6., “Okul ve Meslek Seçimi” metni 7. sözcüğün silinmesiyle boşluk tamamlama testine dönüştürülmüştür. Dönüştürme işleminde ilk ve son cümlelerin bütünlüğüne zarar verilmemiş, özel isimlere denk gelen boşluklar atlanmıştır. Ayrıca boşluklar metin genelinde standart olarak 12 yazı birim uzunluğunda oluşturulmuştur. Açıklanan prosedür ile elde edilen boşluk tamamlama testlerinden bir kesit Tablo 3 içerisinde yer almaktadır.

Tablo 3. Araştırma İçin Hazırlanan Boşluk Tamamlama Testlerinden Örnek Kesitler

Eğer Gençlik Bilseydi	Okumak	Roman Okumak	Okul ve Meslek Seçimi
Fransızların bir sözü vardır: "Gençlik bilseydi, ihtiyarlık yapabilseydi." derler. Ne yazık ki _____ bilmez. Bilmediği için _____ yapabileceği birçok şeyi _____ artık ...	Okumak insan için bir zevk, bir eğlence olduğu kadar, hiç kuşkusuz eğitici bir eylemdir de. Bilgilerimizi artırarak, aklımızı işleterek; _____, görüşümüzü genişleten bir eylemdir. _____ insan olayları değerlendirmede, çevresini _____, yaşamın	Kişileri roman okumayı sevenlerle roman okumayı sevmeyenler diye ikiye ayırabiliriz. Roman okumayı sevmeyenlerden bir hayır _____ demiyorum, büyük işlere asıl onların _____ söyleseler ona da inanırım. Ama _____ hoşlanmam onlardan. ...	Her ders yılı sonunda birçok anneler, babalar, çocuklar, gençler tasalanırlar. Anneler, babalar, öğrenimlerini bir üst derecedeki _____ devam ettirmeye isteyen çocuklarını nereye vereceklerini _____ veya seçtikleri okullara göndermek imkânı bulamazlar. _____ fazla ...

Boşluk tamamlama testlerinde, metinlerden silinen sözcüklerin nitelikleri de hayli önemlidir. Çünkü bu sözcüklerin niteliklerindeki değişimler, katılımcıların başarılarını da etkileyebilmektedir. Araştırma sürecinde oluşturulan boşluk tamamlama testlerinde silinen sözcüklerin nitelikleri aşağıda yer alan Tablo 4'te sunulmuştur:

Tablo 4. Boşluk Tamamlama Testlerinde Silinen Sözcüklerin Türleri

Sözcük Türü	Eğer Gençlik Bilseydi	Okumak	Roman Okumak	Okul ve Meslek Seçimi
İsim (f)	17 (%34)	21 (%42)	17 (%34)	14 (%28)
Fiil (f)	11 (%22)	4 (%8)	10 (%20)	4 (%8)
Fiilimsi (f)	6 (%12)	10 (%20)	4 (%8)	9 (%18)
Sıfat (f)	8 (%16)	3 (%6)	6 (%12)	12 (%24)
Zarf (f)	3 (%6)	1 (%2)	2 (%4)	4 (%8)
Zamir (f)	-	5 (%10)	7 (%14)	6 (%12)
Bağlaç (f)	3 (%6)	4 (%8)	3 (%6)	-
Edat (f)	2 (%4)	2 (%4)	1 (%2)	1 (%2)
Toplam (f)	50	50	50	50

Tablo 4 incelendiğinde tüm testlerde silinen sözcüklerin yoğunluğunu "isim" türündeki sözcüklerin oluşturduğu görülmektedir. Benzer şekilde "edat", "bağlaç" ve "zarf" türündeki sözcükler de testlerde en az silinen sözcük türlerini karşılamaktadır. Diğer yandan "fiil", "fiilimsi" "zamir" ve "sıfat" türünden silinen sözcüklerin yoğunluğu ise metinler arasında farklılaşmaktadır. Silinen sözcük özelliklerinin testlerden alınacak puanları etkileme potansiyelleri göz önüne alındığında ise sözcük niteliklerindeki değişimlerin araştırmanın bir sınırlılığını oluşturduğu ifade edilebilir.

Çalışma kapsamında oluşturulan boşluk tamamlama testleri ile ilgili diğer bir husus ise soru maddelerinin hangi okuduğunu anlama düzeylerinde ölçüm yaptıklarıyla ilgilidir. Söz konusu boşluk tamamlama maddeleri esasında tek bir düzeyde ölçüm yapılmamaktadır. Çünkü bu testlerde öncelikle okuyucunun testi baştan sona okuması, konuya ilişkin fikir sahibi olması, daha sonra yazarın amacını, düşünce yapısını, üslubunu hatta kişiliğini analiz etmesi, metnin diğer parçaları ile ilişkiler kurması, metin içerisindeki yapıları anlamlandırması ve sözcük bilgisini işe koşarak boşluğu doldurması gerekmektedir. Tüm bunların ardından ise okuyucunun yazdığı sözcüğün metne uygunluğunu sorgulaması beklenmektedir. Yani boşluk tamamlama testlerinin Barrett Taksonomisinde yer alan yüzeysel anlama, fark etme, hatırlama, çıkarımsal/derin anlama, değerlendirme düzeylerini; Bloom Taksonomisinde yer alan hatırlama, anlama, uygulama, çözümlenme basamaklarını bünyesinde barındırdığı ifade edilebilir. Diğer yandan boşluk tamamlama testlerinde yer alan maddelerin ölçüm

düzeylerinin birbirlerinden net bir şekilde ayrılması da çok olası değildir. Çünkü maddelerin hepsi aynı formatta, aynı beklenti ile oluşturulmakta; okuyucular da maddelerin birçoğunda benzer süreçleri deneyimlemektedir. Buna rağmen sabit oranlı silme işlemlerinin doğası gereği bazı durumlarda yalnızca yüzeysel/cümle düzeyinde anlama ile çözülebilecek sorularla karşılaşılabilir. Ancak testlerin geneli ele alındığında bahsi geçen türde sorularla pek karşılaşılması ve derin anlamayı gerektiren soruların oldukça büyük bir yüzdeyi kapsamaması bu sorunu göz ardı edilebilir hâle getirmektedir. Örnek olarak bu çalışma içerisinde oluşturulan boşluk tamamlama testlerinde yalnızca cümle düzeyinde anlama yahut sözcük kalıbının bilinmesiyle çözülebilecek maddeler ile derin anlamayı gerektiren maddelerin oranları aşağıdaki Tablo 5 içerisinde verilmiştir.

Tablo 5. Boşluk Tamamlama Testlerindeki Maddelerin Ölçüm Düzeyleri

Anlama Düzeyi	Eğer Gençlik Bilseydi		Okumak		Roman Okumak		Okul ve Meslek Seçimi	
	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
Yüzeysel Anlama	3	6	5	10	7	14	5	10
Derin Anlama	47	94	45	90	43	86	45	90

Tablo 5 incelendiğinde, boşluk tamamlama testlerindeki maddelerin %86-%94 oranlarında metnin genel yapısını anlamayı, çözümlenme yapmayı, cümleler ve paragraflar arası ilişkiler kurmayı ve tahmin etmeyi gerektiren derin anlama düzeyinde ölçüm yaptığı görülmektedir. Metinlerin detayları incelendiğinde ise “Eğer Gençlik Bilseydi” adlı metinde 3 (M28, M32, M47), “Okumak” isimli metinde 5 (M5, M21, M26, M27, M37), “Roman Okumak” metninde 7 (M1, M6, M8, M17, M25, M35, M45), “Okul ve Meslek Seçimi” adlı metinde ise 5 (M11, M14, M15, M36, M39) adet, metnin genelinde anlam kurmayı gerektirmeyen, cümlelerin anlaşılmasıyla ya da sözcük kalıbının bilinmesiyle yanıtlanabilecek madde bulunduğu anlaşılmaktadır.

