



Test Maddelerinin Ölçtüğü Bilişsel Düzey ve Test Yönergesinin Tahminle Yanıtlama Davranışına Etkisi¹

The Effect of Cognitive Level Measured by Test Items and Test Instruction on Guessing Tendency

Ercan ÇOBAN, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, coban.ercan@gmail.com

R. Nükhet DEMİRTAŞLI, Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, rnukhet@yahoo.com

Öz. Yarı deneysel bir araştırma olan bu çalışmanın amacı, test yönergesi ve test maddelerinin ölçtüğü bilişsel düzeyin tahmin davranışı üzerindeki etkisini incelemektir. Araştırmanın çalışma grubu, 2014-2015 bahar döneminde Ankara Üniversitesi Eğitim Bilimleri Fakültesinde öğrenim gören 220 öğrenciden oluşmaktadır. Çalışma grubundaki öğrencilere, bilgi, kavrama ve uygulama düzeylerini ölçen eşit sayıda madde içeren çoktan seçmeli ölçme ve değerlendirme dersi başarı testi uygulanmıştır. Test yönergesinin tahmin davranışı üzerindeki etkisini belirlemek için test dört farklı yönergeyle uygulanmıştır. Veriler; Kruskal-Wallis testi, ilişkili örneklemeler tek faktörlü varyans analizi ve Friedman testiyle analiz edilmiştir. Araştırmadan elde edilen bulgular, tahmin davranışının test yönergesi ve testteki maddelerin ölçtüğü bilişsel düzeye göre manidar olarak değiştiğini göstermiştir.

Anahtar Sözcükler: Test yönergesi, düzeltme formülleri, bilişsel düzey, tahminle yanıtlama davranışı

Abstract. The aim of this study which is a quasi-experimental research is to investigate the effect of test instruction and cognitive level measured by test items on guessing tendency. The study group of the research consisted of 220 undergraduate students studying at different departments of Ankara University Faculty of Educational Sciences in 2014-2015 spring term. An achievement test of measurement and evaluation containing equal number of items measuring knowledge, comprehension and application levels was applied to students in the study group. The test was applied with four different instructions to determine the effect of test instruction on guessing. The data were analyzed by using Kruskal-Wallis test, repeated measure one-way analysis of variance and Friedman test. The findings showed that test instruction and cognitive level measured by test items significantly affected guessing tendency.

Keywords: Test instruction, correction formulas, cognitive level, guessing tendency

¹ Bu çalışma "Test yönergesi ve risk alma eğiliminin seçmeli testlerin psikometrik özelliklerine ve tahminle yanıtlama davranışına etkisi" başlıklı yüksek lisans tez çalışmasının bir bölümünden özetlenerek hazırlanmış ve 31 Mayıs-3 Haziran 2016 tarihleri arasında Muğla'da gerçekleştirilen "3.Uluslararası Avrasya Eğitim Araştırmaları Kongresi (EJER)" başlıklı kongrede sözlü bildiri olarak sunulmuştur.

SUMMARY

Introduction

There is a very important limitation of multiple choice tests which are commonly used in educational measurement. Test takers may be able to give correct answer by random guessing. Some correction formulas have been developed to prevent guessing since guessing diminishes reliability (Zimmerman & Williams, 1965) and validity (Baykul, 2010) of test scores. Test practitioners try to prevent guessing by using test instructions stating that correction formulas will be used. However, there is no consensus in literature about the effect of test instruction on guessing. Moreover, most of the studies have focused on classical correction formula. That is why, it is necessary to conduct a study which examines the effect of different correction formulas on guessing tendency. Besides, there is no empirical research about effect of cognitive level measured by test items on guessing tendency. Therefore, the aim of this study is to investigate the effect of test instruction and cognitive level measured by test items on guessing tendency.

Method

The research was conducted as a quasi-experimental research. An achievement test of measurement and evaluation containing equal number of items measuring knowledge, comprehension and application levels was applied to 220 undergraduate students studying at different departments of Ankara University Faculty of Educational Sciences. The test was applied to four similar groups with different instructions. Instructions only differed in terms of explanation about scoring, other explanations were same. Explanations about scoring were as follow:

- **Instruction 1 (Y₁):** There will be no correction for wrong answers.
- **Instruction 2 (Y₂):** One correct answer will be deleted because of four wrong answers.
- **Instruction 3 (Y₃):** One correct answer will be added for five omitted items.
- **Instruction 4 (Y₄):** One correct answer will be deleted because of each wrong answer.

The number of wrong answers and omitted items were used to compare the guessing tendency of these four groups of students. The data were analyzed by using Kruskal-Wallis test, repeated measure one-way ANOVA and Friedman test.

Results

The results of Kruskal-Wallis test indicated that the effect of test instruction on the number of wrong answers was statistically significant ($\chi^2_{(3)}=11.23$; $p<0.05$). Mann-Whitney U tests revealed that the number of wrong answers of the students who took the test with Y₁ was significantly greater than the number of wrong answers of the students who took the test with Y₂ (U=1186, $r= 0.19$), Y₃ (U = 994, $r = 0.30$) and Y₄ (U = 1091, $r = 0.24$). However, there was no significant difference among the number of wrong answers of the groups Y₂, Y₃ and Y₄. It was found that there was significant difference among the number of omissions of the students who took the test with different instructions ($\chi^2_{(3)}=33.04$; $p<0.05$). The number of omissions of the students who take the test with Y₁ was significantly smaller than the number of omissions of the students who take the test with Y₂ (U = 861, $r= 0.40$), Y₃ (U = 666, $r = 0.50$) and Y₄ (U = 828, $r = 0.41$). Nevertheless, there was no significant difference among the number of omissions of the groups Y₂, Y₃ and Y₄. One way repeated measure analysis of variance results showed that the effect of cognitive level measured by test items on the number of wrong answers in group Y₁ was statistically significant ($F_{(1.69,91.08)}=7.06$; $p<0.05$). Paired comparisons revealed that average number of wrong answers was higher in items measuring application level ($\bar{X} = 3.20$) compared to items measuring knowledge ($\bar{X} = 2.36$) and comprehension ($\bar{X} = 2.36$) levels. According to

Friedman test results, the effect of cognitive level measured by test items on the number of wrong answers in group Y_3 was statistically significant ($\chi^2_{(2)}=18.67$; $p<0.05$). Wilcoxon test results showed that the number of wrong answers was smaller in items measuring application level compared to items measuring knowledge level ($z=-2.57$, $p=0.01$) in this group. Friedman tests were conducted to examine the effect of cognitive level of test item on the number of omitted answers in each group (Y_1 , Y_2 , Y_3 and Y_4). Results indicated that the effect of cognitive level measured by test item was statistically significant in groups Y_1 ($\chi^2_{(2)}=30.75$; $p<0.05$), Y_2 ($\chi^2_{(2)}=61.88$; $p<0.05$), Y_3 ($\chi^2_{(2)}=80.56$; $p<0.05$) and Y_4 ($\chi^2_{(2)}=75.44$; $p<0.05$). In all groups, the number of omissions was higher in items measuring application level compared to items measuring knowledge and comprehension level. In general (except group Y_4), the number of omissions was higher in items measuring comprehension level compared to items measuring knowledge level.

