

Testlerde Hatalı Soru Olması Durumunda Kullanılan Yöntemlerin Puanlamaya ve Sıralamaya Olan Etkisi*

The Effect of the Methods Used in Case of Faulty Items in the Tests on Scoring and Ranking

Ayfer SAYIN¹, Hakan KOĞAR²

¹Gazi Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme ABD, ayfersayin@gazi.edu.tr

²Akdeniz Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme ABD, hakankogar@gmail.com

Makalenin Geliş Tarihi: 18.05.2021

Yayına Kabul Tarihi: 05.08.2021

ÖZ

Bu araştırma kapsamında testlerde hatalı soru olması durumunda soru için kullanılan yöntemlerin bireylerin puanlamalarına ve sıralamalarına olan etkisinin incelenmesi amaçlanmıştır. Bu doğrultuda 2016 TEOG uygulamasına katılan 64.788 öğrencinin Türkçe alt testinde yer alan 20 soruya verdikleri cevaplar üzerinde incelemeler gerçekleştirilmiştir. Araştırma kapsamında öncelikle farklı özelliklere sahip (en düşük ayıricılığa sahip, en çok boş bırakılan, en zor) soru iptal edilmiştir. Ardından da iptal edilen sorulara yönelik dört farklı puanlama yöntemine (herkese doğru kabul etme, diğer sorular üzerinden hesaplama yapma, en az bir doğrusu olan herkese doğru kabul etme, alt testi bazında ağırlıklandırma) dayalı olarak öğrencilere ilişkin dört farklı puan türü (toplam puan, net puanı ve sıralamalar) hesaplanmıştır. Aynı zamanda her bir puan türüne göre öğrenciler sıralanmıştır. Hesaplama sonuçlarında soru iptali durumunda herkese doğru kabul etme ve en az bir doğrusu olan herkese doğru kabul etme yöntemlerinin hem ortalama hem de sıralamada gerçek ortalama ve sıralamadan en uzak değerleri ürettiği belirlenmiştir. Diğer sorular üzerinden hesaplama yapma ve alt testi bazında ağırlıklandırma yöntemlerinin ortalamalar bakımından benzer olduğu, öğrenci sıralamasının değişmesine en dirençli olan yöntemin de alt testi bazında ağırlıklandırma yöntemi olduğu tespit edilmiştir.

Anahtar Sözcükler: Hatalı soru, İptal madde, Puanlama yöntemleri, Test geçerliği, TEOG.

* **Alıntılama:** Sayın, A. ve Koğar, H. (2021). Testlerde hatalı soru olması durumunda kullanılan yöntemlerin puanlamaya ve sıralamaya olan etkisi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 41(3), 1437-1464.

ABSTRACT

Within the scope of this study, it was aimed to examine the effects of the methods used for the question on the scores and rankings of individuals in case of incorrect questions in the tests. Accordingly, 64,788 students who participated in the 2016 TEOG application were examined on the answers they gave to 20 questions in the Turkish subtest. Within the scope of the research, first of all, A question was canceled with different characteristics (lowest separation, most left blank, most difficult question). Then, four different types of scores (total score, net score) for the canceled questions based on four different scoring methods (accepting everyone right, calculating based on other questions, accepting at least one correct one, weighting on the basis of the subtest) score and rankings have been calculated. At the same time, students are ranked according to each score type. In the calculation results, it was determined that in case of a cancellation of the question, the methods of accepting everyone as true and accepting everyone with at least one correct one were actually producing the furthest values in both average and ranking. It has been determined that calculating on other questions and weighting methods on subtest basis are similar in terms of averages, and the method that is most resistant to change in student ranking is the weighting method on the basis of subtest.

Keywords: *Incorrect question, Canceled item, Scoring methods, Test validity, TEOG.*

GİRİŞ

Test sonuçları, bireyler ve grupların devam ettikleri program, devam edecekleri eğitim ve istihdamlarına yönelik temel ve önemli bilgilere ulaşılması anlamında büyük bir öneme sahiptir. Bu nedenle de testlerin yanlış kullanılması, teste dayalı kararlardan olumsuz etkilenir (Çobanoğlu Aktan, Aksu ve Eser, 2018). Test geliştirme; bireylerin bilgi, beceri, yetenek, ilgi alanları, tutumları ve diğer karakteristik özelliklerinin bir plan doğrultusunda maddelerin geliştirilmesi ve test formuna dönüştürülmesi süreçlerini kapsamaktadır. Aynı zamanda testin uygulanması için uygun koşulların belirlenmesi, test performansının puanlanma yönteminin belirlenmesi ve testi alan kişilere puanların nasıl duyurulacağı yöntemlerinin tespit edilmesini de içermektedir (AERA, APA ve NCME, 1985). Test geliştirme; test için genel planlamanın yapılması, test alanının ve testin uygulanacağı hedef kitlenin tanımlanması, içerik belirlenmesi, maddelerin geliştirilmesi, test düzeninin oluşturularak testin oluşturulması, uygulanması, puanlanması, kesme noktalarının belirlenmesi, test sonuçlarının raporlanması ve testin güvenli bir şekilde saklanması aşamaları kapsamında gerçekleştirilmelidir (Lane, Raymond, Haladayna ve Downing, 2016). Ancak Türkiye’de 2010-2012 yılları arasında yayınlanan tezlerde geliştirilen testlerin %68’inin test geliştirme aşamalarının tamamına uymadığı belirlenmiştir (Özdemir ve Yandı, 2012). Test geliştirme süreci incelendiğinde test geliştirmenin sadece madde yazımı ve uygulanmasını değil, puanlamayı da kapsadığı görülmektedir.

Test sonuçlarının puanlanması sürecinde, ölçme kuralları önceden belirlense de, test uygulandıktan sonra hatalı sorularla karşılaşabilmektedir. SAT, NAEP, TIMSS gibi geniş ölçekli sınavlarda uzmanlarının görüşlerinin yanı sıra uygulama öncesinde aday maddelerin denemesini yaparak uygulama için madde seçimi yapılabilmektedir (ETS, 1994; IEA, 2020). Türkiye’de ise ÖSYM, MEB gibi geniş ölçekli test uygulayan kurumlar, güvenlik sorunu nedeniyle aday maddelerine yönelik uzman görüşüne dayalı tercihler yapmaktadır. Bu nedenle de testler uygulandıktan sonra hatalı sorularla karşılaşabilmektedir. Hatalı sorularla karşılaşıldığı durumlarda da sorular iptal

edilmekte ve puanlama yönergesinde değişiklikler meydana gelmektedir. Söz konusu durumda puan hesaplamasına esas olacak yöntem burada ön plana çıkmaktadır. Çünkü testte hatalı soru bulunması durumunda sınav esnasında karışacak hataları kestirmek güçtür. Bireylerin hatalı soruda ne kadar vakit harcadığı, soruyu cevaplandıramadığı için psikolojik düzeyindeki değişimleri (moralinin bozulması, kaygı düzeyinin artması vb.) kestirmek zordur. Bununla birlikte testteki hatalı soruyu hiç görmeyen, hatayı fark etmeyen ya da soru üzerinde fazla vakit/çaba harcamadan diğer sorulara geçen kişiler de bulunmaktadır. Bu durumda uygulama sonrasında hatalı sorunun belirlenip iptal edilmesinin ardından puanlama için kullanılacak yöntemin ne olacağı önem kazanmaktadır.

Alanyazında hatalı bir soruyla, test uygulaması öncesi ve sonrasında karşılaşıldığı durumlarda ne yapılması gerektiği ile ilgili çeşitli görüşler yer almaktadır. Varma'ya (2006) göre, özellikle düşük madde ayırt edicilik katsayısı nedeniyle düzeltilemeyecek durumda olan hatalı maddelerin puanlamadan ve testten tamamen çıkartılmasını önermektedir. Ludlow (2001) ise özellikle negatif madde ayırt edicilik katsayısına sahip olan maddelerin test geliştirici tarafından testten çıkartılması gerektiğini belirtmiştir. Ayrıca özellikle yüksek riskli (high-stakes) sınavlarda, testin en az bir kez alan denemesine tabi tutulması, problemleri maddelerin açığa çıkarılması ve bu maddelerin düzeltilmesi ya da tamamen testten çıkartılması, sonrasında ise esas uygulamaya geçilmesini önermektedir. Yüksek riskli sınavlarda hatalı bir sorunun yer almasının psikometrik açıdan savunulabilir bir yanı olmadığına dikkat çekmektedir.