Tablo 6. Çalışma Kapsamında Oluşturulan Boşluk Tamamlama Testlerinin KR-20 Güvenirlik Katsayıları

Metinler	Orijinal Puanlama	Bağlamsal Puanlama
Eğer Gençlik Bilseydi	0.72	0.87
Okumak	0.75	0.89
Roman Okumak	0.73	0.87
Okul ve Meslek Seçimi	0.71	0.84

Tablo 6 içerisinde, geçerliği ve güvenirligi kanıtlanmış olan “Bilgilendirici Metinleri Anlama Ölçeği”nden hareketle araştırma sürecinde oluşturulan boşluk tamamlama testlerinin KR-20 değerleri sunulmuştur. Bir testin her maddesine verilen cevapların 1 (doğru) ya da 0 (yanlış) olarak puanlanabildiği durumlarda KR-20 formülü kullanılmaktadır (Arastaman ve Kısa, 2020; Büyüköztürk vd., 2023). Söz konusu güvenirlilik katsayıları incelendiğinde ise her biri 50 maddeden oluşan tüm testlere yönelik KR-20 değerlerinin “0.70” katsayısından yüksek olduğu görülmektedir. Daha önce belirtildiği gibi güvenirlilik katsayısının 0.70 ya da daha yüksek olmasının testin güvenirligi için yeterli olduğu (Büyüköztürk, 2006) göz önüne alındığında ise bu araştırma kapsamında oluşturulan boşluk tamamlama testlerinin güvenilir araçlar olduğu sonucuna ulaşılabilir.

Tablo 7. Testlere İlişkin Betimsel Analizler

Test	N	Soru S.	Ortalama	Medyan	Mod	Ss	Varyans	Min	Maks
Eğer Gençlik Bilseydi Metni Çoktan Seçmeli T.	60	14	8.5	9	10	1.7	3.1	3	12
Okumak Metni Çoktan Seçmeli T.	60	14	7	7	8	2	2	4	10
Roman Okumak Çoktan Seçmeli T.	60	14	9	10	11	2.8	7.8	2	14
Okul ve Meslek Seçimi Metni Çoktan Seçmeli T.	60	14	9	10	11	2	6	4	13
Eğer Gençlik Bilseydi Metni Orijinal Boşluk Tamamlama T.	60	50	9.6	9	7	4.3	18.9	1	19
Okumak Orijinal Boşluk Tamamlama T.	60	50	6	6	1	4	18	0	20
Roman Okumak Metni Orijinal Boşluk Tamamlama T.	60	50	5.8	5	5	3.5	12	1	15
Okul ve Meslek Seçimi Metni Orijinal Boşluk Tamamlama T.	60	50	6	6	4	3	12	1	15
Eğer Gençlik Bilseydi Metni Bağlamsal Boşluk Tamamlama T.	60	50	17.1	17	13	8.5	72	2	39
Okumak Metni Bağlamsal Boşluk Tamamlama T.	60	50	15.4	15	18	8.3	69	2	36
Roman Okumak Metni Bağlamsal Boşluk Tamamlama T.	60	50	19	19	15	9	86	1	37
Okul ve Meslek Seçimi Metni Bağlamsal Boşluk Tamamlama T.	60	50	16	16	16	7	54	2	36

Tablo 7 içerisinde 3 farklı kategoride bulunan ölçme araçlarına yönelik betimsel istatistikler sunulmuştur. Bu istatistikler incelendiğinde farklı yapılarıdaki araçların ortalama zorluklarının birbirlerinden farklılaştığı görülmektedir. Testlerin ortalama güçlükleri çoktan seçmeli test olarak yapılandırılmış “Eğer Gençlik Bilseydi” adlı metinde “0.6”, “Okumak” adlı metinde “0.5”, “Roman Okumak” metninde “0.64” “Okul ve Meslek Seçimi” isimli metinde “0.64”; orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testi olarak yapılandırılmış “Eğer Gençlik Bilseydi” adlı metinde “0.19”, “Okumak” adlı metinde “0.12”, “Roman Okumak” metninde “0.11” “Okul ve Meslek Seçimi” isimli metinde “0.12”; bağlamı koruyan sözcüklerin doğru kabul edildiği boşluk tamamlama testi olarak yapılandırılmış “Eğer Gençlik Bilseydi” adlı metinde “0.34”, “Okumak” adlı metinde “0.3”, “Roman Okumak” metninde “0.38” “Okul ve Meslek Seçimi” isimli metinde “0.32” olarak tespit edilmiştir. Bu durum, literatürde karşılığını bulduğu üzere boşluk tamamlama testlerinin alışılmış ve standartlaştırılmış okuduğunu anlama testlerinden daha zor yapıda olduğunu göstermektedir. Diğer yandan tüm testlerdeki tüm maddeler aynı puan değerine sahiptir. Bu doğrultuda 14 sorudan oluşan çoktan seçmeli testlerdeki toplam doğru sayısı “7.14” değeriyle, 50 sorudan oluşan boşluk tamamlama testlerindeki toplam doğru sayısı ise “2” değeri ile çarpılmış ve tüm testler “100” tam puan üzerinden değerlendirilmiştir.

Açıklanan veri toplama araçlarıyla birlikte sonuç olarak 4 farklı metne yönelik, her biri 14 çoktan seçmeli okuduğunu anlama sorusuna sahip 4 farklı okuduğunu anlama testi ve her biri farklı prosedürlerle oluşturulmuş 4 farklı boşluk tamamlama testi oluşturulmuştur. Bu bağlamda araştırma sürecinde 8 farklı veri toplama aracı vasıtasıyla araştırmanın verileri toplanmıştır.

Verilerin Toplanması ve Analizi

Veriler, araştırmacılar tarafından 2023-2024 eğitim/öğretim döneminde öğrencilerle gerçekleştirilen yüz yüze uygulamalar ile yaklaşık 4 aylık bir süreçte toplanmıştır. Öncelikli olarak öğrencilerle ikişer ders saati boyunca ölçme aracı olarak kullanılmayan boşluk tamamlama örnekleri yapılmış ve öğrencilerin uygulamaya yönelik farkındalık kazanmaları amaçlanmıştır. Ardından öğrencilere, metinlerin boşluk doldurma testi hâline dönüştürülmüş versiyonları sunulmuştur. Her bir boşluk tamamlama testi için öğrencilere 1 ders saati verilmiştir. Bu işlemin ardından yaklaşık 2 hafta boyunca öğrencilerle bu araştırmaya yönelik hiçbir uygulama gerçekleştirilmemiştir. 2 haftanın ardından ise sırayla öğrencilere “Bilgilendirici Metinler Okuduğunu Anlama Ölçeği” içerisinde yer alan testler uygulanmıştır. Bu uygulamanın da sona ermesiyle veri toplama süreci tamamlanmıştır.

İki test türünden de verilerin elde edilmesiyle veri analizi sürecine geçiş yapılmıştır. Veri analiz sürecinde ilk olarak testlerden alınan puanların normallik dağılımları ele alınmıştır. Boşluk tamamlama testlerinin puanlama sürecinde iki farklı puanlama yöntemi kullanıldığı için bu süreçte 8 farklı ölçme aracına yönelik 12 farklı sonuç listesi normallik analizine tabi tutulmuştur. Testlerin normallik analizlerinden elde edilen sonuçlar ise Tablo 8 içerisinde sunulmuştur.

Tablo 8. Testlerden Elde Edilen Puanların Normallik Dağılımları

Testler	Çarpıklık (Skewness)	Basıklık (Kurtosis)
“Eğer Gençlik Bilseydi” Çoktan Seçmeli Test	-.646	.539
“Okumak” Çoktan Seçmeli Test	-.364	-.551
“Roman Okumak” Çoktan Seçmeli Test	-.635	-0.63
“Okul ve Meslek Seçimi” Çoktan Seçmeli Test	-.478	-.618
“Eğer Gençlik Bilseydi” Orijinal Boşluk Tamamlama	.162	-.655
“Okumak” Orijinal Boşluk Tamamlama	.738	.665
“Roman Okumak” Orijinal Boşluk Tamamlama	.577	-.345
“Okul ve Meslek Seçimi” Orijinal Boşluk Tamamlama	.561	-.293
“Eğer Gençlik Bilseydi” Bağlamsal Boşluk Tamamlama	.308	-.593
“Okumak” Bağlamsal Boşluk Tamamlama	.082	-.624
“Roman Okumak” Bağlamsal Boşluk Tamamlama	.355	-.218
“Okul ve Meslek Seçimi” Bağlamsal Boşluk Tamamlama	.428	.065

Tablo 8 incelendiğinde tüm testlere ait çarpıklık ve basıklık değerlerinin -0.7 ile 0.8 arasında olduğu görülmektedir. Çarpıklık değerlerinin -1 ile +1 aralığının dışında kalmasının büyük ölçüde çarpık bir dağılıma işaret ettiği (Hair, Black, Babin, Anderson ve Tatham, 2013) ve basıklık değerinin ± 1.0 aralığında olmasının çoğu amaç için mükemmel olduğu (George ve Mallery, 2010) dikkate alındığında bu araştırmada testler vasıtasıyla elde edilen verilerin normal dağılım gösterdiklerini söylemek mümkündür. Dolayısıyla çalışmanın amaçları doğrultusunda gerçekleştirilecek olan korelasyon analizleri normal dağılımlara yönelik hesaplamalarla gerçekleştirilmiştir. Araştırma sürecinde hesaplanan korelasyon katsayılarının değerlendirilmesinde dikkate alınan sınırlar ise Tablo 9 (Köklü, Büyüköztürk ve Çokluk, 2007; Taşpınar, 2017; Taylor, 1990) içerisinde gösterilmiştir.