Discussion and Conclusion

According to results of the study, the number of wrong answers was higher when the instruction stating that there would be no correction for guessing (Y_1) was used. This finding which is similar to the finding of Prieto and Delgado (1999a) shows that instructions containing explanation about the use of correction formulas may decrease guessing tendency. As parallel to this finding, the number of omitted items was smaller in group Y_1 compared to groups Y_2 , Y_3 and Y_4 . Being consistent with many studies (Albanese, 1988; Angoff & Schrader, 1984; Prieto & Delgado, 1999a), this finding also indicates that the use of correction formulas may prevent guessing tendency. However, the type of correction formula did not significantly change either the number of wrong answers or the number of omissions. There were not significant differences in either the number of wrong answers or the number of omissions between groups Y_2 , Y_3 and Y_4 . That is why; it may be concluded that there is no difference in decreasing guessing tendency among correction formulas. It is seen that guessing tendency decreases when the cognitive level measured by test items increases. Therefore, test developers should give weight to the items measuring high-level skills in multiple-choice tests in order to reduce guessing tendency. In this study, a multiple-choice test consisting of items measuring only knowledge, comprehension and application levels was used. Researchers may reach different conclusions in future studies if they use a multiple-choice test consisting of items measuring high-level skills such as analysis, synthesis and evaluation.

GİRİŞ

Çoktan seçmeli testler, eğitimdeki ölçme ve değerlendirme işlemlerinde yaygın bir şekilde kullanılmaktadır. Puanlamasının objektif ve kolay olması, çoktan seçmeli testlerin sıklıkla tercih edilmesine yol açmaktadır (Haladyna, 1997). Bununla birlikte çoktan seçmeli testlerin çok önemli bir dezavantajı vardır. Testi alan bireyler tahmine giderek doğru yanıt bulabilmektedirler (Baykul, 2010; Crocker & Algina, 1986; Dirkszwaner, 2003; Zimmerman & Williams, 1965). Doğru yanıtın tahminle bulunması sonucunda test puanlarına hata karışmaktadır. Test puanlarına karışan bu hata, testin geçerliğini (Baykul, 2010) ve güvenilirliğini (Zimmerman & Williams, 1965) düşürmektedir. Bu nedenle, test uygulayıcılar tahminle yanıtlama davranışını engellemenin yollarını aramışlardır. Araştırmacılar tahmin davranışını engellemek ve test puanlarına karışan hatayı ortadan kaldırmak için değişik puanlama formülleri geliştirmişlerdir (Crocker & Algina, 1986; Kurz, 1999).

Tahmin davranışı sonucu test puanlarına karışan hata puanlarını düzeltmek için kullanılan en yaygın puanlama formüllerinden biri klasik düzeltme formülüdür. Bu formül, DP, düzeltilmiş test puanını; D, Doğru yanıtlanan soru sayısını; Y, yanlış yanıtlanan soru sayısını ve k, seçenek sayısını göstermek üzere,

$$DP = D - \frac{Y}{k - 1}$$

şeklindedir (Crocker & Algina, 1986; Reid, 1976). Tüm seçeneklerin seçilme olasılığının eşit olduğunu ve bütün yanlış yanıtların tahmin sonucu oluştuğunu varsayan bu formülün kullanılmasının amacı yanlış cevapları cezalandırarak, cevaplayıcıların tahminle yanıtlama davranışlarını engellemektir (Thorndike, 2005). Bu düzeltme yöntemi yanıtlayıcıların bazen kısmi bilgilerini kullanarak ya da yanlış öğrenmelerinden dolayı yanıt verebileceklerini dikkate almamaktadır (Rowley & Traub, 1977). Bu nedenden dolayı, klasik düzeltme formülü bazı yanıtlayıcıların puanının gereğinden fazla düzeltirken (overcorrection), bazılarınınkini yeterince düzeltmemektedir (undercorrection) (Zimmerman & Williams, 1965).

Tahminde bulunan cevaplayıcıları cezalandırmak yerine, tahminde bulunmayan cevaplayıcıları ödüllendirmek, şans başarısını düzeltmek için kullanılan diğer bir yöntemdir. Bu yöntem, tahmine gitmeyen cevaplayıcıların tahmine gitmeleri durumunda boş bırakılan maddelerin bir kısmını doğru cevaplayacakları varsayımı üzerine kuruludur. Seçenek sayısı k olmak üzere cevaplayıcıların rastgele tahminle soruların 1/k kadarını doğru cevaplayacaklarını varsayar. Bundan dolayı, tahminde bulunmayan cevaplayıcılara boş bırakılan soru sayısının seçenek sayısına oranı kadar puan ödül olarak verilir. Bu puanlama yönteminde, DP, düzeltilmiş puan; D, doğru yanıtlanan soru sayısı; B, boş bırakılan soru sayısı ve k seçenek sayısı olmak üzere,

$$DP = D + \frac{B}{k}$$

formülü kullanılmaktadır (Ebel, 1965).

Düzeltilmiş test puanını hesaplamak için doğru sayısından yanlış sayısının çıkarıldığı düzeltme formülü, alan yazında kullanılan diğer bir yöntemdir. Bu yöntemde, DP, düzeltilmiş puan; D, doğru sayısı ve Y, yanlış sayısı olmak üzere,

$$DP = D - Y$$

formülüyle hesaplanmaktadır (Prieto & Delgado, 1999a).

Prieto ve Delgado (1999a), test yönergesinin cevaplayıcıların tahminle yanıtlama davranışı üzerindeki etkisini belirlemek için yaptıkları çalışmada, bu yöntemin yanlış cevap sayısını azaltmada ve boş sayısını artırmada diğer iki yönteme göre daha etkili olduğunu bulmuşlardır.