Ülkemizde birçok farklı amaçla test uygulanmaktadır ve bu testler içerisinde en çok öne çıkanlar kademeler arası geçiş sisteminde uygulanan testlerdir. Türkiye'de liselere giriş sürecinde Milli Eğitim Bakanlığı (MEB), üniversitelere giriş sürecinde de Ölçme, Seçme ve Yerleştirme (ÖSYM) tarafından geniş ölçekli testler uygulanmaktadır. Bu testlerde de hatalı sorunun iptal edilmesi sonrasında puanlamaya esas olacak yöntemle ilgili değişiklikler bulunmaktadır. Örneğin öğretmen adaylarının katıldığı ve puanlarına dayalı olarak sıralandığı KPSS uygulamasında hatalı soru olması durumunda 2016 yılına kadar, sorunun iptal edildiği ya da geçersiz sayıldığı durumda ilgili testte en az bir

soruyu işaretleyen herkese sorunun doğru kabul edilmesi öngörülüyordu: “Sınavdan sonra yapılan analizlerde bir testte sorulardan herhangi biri, biçimsel veya bilimsel bir nedenle ÖSYM tarafından ‘geçersiz’ sayıldığı takdirde, bu soru, ilgili testi işaretleyen tüm adaylar için doğru cevaplanmış kabul edilecektir.” (ÖSYM, 2016, 17). Sarıtaş ve Bağcı (2017) ise ÖSYM’ye açılan davalarda soru iptal edildikten sonra sorunun iptal edildiği testte en az bir doğrusu bulunan tüm adaylar için sorunun doğru kabul edildiğine dikkat çekmektedir. Ancak 2017 yılında bu uygulama değiştirilerek iptal edilen sorunun puanının diğer sorulara verilerek hesaplama yapılmasına karar verilmiştir: “Sınavdan sonra ÖSYM Başkanlığı veya yargı mercileri tarafından iptaline karar verilen sorular değerlendirme dışı bırakılarak geçerli soruların puan değerinin yeniden saptanması suretiyle puanlama yapılacaktır.” (ÖSYM, 2017, 20). Başka bir anlatımla testte yer alan bir sorunun iptal edilmesi durumunda puanlama için zaman içinde farklı yöntemler denendiği görülmektedir.

MEB tarafından liselere giriş sürecinde uygulanan merkezi sınav uygulamasında 2015-2016 ve 2016-2017 eğitim-öğretim dönemlerinde soru iptaline yönelik “Soru iptali olması durumunda ilgili soru o ders sınavına katılan bütün öğrenciler için doğru kabul edilerek değerlendirilecektir.” uygulamalarına yer verilmiştir. 2018 Merkezi Sınav Uygulama kılavuzunda ise “Sınavlarda iptaline karar verilen sorular değerlendirme dışı bırakılarak geçerli soruların puan değerinin yeniden saptanması suretiyle puanlama yapılır.” açıklaması yer almaktadır (MEB, 2018). Söz konusu uygulamalardan da anlaşıldığı gibi soruların iptal edildikten sonra puanlamanın nasıl gerçekleştirilmesi gerektiğine yönelik henüz kesin bir yöntem bulunmamaktadır. Ancak özellikle yerleştirme testlerinde puanlama sürecinde bir puanın bile adaylar için önem taşıdığı düşünüldüğünde iptal sorularının puanlamasının da önemli olduğu ortaya çıkmaktadır.

2014 Kasım ayında yapılan TEOG’da fen testinde 1, matematik testinde ise 2 soru iptal edilmiştir. 28-29 Kasım 2013’te yapılan TEOG sınavlarındaki soruların tartışılması ve doğurduğu sorunlar, puanların hesaplanacağı Temmuz 2014’e kadar devam etmiştir. Ayrıca, idarî mahkeme daha önce iptal edilen 3 soruya ek olarak 4 sorunun daha iptal edilmesine karar vermiştir. Kamuoyunda sıklıkla yer bulan TEOG ile ilgili tartışmalar

tercih ve yerleřtirme s¼recinde de devam etmiřtir. ¼ğrencilerin TEOG'da yaptıkları yanıřların puan sistemine dâhil edilmemesi ve puanların sadece ¼ğrencilerin doęru yanıřları ¼zerinden hesaplanması, aynı puandaki ¼ğrenci sayısında oldukça y¼ksek bir artış yařanmasına sebep olmuřtur (TEDMEM, 2014).

Bu alıřma ile ¼lme sonularının geerlięi ve g¼venirlięi ¼zerinde doęrudan etkisi bulunan hatalı sorunun testte bulunması durumunda farklı y¼ntemlere dayalı geekleřtirilen hesaplama sonuları incelenmiř olacak, bu sayede y¼ntemsel deęiřiklięin pratikte oluřturduęu sonular g¼r¼lebilecektir.

Arařtırmanın amacı

Bu alıřmada testlerde hatalı soru olduęu durumlarda hesaplama iin kullanılan y¼ntemlerin bireylerin ortalamalarında ve sıralamalarında nasıl bir deęiřim g¼sterdięinin incelenmesi amalanmıřtır. Arařtırma kapsamında ayrıca hatalı soru olduęu durumlarda puanlamaya y¼nelik yeni bir hesaplama y¼ntemi ¼nerilerek bu baęlamda da bireylerin sıralamalarındaki farklılıęın incelenmesi amalanmıřtır.

Y¼NTEM

Arařtırma deseni

Bu alıřma kapsamında testlerde hatalı soru olması durumunda kullanılan y¼ntemlerin bireylerin sıralamalarına olan etkisi incelenmiř; aynı zamanda yeni bir hesaplama y¼ntemi ¼nerilmiřtir. Bu doęrultuda arařtırmanın temel bir arařtırma nitelięinde olduęu s¼ylenebilir.

Arařtırma grubu

Bu arařtırma, 2016 TEOG uygulamasına katılan 64.788 ¼ğrenci ile y¼r¼t¼lm¼řt¼r. 2016 yılında TEOG uygulamasının T¼rke testine 1,131,588 ¼ğrenci katılmıřtır. Arařtırma kapsamında iptal edilen soruların ¼ğrenciler tarafından aynı sırada g¼r¼lmesi amalandıęından TEOG uygulamasında A kitapıęını alan 64,788 ¼ğrenci alıřmaya

dâhil edilmiştir. Öğrencilerin verileri, araştırmacılar tarafından doğrudan toplanmamış olup verilerin kullanım izni MEB'den alınmıştır.

Veri Toplama Araçları

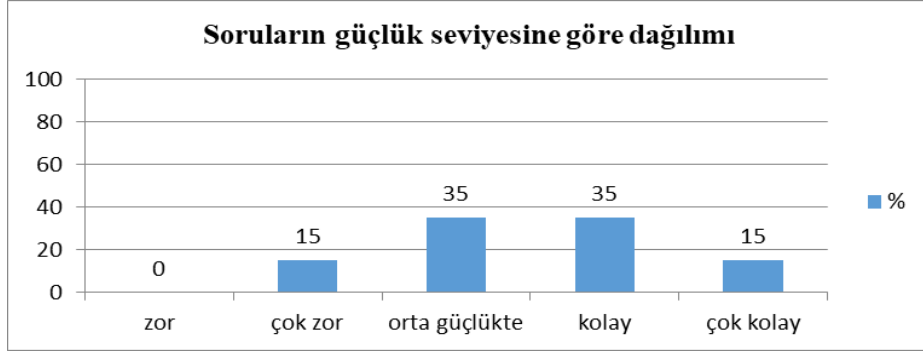
Araştırmada tanıma ve yerleştirme amacıyla kullanılan testlerdeki hatalı sorunun öğrencilerin sıralamalarını ne düzeyde değiştirdiğinin incelenmesi amaçlanmaktadır. Bu doğrultuda Türkiye'de liselere geçiş amacıyla kullanılan TEOG uygulaması içerisinde yer alan Türkçe testi kullanılmıştır. Türkçe testinde toplam 20 madde yer almaktadır. Testte yer alan maddeler; sözcükte anlam, cümlede anlam, parçada anlam, anlatım, dil bilgisi, yazım ve noktalama konularından oluşmaktadır. Bu konuların belirlenme sürecinde Türkçe eğitimi alanında görev yapan beş öğretim üyesinin görüşleri alınmıştır. Uygulamaya katılan öğrenciler içerisinde 64,788 öğrencinin cevapları doğrultusunda madde istatistikleri hesaplanmış, sonuçlar Tablo 1'de gösterilmiştir.