Tablo 9. Korelasyon Katsayılarının Yorumlanması

Değer	Yorum
0.00 – 0.19	Çok zayıf ilişki
0.20 – 0.39	Zayıf ilişki
0.40 – 0.69	Orta düzeyde ilişki
0.70 – 0.89	Yüksek düzeyde ilişki
0.90 – 1.00	Çok yüksek düzeyde ilişki

Tablo 9 doğrultusunda çalışma boyunca 0.20–0.39 arasındaki korelasyon değerleri, zayıf ilişki; 0.40- 0.69 arasındaki korelasyon değerleri, orta düzey ilişki ve 0.70–0.89 arasındaki korelasyon değerleri yüksek düzeyde ilişki olarak kabul edilmiş ve yorumlanmıştır.

Etik Kurul Onayı

Bu araştırma kapsamında belirlenen veri toplama prosedürünün ve toplanan verilerin herhangi bir etik problem barındırmadığı, 2024-SBB-0074 protokol koduyla Bartın Üniversitesi Sosyal ve Beşerî Bilimler Etik Kurulu tarafından onaylanmıştır. Gönüllülük esasına dayalı olarak çalışmaya katılanlardan hiçbir kişisel veri toplanmamış ve araştırma boyunca tüm etik kurallara uyulmuştur.

Bulgular

Araştırma sürecinde elde edilen bulgular, araştırma sorularının sırası ile uyumlu bir şekilde aşağıda sunulmuştur:

Tablo 10. Aynı Metne Yönelik Çoktan Seçmeli Okuduğunu Anlama Testi ile Düzenli Olarak 4. Kelimenin Silindiği Boşluk Tamamlama Testinin Pearson Korelasyon Analizi Sonuçları

	Ortalama	Ss	N	1	2
1 Çoktan Seçmeli Test	61.3	12.63	60	1	
2 Orijinal Puanlanan BTT	19.2	8.71	60	.538**	1
3 Bağlamsal Puanlanan BTT	34.3	17	60	.587**	.850**

**p<0.01; BTT: Boşluk Tamamlama Testi

Tablo 10 incelendiğinde aynı metne yönelik çoktan seçmeli okuduğunu anlama testi ile düzenli olarak 4. kelimenin silindiği boşluk tamamlama testi arasında orta düzeyde pozitif bir ilişki olduğu görülmektedir. Bu ilişki düzeyi, boşluk tamamlama testlerinin puanlanma sürecinde orijinal kelimelerin kabul edildiği puanlamada olduğu gibi bağlamı koruyan kelimelerin doğru kabul edildiği puanlama yönteminde de kendisini göstermektedir. Ancak korelasyon katsayıları incelendiğinde, bağlamı koruyan sözcüklerin doğru kabul edilmesinin, çoktan seçmeli testler ile orijinal sözcüklerin doğru kabul edildiği puanlamadan daha tutarlı sonuçlar verdiği görülmektedir. Diğer yandan bağlamı koruyan kelimelerin doğru kabul edildiği puanlama yönteminden elde edilen boşluk tamamlama testi puanları ile orijinal kelimelerin doğru kabul edildiği puanlama yöntemi ile ulaşılan boşluk tamamlama puanları arasında .850 korelasyon katsayısı ile yüksek düzeyde pozitif bir ilişki olduğu sonucuna ulaşılmıştır.

Tablo 11. Aynı Metne Yönelik Çoktan Seçmeli Okuduğunu Anlama Testi ile Düzenli Olarak 5. Kelimenin Silindiği Boşluk Tamamlama Testinin Pearson Korelasyon Analizi Sonuçları

	Ortalama	Ss	N	1	2
1 Çoktan Seçmeli Test	50.11	10.8	60	1	
2 Orijinal Puanlanan BTT	11.97	8.42	60	.383**	1
3 Bağlamsal Puanlanan BTT	37.87	18.57	60	.322*	.706**

**p<0.01; *p<0.05; BTT: Boşluk Tamamlama Testi

Tablo 11 incelendiğinde, Tablo 10'te görülen orta düzeydeki anlamlı ilişkiden farklı olarak 5. sözcüğün düzenli olarak silindiği boşluk tamamlama testleri ile çoktan seçmeli testler arasında zayıf düzeyde anlamlı bir ilişki olduğu görülmektedir. Bu ilişkinin niteliği, boşluk tamamlama testlerinde işe koşulan her iki puanlama yönteminde de aynı olmakla birlikte orijinal sözcüklerin doğru kabul edildiği puanlama prosedüründe .383 korelasyon katsayısı ile daha yüksek konumdadır. Bu puanlama yöntemlerinin birbirleri ile ilişkisi ele alındığında ise kullanılan yöntemlerin .706 korelasyon katsayısı ile yüksek düzeyde bir tutarlılığa sahip olduğu sonucuna ulaşılmaktadır.

Tablo 12. Aynı Metne Yönelik Çoktan Seçmeli Okuduğunu Anlama Testi ile Düzenli Olarak 6. Kelimenin Silindiği Boşluk Tamamlama Testinin Pearson Korelasyon Analizi Sonuçları

	Ortalama	Ss	N	1	2
1 Çoktan Seçmeli Test	64.76	20.06	60	1	
2 Orijinal Puanlanan BTT	11.6	6.92	60	.639**	1
3 Bağlamsal Puanlanan BTT	30.7	16.7	60	.637**	.707**

**p<0.01; BTT: Boşluk Tamamlama Testi

Tablo 12’de aynı metne yönelik düzenli olarak 6. sözcüğü silinen ve farklı şekillerde puanlanan boşluk tamamlama testi ile çoktan seçmeli test arasındaki korelasyonlar yer almaktadır. Tablo 12 incelendiğinde, düzenli olarak 4. sözcüğün silindiği boşluk tamamlama testlerinde olduğu gibi düzenli olarak 6. sözcüğün silindiği boşluk tamamlama testleri ile çoktan seçmeli testler arasında orta düzeyde anlamlı bir ilişki olduğu sonucuna ulaşılmaktadır. Ayrıca söz konusu ilişkinin kullanılan her iki puanlama yönteminde de oldukça benzer olması dikkat çekmektedir. Keza bu iki puanlama yöntemi arasında da .707 korelasyon katsayısı ile yüksek düzey bir ilişki bulunmaktadır. Ancak düzenli olarak 6. sözcüğün silindiği boşluk tamamlama testi, çoktan seçmeli testler ile 4 ve 5. kelimelerin düzenli olarak silindiği boşluk tamamlama testlerinden daha yüksek bir tutarlılık sergilemektedir.

Tablo 13. Aynı Metne Yönelik Çoktan Seçmeli Okuduğunu Anlama Testi ile Düzenli Olarak 7. Kelimenin Silindiği Boşluk Tamamlama Testinin Pearson Korelasyon Analizi Sonuçları

	Ortalama	Ss	N	1	2
1 Çoktan Seçmeli Test	65.47	17.25	60	1	
2 Orijinal Puanlanan BTT	12.97	6.89	60	.416**	1
3 Bağlamsal Puanlanan BTT	32.93	14.66	60	.405**	.860**

**p<0.01; BTT: Boşluk Tamamlama Testi

Tablo 13 ele alındığında, düzenli olarak 7. sözcüğün silindiği boşluk tamamlama testlerinin de çoktan seçmeli testler ile orta düzeyde anlamlı bir tutarlılık gösterdiği görülmektedir. Keza bu testte de farklı puanlama yöntemleriyle ulaşılan sonuçların çoktan seçmeli testler ile gösterdiği tutarlılıkların benzer olduğu dikkat çekmektedir. Ancak 5. ve 6. sözcüklerin düzenli olarak silinmesiyle elde edilen boşluk tamamlama testlerinde olduğu gibi bu testte de orijinal kelimelerin doğru kabul edildiği puanlama yönteminin daha yüksek bir korelasyon katsayısına sahip olduğu elde edilen bulgular arasındadır.