Çoktan seçmeli test uygulanırken test yönergesinde düzeltme formülünün kullanılacağına açıklanmasının, tahminle yanıtlama davranışında nasıl bir değişikliğe sebep olduğu konusunda değişik görüşler vardır. Ebel (1965)'e göre, düzeltme formülünün faydası matematiksel

olmaktan çok psikolojiktir. Eğer düzeltme formülünün uygulanacağı cevaplayıcılara duyurulursa, bu durum test puanlarının geçerlik ve güvenilirliğini artırabilir. Cronbach (1960)'a göre, şüphe içinde kalınan sorulara tahmin ile cevap verme davranışı şans başarısına karşı düzeltme uygulanacağını belirten yönergelerle tamamen ortadan kaldırılamaz. Bu davranış kişilik özellikleri tarafından belirlenmektedir. Dolayısıyla tahmine gitme eğilimindeki farklılıkları ortadan kaldırmanın tek yolu, cevaplayıcılardan tüm soruları cevaplamalarını istemektir. Böyle bir uygulamayla şans başarısından hiçbir öğrenci avantaj sağlamamış olur. Benzer bir şekilde, klasik düzeltme yönteminin çekingen ve temkinli bireyler için dezavantajlı olduğunu savunan Budescu ve Bar-Hillel (1993)'e göre, bu durumu ortadan kaldırmak için, testi alan bireyler testteki tüm soruları yanıtlamaları için teşvik edilmelidir.

Testi farklı yönergelerle uygulamanın tahminle yanıtlama davranışına etkisi üzerine yapılan çeşitli çalışmalarda ödül ve ceza kavramları üzerinde durulmuştur. Bereby-Meyer, Meyer ve Flascher (2002), yanlış yanıtların cezalandırılacağını belirten test yönergesi verildiği durumda maddeleri yanıtı bırakma eğiliminin, yanıtı bırakılan maddeler için ödül verileceğini belirten test yönergesi verildiği duruma göre daha fazla olduğunu bulmuşlardır. Avila ve Torrubia (2004), üniversite öğrencilerine cezaya duyarlılık (sensitivity to punishment) ve ödüle duyarlılık (sensitivity to reward) ölçeklerini uyguladıkları çalışmada, cezaya ve ödüle duyarlılığın tahminle yanıtlama davranışını etkilediği sonucuna ulaşmışlardır. Güçlük düzeyi yüksek olan testlerde, öğrencilerin başarısızlık beklentisinin daha fazla olduğunu ve bu durumunun ceza duyarlılığı yüksek öğrencilerin çoktan seçmeli test maddelerini boş bırakma olasılığını arttırdığını bulmuşlardır. Güçlük düzeyi düşük testleri ise öğrencilerin bir tür ödül olarak algıladıklarını ve bu durumun ödül duyarlılığı yüksek öğrencilerin yanlış cevap verme olasılığını arttırdığını bulmuşlardır. Espinosa ve Gardeazabal (2010), yaptıkları çalışmada yanlış cevapları cezalandıran puanlama yönteminin, boş bırakılan madde sayısını değiştirdiğini bulmuşlardır. Yanlış yanıtlara verilen cezayı arttırdıklarında, boş bırakılan madde sayısının da arttığını gözlemlemişlerdir.

Alan yazında çoktan seçmeli testleri tahminle cevaplama davranışını engellemek için test yönergesine düzeltme yapılacağına dair açıklama konulması dışında başka öneriler de bulunmaktadır. Haladyna (1997), testteki maddelerin ölçtüğü zihinsel düzey arttıkça şansa doğru cevaplama olasılığının azalacağını ve testin güvenilirliğinin artacağını belirtmiştir. Swineford ve Miller (1953) düzeltme uygulanacağını belirten yönergeyle uygulanan testte özellikle zor sorular yanıtlanırken, tahmin davranışının azaldığını bulmuşlardır. Burton (2001), test uzunluğunun ve test maddelerinin seçenek sayısının artırılmasıyla, tahminle yanıtlama davranışının test güvenilirliği üzerindeki etkisinin azalacağını bulmuştur. Benzer bir şekilde, Valadares (2011) hazırlanan testte kazanım başına düşen soru sayısı ve seçenek sayısı artırıldığında düzeltme formülü kullanmaya gerek kalmayacağını belirtmiştir.

Sonuç olarak, tahminle yanıtlama davranışını engellemek için test yönergesine düzeltme yapılacağını belirten açıklama konulmaktadır. Fakat bu açıklamaların tahmin davranışını engellediğini gösteren ortak bir bulgu yoktur. Ayrıca yapılan araştırmalarda genellikle klasik düzeltme formülünün kullanılacağını belirten yönergeler kullanılmıştır. Bu sebeple farklı düzeltme yöntemlerinin kullanılacağını belirten yönergelerin tahmin davranışı üzerindeki etkisinin araştırılması gerekmektedir. Test maddelerinin ölçtüğü bilişsel düzeyin tahmin davranışı üzerindeki etkisi hakkında görgül araştırma bulunmamaktadır. Bu nedenle, bu araştırmada test maddelerinin ölçtüğü bilişsel düzey ve test yönergesinin tahminle yanıtlama davranışı üzerindeki etkisini incelemek amaçlanmıştır. Araştırmanın amacı doğrultusunda aşağıdaki sorulara cevap aranmıştır:

1. Öğrencilerin tahminle yanıtlama davranışları (*yanlış cevaplanan madde sayısı ve boş bırakılan madde sayısı*), kullanılan düzeltme yönergesine göre farklılaşmakta mıdır?
2. Öğrencilerin tahminle yanıtlama davranışları (*yanlış cevaplanan madde sayısı ve boş bırakılan madde sayısı*), test maddelerinin ölçtüğü bilişsel düzeye göre farklılaşmakta mıdır?

YÖNTEM

Araştırma Modeli

Bu araştırmada, test yönergesinin ve test maddelerinin ölçtüğü bilişsel düzeyin tahminle yanıtlanma davranışı üzerindeki etkisini belirlemek amaçlandığı için farklı bilişsel düzeyleri ölçen maddelerden oluşan ölçme ve değerlendirme dersi başarı testi dört farklı yönergeyle uygulanmıştır. Yönergelerde sadece testin nasıl puanlanacağı hakkındaki açıklamalar farklı olup, diğer açıklamalar aynıdır. Yönergelerde puanlamaya yönelik açıklamalar şu şekildedir:

- **Yönerge 1 (Y₁):** Puanınız hesaplanırken, yanlış yanıtlarınız doğru yanıtlarınızı götürmeyecektir.
- **Yönerge 2 (Y₂):** Puanınız, dört yanlış yanıtınız bir doğru yanıtınızı götürecektir.
- **Yönerge 3 (Y₃):** Puanınız hesaplanırken, yanıtını boş bıraktığınız her beş soru bir doğru yanıt olarak sayılacaktır.
- **Yönerge 4 (Y₄):** Puanınız, doğru yanıtlarınızın sayısından yanlış yanıtlarınızın sayısı çıkarılarak hesaplanacaktır.