Tablo 1. Madde İstatistikleri

Maddeler	Güçlük indeksi	Nokta çift serili korelasyon	Ayrıcalık (discriminat) indeksi	Konu dağılımı
m1	0.62	0.52	0.63	Sözcükte anlam
m2	0.61	0.57	0.70	Sözcükte anlam
m3	0.38	0.32	0.39	Cümlede anlam
m4	0.75	0.50	0.54	Cümlede anlam
m5	0.36	0.46	0.56	Parçada anlam
m6	0.81	0.48	0.46	Parçada anlam
m7	0.67	0.61	0.72	Parçada anlam
m8	0.68	0.46	0.54	Parçada anlam
m9	0.87	0.44	0.36	Parçada anlam
m10	0.53	0.52	0.65	Parçada anlam
m11	0.46	0.60	0.76	Anlatım
m12	0.49	0.34	0.42	Öğretici tür
m13	0.76	0.48	0.50	Parçada anlam
m14	0.68	0.56	0.65	Anlatım
m15	0.60	0.52	0.64	Anlatım
m16	0.54	0.53	0.67	Dil bilgisi
m17	0.60	0.58	0.72	Dil bilgisi
m18	0.35	0.47	0.58	Dil bilgisi
m19	0.45	0.38	0.47	Yazım
m20	0.81	0.51	0.53	Noktalama

Tablo 1'de yer alan bilgiler incelendiğinde testte yer alan maddelerin güçlük indeksleri 0.35 ile 0.87 arasında değişiklik göstermektedir.

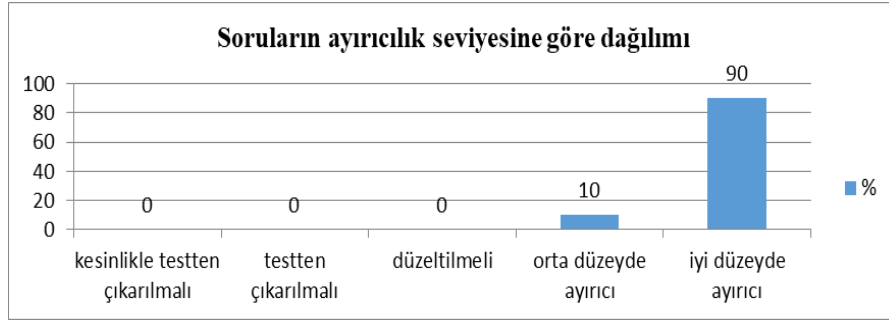
Testte bulunan soruların güçlük indeksleri 0.00-0.20 çok zor; 0.21-0.40 zor; 0.41-0.60 orta güçlükte; 0.61-0.80 kolay; 0.81-1.00 çok kolay olacak şekilde kategorileştirildiğinde (Güler, 2011) soruların dağılımı Şekil 1’de gösterilmiştir.



Şekil 1. Testte Yer Alan Soruların Güçlük Seviyesine Göre Dağılımı

Şekil 1’de de görüldüğü gibi testte bulunan sorulardan hiçbiri öğrencilere çok zor gelmemiştir. Soruların 3’ü öğrencilere zor, 7’si orta güçlükte, 7’si kolay ve 3’ü de çok kolay gelmiştir.

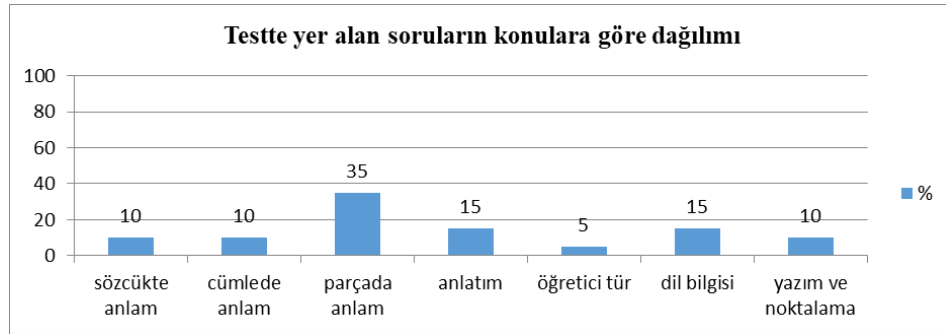
Soruların ayırt edicilik indeksleri incelendiğinde nokta çift serili korelasyon katsayılarının 0.32 ile 0.61 arasında farklılık göstermektedir. Alt-üst grupta bulunan öğrencilerin cevaplarına dayalı olarak hesaplanan ayırtıcılık indeksinin 0.36 ile 0.76 arasında değişiklik gösterdiği tespit edilmiştir. Negatif ayırtıcılık kesinlikle testten çıkarılmalıdır; 0.00-0.19 testten çıkarılmalıdır; 0.20-0.29 düzeltilmeli; 0.30-0.39 orta düzeyde ayırtıcı; 0.40 ve üzeri iyi düzeyde ayırtıcı maddeler olacak şekilde bir kategorileştirme yapılmıştır. Bu doğrultuda testte yer alan soruların dağılımı Şekil 2’de gösterilmiştir.



Şekil 2. Testte Yer Alan Soruların Ayırıcılık Seviyesine Göre Dağılımı

Şekil 2’de yer alan bilgiler incelendiğinde testte bulunan soruların 2’sinin orta düzeyde, 18’inin ise çok iyi düzeyde ayırıcı maddeler olduğu tespit edilmiştir.

Testteki soruların konu dağılımları incelendiğinde de soruların konulara göre dağılımı Şekil 3’te gösterilmiştir.



Şekil 3. Testte Yer Alan Soruların Konularına Göre Dağılımı

Şekil 3 incelendiğinde testte bulunan soruların 2’sinin sözcükte anlam, 2’sinin cümlede anlam, 7’sinin parçada anlam, 3’ünün anlatım, 1’inin öğretici tür, 3’ünün dil bilgisi, 2’sinin de yazım ve noktalama konularına ait olduğu belirlenmiştir.

TEOG uygulamasına katılan öğrenciler içerisinde araştırmaya dâhil edilen 64,788 öğrencinin cevapları doğrultusunda hesaplanan test istatistikleri Tablo 2’de gösterilmiştir.

Tablo 2. Test İstatistikleri

Test istatistikleri	Değerler
Testten alınabilecek en yüksek puan (madde sayısı)	20
Testi alan öğrenci sayısı	64,788
Testten alınan en düşük puan	0.00 (%0,1)
Testten alınan en yüksek puan	20.00 (%3)
\bar{X} (Ortalama)	12.01
SD (Standart sapma)	4.56
Ortalama güçlük (p _{ij})	0.60
Ortalama ayıricılık (r _{ij})	0.57

Tablo 2’de yer alan değerler incelendiğinde testte yer alan sorulardan tamamına yanlış cevap veren ve/veya boş bırakan öğrenciler (%0.1; n=38) olduğu gibi tamamını doğru cevaplandıran öğrenciler de (%3; n=1941) bulunduğu görülmektedir. Testi alan çocuklar soruların ortalama %60’ını doğru cevaplandırmışlardır. Testteki soruların ortalama ayırt ediciliklerinin yüksek düzeyde olduğu ($r_{ij}=0.57$) tespit edilmiştir. TEOG 2016 uygulamasına katılan öğrencilerin (N=1,131,588) Türkçe alt testinden almış oldukları puanların 0-100 arasında değişiklik gösterdiği, ortalamasının 60 olduğu raporlanmıştır (MEB, 2016). Bu doğrultuda TEOG uygulamasında A kitapçığını alan öğrencilerin sonuçlarının tüm öğrenci sonuçlarıyla benzerlik gösterdiği görülmektedir.

Testin güvenilirliğine ilişkin hesaplanan değerler Tablo 3’te gösterilmiştir.

Tablo 3. Güvenirlik Analizi Sonuçları

Güvenirlik katsayısı	Değerler
KR-20	0.833
İki yarı güvenirlik katsayısı	0.823
Guttman güvenirlik katsayısı	0.822

Tablo 3 incelendiğinde test maddelerine verilen cevapların iç tutarlılığının, başka bir anlatımla güvenirliklerinin, yüksek düzeyde olduğu saptanmıştır.