Sonuç ve Tartışma

Boşluk tamamlama testlerinin Türkçe özelinde okuduğunu anlama düzeyini ölçüp ölçmediğini belirlemek ve boşluk tamamlama testlerinin uygulanışına yönelik öne sürülen prosedürleri test etmek amacıyla yürütülen bu araştırma özelinde ulaşılan sonuçlar şu şekildedir:

Araştırmada ilk olarak veri toplama araçları elde edilmiş, ardından veri toplama sürecine geçilmiştir. Bu süreçte, tüm uygulamalar için aynı sistematik takip edilmiştir. Bu sistematik, öğrencilere öncelikle metinlerin boşluk tamamlama versiyonlarının sunulması, 2 haftalık bekleme süresinin ardından ise aynı metinlere yönelik on dörder çoktan seçmeli soruya sahip testlerin uygulanmasından oluşmaktadır. Açıklanan uygulama süreci doğrultusunda ilk olarak 4. sözcüğü silinen ve farklı şekillerde puanlanan boşluk tamamlama testi ile aynı metne yönelik çoktan seçmeli test arasındaki korelasyonlar incelenmiştir. İnceleme sonrası boşluk tamamlama testinin puanlanma prosedürü fark etmeksizin iki test arasında orta düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmüştür. Daha sonra aynı işlemler, düzenli olarak 5. sözcüğü silinen ve farklı şekillerde puanlanan boşluk tamamlama testleri için gerçekleştirilmiş ve çoktan seçmeli okuduğunu anlama testi ile boşluk tamamlama testleri arasında zayıf düzeyde pozitif bir ilişki bulunmuştur. Aynı işlemlerin devam ettirilmesiyle ise 6. ve 7. sıradaki sözcüklerin düzenli olarak silindiği boşluk tamamlama testleri ile çoktan seçmeli testler arasında da puanlama yöntemi fark etmeksizin orta düzeyde pozitif bir ilişki tespit edilmiştir. Nitekim genel olarak boşluk tamamlama testleri ile çoktan seçmeli testler arasında orta düzeyde anlamlı bir ilişki olduğu, orijinal kelimelerin doğru kabul edildiği puanlama yönteminin çoktan seçmeli testler ile nispeten daha tutarlı sonuçlar verdiği ve çoktan seçmeli testler ile en yüksek düzeyde tutarlılığın düzenli olarak 6. sözcüğün silindiği ve orijinal sözcüklerin doğru kabul edildiği puanlama yöntemiyle elde edildiği sonucuna ulaşılmıştır. Ancak en yüksek korelasyonların elde edildiği 6. sözcüğün düzenli olarak silindiği uygulamada farklı puanlama yöntemleriyle elde edilen sonuçlar arasındaki farkın oldukça küçük olduğu da dikkate değerdir.

Açıklanan süreçler kapsamında, işe koşulan tüm boşluk tamamlama testleri ile çoktan seçmeli okuduğunu anlama testleri arasında anlamlı bir ilişki olduğu tespit edilmiştir. Bu ilişkinin niteliği ise testlere göre değişkenlik göstermekle birlikte genel olarak orta düzeyde pozitifdir. Bu doğrultuda genel olarak boşluk tamamlama testlerinin okuduğunu anlamayı ölçme konusunda oldukça sık başvurulan ve ulusal sınavlarda başat soru tarzı olarak kullanılan çoktan seçmeli testlerle tutarlı olduğunu ve okuduğunu anlamayı çoktan seçmeli testlerle benzer düzeyde ölçtüğünü söylemek mümkündür. Bu bağlamda araştırma sürecinde ulaşılan sonuçların; boşluk tamamlama testlerinin farklı diller için geçerli sonuçlar vereceğini/verdiğini (Klare, 1974; Suehiro ve Santos, 2015; Sukarni, 2021; Taylor, 1956; Tremblay ve Garrison, 2010), metinlerin okunabilirliğinin (Bormuth, 1963; Bormuth, 1967; Klare, Simaiko ve Stolurow, 1972) ve öğrencilerin okuduğunu anlama düzeyinin (Brown, 1982; Brown ve Grüter, 2020; Febriyanti, 2017; Huensch, 2013; Şahindokuyucu, 2006; Lu, 2006; Oller, 2006) tespit edilmesinde boşluk tamamlama testlerinin kullanılabilirliğini destekleyen çalışmalarla tutarlı olduğu ifade edilebilir. Keza bu sonuçlar, Türkçe özelinde ulusal alan yazında gerçekleştirilen ve boşluk tamamlama testlerinin öğrencilerin okuma seviyesinin tespitinde (Tunçer ve Erden, 2015; Ulusoy, 2009; Uyanık, 2012), okuduğunu anlama düzeyinin ölçülmesinde (Akyol, 2020; Çetinkaya, 2010; Hızarcı, 2009) ve metinlerin okunabilirliğinin belirlenmesinde (Çepni, Gökdere ve Küçük, 2002; Keskin ve Akıllı, 2013; Köse, 2009) geçerli sonuçlar verdiği/vereceğini belirten araştırmalarla uyumlu olduğu söylenebilir. Dolayısıyla boşluk tamamlama testlerinin, sahip olduğu pratik ve nesnel uygulama ve hemen hemen her metne başka bir araca ihtiyaç duymaksızın doğrudan uygulanabilme özellikleri ile okuduğunu anlama düzeyinin ölçümünde ve metinlerin okunabilirliğini belirlemede kullanılabilirliği sonucuna ulaşılmaktadır.

Boşluk tamamlama testleri, her ne kadar okuduğunu anlama düzeyini ölçülmesinde oldukça işlevsel araçlar olarak nitelendirilse de ilgili alan yazında testlere yönelik belirli tartışmalı konular bulunmaktadır. Bu tartışmaların yoğunlaştığı konulardan birisi ise boşluk tamamlama testlerinin hazırlanma sürecinde kaçınıcı sıradaki sözcüğün silinmesi gerektiğidir. Bormuth (1964) bu hususta ana dili için geliştirilen bir boşluk tamamlama testinde silinen n. sıradaki sözcüğün sağındaki ya da solundaki kelimelerin silinmesiyle eş değer testler elde edilebileceğini savunmaktadır. Benzer şekilde Bachman (1982), boşluk tamamlama testlerinin detaylarının test edilmesine gerek olmadığını, düzenli bir silme işlemiyle geçerli sonuçlar elde edileceğini belirtmektedir. Oller ve diğerleri (1972) ise silme işleminin düzensiz olsa bile geçerli sonuçlar verdiğini ancak kullanıma uygun olmadığını ifade etmektedir. Diğer yandan Potter (1968), bir metnin %20'sinden fazlasının silinmesinin metni fazla zorlaştırdığını öne sürmektedir. Uluslararası alan yazında süregelen bu ifadelerin yanı sıra benzer tartışmalar ulusal alan yazında Türkçe için de yaşanmaktadır. Örnek olarak “yanlış envanteri analizi” (Akyol, 2020) vasıtasıyla belirlenen okuma düzeyi ile boşluk tamamlama testleri arasındaki tutarlılığı ölçen Kaplan ve Çiftçi (2021), düzenli olarak 4., 5. ve 6. sıradaki sözcüklerin silindiği testler arasından en tutarlı sonuçları düzenli olarak 4. sözcüğün silindiği boşluk tamamlama testinden aldıklarını belirtmişlerdir. Benzer şekilde, boşluk tamamlama testleri ile 6 açık uçlu okuduğunu anlama sorusundan oluşan test arasındaki korelasyonu inceleyen Uyanık (2011), 7. sıradaki sözcüğün düzenli olarak silindiği boşluk tamamlama testinin 6. sıradaki sözcüğün düzenli olarak silindiği boşluk tamamlama testinden daha uygun olduğunu; yine bir başka 6 açık uçlu okuduğunu anlama sorusundan oluşan test ile boşluk tamamlama testleri arasındaki korelasyonu inceleyen Tunçer ve Erden (2015) ise 10. sıradaki sözcüğün silinmesinin ve sözcüğün ilk harfinin ipucu olarak verilmesinin daha tutarlı olduğunu savunmaktadır. Diğer yandan Keskin ve Akıllı (2013) aynı metin üzerinde farklı sıradaki sözcüklerin silinmesiyle oluşturdukları metinleri öğrencilere uygulamış ve anlamlı bir fark tespit edememiştir.