Araştırma, son-test kontrol gruplu desende yarı deneysel bir çalışmadır. Yarı deneysel çalışmalar, seçkisiz atamayı gerektirmediği için eğitim araştırmalarında sıklıkla kullanılmaktadır. Bu çalışmalarda seçkisiz atama yapılmadığı için, araştırmacının olası dış etkiler üzerindeki kontrolü gerçek deneysel çalışmalara göre daha azdır (Creswell, 2012).

Çalışma Grubu

Bu araştırmanın çalışma grubunu, 2014-2015 Eğitim Öğretim Yılı Bahar Dönemi'nde Ankara Üniversitesi Eğitim Bilimleri Fakültesi'nin lisans programlarında öğrenim gören 220 öğrenci oluşturmaktadır. Çalışma grubunu oluşturan öğrencilerin cinsiyet ve uygulanan yönergeye göre dağılımları Tablo 1'de verilmiştir.

Tablo 1. Çalışma Grubunun Uygulanan Yönerge ve Cinsiyete Göre Dağılımı

	Cinsiyet		Toplam	
	Kadın	Erkek		
Yönerge	Y ₁	46	9	55
	Y ₂	48	7	55
	Y ₃	43	12	55
	Y ₄	42	13	55
	Toplam	179	41	220

Veri Toplama Aracı

Araştırmada test yönergesi ve test maddelerinin ölçtüğü bilişsel düzeyin tahminle yanıtlanma davranışı üzerindeki etkisini belirlemek için ölçme ve değerlendirme dersi başarı testi kullanılmıştır. Araştırma kapsamında kullanılan ölçme ve değerlendirme dersi başarı testi, araştırmacılar tarafından test geliştirme aşamalarına uygun olarak geliştirilmiştir. Testin kapsam geçerliğini sağlamak için belirtke tablosu hazırlanmış ve uzman görüşü alınmıştır. Testteki maddeler yazılırken, Ulutaş (2003) tarafından geliştirilen ölçme ve değerlendirme dersi başarı testindeki maddelerden yararlanılmıştır. Teste toplam 24 madde bulunmaktadır. Teste bilgi, kavrama ve uygulama düzeylerini ölçen eşit sayıda madde bulunmaktadır. Testin A, B, C ve D olmak üzere her biri farklı bir yönerge içeren dört formu oluşturulmuştur. Bu formlar sırasıyla Y₁, Y₂, Y₃ ve Y₄ yönergeleriyle çalışma grubundaki öğrencilere uygulanmıştır. Bu formların ham

puanlara göre hesaplanan KR-20 güvenilirlik katsayıları sırasıyla 0.61, 0.61, 0.59 ve 0.56 değerlerine eşittir.

Verilerin Toplanması

Veriler toplanırken, öğrencilere öncelikle başarı testinin yanıt kâğıdı dağıtılmıştır. Öğrenciler yanıt kâğıdını doldurduktan sonra, başarı testini yanıtlamışlardır. Başarı testleri yanıtlama yönergeleri farklılaşan dört farklı form halinde (A, B, C ve D) katılımcılara seçkisiz olarak dağıtılmıştır. Test dağıtılırken A, B, C ve D formlarının eşit sayıda uygulanmasına dikkat edilmiştir. Test dağıtıldıktan sonra, öğrencilerden test yönergelerini dikkatle okumaları istenmiş ve uygulama başlatılmıştır. Her uygulama yaklaşık 40 dakika sürmüştür.

Verilerin Analizi

Araştırma kapsamında toplanan veriler, SPSS 17.0 paket programı kullanılarak analiz edilmiştir.

Testi farklı yönergelerle alan öğrencilerin yanlış yanıtladıkları ve boş bıraktıkları madde sayılarını karşılaştırmak için Kruskal-Wallis testi kullanılmıştır. Öğrencilerin farklı bilişsel düzeyleri ölçen maddelerdeki yanlış sayılarını karşılaştırmak için ilişkili örneklemeler tek faktörlü varyans analizi kullanılmıştır. Fakat testi Y_3 ile alan öğrencilerin farklı bilişsel düzeydeki yanlış sayıları normal dağılım göstermediği için Friedman testi kullanılmıştır. Öğrencilerin farklı bilişsel düzeyleri ölçen maddelerdeki boş sayılarını karşılaştırmak için Friedman testi kullanılmıştır.

BULGULAR

Bu bölümde toplanan verilerin analizi sonucunda elde edilen bulgulara yer verilmiştir. Bulgular araştırmanın alt problemleriyle aynı sırada verilmiştir.

Test Yönergelerinin Tahminle Yanıtlama Davranışı Üzerindeki Etkisi Hakkındaki Bulgular

Test yönergelerinin tahminle yanıtlama davranışı üzerindeki etkisini belirlemek için farklı yönergelerle uygulanan testlerde öğrencilerin yanlış yanıtladıkları ve boş bıraktıkları madde sayıları belirlenmiş ve belirlenen bu değerler arasında manidar fark olup olmadığı sınıanmıştır.

Test dört farklı yönergeyle uygulandığında öğrencilerin yanlış ve boş sayılarına ilişkin betimsel istatistikler Tablo 2'de verilmiştir. Y_i 'ler ($i=1, 2, 3,4$) sırasıyla testi yönerge 1, yönerge 2, yönerge 3 ve yönerge 4 ile alan öğrencileri simgelemektedir.

Tablo 2. Öğrencilerin Yanlış ve Boş Sayılarına Ait Betimsel İstatistikler

Betimsel İstatistik	Yanlış				Boş			
	Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
N	55	55	55	55	55	55	55	55
Ortalama	7.84	6.76	6.11	6.27	1.51	3.93	4.95	3.65
Ortanca	8	6	5	6	0	4	5	3
Tepe Değeri	9	4	5	7	0	0	5	0
Standart Sapma	3.39	3.58	3.20	3.26	2.80	3.35	3.41	3.10
Çarpıklık	0.37	1.04	1.22	0.75	1.84	0.41	0.31	0.32
Basıklık	-0.28	1.13	0.81	0.86	2.45	-0.55	-0.15	-1.23
En Düşük	1	1	2	1	0	0	0	0
En Yüksek	15	18	16	16	11	13	14	9

Normallik varsayımı sağlanmadığından testi farklı yönergelerle alan öğrencilerin yanlış sayıları arasında manidar fark olup olmadığını belirlemek için Kruskal-Wallis testi kullanılmıştır. Analiz sonucu elde edilen bulgular Tablo 3'te verilmiştir.