Etik Kurallara Uygunluk

Araştırma kapsamında ülke genelinde merkezî bir sınav sistemine katılan öğrencilerin verisi olan 2016 TEOG verileri kullanılmıştır. Veriler, MEB tarafından toplanmış olup 06.02.2018 tarihli 57750415-480.01-E.2509280 sayılı yazı kapsamında verilerin bilimsel amaçlı kullanılması izni alınarak veriler temin edilmiştir. Araştırmacılar,

verileri kendileri toplamamışlar, mevcut veriler üzerinde incelemeler gerçekleştirmişlerdir.

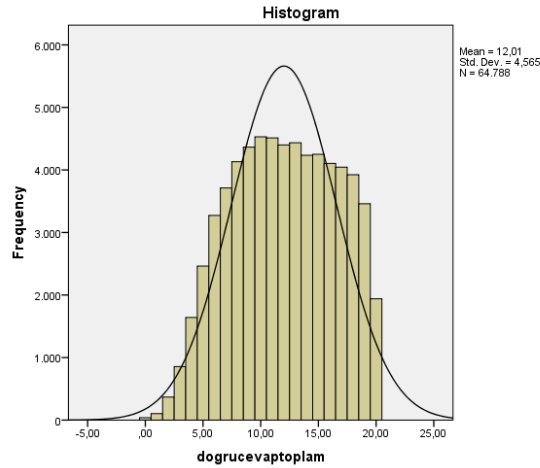
Verilerin çözümlenmesi ve yorumlanması

Bu aşamada öncelikle veri setinin incelemesi gerçekleştirilmiş, ardından verileri düzenlenmiş ve veri analizi-yorumlama süreci gerçekleştirilmiştir.

Veri Setinin İncelenmesi

Verilerin analizi gerçekleştirilmeden önce veri seti incelenmiştir. Öğrencilerin 20 madde üzerinden doğru cevapları 1; yanlış cevapları 0 olacak şekilde puanlaması yapılmıştır. Maddeyi boş bırakan öğrencilerin cevapları için 0 ya da 1 olacak şekilde bir kodlama işlemi gerçekleştirilmemiştir.

Öğrencilerin Türkçe alt testindeki doğru sayılarının dağılımının incelenmesi amacıyla öncelikle çarpıklık ve basıklık katsayısı hesaplanmıştır. Hesaplama sonucunda çarpıklık katsayısı -0.094; basıklık katsayısı -0.950 olarak bulunmuştur. Öğrencilerin doğru cevap dağılımlarına yönelik oluşturulan histogram grafiği de Şekil 4'te gösterilmiştir.



Şekil 4. Öğrencilerin Türkçe Alt Testindeki Doğru Cevaplarının Dağılımı

Veri Setinin Düzenlenmesi

Verilerin analizi sürecinde öncelikle hiçbir sorunun hatalı olmadığı durumdaki "ölçüt" değerler hesaplanmıştır.

Ölçüt Puanlar

- i. *Toplam doğru cevap:* Öğrencilerin doğru cevap verdikleri toplam sayı hesaplanmıştır.
- ii. *Net:* Düzeltme formülü uygulanarak (3 yanlışın 1 doğruyu götürmesi) öğrencilerin netleri hesaplanmıştır.
- iii. *Sıralama:* Öğrencilerin doğru cevap sayılarına göre sıraları hesaplanmıştır.

Soru İptal Koşulları

- i. En düşük ayıcılığa sahip soru
- ii. En çok boş bırakılan soru
- iii. En zor soru
- iv. En düşük ayırıcılık, en çok boş bırakılan ve en zor sorunun birlikte iptali

Hesaplamalar

- İptal edilen soru için:
 - i. Herkese doğru kabul etme, (2016 yılına kadar merkezi sınavların kabulü)
 - ii. Diğer sorular üzerinden hesaplama yapma, (2017 yılından itibaren merkezi sınavlarında kullanma)
 - iii. En az bir doğrusu olan herkese doğru kabul etme, (2016 yılına kadar merkezi sınavların kabulü)
 - iv. Alt testi bazında ağırlıklandırma (araştırmacıların önerisi)

dört yöntemle dayalı olarak iptal edilen soru için değer atanarak hesaplamalar yapılmıştır.

Verilerin Çözümlemesi

Araştırmada ölçüt puanlar ile koşullardan elde edilen puanlar arasında anlamlı bir farklılığın olup olmadığının belirlenmesi amacıyla öncelikle betimsel istatistikler hesaplanmış, farklılıklara yönelik eta-kare değerleri belirlenmiştir. Öğrenci sıralarının karşılaştırılması amacıyla da puanı artan, azalan ve aynı kalan öğrenci sayıları ve yüzdeleri hesaplanmıştır. Araştırmada gerçekleştirilen işlemler Tablo 4'te özetlenmiştir.

Tablo 4. Verilerin Çözümleme Aşamaları

İptal nedeni	İptal soru sayısı	İptal sonrası hesaplama yöntemi	Hesaplama
En düşük ayrırcılıktaki soru	1	Herkese doğru kabul etme Diğer sorular üzerinden hesaplama yapma En az bir doğrusu olan herkese doğru kabul etme Alt testi bazında ağırlıklandırma	Toplam doğru cevap Toplam net sayısı Sıralama
En çok boş bırakılan soru	1	Herkese doğru kabul etme Diğer sorular üzerinden hesaplama yapma En az bir doğrusu olan herkese doğru kabul etme Alt testi bazında ağırlıklandırma	Toplam doğru cevap Toplam net sayısı Sıralama
En zor soru	1	Herkese doğru kabul etme Diğer sorular üzerinden hesaplama yapma En az bir doğrusu olan herkese doğru kabul etme Alt testi bazında ağırlıklandırma	Toplam doğru cevap Toplam net sayısı Sıralama
Tüm koşulları sağlayanlar	3	Herkese doğru kabul etme Diğer sorular üzerinden hesaplama yapma En az bir doğrusu olan herkese doğru kabul etme Alt testi bazında ağırlıklandırma	Toplam doğru cevap Toplam net sayısı Sıralama

BULGULAR

Çeşitli İptal Nedenleri ve Tekrar Puan Hesaplama Yöntemlerine Dayalı Gerçekleştirilen Hesaplamalarda Mutlak Puanların Değişimi Nasıldır?

Araştırma kapsamında öncelikle herhangi bir soru iptal edilmeden önce öğrencilerin ham puanlarına yönelik betimsel istatistikler hesaplanmış, bulgular Tablo 5'te gösterilmiştir.

Tablo 5. Öğrencilerin Tüm Testten Aldıkları Puanlara (Ham Puanlara) İlişkin Hesaplanan Betimsel İstatistikler

Puanlar	İstatistikler	Değerler
Doğru cevap	Ortalama	12.01
	Ortanca	12.00
	Minimum	0.00
	Maksimum	20.00
Net	Ortalama	9.39
	Ortanca	9.70
	Minimum	-6.70
	Maksimum	20.00

Tablo 5'te yer alan bilgiler incelendiğinde testteki herhangi bir soru iptal edilmediği durumda öğrencilerin doğru cevap sayıları 0.00 ile 20.00 arasında değişiklik göstermektedir. Öğrencilerin doğru cevaplarının ortalaması 12.01; ortancası da 12.00'dir. Herhangi bir sorunun iptal edilmediği testte şans başarısı arıtma yapıldığında öğrenci netlerinin -6.70 ile 20.00 arasında değişiklik gösterdiği, net doğru cevaplarının ortalamasının 9.39; ortancasının da 9.70 olarak hesaplandığı belirlenmiştir.

Araştırmada dört farklı koşula bağlı olarak (en düşük ayıricılıkta madde, en çok boş bırakılan madde, en zor madde) maddeler iptal edilerek dört farklı puanlama yöntemine (herkese doğru kabul etme, diğer sorular üzerinden hesaplama yapma, en az bir doğrusu olan herkese doğru kabul etme, alt testi bazında ağırlıklandırma) dayalı olarak öğrencilere ilişkin doğru cevap ve net olmak üzere mutlak puanları hesaplanmıştır. Öğrencilerin mutlak puanlarına ilişkin betimsel istatistikler Tablo 6'da gösterilmiştir.