Bu çalışmada ise Türkçe özelinde en tutarlı sonuçların düzenli olarak 6. sözcüğün silindiği boşluk tamamlama testlerinden elde edildiği tespit edilmiştir. Bu sonucun, Ulusoy'un (2009) ve Hızarcı'nın (2009) 6. sıradaki sözcüklerin silinmesinin araştırmayı 5. sözcüğün silinmesinden daha geçerli sonuçlara ulaştıracağına dair tespiti ile tutarlı olduğunu söylemek mümkündür. Ancak farklı sıradaki sözcüklerin silindiği testlerin farklı metinlere yönelik olduğu dikkate alındığında, bu karşılaştırmanın sağlıklı bir şekilde yapılamayacağı ve kesin sonuçlar için daha detaylı araştırmalara ihtiyaç duyulduğu ifade edilebilir. Prosedürler arasında net bir karşılaştırma yapılabilmesi için farklı prosedürler dışında okuduğunu anlamayı etkileyebilecek unsurların en aza indirilmesi gerekmektedir. Bu çalışmada her ne kadar gereken özen gösterilmiş olsa da farklı prosedürlerle oluşturulan testlerden alınan puanları etkileyen tek faktör değişen prosedürler değildir. Metinlerin ve silinen kelimelerin birbirlerinden farklılaşması bu değerlendirmeyi yapmayı zorlaştırmaktadır. Diğer yandan konuyla ilgili gerçekleştirilecek gelecek araştırmaların objektif ve nesnel bir düzlemde gerçekleştirilmesi önem arz etmektedir.

Boşluk tamamlama testlerinde silme işlemine yönelik ön plana çıkan tartışmaların bir diğer boyutunu da silme prosedüründe benimsenen yaklaşım oluşturmaktadır. Daha önce açıklandığı üzere silme prosedürü sabit oranlı ya da rasyonel olarak iki farklı şekilde yapılmaktadır. Bahsi geçen iki yöntemin de belirli avantajları ve dezavantajları bulunmaktadır. Örnek olarak sabit oranlı silme işlemiyle standart bir teknik elde edilmekte, her metne anında ve zahmetsizce uygulanabilmektedir. Diğer yandan rasyonel silme işlemiyle ise metnin anlam boyutuna daha derin bir şekilde odaklanılabilmekte, anlamayı daha iyi ölçen testler oluşturulabilmektedir. Nitekim Türkçe gibi söz dizimindeki esnekliklerin ve değişikliklerin anlamı derinden etkileyebileceği dillerde rasyonel silme işleminin daha uygun olacağı söylenebilir. Ancak bu çalışmada rasyonel silme işlemi ile ilgili herhangi bir analiz gerçekleştirilmemiştir. Bu nedenle rasyonel silme işlemine yönelik herhangi bir yorum yapılamamaktadır. Bu çerçevede rasyonel silme prosedürlerinin Türkçe için geçerli sonuç verip vermediği ulusal alan yazında sorgulanabilir.

Yukarıdaki tartışmalara ek olarak boşluk tamamlama testlerinin puanlanma süreci de ulusal ve uluslararası alan yazında bir tartışma konusu olarak göze çarpmaktadır. Bu konuya yönelik tartışmalar ise yalnızca orijinal sözcüklerin mi yoksa cümlenin anlamını koruyan tüm sözcüklerin doğru kabul edileceği ile ilgilidir. Esasında boşluk tamamlama testlerinin puanlanmasında kullanılacak üç farklı yaklaşım bulunmaktadır ve bu yaklaşımlar “Roman Okumak” adlı metinden alınan örnek cümle üzerinden aktarılmaya çalışılacaktır:

“Roman ise gerçekten uzaklaşmaz, _____ gerçeği kavratmaya, hep gerçeği anlatmaya _____.”

Yukarıdaki örnekte ilk boşluğun metindeki orijinal karşılığı “hep” sözcüğüdür. Burada “hep” sözcüğü dışındaki tüm sözcüklerin yanlış kabul edilmesi sadece metindeki birebir sözcüklerin doğru kabul edilmesi anlamına gelmektedir. Diğer yandan bu boşluğa “daima” kelimesinin yazılması da metnin genel yapısını ve anlamını uygun şekilde korumaktadır. Bu şekilde boşlukta yer alması gereken sözcükle birebir eşleşen sözcüklerin doğru kabul edilmesi ikinci bir yöntemdir. Üçüncü yöntem ise cümlenin anlamını koruyan ancak metnin genel anlamından nispeten uzaklaşan sözcüklerin de doğru kabul edilmesidir. Bu duruma örnek olarak birinci boşluğa “roman” sözcüğünün getirilmesidir. Boşluğa “roman” sözcüğünün getirilmesi ile cümlenin anlamı korunsa da metnin genel anlamından ve amacından uzaklaşmaktadır. Benzer örnekler ikinci boşluk için de verilebilir. İkinci boşlukta yer alan orijinal sözcük orijinal sözcük, “çalışır” kelimesidir. Ancak boşluğa yazılabilecek olan “uğraşır” sözcüğü de metnin anlamını “çalışır” kelimesiyle benzer şekilde karşılamaktadır. Diğer yandan boşluğa yazılabilecek “özenir” yahut “yoğunlaşır” sözcükleri ise cümleyi anlamlı hâle getirmekte ancak metnin hedeflediğinden farklı bir anlam oluşturmaktadır. Bu çalışmada ise yalnızca orijinal sözcüklerin doğru kabul edilmesi ve bağlamı koruyan sözcüklerin doğru kabul edilmesi üzerine bir inceleme gerçekleştirilmiştir. Nitekim araştırma sürecinde; boşluklara orijinal kelimeleri yazan öğrencilerin testin genelinde daha başarılı olduğu, orijinal kelimeleri yazamayan öğrencilerin ise genel olarak diğer boşlukları doldurmada da başarısız olduğu görülmüştür. Dahası, bağlamı koruyan sözcüklerin doğru kabul edildiği puanlamalarda genelde çoktan seçmeli testler ile daha düşük korelasyonlar elde edilmiştir. Bu sonucun, orijinal kelimelerin doğru kabul edilmesini savunan araştırmacıların (Taylor, 1953; Ulusoy, 2009; Uyanık, 2012) görüşleriyle tutarlı olduğu ifade edilebilir. Diğer yandan bağlamı koruyan kelimeleri doğru kabul etmenin gerekli olduğunu savunan araştırmacılar da (Kaplan ve Çiftçi, 2021; Oller vd., 1972) bulunmaktadır. Söz konusu ilişkiye ek olarak iki puanlama yöntemiyle de elde edilen puanların çoktan seçmeli okuduğunu anlama testleri ile benzer korelasyonlara sahip olduğunu söylemek mümkündür. Ancak bu noktada, bir ölçme aracı ile elde edilen verilerin objektif olarak puanlanabilmesi gerekliliği (Turgut ve Baykul, 2019), orijinal kelimelerin doğru kabul edildiği puanlamanın esas alınmasını daha makul hâle getirmektedir. Benzer şekilde bu prosedürle puanlanan boşluk tamamlama testleri oldukça nesnel ölçme araçları olarak kendilerini göstermektedir. Keza bu araştırmada da hâlihazırda çoğunlukla en yüksek korelasyon düzeyleri orijinal sözcüklerin doğru kabul edildiği puanlama yöntemiyle elde edilmiştir. Ancak burada Türkçenin eklemeli bir dil olması sebebiyle yalnızca orijinal kelimelerin dikkate alındığı puanlama sürecinde sözcüklerin öğrenciler tarafından hangi zamanda ya da durumda çekimlendiği dikkate alınmamıştır. Çünkü bilindiği üzere Türkçede bir yargı, birden çok şekilde çekimlenerek ifade edilebilmektedir. Örnek olarak Ahmet kişinin yarın bir yere geleceğini bildiren cümle; “Ahmet yarın buraya gelecek.”, “Ahmet yarın buraya gelecektir.”, “Ahmet yarın buraya geliyor.”, “Ahmet yarın buraya gelir.”, “Ahmet yarın buraya gelebilir.”, “Ahmet yarın buraya gelmeli.” şekillerinde gösterilebilir. Bu durumda “gelmek” ifadesinin çekimleniş farklılığından dolayı yanlış kabul edilmesinin doğru olmayacağı düşünülmektedir. Dolayısıyla Türkçe özelinde bu husustaki standart, uluslararası alan yazından hareketle çekimli bir dil olan ve kelimenin yapısındaki değişimlerle anlamın güçlü bir şekilde değiştiği İngilizce için geçerli standart üzerinden tayin edilmemelidir. Bu yönüyle, bu araştırmada orijinal kelimelerin doğru kabul edildiği puanlama metodolojisinin, sadece orijinal kelimelerle birebir sözcüklerin doğru kabul edildiği ve farklı hecelere sahip sözcüklerin yanlış olarak ele alındığı yöntemden farklılaştığı söylenebilir.