Tablo 3. Yanlış Sayısının Uygulanan Yönergeye Göre Karşılaştırılmasına İlişkin Sonuçlar

Yönerge	N	Sıra Ortalaması	sd	χ^2	p	Manidar Fark
Y ₁	55	133.53	3.00	11.23	0.01	Y ₁ -Y ₂ , Y ₁ -Y ₃ , Y ₁ -Y ₄
Y ₂	55	109.63				
Y ₃	55	94.86				
Y ₄	55	103.98				

Tablo 3'de görüldüğü üzere, öğrencilerin yanlış sayısının yönergeye göre manidar bir farklılık gösterdiği bulunmuştur ($\chi^2_{(3)}=11.23$; $p<0.05$). Farkın hangi yönergeleri alan yanıtlayıcılar arasında olduğu araştırmak için Mann Whitney U-testi ile çoklu karşılaştırmalar yapılmıştır. Mann Whitney U-testi sonuçlarına göre, Y₁ ile alan öğrencilerin yanlış sayısının , Y₂ (U = 1186, r = 0.19), Y₃ (U = 994, r = 0.30) ve Y₄ (U = 1091, r = 0.24) ile alanlardan manidar olarak yüksek olduğu bulunmuştur. Testi yönerge 2, yönerge 3 ve yönerge 4 ile alan öğrencilerin yanlış sayıları arasında manidar fark bulunmamıştır.

Testi farklı yönergelerle alan öğrencilerin boş sayılarını karşılaştırmak için Kruskal-Wallis testi kullanılmıştır. Analiz sonucu elde edilen bulgular Tablo 4'te verilmiştir.

Tablo 4. Boş Sayısının Uygulanan Yönergeye Göre Karşılaştırılmasına İlişkin Sonuçlar

Yönerge	N	Sıra Ortalaması	sd	χ^2	p	Manidar Fark
Y ₁	55	70.82	3.00	33.04	0.00	Y ₁ -Y ₂ , Y ₁ -Y ₃ , Y ₁ -Y ₄
Y ₂	55	118.42				
Y ₃	55	135.84				
Y ₄	55	116.93				

Tablo 4'te görüldüğü üzere, öğrencilerin boş sayısının yönergeye göre manidar bir farklılık gösterdiği bulunmuştur ($\chi^2_{(3)}=33.04$; $p<0.05$). Farkın hangi yönergeleri alan öğrenciler arasında olduğu araştırmak için Mann Whitney U-testi ile çoklu karşılaştırmalar yapılmıştır. Mann Whitney U-testi sonuçlarına göre testi Y₁ ile alan öğrencilerin boş sayısının , Y₂ (U = 861, r = 0.40), Y₃ (U = 666, r = 0.50) ve Y₄ (U = 828, r = 0.41) ile alanlardan manidar olarak düşük olduğu bulunmuştur. Testi yönerge 2, yönerge 3 ve yönerge 4 ile alan yanıtlayıcıların boş sayıları arasında manidar fark bulunmamıştır.

Tahminle Yanıtlama Davranışının Maddelerin Ölçtüğü Bilişsel Düzeyle İlişkinine Dair Bulgular

Test maddelerinin ölçtüğü bilişsel düzeyin tahminle yanıtlama davranışı üzerindeki etkisini belirlemek için öğrencilerin bilgi, kavrama ve uygulama düzeyini ölçen maddelerdeki yanlış yanıtladıkları ve boş bıraktıkları madde sayıları belirlenmiş ve belirlenen bu değerler arasında manidar fark olup olmadığı sınanmıştır.

Öğrencilerin bilgi, kavrama ve uygulama düzeyini ölçen maddelerdeki yanlış sayılarına ilişkin betimsel istatistikler Tablo 5'te verilmiştir. Tablo 5'te yer alan yanlış sayısı ortalamalarının, test maddelerinin ölçtüğü bilişsel düzeye göre manidar bir şekilde farklılaşp farklılaşmadığı belirlemek için ilişkili örneklem için tek faktörlü varyans analizi kullanılmıştır. Normallik varsayımını sağlayan gruplar (Y₁ , Y₂ ve Y₄) için ilişkili örneklem için tek faktörlü varyans analizi uygulanmıştır. Mauchly test sonuçlarına göre, normallik varsayımını sağlayan gruplardan sadece testi Y₄ ile alan grubun verileri küresellik (sphericity) varsayımını sağlamaktadır. Testi Y₁ ve Y₂ ile alan öğrencilerin verileri küresellik varsayımını sağlamadığından,

epsilon değerlerine ($\epsilon > 0.75$) bakılarak Huynh-Feldt değerleri kullanılmıştır (Can, 2013). Analiz sonucu elde edilen bulgular Tablo 6'da verilmiştir.

Tablo 5. Öğrencilerin Farklı Bilişsel Düzeyleri Ölçen Maddelerdeki Yanlış Sayılarına İlişkin Betimsel İstatistikler

Yönerge	Bilişsel Düzey	Betimsel İstatistik						
		N	\bar{X}	Ortanca	Tepe Değeri	Standart Sapma	Çarpıklık	Basıklık
Y ₁	Bilgi	55	2.36	2.00	1.00	1.27	0.46	-0.18
	Kavrama	55	2.27	2.00	2.00	1.16	0.17	-0.12
	Uygulama	55	3.20	3.00	2.00	2.22	0.26	-0.84
Y ₂	Bilgi	55	2.45	2.00	3.00	1.32	0.71	1.46
	Kavrama	55	2.11	2.00	2.00	1.29	0.71	0.74
	Uygulama	55	2.20	1.00	0.00	2.14	0.82	-0.51
Y ₃	Bilgi	55	2.42	2.00	2.00	1.30	0.37	0.07
	Kavrama	55	2.11	2.00	2.00	1.26	0.37	-0.06
	Uygulama	55	1.58	1.00	1.00	2.09	1.72	2.00
Y ₄	Bilgi	55	2.31	2.00	1.00	1.32	0.46	-0.61
	Kavrama	55	1.98	2.00	2.00	1.28	0.20	-0.38
	Uygulama	55	1.98	2.00	.00	1.87	0.96	0.33