Tablo 6. Çeşitli İptal Nedenleri ve Tekrar Puan Hesaplama Yöntemleri Sonucunda Elde Edilen Betimsel İstatistikler

Yöntem	Puan	Betimsel Değer	E.D.A.*	E.B.B.*	E.Z.S.*	T.K.S.*
Herkes doğru kabul etme	D.C.	Ortalama	12.63	12.66	12.56	13.83
		Ortanca	13.00	13.00	13.00	14.00
		Minimum	1.00	1.00	1.00	3.00
	Net	Maksimum	20.00	20.00	20.00	20.00
		Ortalama	10.19	10.23	10.10	11.79
		Ortanca	10.67	10.67	10.67	12.00
Diğer sorular üzerinden hesaplama yapma	D.C.	Minimum	-5.33	-5.33	-5.33	-2.67
		Maksimum	20.00	20.00	20.00	20.00
		Ortalama	12.24	12.28	12.17	12.75
	Net	Ortanca	12.63	12.63	12.63	12.94
		Minimum	0.00	0.00	0.00	0.00
		Maksimum	20.00	20.00	20.00	20.00
En az bir doğrusu olan herkese doğru kabul etme	D.C.	Ortalama	9.80	9.84	9.70	10.70
		Ortanca	10.30	10.30	10.30	10.94
		Minimum	-6.33	-6.33	-6.33	-5.67
	Net	Maksimum	20.00	20.00	20.00	20.00
		Ortalama	12.64	12.66	12.56	13.83
		Ortanca	13.00	13.00	13.00	14.00
Alt test bazında ağırlıklandırma	D.C.	Minimum	0.00	0.00	0.00	0.00
		Maksimum	20.00	20.00	20.00	20.00
		Ortalama	10.18	10.23	10.09	11.79
	Net	Ortanca	10.67	10.67	10.67	12.00
		Minimum	-6.33	-6.33	-6.33	-2.67
		Maksimum	20.00	20.00	20.00	20.00
Alt test bazında ağırlıklandırma	D.C.	Ortalama	12.38	12.24	12.38	12.96
		Ortanca	13.00	12.50	13.00	13.50
		Minimum	0.00	0.00	0.00	0.00
	Net	Maksimum	20.00	20.00	20.00	20.00
		Ortalama	9.85	9.80	9.90	10.92
		Ortanca	10.67	10.17	10.67	11.50
Net	Minimum	-6.33	-6.33	-6.33	-5.67	
	Maksimum	20.00	20.00	20.00	20.00	

* E.D.A.: En düşük ayıricılıktaki maddeyi çıkarma, E.B.B.: En çok boş bırakılan soruyu çıkarma, E.Z.S.: En zor soruyu çıkarma, T.K.S.: Tüm koşulları sağlayan maddeleri çıkarma

Tablo 6’da görüldüğü gibi *iptal edilen sorunun herkese doğru kabul edildiği durum için* inceleme yapıldığında; sorunun hatalı olmaması durumunda alınacak ortalama puandan daha yüksek bir ortalama hesaplanmaktadır. İptal edilen soru sayısı arttıkça da ortalamadaki benzer artış söz konusu olmaktadır. Benzer durum şans başarısı artırma sonucunda hesaplanan net puanlar için de geçerlidir.

Hiçbir sorunun hatalı olmadığı ve iptal edilmediği durumdaki ortalama değerin 12.01 olarak hesaplandığı göz önüne alındığında; sorunun iptal edilerek diğer sorular ağırlıklandırılarak hesaplama yapıldığında ortalama değerlerin olması gerekenden yüksek hesaplandığı görülmektedir. Bu değer, koşullara bağlı farklılık gösterse de iptal edilen soru sayısının artması durumunda ortalama değerin de olması gerekenden daha yüksek hesaplama eğilimi gösterdiği görülmektedir. Benzer durum şans başarısı arıtma sonucunda hesaplanan net puanlar için de geçerlidir.

En az bir doğrusu olan herkese sorunun doğru kabul edildiği durum için inceleme yapıldığında; hiçbir sorunun hatalı olmadığı ve iptal edilmediği durumdaki ortalama değerin 12.01 olarak hesaplandığı göz önüne alındığında; diğer yöntemlerde olduğu gibi değerlerin olması gerekenden yüksek hesaplandığı belirlenmiştir. Bu değer, koşullara bağlı farklılık gösterse de iptal edilen soru sayısının artması durumunda ortalama değerin de olması gerekenden daha yüksek hesaplama eğilimi gösterdiği görülmektedir. Benzer durum şans başarısı arıtma sonucunda hesaplanan net puanlar için de geçerlidir.

Alt test bazında ağırlıklandırma yapılan durum için inceleme yapıldığında; hiçbir sorunun hatalı olmadığı ve iptal edilmediği durumdaki ortalama değerin 12.96 olarak hesaplandığı göz önüne alındığında; diğer yöntemlerde olduğu gibi değerlerin olması gerekenden yüksek hesaplandığı belirlenmiştir. Bu değer, koşullara bağlı farklılık gösterse de iptal edilen soru sayısının artması durumunda ortalama değerin de olması gerekenden daha yüksek hesaplama eğilimi gösterdiği görülmektedir. Benzer durum şans başarısı arıtma sonucunda hesaplanan net puanlar için de geçerlidir.

İptal soru ya da soruların olması durumunda farklı puan hesaplama yöntemlerine dayalı olarak hesaplanan doğru cevap ve net puanları ile ham puanlar arasındaki fark olup olmadığının belirlenmesi amacıyla ilişkili ölçümlerde t testi yapılmış ve sonuçlar Tablo 7'de gösterilmiştir.

Tablo 7. Çeşitli İptal Nedenleri ve Tekrar Puan Hesaplama Yöntemleri Sonucunda Hesaplanan Doğru Cevap Sayılarının Ham Puanlar Arasında Anlamlı Bir Farklılık Gösterip Göstermediğine Yönelik Hesaplanan t Testi Sonuçları

Puan	E.D.A.		E.B.B.		E.Z.S.		T.K.S.		
	O.F.	E.K.	O.F.	E.K.	O.F.	E.K.	O.F.	E.K.	
D.C.	A-G	0.619*	0.62	0.651*	0.65	0.552*	0.55	1.821*	0.79
	B-G	0.231*	0.21	0.265*	0.25	0.160*	0.10	0.734*	0.41
	C-G	0.618*	0.62	0.650*	0.65	0.551*	0.55	1.820*	0.79
	D-G	0.365*	0.27	0.221*	0.21	0.361*	0.27	0.948*	0.43
Net	A-G	0.797*	0.60	0.840*	0.62	0.707*	0.53	2.399*	0.79
	B-G	0.409*	0.29	0.454*	0.30	0.316*	0.24	1.311*	0.58
	C-G	0.797*	0.60	0.840*	0.62	0.707*	0.53	2.399*	0.79
	D-G	0.459*	0.30	0.411*	0.29	0.517*	0.40	1.526*	0.60

* p<0.05 ** O.F.: Ortalama fark, E.K.: Eta-kare

A: Herkese doğru kabul etme; B: Diğer sorular üzerinden hesaplama yapma; C: En az bir doğrusu olan herkese doğru kabul etme; D: Alt test bazında ağırlıklandırma

* E.D.A.: En düşük ayırıcılıktaki maddeyi çıkarma, E.B.B.: En çok boş bırakılan soruyu çıkarma, E.Z.S.: En zor soruyu çıkarma, T.K.S.: Tüm koşulları sağlayan maddeleri çıkarma

Tablo 7 incelendiğinde, en düşük ayırıcılıkta maddenin iptal edilmesi durumunda gerçekleştirilen hesaplamalarda gerçek puan ile hesaplanan puan ortalaması farkının en yüksek olduğu yöntemin “herkese doğru kabul etme” ve “en az bir doğrusu olan herkese doğru kabul etme” olduğu görülmektedir. Gerçek puana en yakın değerlerin ise “diğer sorular üzerinden hesaplama yapma” ve “alt test bazında ağırlıklandırma” yöntemlerine dayalı hesaplamalarda olduğu tespit edilmiştir. En çok boş bırakılan sorunun iptal edilmesi durumunda “alt test bazında ağırlıklandırma” yönteminin en iyi sonuçları verdiği; diğer iptal durumlarında ise en iyi sonucu “diğer sorular üzerinden hesaplama yapma” yönteminin verdiği saptanmıştır.