Boşluk tamamlama testleriyle ilgili gündeme getirilen diğer bir eleştiri konusu ise ilgili testlerin standartlaştırılmış okuduğunu anlama testleri ile “yüksek” düzeylerde korelasyon göstermemesiyle ilgilidir. Alan yazında her ne kadar boşluk tamamlama testleri ile standartlaştırılmış okuduğunu anlama testleri ile “.90” üzeri korelasyonlar tespit edilse de (Bormuth, 1967) genel olarak araştırmalarda bu denli yüksek korelasyonlara erişilememektedir. Boşluk tamamlama testleri ile diğer standartlaştırılmış testlerin tamamen farklı yapılarda olmalarından kaynaklı farklı stratejilerin işe koşulmasını gerektirmesi (Retorta, 2001, s. 137), test yapılarına maruziyetin test başarısını etkilemesi (Ruddell, 2005) ve öğrencilerin diğer standartlaştırılmış okuduğunu anlama testlerinin yapısı kadar boşluk tamamlama testlerine hâkim olmaması, “.85” ve üzeri korelasyonların çoğunlukla “test-tekrar test” yöntemiyle elde edilmesi ve bu denli yüksek korelasyonlara nadiren ulaşılması (Creswell, 2012/2020) bu durumun sebepleri olarak ele alınabilir. Bu sebepler aynı zamanda boşluk tamamlama testlerine yönelik yürütülen çalışmalarda ortaya çıkan sınırlılıkları da temsil etmektedir. Diğer yandan aynı beceriyi ölçen araçlar arasında daima yüksek korelasyonlar elde edilememektedir. Kleijn, Pander Maat ve Sanders’in (2019) belirttiği üzere iyi yapılandırılmış ve standartlaştırılmış okuduğunu anlama testleri dahi birbirleriyle “.31”-“.79” değerleri arasında korelasyonlar gösterirken (Cutting ve Scarborough, 2006; Keenan, Betjemann ve Olson, 2008) boşluk tamamlama testleri ile standartlaştırılmış okuduğunu anlama testleri arasında “0.6” dolaylarında elde edilen korelasyonların boşluk tamamlama testlerinin standartlaştırılmış okuduğunu anlama testlerinden daha düşük performans göstermediğini ortaya koymaktadır.

Sonuç olarak bu çalışmada yapılan açıklamalar ve değerlendirmelerle birlikte ulaşılan yargılar; boşluk tamamlama testleri ile çoktan seçmeli okuduğunu anlama testleri arasında anlamlı bir ilişki olduğu, boşluk tamamlama testlerinin okuduğunu anlama düzeyinin ölçümde kullanılabileceği, daha geçerli bir boşluk tamamlama testi oluşturmak için düzenli olarak 6. sıradaki boşluğun silinmesi gerektiği, daha güvenilir ve nesnel bir puanlama için boşluk tamamlama testlerinde orijinal kelimelerin kabul edilmesi gerektiği şeklinde özetlenebilir. Bu kapsamda boşluk tamamlama testlerini okuduğunu anlama düzeyinin ölçülmesine yönelik bir araç olarak kullanmanın mümkün olduğu ifade edilebilir. Ayrıca bu araçların, açık uçlu soruların sahip olduğu puanlayıcının özelliklerinden etkilenme, puanlama sürecinin objektiflik açısından zorluğu gibi öznellik sorunlarını ve puanlamanın zaman açısından ekonomik olmaması problemlerini bünyesinde barındırmadığı görülmektedir. Çünkü boşluk tamamlama testleri oldukça nesnel ve kolay bir şekilde puanlanabilmektedir. Benzer şekilde boşluk tamamlama testlerinin; çoktan seçmeli soruların sahip olduğu şans faktörü, öğrencilerin yaratıcılıklarını gölgeleme, yorumlama güçlerini kısıtlama, oluşturulma süreçlerinde uzmanlık ve tecrübe gerektirme gibi sorunlara da iye olmadığını söylemek mümkündür. Çünkü boşluk tamamlama testlerinde, okuyucunun metinde anlatıları yorumlayarak, sözcük bilgisini ve yaratıcılığını işe koşarak cevabı kendisinin üretmesi gerekmektedir. Keza daha önce bahsedildiği üzere boşluk tamamlama testleri sabit oranlı silme işlemiyle hazırlandığında herhangi bir uzmanlık gerektirmemektedir. Diğer yandan ölçme değerlendirme uygulamalarının çeşitli olması da önem arz etmektedir. Aksi hâlde devamlı olarak aynı türdeki ölçme araçlarıyla değerlendirilen öğrencilerin kalıplaşmış zihin yapılarına sahip olmaları (Üstüner ve Şengül, 2004) ve bu kalıpların dışına çıkmakta zorlanmaları muhtemeldir. Bu durum da eleştirel düşünmeye, girişimciliğe ve yaratıcılığa önem verilen modern dünyada ve eğitim sistemimizde ciddi bir sorundur. Bu bağlamda ölçme değerlendirme uygulamalarında işe koşulan araçların çeşitlendirilmesi, hâlihazırda kullanımda olan araçların zayıf yönlerini telafi edebilecek uygulamaların eğitime dâhil edilmesinin bir gereklilik olduğu söylenebilir.

Öneriler

- Okuduğunu anlama düzeyini ölçme konusunda pratik ve nesnel araçlar olan boşluk tamamlama testleri, eğitim süreçlerine sistematik bir şekilde dâhil edilmelidir.
- Boşluk tamamlama testlerinin puanlama sürecinde nesnelliğin ön planda olduğu, orijinal kelimelerin doğru kabul edildiği puanlama sistemi, daha geçerli sonuçlar elde etmek için kullanılmalıdır. Ancak Türkçenin yapısal özellikleriyle ortaya çıkan, bir ifadenin farklı şekillerde çekimlenerek sunulabilme potansiyeli doğrultusunda farklı şekillerde çekimlenmiş sözcüklerin yanlış kabul edilmemesi makul olacaktır.

- Bořluk tamamlama testlerinin farklı diller için ortaya konulan diđer iřlevleri Trke için de test edilmelidir.
- Trke için bořluk tamamlama testlerinde bulunması gereken ideal bořluk sayısı farklı alıřmalarla test edilmelidir.
- Nesnel özellikler taşıyan okuduđunu anlama testleri ile bořluklara gelmesi gereken szcklerin ilk harfinin ipucu olarak verildiđi bořluk tamamlama testleri arasındaki korelasyonlar incelenmelidir.
- Farklı sınıf seviyelerinde ve yař gruplarında bořluk tamamlama testlerinin geerliliđi sorgulanmalıdır.
- Dnřtrme iřleminin rasyonel řekilde gerekleřtirildiđi bořluk tamamlama testlerinin geerliliđi Trke zelinde ele alınmalıdır.
- Bořluk tamamlama testlerinin hazırlanmasına ynelik farklı prosedrler, aynı okunabilirlik dzeyindeki metinlerden silinen benzer nitelikli szcklerle test edilmeli ve bylece en uygun bořluk tamamlama testi oluřturma prosedr tespit edilmelidir.
- Bořluk tamamlama testlerinin geerliliđi, daha kalabalık ve ilgili test trne ynelik gerekli tecbeyi edinmiř alıřma grupları ile test edilmelidir.
- Bořluk tamamlama testlerinin geerliliđi farklı trlerdeki standartlařtırılmıř okuduđunu anlama testleri ile karřılařtırılmalıdır.