Tablo 6. Farklı Bilişsel Düzeyi Ölçen Maddelerdeki Yanlış Sayılarına İlişkin İlişkili Örneklemeler Tek Faktörlü Varyans Analizi Sonuçları

Yönerge	Varyansın Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	Anlamlı Fark
Y ₁	Öğrenciler Arası	206.51	54.00	3.82	7.06	0.00	B-U,K-U
	Bilişsel Düzey*	28.74	1.69	17.04			
	Hata	219.93	91.08	2.41			
	Toplam	455.18	146.77				
Y ₂	Öğrenciler Arası	230.64	54.00	4.27	0.96	0.38	
	Bilişsel Düzey*	3.53	1.80	1.96			
	Hata	199.14	97.41	2.04			
	Toplam	433.31	153.21				
Y ₄	Öğrenciler Arası	191.64	54.00	3.55	1.18	0.31	
	Bilişsel Düzey	3.93	2.00	1.96			
	Hata	180.07	108.00	1.67			
	Toplam	375.64	164.00				

*Huynh-Feldt değerleri

B:Bilgi, K:Kavrama, U:Uygulama

Tablo 6'da görüldüğü üzere, testi Y₁ ile alan yanıtlayıcıların farklı bilişsel düzeyleri ölçen maddelerdeki yanlış sayısı ortalamalarının en az ikisi arasında manidar fark bulunmuştur ($F_{(1.69,91.08)}=7.06$; kısmi $\eta^2= 0.12$, $p<0.05$). İkili karşılaştırmalar, yönerge 1 ile testi alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki yanlış sayısı ortalamasının ($\bar{X} = 3.20$), bilgi ($\bar{X} = 2.36$) ve kavrama ($\bar{X} = 2.27$), düzeyini ölçen maddelerdeki yanlış sayısı ortalamalarından manidar olarak yüksek olduğunu göstermektedir. Testi Y₂ ($F_{(1.80,97.41)}=0.96$; $p>0.05$) ve Y₄ ($F_{(2,108)}=1.18$; $p>0.05$) ile alan yanıtlayıcıların farklı bilişsel düzeyleri ölçen maddelerdeki yanlış sayısı ortalamaları arasında manidar fark bulunmamıştır.

Testi yönerge 3 ile alan yanıtlayıcıların yanlış sayıları dağılımı normal olmadığı için, yanlış sayılarını karşılaştırmak için Friedman testi kullanılmıştır. Analiz sonucu elde edilen bulgular Tablo 7’de verilmiştir.

Tablo 7. Testi Y_3 ile Alan Öğrencilerin Farklı Bilişsel Düzeyi Ölçen Maddelerdeki Yanlış Sayılarına İlişkin Friedman Testi Sonucu

Bilişsel Düzey	N	Sıra Ortalaması	sd	χ^2	p	Anlamlı Fark
Bilgi	55	2.34	2	18.67	0.00	B-U
Kavrama	55	2.08				
Uygulama	55	1.58				

B:Bilgi, K:Kavrama, U:Uygulama

Tablo 7’de görüldüğü üzere, testi Y_3 ile alan öğrencilerin yanlış sayısının test maddelerinin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği bulunmuştur ($\chi^2_{(2)}=18.67$; $p<0.05$). Wilcoxon Testi sonuçları, testi yönerge 3 ile alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki yanlış sayısının, bilgi ($z=-2.57$, $p=0.01$) düzeyini ölçen maddelerdeki yanlış sayısından manidar olarak düşük olduğunu göstermektedir.

Öğrencilerin bilgi, kavrama ve uygulama düzeyini ölçen sorulardaki boş sayılarına ilişkin betimsel istatistikler Tablo 8’de verilmiştir

Tablo 8. Öğrencilerin Farklı Bilişsel Düzeyleri Ölçen Maddelerdeki Boş Sayılarına İlişkin Betimsel İstatistikler

Yönerge	Bilişsel Düzey	Betimsel İstatistik						
		N	\bar{X}	Ortanca	Tepe Değeri	Standart Sapma	Çarpıklık	Basıklık
Y_1	Bilgi	55	0.04	0.00	0.00	0.19	5.09	24.85
	Kavrama	55	0.24	0.00	0.00	0.54	2.28	4.33
	Uygulama	55	1.24	0.00	0.00	2.25	1.71	1.66
Y_2	Bilgi	55	0.11	0.00	0.00	0.37	3.65	13.95
	Kavrama	55	0.45	0.00	0.00	0.79	1.80	2.66
	Uygulama	55	3.36	4.00	0.00	2.81	0.10	-1.38
Y_3	Bilgi	55	0.16	0.00	0.00	0.50	4.00	19.15
	Kavrama	55	0.78	0.00	0.00	1.18	1.35	0.67
	Uygulama	55	4.00	4.00	4.00	2.49	-0.20	-0.71
Y_4	Bilgi	55	0.16	0.00	0.00	0.54	3.88	16.37
	Kavrama	55	0.44	0.00	0.00	0.83	1.80	2.15
	Uygulama	55	3.05	3.00	0.00	2.41	0.08	-1.32

Normallik varsayımı sağlanmadığından, boş sayılarını karşılaştırmak için Friedman Testi kullanılmıştır. Yanıtlayıcıların bilgi, kavrama ve uygulama düzeyini ölçen maddelerdeki boş sayılarının karşılaştırılmasına ilişkin analiz sonuçları Tablo 9’da verilmiştir.