Çeşitli İptal Nedenleri ve Tekrar Puan Hesaplama Yöntemlerine Dayalı Gerçekleştirilen Hesaplamalarda Sıralamaların Değişimi Nasıldır?

Araştırmada TEOG Türkçe alt testinde bulunan 20 soruyu cevaplandıran 64,788 öğrencinin cevapları incelenmiştir. Öğrencilerin doğru cevap sayılarına göre dağılımları ve sıraları Tablo 8’de gösterilmiştir.

Tablo 8. Öğrencilerin Doğru Cevap Sayısına Göre Dağılımları ve Sıralamaları

Doğru cevap	f	%	Sıra
20.0	1941	3.0	1
19.0	3459	5.3	2
18.0	3923	6.1	3
17.0	4045	6.2	4
16.0	4105	6.3	5
15.0	4250	6.6	6
14.0	4234	6.5	7
13.0	4434	6.8	8
12.0	4400	6.8	9
11.0	4513	7.0	10
10.0	4531	7.0	11
9.0	4365	6.7	12
8.0	4132	6.4	13
7.0	3712	5.7	14
6.0	3274	5.1	15
5.0	2462	3.8	16
4.0	1641	2.5	17
3.0	856	1.3	18
2.0	369	.6	19
1.0	104	.2	20
0.0	38	.1	21

Tablo 8’de yer alan bilgiler incelendiğinde öğrencilerin doğru cevapları üzerinden 21 sıraları bulunduğu görülmektedir.

Araştırma kapsamında dört farklı koşulda teste yer alan madde(ler) iptal edilmiş ve dört farklı puanlama yöntemine dayalı olarak hesaplamalar gerçekleştirilmiştir. Hesaplamalar sonucunda öğrencilerin sıralamalarındaki değişimler Tablo 9’da gösterilmiştir. Tablo 9’da N: sırası azalan öğrencileri, P: sırası artan öğrencileri, E: sırası değişmeyen öğrencileri göstermektedir.

Tablo 9. Çeşitli İptal Nedenleri ve Tekrar Puan Hesaplama Yöntemleri Sonucunda Sırası Değişen Öğrenci Sayısı

Puan		E.D.A.			E.B.B.			E.Z.S.			T.K.S.		
		N	P	E	N	P	E	N	P	E	N	P	E
D.C.	A-G	0	40096	24692	0	42167	22621	0	35741	29047	0	58436	6352
	B-G	22751	40058	1979	20680	42129	1979	27106	35703	1979	13306	49503	1979
	C-G	0	40057	24731	0	42129	22659	0	35703	29085	0	58398	6390
	D-G	4329	27973	32486	7695	28069	29024	3744	27143	33901	6622	46620	11546
Net	A-G	10084	46070	8634	8918	48194	7676	11796	43040	9952	2310	59676	2802
	B-G	22752	40072	1964	20681	42143	1964	27107	35717	1964	9167	53657	1964
	C-G	10085	46045	8658	8919	48170	7699	11797	43016	9975	2310	59653	2825
	D-G	19242	38934	6612	14193	45396	5199	14402	41659	8727	7670	54724	2394

A: Herkese doğru kabul etme; B: Diğer sorular üzerinden hesaplama yapma; C: En az bir doğrusu olan herkese doğru kabul etme; D: Alt test bazında ağırlıklandırma; G: Ham sıralama

* E.D.A.: En düşük ayıricılıktaki maddeyi çıkarma, E.B.B.: En çok boş bırakılan soruyu çıkarma, E.Z.S.: En zor soruyu çıkarma, T.K.S.: Tüm koşulları sağlayan maddeleri çıkarma

Tablo 9'da görüldüğü gibi *testte en düşük ayıricılıktaki sorunun iptal edilmesi durumunda* soruyu herkese doğru kabul etme yöntemine dayalı olarak 40,096 öğrencinin (soruyu yanlış yapan ya da boş bırakanlar) sırasının yükseldiği görülmektedir. Başka bir anlatımla öğrencilerin %62'sinin sırasının yükseldiği belirlenmiştir.

Tablo 9 incelendiğinde en düşük ayırt edicilik indeksine sahip sorunun iptal edilmesi durumunda sorunun herkese kabul edildiği yöntemde öğrencilerin %38'inin, diğer sorular üzerinde hesaplama yönteminde %3'ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme durumunda %38'inin, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %50'sinin sırasının değişmediği tespit edilmiştir. Sıralamaların değişmesine en dayanıklı yöntemin alt testler bazında ağırlıklandırma yöntemi olduğu görülmektedir.

Testte en çok boş bırakılan sorunun iptal edilmesi durumunda sorunun herkese kabul edildiği yöntemde öğrencilerin %35'inin, diğer sorular üzerinde hesaplama yönteminde %3'ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme durumunda %34'ünün, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %44'ünün sırasının değişmediği tespit edilmiştir.

Testin en zor sorusunun iptal edilmesi durumunda sorunun herkese kabul edildiği yöntemde öğrencilerin %45'inin, diğer sorular üzerinde hesaplama yönteminde %3'ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme durumunda %44'ünün, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %52'sinin sırasının değişmediği tespit edilmiştir.

Testin en zor sorusunun iptal edilmesi durumunda da 3.744 öğrencinin sırasının düştüğü (%6), 27.143 öğrencinin (%42) ise sırasının arttığı görülmektedir.

Tablo 9'da görüldüğü gibi testin en zor sorusunun iptal edilmesi durumunda sorunun herkese kabul edildiği yöntemde öğrencilerin %45'inin, diğer sorular üzerinde hesaplama yönteminde %3'ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme durumunda %44'ünün, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %52'sinin sırasının değişmediği tespit edilmiştir.

Testteki söz konusu şartları sağlayan üç sorunun iptal edilmesi durumunda soruyu herkese doğru kabul etme yöntemine dayalı olarak 58,436 öğrencinin (soruyu yanlış yapan ya da boş bırakanlar) sırasının yükseldiği görülmektedir. Diğer bir anlatımla öğrencilerin %90'ının sırasının yükseldiği tespit edilmiştir. Diğer sorular üzerinden hesaplama yapma yönteminde 13,306 öğrencinin (%21) sırasının düştüğü, 49,503 öğrencinin (%76) sırasının ise yükseldiği belirlenmiştir. En az bir doğrusu bulunan öğrencilere iptal edilen sorunun doğru kabul edilmesi durumunda da 58,398 öğrencinin (%90) sırasının yükseldiği saptanmıştır. Alt testler bazında ağırlıklandırma yapılması durumunda da 6,622 öğrencinin sırasının düştüğü (%10), 46,620 öğrencinin (%72) ise sırasının arttığı görülmektedir.

Tablo 9'da görüldüğü gibi üç sorunun iptal edilmesi durumunda sorunun herkese kabul edildiği yöntemde öğrencilerin %10'unun, diğer sorular üzerinde hesaplama yönteminde %3'ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme durumunda %10'unun, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %18'inin sırasının değişmediği belirlenmiştir.

TARTIŞMA ve SONUÇ

Bireylerin hakkında karar verilmesine dayanak oluşturan geniş ölçekli testler, maksimum performans testi özelliği göstermektedir. Bireylerin akademik başarıları, zekâları gibi bilişsel davranışlara odaklanan bu testlerdeki (Büyüköztürk, Kılıç Çakmak, Ekgün, Karadeniz ve Demirel, 2018) maddelerin doğru bir cevabı bulunmaktadır. Testte yer alan maddenin doğru cevabının net olmaması, birden fazla seçeneğin doğru cevap niteliği taşıması, sorudaki verilen açıklama ya da soru cümlesinin farklı anlamlara gelecek şekilde kurgulanması vb. gerekçelerle hatalı olduğu belirlenen soruların iptal edilmesi gerekmektedir. Bu nedenle de test geliştirme sürecinde ön uygulama çalışmaları olması gerektiğinden söz edilmektedir ki (Crocker ve Algina, 1986) sorular üzerinde yapılan incelemeler doğrultusunda sorudaki hatalar önceden fark edilebilir. Ancak ÖSYM, MEB gibi kurumlar tarafından gerçekleştirilen geniş ölçekli testlerin güvenlik nedeniyle ön uygulama çalışmaları yapılamamaktadır. Anadolu Üniversitesi Açıköğretim Fakültesi tarafından gerçekleştirilen testlere yönelik de (Karadağ, Akıldız, Kumtepe ve Akgün, 2017) soruların alan uzmanları tarafından test tekniği açısından incelenerek redaksiyonunun yapıldığı ve test içeriğine uygunluğunun kontrol edildiği belirtilmektedir. Bu nedenle de sınav sonrasında uzman-aday bakış açısı farklılığı nedeniyle kimi sorular hatalı görülerek iptal edilebilmektedir. 2019 yılında TYT temel soru kitapçığındaki matematik testinin 7. sorusu ile AYT'nin Türk dili ve edebiyatı sosyal bilimler-1 temel kitapçığının 33. Sorusu iptal edilmiştir (ÖSYM, 2019). Burada karşılaşılan sorun ise iptal edilen soru sonrasında puanlamanın nasıl olacağıdır. Bu araştırma kapsamında farklı koşullarda iptal edilen sorular olması durumunda gerçekleştirilen hesaplama yöntemlerinin test sonuçlarına etkisi, ham ve net ve puanlar ile öğrenci sıralaması bazında incelenmiştir.