Sınırlılıklar

Arařtırma srecinde, hangi sıradaki szcđn dzenli olarak silinmesi gerektiđini belirlemeye ynelik alıřmalarda farklı metinler kullanılmıřtır. Metinlerin birok zelliđi, testlerden elde edilen puanları etkileyebildiđi için bu srete korelasyonlara ynelik tek etken testlerin oluřturma metodolojisindeki farklılık olmamıřtır. Diđer yandan đrencilerin sınav trne ynelik tecbelerinin bařarılarını etkilediđi bilinmektedir. alıřma grubundaki đrencilerin ise bořluk tamamlama testlerine oktan semeli testlerden ok daha az dzeyde maruz kalmıř olması da arařtırmanın sınırlılıklarından birini oluřturmaktadır. Arařtırmaya ynelik bir diđer sınırlılık ise alıřma grubunun niceliđi ile ilgilidir. Daha genellenebilir sonular alınması için daha byk katılımcı gruplarıyla alıřılması yerinde olacaktır.

Kaynakça

- Abanami, A. A. (1982). *Readability analysis of the 11th and 12th grade earth science textbooks used in the public schools in Saudi Arabia* (Doktora tezi). Houston Üniversitesi, Houston
- Abraham, R. G. ve Chapelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal*, 76(4), 468-479.
- Akyol, H. (2020). *Türkçe ilk okuma yazma öğretimi*. Ankara: Pegem Akademi.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Arastaman, G. ve Kısa, N. (2020). Geçerlik ve güvenilirlik. N. Cemaloğlu (Ed.), *Bilimsel araştırma teknikleri ve etik içinde* (s. 193-205). Ankara: Pegem Akademi.
- Arıcı, A. F. (2018). *Okuma eğitimi*. Ankara: Pegem Akademi.
- Aşlıoğlu, B. (1993). *Ortaokullarda Türkçe öğretimi* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Ateşman, E. (1997). Türkçede okunabilirliğin ölçülmesi. *Dil Dergisi*, 58, 71-74.
- Bachman, L. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16(1), 61-70.
- Balcı, A. (2016). *Okuma ve anlama eğitimi*. Ankara: Pegem Akademi.
- Başaran, M. (2013). Okuduğunu anlamamanın ölçülmesinde paragraftan anlam kurmaya dayalı çoktan seçmeli sorular. *Eğitim Bilimleri Araştırmaları Dergisi*, 3(2), 107-121.
- Booth, D. (1998). *Guiding the reading process*. Portland Maine: Stenhouse Publishers.
- Bormuth, J. (1963). Cloze as a measure of readability. *Proceedings of the International Reading Association*, 1, 131-134.
- Bormuth, J. R. (1964). Mean word depth as a predictor of comprehension difficulty. *California Journal of Educational Research*, 15, 226-231.
- Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, 1, 79-132.
- Bormuth, J. R. (1967). *Cloze readability procedure*. Kaliforniya: Kaliforniya Üniversitesi.
- Brown, J. D. (1982). *Testing EFL reading comprehension in engineering English* (Doktora tezi). Kaliforniya Üniversitesi, Kaliforniya.
- Brown, J. D. ve Grüter, T. (2020). The same cloze for all occasions?: Using the Brown (1980) cloze test for measuring proficiency in SLA research. *International Review of Applied Linguistics in Language Teaching*, 60, 1-26.
- Büyüköztürk, Ş. (2006). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2023). *Eğitimde bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Cain, K. ve Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708.
- Carlisle, J. ve Rice, M. (2004). Assessment of reading comprehension. A. Stone, E. Silliman, B. Ehren ve K. Apel (Ed.), *Handbook of language and literacy içinde* (s. 521-555). New York, NY: Guilford.
- Carvalho, M. ve Souza, A. (2023). Reading assessment in Brazil between the years 2014-2020: Instruments and skills. *Educação e Pesquisa*, 49, 1-19.
- Coleman, E. B. (1965). *On understanding prose: Some determiners of its complexity*. (NSF Final Report GB-2604). Washington D.C.: National Science Foundation.
- Craker, H. V. (1971). *Clozentropy procedure as an instrument for measuring oral English competencies of first grade children* (Doktora tezi). New Mexico Üniversitesi, New Mexico.
- Creswell, J. W. (2020). *Eğitim araştırmaları* (H. Ekşi, Ed. ve Çev., 5. bs.). İstanbul: EDAM. (Orijinal yayın tarihi 2012)

- Cutting, L. E. ve Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277-299.
- Çelenk, S. (2006). *Etkinlik temelli ilköğretim ve yazma öğretimi*. İstanbul: Morpa Kültür Yayınları.
- Çepni, S., Gökdere, M. ve Küçük, M. (2002). Adaption of the readability formulas into the Turkish science textbooks. *Energy Education Science and Technology*, 10(1), 49-58.
- Çetinkaya, G. (2010). *Türkçe metinlerin okunabilirlik düzeylerinin tanımlanması ve sınıflandırılması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Dağ, N. (2010). Okuma güçlüğü'nün giderilmesinde 3P metodu ile boşluk tamamlama (cloze) tekniğinin kullanımı üzerine bir çalışma. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 11(1), 63-74.
- Darnell, D. K. (1970). Clozentropy: A procedure for testing English language proficiency of foreign students. *Speech Monographs*, 37, 36-46.
- Dubay, W. H. (2007). *Smart language: Readers, readability, and the grading of text*. Impact Information.
- Febriyanti, P. (2017). The correlation between reading comprehension and students' ability in answering cloze test of the seventh grade students at SMPN I Kalipuro Banyuwangi in the 2014/2015 Academic Year. *Language and Art Journal*, 1(2), 36-47.
- Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221-233.
- Froese, V. (1971). Cloze readability versus the Dale-Chall formula. *International Reading Association*, 1, 19-23.
- George, D. ve Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference, 17.0 update*. Boston: Pearson.
- Geske, A. ve Ozola, A. (2008). Factors influencing reading literacy at the primary school level. *Problem of Education In 21st Century*, 6, 71-77.
- Güler, N. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. ve Tatham, R. L. (2013). *Multivariate data analysis*. Harlow: Pearson.
- Harmer, J. (2002). *The practice of English language teaching* (3. bs.). England: Longman.
- Hızarcı, S. H. (2009). *İlköğretim 6. sınıf yeni sosyal bilgiler ders kitaplarının okunabilirlik düzeylerinin incelenmesi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. J. W. Oller Jr. ve K. Perkins (Ed.), *Research in language testing* içinde (s. 121-128). Rowley, MA: Newbury House.
- Huensch, A. (2013). *The perception and production of palatal codas by Korean L2 learners of English* (Doktora tezi). Illinois Üniversitesi, Illinois.
- Irvine, P., Atai P. ve Oller J. W. (1974). Cloze, dictation, and the test of English as a foreign language. *Language Learning*, 24(2), 245-252.
- James, W. (2004). *Special education and social development*. New Delhi: Anmol Publications PVT. LTD.
- Kalyoncu, R. ve Memiş, M. (2024). Türkçe için oluşturulmuş okunabilirlik formüllerinin karşılaştırılması ve tutarlılık sorgusu. *Ana Dili Eğitimi Dergisi*, 12(2), 417-436.
- Kaplan, K. ve Çiftçi, M. (2021). Okuma seviyesinin belirlenmesinde ikame kelime uygulaması. *Türk Dili Araştırmaları Yıllığı-BELLETEN*, 72, 209-236.
- Karakaya, İ. (2014). Bilimsel araştırma yöntemleri. A. Tanrıoğen (Ed.), *Bilimsel araştırma yöntemleri* içinde (s. 57-82). Ankara: Anı Yayıncılık.
- Karatay, H. (2018). *Okuma eğitimi kuram ve uygulama*. Ankara: Pegem Akademi.