Tablo 9. Farklı Bilişsel Düzeyi Ölçen Maddelerdeki Boş Sayılarına İlişkin Friedman Testi Sonuçları

Yönerge	Bilişsel Düzey	N	Sıra Ortalaması	sd	χ^2	p	Anlamlı Fark
Y ₁	Bilgi	55	1.76	2	30.75	0.00	B-K, B-U, K-U
	Kavrama	55	1.94				
	Uygulama	55	2.30				
Y ₂	Bilgi	55	1.55	2	61.88	0.00	B-K, B-U, K-U
	Kavrama	55	1.80				
	Uygulama	55	2.65				
Y ₃	Bilgi	55	1.44	2	80.56	0.00	B-K, B-U, K-U
	Kavrama	55	1.75				
	Uygulama	55	2.82				
Y ₄	Bilgi	55	1.55	2	75.44	0.00	B-U, K-U
	Kavrama	55	1.69				
	Uygulama	55	2.76				

B:Bilgi, K:Kavrama, U:Uygulama

Tablo 9'da görüldüğü üzere, testi Y₁ ile alan yanıtlayıcıların boş sayısının test maddelerinin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği bulunmuştur ($\chi^2_{(2)}=30.75$; $p<0.05$). Wilcoxon Testi sonuçlarına göre testi Y₁ ile alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki boş sayısının, bilgi ($z=3.63$, $p<0.017$) ve kavrama ($z=3.53$, $p<0.017$) düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksek olduğu görülmektedir. Kavrama düzeyini ölçen maddelerdeki boş sayısı, bilgi düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksektir ($z=2.81$, $p<0.016$). Testi Y₂ ile alan yanıtlayıcıların boş sayısının test maddelerinin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği bulunmuştur ($\chi^2_{(2)}=61.88$; $p<0.05$). Wilcoxon Testi sonuçlarına göre, testi yönerge 2 ile alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki boş sayısının, bilgi ($z=5.43$, $p<0.017$) ve kavrama ($z=5.40$, $p<0.017$) düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksek olduğu görülmektedir. Kavrama düzeyini ölçen maddelerdeki boş sayısı, bilgi düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksektir ($z=3.21$, $p<0.017$). Testi Y₃ ile alan yanıtlayıcıların boş sayısının test maddelerinin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği bulunmuştur ($\chi^2_{(2)}=80.56$; $p<0.05$). Wilcoxon Testi sonuçlarına göre, testi yönerge 3 ile alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki boş sayısının, bilgi ($z=5.87$, $p<0.017$) ve kavrama ($z=5.86$, $p<0.017$) düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksek olduğu görülmektedir. Kavrama düzeyini ölçen maddelerdeki boş sayısı, bilgi düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksektir ($z=3.64$, $p<0.017$). Testi Y₄ ile alan yanıtlayıcıların boş sayısının test maddelerinin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği bulunmuştur ($\chi^2_{(2)}=75.44$; $p<0.05$). Wilcoxon Testi sonuçlarına göre, testi Y₄ ile alan yanıtlayıcıların uygulama düzeyini ölçen maddelerdeki boş sayısının, bilgi ($z=5.66$, $p<0.017$) ve kavrama ($z=5.67$, $p<0.017$) düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksek olduğu görülmektedir.

TARTIŞMA ve SONUÇ

Araştırmada testi herhangi bir düzeltme yapılmayacağını belirten Y₁ yönergesi ile alan öğrencilerin yanlış sayısının diğer öğrencilerden fazla olduğu bulunmuştur. Prieto ve Delgado (1999a)'nın bulgusuyla paralel olan bu bulgu, düzeltme yapılacağı belirtildiğinde yanlış sayısının manidar olarak azaldığını göstermektedir. Wood (1976) da Y₁, Y₂ ve Y₃'ü karşılaştırdığı çalışmada, benzer bir sonuca ulaşmıştır. Düzeltme yapılacağını belirten yönergeler kullanıldığında yanlış sayısının manidar şekilde azalması, bu yönergelerin tahminle yanıtlama davranışı azalttığını göstermektedir. Fakat tahminle yanıtlama davranışını azaltmada düzeltme

yapılacağını belirten yönergeler (Y_2 , Y_3 ve Y_4) arasında manidar fark yoktur. Testi Y_3 ile alan öğrencilerin yanlış sayısı daha az olmasına rağmen, fark manidar değildir.

Testi Y_1 ile alan öğrencilerin boş sayısı, diğer öğrencilerin boş sayısından manidar olarak düşüktür. Alan yazındaki araştırmalarla (Albanese, 1988; Angoff & Schrader, 1984; Prieto & Delgado, 1999a) uyumlu olan bu bulgu, düzeltme yapılmayacağını belirten yönerge (Y_1) kullanıldığında tahminle yanıtlanma davranışının arttığına diğer bir göstergesidir. Test yönergesinde düzeltme uygulanacağı belirtildiğinde öğrenciler cevabını bilmedikleri maddeleri boş bırakma eğilimi göstermektedirler. Testi yanıtsız bırakılan maddeler için ödül verileceğini belirten Y_3 ile alan öğrencilerin boş sayısı, yanlış yanıtların cezalandırılacağını belirten Y_2 ve Y_4 ile alan öğrencilerin boş sayısından fazla olmasına rağmen, aralarında manidar fark yoktur. Bu durum ödül ve ceza yöntemlerinin tahmin davranışı üzerindeki etkileri arasında farklılık olmadığını bir göstergesi olarak düşünülebilir. Bereby-Meyer, Meyer ve Budescu (2003)'nin bulgusu bu durumla uyumsuzdur. Bu araştırmacılar ceza verilen durumda (Y_2) boş sayısının ödül verilen duruma göre (Y_3) daha fazla olduğunu bulmuşlardır.

Öğrencilerin farklı bilişsel düzeyleri ölçen maddelerdeki yanlış sayıları incelendiğinde sadece testi Y_1 ve Y_3 ile alan öğrencilerin yanlış sayılarının maddelerin ölçtüğü bilişsel düzeye göre manidar farklılık gösterdiği görülmektedir. Herhangi bir düzeltme yapılmayacağını belirten Y_1 'in uygulandığı grupta, öğrencilerin uygulama düzeyindeki yanlış sayısı, bilgi ve kavrama düzeyindeki yanlış sayısından manidar olarak yüksektir. Bu duruma bakılarak, öğrencilerin uygulama düzeyindeki maddeleri yanıtlamakta zorlandıkları ve bu maddeleri yanıtlarken daha fazla tahmin davranışı gösterdikleri şeklinde yorum yapılabilir. Yanıtsız bırakılan maddeler için ödül verileceğini belirten Y_3 'ün uygulandığı grupta ise uygulama düzeyindeki yanlış sayısı, bilgi düzeyindeki yanlış sayısından manidar olarak düşüktür. Bu durum uygulama düzeyindeki maddeleri doğru yanıtlamakta zorlanan öğrencilerin, bu maddeleri tahminle yanıtlamak yerine ödül kazanmak amacıyla yanıtsız bırakmalarından kaynaklanmış olabilir. Testi yanlış yanıtlar için ceza uygulanacağına belirten Y_2 ve Y_4 ile alan öğrencilerin yanlış sayıları, maddelerin ölçtüğü bilişsel düzeye göre manidar farklılık göstermemektedir. Bu bulgu, öğrencilerin bilişsel düzey fark etmeksizin yanıtını bilmedikleri maddeleri yanıtsız bırakma eğiliminde olduklarının bir göstergesi olarak düşünülebilir. Yanlış yanıtlar cezalandırılacağı için, öğrenciler bilmedikleri madde hangi bilişsel düzeyde olursa olsun yanıtlamak yerine boş bırakmış olabilirler.