Farklı koşullarda iptal edilen sorular sonrasında gerçekteki puan ile hesaplanan puan arasındaki en yüksek farkın “herkese doğru kabul etme” yöntemiyle hesaplandığı belirlenmiştir. Bir sorunun iptal edilmesi durumunda sorunun herkese doğru kabul edilmesi ve onunla neredeyse aynı sonuçlar üreten “en az bir doğrusu olan herkese

kabul etme” yöntemleri, sorunun ölçmeyi amaçladığı beceriye sahip olmayan kişilerin tam puan almasına ve söz konusu beceriye sahip olanlarla aynı seviyede ölçüm yapılmasına neden olmaktadır. Çünkü gerçek durumda hatalı olan sorunun yokladığı davranışa sahip öğrenciler soruda hata olmasaydı soruyu zaten doğru cevaplandırıcaklardı bu nedenle “bilen” olarak adlandırılan bu öğrenciler için herkese doğru kabul etme yöntemi bir dezavantaj oluşturmaktadır. Aynı zamanda hatalı soru ya da soruların iptal edilmesi durumunda sorunun herkes için doğru kabul edilerek puanlanması; sorunun ölçmeyi amaçladığı beceriye sahip olmadığı için soruyu yanlış cevaplandırarak ya da boş bırakacak kişilere de soru doğru kabul edilerek puan verildiği için testin ortalamasının gerçek değerden daha yüksek hesaplanmasına neden olmaktadır. Başka bir anlatımla soru iptali test geliştiren, testi kullanan ve ilgili kurumlar için istenmeyen bir durum olmasının yanı sıra testin psikometrik özelliklerini olumsuz etkilemektedir (Sarıtaş ve Bağcı, 2017). İptal edilen sorunun ölçmek istediği özelliği gerçekten bilen ve bilmeyen kişilerin ayrımı yapılamamaktadır. İptal edilen soru sayısı arttıkça herkese doğru kabul etme yöntemine göre hesaplama sonucunda ortalamalar da gerçek değerden uzaklaşmaktadır. Bu durum öğrencilerin sıralamasının da değişmesine neden olmaktadır.

Benzer şekilde farklı koşullara bağlı olarak soruların iptal edilmesi sonrasında ve “en az bir doğrusu olan herkese doğru kabul etme” yöntemine dayalı olarak gerçekleştirilen hesaplama sonuçlarında hesaplanan ortalama puanların, gerçek puanlardan yüksek olduğu belirlenmiştir. Test uzunluğuna da bağlı olarak en az 1 sorunun doğru işaretlenmesi, iptal edilen sorunun ölçmeyi amaçladığı beceriye sahip olduklarının doğrudan bir göstergesi sayılamaz. Bu nedenle de söz konusu hesaplama yöntemi ile gerçek ortalama arasında farklılıklar gözlenmektedir. Bu durum, puanların da değiştiğinin bir göstergesidir. Geniş ölçekli testlerden alınacak virgüllü puanların bile sıralamada farklılık yaratması nedeniyle bir öğrencinin belki istediği bir okul veya programa yerleşmemesi veya istemediği bir okul veya programa yerleşmesi durumları ortaya çıkabilmektedir (Baş ve Kıvılcım, 2019). Bu da iptal edilen sorular sonrasında puanların değişimindeki etkiyi göstermektedir.

Farklı koşullarda soru iptaline yönelik gerçekleştirilen hesaplamalarda “diğer sorular üzerinden hesaplama yapma” yöntemine göre o sorunun puan değeri puanlamaya dâhil edilmez. Diğer soruların ağırlıkları artırılarak yeniden hesaplama yapılır. Eğer test, tek boyutluluk özelliği gösteriyor ve iç tutarlığı yüksek bir test ise bu durumda söz konusu yönteme göre yapılan hesaplamaların gerçek değere yakın olması beklenir. Araştırma kapsamında incelenen Türkçe testinde farklı alt testlerden (sözcükte anlam, cümlede anlam vb.) bulunmasına karşın diğer sorular üzerinden hesaplama yöntemine dayalı hesaplamaların herkese doğru kabul etme ve en az bir doğrusu olan herkese doğru kabul etme yöntemlerinden gerçek değere çok yakın sonuçlar ürettiği tespit edilmiştir. TEDMEM’de (2014) TEOG 2014 yılında ilk uygulama sınavında 2 soru iptal edilmesine karşın mazeret sınavında iptal soru olmadığından asıl sınava giren öğrencilerin 1 soruyu yanlış yapmaları durumunda 0.56 puanlık, telafi sınavına giren öğrencilerde ise 0.5 puanlık kayıp yaşamasına neden olduğu belirtilmiştir.

Geniş ölçekli testler uygulayan ÖSYM ve MEB’in yönetmeliklerinde soru iptaline yönelik uygulanmış/uygulanan yöntemlerin yanı sıra araştırmacılar tarafından “alt test bazında ağırlıklandırma” yöntemi önerilerek sonuçlar incelenmiştir. Hesaplama sonuçları, bütün iptal koşullarında alt test bazından ağırlıklandırma yönteminin herkese doğru kabul etme ve en az bir doğrusu olan herkese doğru kabul etme yöntemlerinden gerçek değere çok yakın sonuçlar ürettiği belirlenmiştir. Diğer sorular üzerinden hesaplama yapma yöntemine dayalı hesaplama sonuçları ile alt test bazında ağırlıklandırma sonuçlarının genel olarak benzerlik gösterdiği tespit edilmiştir. En çok boş bırakılan sorunun iptal edilmesi durumunda “alt test bazında ağırlıklandırma” yönteminin en iyi sonuçları verdiği; diğer iptal durumlarında ise en iyi sonucu diğer sorular üzerinden hesaplama yapma yönteminin verdiği saptanmıştır.

Öğrencilerin sıralamaları bazında inceleme yapıldığında da farklı iptal koşullarına bağlı olarak kimi öğrencilerin sıralarının yükseldiği kimi öğrencilerin sıralarının düştüğü, kimi öğrencilerin sırasının ise gerçek değerle aynı olduğu görülmektedir. Herkese kabul edildiği yöntemde öğrencilerin %10-45’inin, diğer sorular üzerinde hesaplama yönteminde %3’ünün, en az bir doğrusu olan herkese soruyu doğru kabul etme

durumunda %10-44'inin, alt testler bazında ağırlıklandırma yönteminde ise öğrencilerin %18-52'sinin sırasının gerçek sıralamalarıyla aynı olduğu belirlenmiştir. Test puanlarının anlamı ve yorumlanması ile toplumsal yaşamda insan hayatı için önemi yüksek olan birçok kararın alınmasında (Ör; işe girme, okulu bitirme, zekâ düzeyinin belirlenmesi, hasta olup olmama gibi.) hem testten alınan puanlar hem de sıralamalar kullanıldığı için (Sarıtaş ve Bağcı, 2017) hatalı soru olması durumunda gerçekleştirilecek hesaplama yöntemine dayalı sıralamaların da korunması büyük önem taşımaktadır.