- Karatay, H., Bolat, K. K. ve Güngör, H. (2013). Türkçe ders kitaplarındaki metinlerin okunabilirlik ve anlaşılabilirliği. *The Journal Academic Social Science Studies*, 6(6), 603-623.
- Katz, S. ve Lautenschlager, G. J. (1994). Answering reading comprehension items without passages on the SAT-I, the ACT, and the GRE. *Educational Assessment*, 2(4), 295-308.
- Keenan, J. M., Betjemann, R. S. ve Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300.
- Keskin, H. K. ve Akıllı, M. (2013). Fen ve teknoloji ders kitaplarının okunabilirliğinin farklılaştırılmış boşluk doldurma testleri ile ölçülmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 27, 47-66.
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, 10, 62-102.
- Klare, G. R., Simaiko, H. W. ve Stolurow, L. M. (1972). The cloze procedure. A convenient readability test for training materials and translations. *International Review of Applied Psychology*, 21(2), 77-106.
- Kleijn, S., Pander Maat, H. ve Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing*, 36(4), 553-572.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press.
- Köklü, N., Büyüköztürk, Ş. ve Çokluk, Ö. (2007). *Sosyal bilimler için istatistik*. Ankara: Pegem Akademi
- Köse, E. Ö. (2009). Biyoloji 9 ders kitabında hücre ile ilgili metinlerin okunabilirlik düzeyleri. *Journal of Arts and Sciences*, 12, 141-150.
- Kurudayıoğlu, M. ve Karadağ, Ö. (2005). Kelime hazinesi çalışmaları açısından kelime kavramı üzerine bir değerlendirme. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 25(2), 293-307.
- Loewe, E. E. (1983). *The effect of using cloze data for revising instructional materials* (Doktora tezi). Florida Üniversitesi, Florida.
- Lu, G. (2006). *Cloze Test and reading strategies in English language teaching in China* (Yüksek lisans tezi). Western Cape Üniversitesi, Cape Town.
- Manis, M. ve Dawes R. M. (1961). Cloze scores as a function of attitude. *Psychological Reports*, 9, 79-84.
- Mariotti, A. S. ve Homan, S. P. (2001). *Linking reading assesment to intruction: An application worktext for elementary classroom teachers*. New Jersey: Lawrence Erlbaum Associates.
- Mariotti, A. S. ve Homan, S. P. (2009). *Linking reading assesment to instruction: An application worktext for elementary classroom teachers*. New York: Routledge.
- Mertens, D. M. (2015). *Research and evaluation in education and psychology*. Thousand Oaks, CA: Sage.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York: Routledge.
- Oller, J. W. (2006). Close tests of the second language proficiency and what they measure. *Language Learning*, 23(1), 105-118.
- Oller, J. W., Bowen, D. J., Dien, T. T. ve Mason, V. W. (1972). Cloze tests in English, Thai, and Vietnamese: Native and non-native performance. *Language Learning*, 22(1), 1-13.
- Ömeroğlu, E. (2018). Açık uçlu sınavlarla çoktan seçmeli test sınavlarının karşılaştırılması test sınavlarının yazma becerisine etkisi. *International Journal of Language Academy*, 6(26), 548-570.
- Özbay, M. (1997). Test türü imtihanların Türkçe öğretimindeki yeri. *Bilge*, 11, 13-16.
- Özdemir, M. ve Doğruöz, E. (2020). Bilimsel araştırma desenleri. N. Cemaloğlu (Ed.), *Bilimsel araştırma teknikleri ve etik içinde* (s. 65-98). Ankara: Pegem Akademi.
- Özyılmaz, G. (2010). *İlköğretim 7. sınıf öğrencilerine okuduğunu anlama stratejilerinin öğretiminin okuduğunu anlama başarısı üzerine etkisi* (Yayımlanmamış yüksek lisans tezi). Yıldız Teknik Üniversitesi, İstanbul.

- Potter, T. C. (1968). *A taxonomy of cloze research, part I: Readability and reading comprehension* (Rapor No. TR1). Kaliforniya: Southwest Regional Laboratory for Educational Research and Development.
- Retorta, M. S. (2001). Multiple-choice and cloze procedures in reading tests: What do they really measure?. *ESpecialist*, 22(2), 127-154.
- Ruddell, M. R. (2005). *Teaching content reading and writing*. New York: John Wiley & Sons.
- Shahnazari-Dorcheh, M., Roshan, S. ve Hesabi, A. (2012). What is the optimum length of a cloze test?. *International Journal of English Linguistics*, 2(5), 142-153.
- Shiba, S. A. (1957). A study of the measurement of readability- application of the cloze procedure to the Japanese language. *Japanese Journal of Psychology*, 28, 67-73.
- Stubbs, J. B. ve Tucker, G. R. (1974). The cloze test as a measure of ESL proficiency for Arab students. *Modern Language Journal*, 58(5), 239-241.
- Suehiro, A. C. B. ve Santos, A. A. A. (2015). Reading comprehension and phonological awareness: Evidence of validity of their measures. *Estudos de Psicologia*, 32(2), 201-211.
- Sukarni, S. (2021). The use of cloze test to test reading comprehension of non-English department students. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP*, 8(1), 74-82.
- Şahindokuyucu, A. (2006). *A study of cloze and multiple-choice tests for measuring reading comprehension of preparatory students* (Yayımlanmamış yüksek lisans tezi). Bolu Abant İzzet Baysal Üniversitesi, Bolu.
- Taşpınar, M. (2017). *Sosyal bilimlerde SPSS uygulamalı nicel veri analizi*. Ankara: Pegem Akademi.
- Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6(1), 35-39.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30(1), 415-433.
- Taylor, W. L. (1956). Recent developments in the use of cloze procedure. *Journalism Quarterly*, 33(1), 42-48.
- Taylor, W. L. ve Lee, K. W. (1954). KM readers lend hand to science: Cloze method works in written Korean and may serve as a tool for Korean language reform. *Korean Messenger*, 3, 4-5.
- Tekin, H. (1982). *Eğitimde ölçme ve değerlendirme*. Ankara: Daily News Ofset Tesisleri.
- Temizkan, M. (2007). *İlköğretim ikinci kademe Türkçe derslerinde okuma stratejilerinin okuduğunu anlama üzerindeki etkisi* (Yayımlanmamış doktora tezi). Gazi Üniversitesi, Ankara.
- Temizkan, M. ve Sallabaş, M. E. (2011). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 30, 207-220.
- Tremblay, A. ve Garrison, M. D. (2010). Cloze tests: A tool for proficiency assessment in research on L2 French. M. T. Prior, Y. Watanabe ve S. K. Lee (Ed.), *Selected proceedings of the second language research forum 2008* içinde (s. 73-88). MA: Cascadilla Press.
- Tuncer, M. (2020). Nicel araştırma desenleri. B. Oral ve A. Çoban (Ed.), *Kuramdan uygulamaya eğitimde bilimsel araştırma yöntemleri* içinde (s. 205-227). Ankara: Pegem Akademi.
- Tunçer, B. ve Erden, G. (2015). Boşluk doldurma testlerinin ilköğretim 4. sınıf öğrencilerinin okuduğunu anlama düzeylerini belirlemede kullanılabilirliği [Özel sayı]. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 318-324.
- Turgut, F. ve Baykul, Y. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Ulusoy, M. (2009). Boşluk tamamlama testinin okuma düzeyini ve okunabilirliği ölçmede kullanılması. *Türk Eğitim Bilimleri Dergisi*, 7(1), 105-126.

- Uyanık, G. (2011). *İlköđretim 5. sınıf öđrencilerinin boşluk tamamlama tekniđiyle belirlenen okuma seviyeleri ile okuduđunu anlama düzeylerinin karşılaştırılması* (Yayımlanmamıř yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Uyanık, G. (2012). İlköđretim 5. sınıf öđrencilerinin okuma seviyelerinin farklı boşluk tamamlama testleri ile belirlenmesi. *Gazi Üniversitesi Gazi Eđitim Fakóltesi Dergisi*, 32(3), 657-672.
- Üstüner, A. ve řengöl, M. (2004). Çoktan seçmeli test tekniđinin Türkçe öđretimine olumsuz etkileri. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 14(2), 197-208.
- Vacca, R. T. ve Vacca, J. A. L. (2005). *Content area reading: Literacy and learning across the curriculum*. Londra: Pearson Education.
- Vari-Cartier, P. (1980). *The readability and comprehensibility of Spanish prose as determined by the frase graph and the cloze procedure* (Doktora tezi). Rutgers Üniversitesi, New Jersey.
- Wahdaniah, Marbun, R. ve Husin, S. (2013). The use of cloze test in increasing the students' reading comprehension. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 2(1), 1-12.
- Williams, R. S., Ari, O. ve Santamaria, C. N. (2011). Measuring college students' reading comprehension ability using cloze tests. *Journal of Research in Reading*, 34(2), 215-231.
- Yılmaz, H. (1998). *Eđitimde ölçme ve deđerlendirme*. Konya: Mikro Basım-Yayın-Dađıtım.