Öğrencilerin farklı bilişsel düzeyleri ölçen maddelerdeki boş sayılarına ilişkin bulgular incelendiğinde, tüm durumlarda öğrencilerin bilgi ve kavrama düzeyini ölçen maddelerdeki boş sayılarının uygulama düzeyini ölçen maddelerdeki boş sayılarından manidar olarak düşük olduğu görülmektedir. Ayrıca öğrencilerin kavrama düzeyini ölçen maddelerdeki boş sayısı bilgi düzeyini ölçen maddelerdeki boş sayısından manidar olarak yüksektir. Maddelerin ölçtüğü bilişsel düzey yükseldikçe boş sayılarının arttığı görülmektedir. Bu bulguya bakılarak, maddelerin ölçtüğü bilişsel düzey yükseldikçe tahminle yanıtlanma davranışının azaldığı yorumu yapılabilir. Herhangi bir düzeltme yapılmayacağını belirtildiği durumda (Y_1) bile maddelerin ölçtüğü bilişsel düzey arttıkça tahmin davranışının azalması, bu çıkarımı güçlendirmektedir. Bu bulguların, maddelerin ölçtüğü bilişsel düzey arttıkça tahmin davranışının azalacağını belirten Haladyna (1997) ile uyumlu olduğu yorumu yapılabilir.

Sonuç olarak, test yönergesi tahminle yanıtlanma davranışını değiştirmektedir. Herhangi bir düzeltme yapılmayacağını belirten yönerge kullanıldığında tahminle yanıtlanma eğilimi daha fazladır. Düzeltme yöntemi fark etmeksizin, düzeltme yapılacağını belirtilmesi tahminle yanıtlanma davranışını azaltmaktadır. Yönerge fark etmeksizin maddelerin ölçtüğü bilişsel düzey arttıkça tahminle yanıtlanma davranışı azalmaktadır. Bu yüzden test geliştiriciler, tahminle yanıtlanma davranışını azaltmak için testlerde üst düzey becerileri ölçen maddelere ağırlık vermelidirler. Bu çalışmada sadece bilgi, kavrama ve uygulama düzeyindeki maddelerden oluşan çoktan seçmeli bir test kullanılmıştır. Gelecekteki çalışmalarda analiz, sentez ve değerlendirme gibi üst düzey becerileri ölçen maddelerden oluşan çoktan seçmeli test kullanılarak tahmin davranışı incelendiğinde farklı sonuçlara ulaşılabilir.

KAYNAKÇA

- Albanese, M. A. (1988). The projected impact of the correction for guessing on individual scores. *Journal of Educational Measurement*, 25(2), 149-157.
- Angoff, W. H., & Schrader, W. B. (1984). A study of hypotheses basic to the use of rights and formula scores. *Journal of Educational Measurement*, 21(1), 1-17.
- Avila, C., & Torrubia, R. (2004). Personality, expectations, and response strategies in multiple-choice question examinations in university students: A test of Gray's hypotheses. *European Journal of Personality*, 18(1), 45-59.
- Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 30(4), 277-291.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2. Baskı). Ankara: Pegem Akademi Yayınları.
- Bereby-Meyer, Y., Meyer, J., & Flascher, O. M. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, 15(4), 313-327.
- Bereby-Meyer, Y., Meyer, J., & Budescu, D. V. (2003). Decision making under internal uncertainty: The case of multiple-choice tests with different scoring rules. *Acta Psychologica*, 112(2), 207-220.
- Burton, R. F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: Question selection and guessing of answers. *Assessment ve Evaluation in Higher Education*, 26(1), 41-50.
- Can, A. (2013). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi*. Ankara: Pegem Akademi Yayınları.
- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating qualitative research* (4th Edition). Boston: Pearson Education.
- Crocker, L., & Algina, J. (1986). *Introduction to classical ve modern test theory*. Belmont: Wadsworth Group.
- Cronbach, L. J. (1960). *Essentials of psychological testing*. (2nd Edition). New York: Harper & Brothers Publishers.
- Dirkzwager, A. (2003). Multiple evaluation: A new testing paradigm that exorcizes guessing. *International Journal of Testing*, 3(4), 333-352.
- Ebel, R. L. (1965). *Measuring educational achievement*. New Jersey: Prentice-Hall.
- Espinosa, M. P., & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, 54(5), 415-425.
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. USA: Ally ve Bacon.
- Kurz, T. B. (1999). "A review of scoring algorithms for multiple-choice tests." *Paper presented at the annual meeting of Southwest Educational Research Association, San Antonio, TX*.
- Prieto, G., & Delgado, A. R. (1999a). The effect of instructions on multiple-choice test scores. *European Journal of Psychological Assessment*, 15(2), 143-150.
- Reid, J. F. (1976). Scoring multiple-choice exams. *The Journal of Economic Education*, 8, 55-59.
- Rowley, G. L., & Traub, R. E. (1977). Formula scoring, number right scoring, and test-taking strategy. *Journal of Educational Measurement*, 14(1), 15-22.
- Swineford, F., & Miller, P. M. (1953). Effects of directions regarding guessing on item statistics of a multiple-choice vocabulary test. *Journal of Educational Psychology*, 44(3), 129-139.
- Thorndike, R. M. (2005). *Measurement and evaluation in psycholgy and education* (7th Edition). New Jersey: Pearson Education.
- Ulutaş, S. (2003). *Genel liselerdeki öğretmenlerin ölçme ve değerlendirme alanındaki yeterlikleri ile ölçme ve değerlendirme ilkelerini uygulama düzeylerinin araştırılması* (Yayımlanmamış yüksek lisans tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Valadares, J. (2011). Correcting scores of tests taking into account the guessing factor: Yes or no? *International Journal for Cross-Disciplinary Subjects in Education*, 2(92), 381-387.
- Wood, R. (1976). Inhibiting blind guessing: The effect of instructions. *Journal of Educational Measurement*, 13(4), 297-307.
- Zimmerman, D. W., & Williams, R. H. (1965). Effect of chance success due to guessing on error of measurement in multiple-choice tests. *Psychological Reports*, 16(3), 1193-1196.