Soruların iptal edilmesi durumunda herkese doğru kabul etme ve en az bir doğrusu olan herkese doğru kabul etme yöntemleri hem ortalama hem de sıralama bazında gerçek değerlerden uzak kestirimler yaptığı için de kullanılması önerilmemektedir.

Test uzunluğu ve test içinde yer alan alt testlerin uzunluğuna bağlı olarak diğer sorular üzerinden hesaplama yapma ve alt test bazında ağırlıklandırarak hesaplama yapma yöntemleri kullanılabilir. Özellikle test sonuçlarından alınan puanların sıralama ölçeğine dönüştürüldüğü durumlarda öğrencilerin gerçek sıralamasına en yakın değerlerin alt test bazında ağırlıklandırarak hesaplama yöntemine göre elde edildiği göz önünde bulundurulmalıdır. Alanyazında testlerde hatalı madde bulunması durumunda ya madde istatistikleri incelenmiş ya da hatalı soru incelemesi için uzman görüşlerine dayalı çalışmalar yapılmıştır. Bu nedenle sonuçların alanyazına dayalı tartışılmaması, araştırmanın bir sınırlılığını oluşturmaktadır.

Bu çalışmada elde edilen sonuçlar, farklı test özelliklerinde sonuçlarla karşılaştırılarak incelenebilir. MEB, ÖSYM gibi özellikle kademeler arası geçiş sistemleri için test uygulayan kurumların testlerinde soru iptaline yönelik geniş önlemler alması önerilmektedir. Bilindiği ve araştırma sonuçlarında da görüldüğü üzere testlerde iptal edilen soru sayısı artış gösterdikçe adayların doğru sayıları, puanları ve sıralamalarındaki değişiklik de artış göstermektedir. Testlerde hatalı maddenin olması durumunda -özellikle sıralamanın ön planda olduğu durumlarda- alt test bazında ağırlıklandırma yöntemlerini seçmesi önerilir. Başka çalışmalarda geliştirilecek yöntemlerin sonuçlarının da bu araştırma sonuçlarıyla karşılaştırılması önerilmektedir.


KAYNAKLAR

- AERA, APA, ve NCME (1985). *Standart for educational and psychological tests*. Washington: American Psychological Association.
- Baş, G., & Kıvılcım, Z. S. (2019). Türkiye’de öğrencilerin merkezi sistem sınavları ile ilgili algıları: bir metafor analizi çalışması. *Eğitimde Nitel Araştırmalar Dergisi*, 7(2), 639-667.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Ekgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2018). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Yayıncılık.,
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, & Winston.
- Çobanoğlu Aktan, D., Aksu, G., & Eser, T. (2018). Türkiye ve Amerika’da engelli öğrenciler için yapılan geniş ölçekli sınavların yasal sorumluluklar, uygulama yöntemleri ve geçerlik açısından incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 14(1): 69-83.
- ETS Report (1994). Dorans, Neil J., Feigenbaum, Miriam D., Feryok, Nancy J., Lawrence, Ida M., Schmitt, Alicia P., & Wright, Nancy K. (1994). *Technical Issues Related to the Introduction of the New SAT and PSAT/NMSQT*. ETS Research Memorandum, RM-94-10.
- Güler, N. (2011). *Eğitimde Ölçme ve Değerlendirme*. Ankara: PegemA.
- IEA (2020). Methods and Procedures: *TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center: Chestnut Hill.
- Karadağ, N., Akıldız, M., Kumtepe, A.T., & Akgün, H.R. (2017). Anadolu Üniversitesi Açıköğretim Sistemi soru yazarlarının ölçme ve değerlendirme seminerlerine ilişkin görüşleri. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 3(3), 9-46.
- Lane, S., Raymond, M.R., & Haladayna, T.M. (2016). *Handbook of test development*. Routledge: New York.
- Ludlow, L. H. (2001). *Teacher test accountability*. Education Policy Analysis Archives, 9, 6.
- MEB (2016). *2015-2016 2. Dönem ortak sınavlar sayısal bilgiler*. MEB Yayınları.
- MEB (2018). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezî sınav başvuru ve uygulama kılavuzu*. MEB Yayınları.
- ÖSYM (2016). *2016 Öğrenci seçme ve yerleştirme sistemi (ÖSYS) kılavuzu*. ÖSYM Başkanlığı.
- ÖSYM (2017). *2017 Öğrenci seçme ve yerleştirme sistemi (ÖSYS) kılavuzu*. ÖSYM Başkanlığı.
- ÖSYM (2019). Duyurular: <https://www.osym.gov.tr/TR,16792/2019-yuksekokretim-kurumlari-sinavi-2019-yks-oturumlari-temel-soru-kitapciklari-ve-cevap-anahtarlari-yayimlandi-16062019.html>.

- Özdemir, C., & Yandı, A. (2012). Türkiye'deki üniversitelerde 2010-2012 yılları arasında yayımlanan tezlerdeki yazarlar tarafından geliştirilmiş başarı testlerinin "test geliştirme aşamaları" bakımından incelenmesi.
Downloads/Turkiye_deki_Universitelerde_2010_2012_Y%20.pdf.
- Sarıtaş, S. & Bağcı, V. (2017). *Test puanlarının kullanım gerekliliklerine yönelik izlenen yasal süreç: ÖSYM örneği*. 26. Uluslararası Eğitim Bilimleri Kongresi, Antalya, Türkiye, 20 - 23 Nisan 2017.
- TEDMEM (2014). *2014 Eğitim değerlendirme raporu. TEDMEM Değerlendirme Dizisi 1: Ankara*.
- Varma, S. (2006). *Preliminary item statistics using point-biserial correlation and p-values*. Educational Data Systems Inc.: Morgan Hill CA. Retrieved, 16 (07).

ORCID

Ayfer Sayın  <https://orcid.org/0000-0003-1357-5674>

Hakan Koğar  <https://orcid.org/0000-0001-5749-9824>

SUMMARY

Introduction

In this study, it is aimed to examine how the methods used for calculation change in the averages and rankings of the individuals when there are faulty questions in the tests. Within the scope of the research, it was also aimed to examine the differences in the rankings of individuals by proposing a new calculation method for scoring in cases where there are erroneous questions.

Method

Within the scope of this study, the effect of the methods used on the ranking of the individuals in case of erroneous questions in the tests was examined; at the same time, a new calculation method has been proposed. In this direction, it can be said that the research is a basic research.

This research was conducted with 64,788 students who participated in the 2016 TEOG application.

In the study, it is aimed to examine to what extent the error problem in the tests used for recognition and placement changes the rankings of the students. In this direction, the Turkish test included in the TEOG application, which is used for the transition to high schools in Turkey, was used. There are a total of 20 items in the Turkish test. The items in the test; meaning in words, meaning in sentences, meaning in parts, expression, grammar, spelling and punctuation. In the process of determining these issues, the opinions of five faculty members working in the field of Turkish education were taken.

In order to determine whether there is a significant difference between the criterion scores and the scores obtained from the conditions in the study, descriptive statistics were calculated first and eta-square values for the differences were determined. In order to compare the student ranks, the number and percentages of students whose scores increased, decreased and remained the same were calculated.

Findings

This situation causes a change in the ranking of the students. In case the questions are canceled, it is not recommended to use the methods of accepting everyone as correct and accepting everyone who has at least one truth as correct, as they make estimates far from real values on both average and ranking basis.

Conclusion

After the questions canceled under different conditions, it was determined that the highest difference between the actual score and the calculated score was calculated after the "acceptance right for everyone" method. If a problem is canceled, the problem is accepted as true for everyone and the "acceptance to everyone with at least one truth" methods that produce almost the same results as it causes people who do not have the skills to measure the problem and measure at the same level as those who have the said skill . Because the students who have the behavior of the wrong question in the real situation would have answered the question correctly if there was no

error in the question. Therefore, the method of accepting everyone right for these students who are called "knowing" creates a disadvantage. At the same time, if the wrong question or questions are canceled, the question is considered correct for everyone and scored. Since the question does not have the skill aimed to measure, people who answer the question incorrectly or leave it blank are given points by accepting the question as correct, causing the average of the test to be calculated higher than the real value. In case the questions are canceled, calculations based on the test length and the length of the subtests included in the test, and calculations based on other questions can be used. It should be kept in mind that the values closest to the actual ranking of the students are obtained according to the calculation method by weighting on the basis of the subtest, especially in cases where the scores obtained from the test results are transformed into a ranking scale.