

DÜZELTME YÖNERGESİNİN TESTİN PSİKOMETRİK ÖZELLİKLERİNE ETKİSİ

THE EFFECT OF FORMULA SCORING INSTRUCTION ON PSYCHOMETRIC CHARACTERISTICS OF THE TEST

Ümit ÇELEN*, Nükhet Çıkrıkçı Demirtaşlı**

ÖZET: Bu çalışma, çoktan seçmeli sorulardan oluşan bir matematik başarı testinin, şans başarısı için düzeltme formülü kullanılacağını işaret eden bir belirlemenin test yönergesinde yer aldığı ve yer almadığı koşullarda uygulandığında, test ve madde istatistiklerinde anlamlı farklılık meydana gelip gelmediğini saptamayı amaçlamış yarı deneysel bir araştırmadır. Araştırma, 2000-2001 öğretim yılında, 14 sağlık meslek lisesinde öğrenim gören 452 birinci sınıf öğrencisi ile yürütülmüştür. Kullanılan ölçme aracı, beş seçenekli 40 sorudan oluşan ve araştırmacı tarafından geliştirilen "Lise 1. Sınıf Matematik Başarı Testi"dir. Araştırmadan elde edilen bulgular, düzeltme yönergesi vermenin, test puanları ortalamasını, geçerliliğini ve madde güçlük indekslerini etkilemediğini, bununla beraber, KR-20 iç tutarlılık güvenilirliğini ve madde ayırıcılık gücü değerlerini artırdığını göstermiştir. Bulgular ilgili literatür ışığında tartışılmıştır.

Anahtar Sözcükler: şans başarısı, tahmin, düzeltme formülü

ABSTRACT: This study is a quasi-experimental research which aims to find out if there is a statistically significant difference on the test and item statistics when a multiple choice mathematics achievement test applied with or without a test instruction that indicate chance attenuation formula will be used. The research has been carried on 452 first year students of 14 vocational health schools. The measurement instrument developed by the researcher which is constitute 40 five-choice questions in a mathematics achievement test for first year classes of high school. Findings point out that giving formula scoring instruction does not influence mean test scores, test validity and item difficulty indices but increases the KR-20 internal consistency reliability and item discrimination values. Findings were discussed in context of related literature.

Keywords: chance success, guessing, correction formula for guessing

1.GİRİŞ

Çoktan seçmeli maddeler, eğitimde sürdürülen ölçme ve değerlendirme faaliyetlerinde sık kullanılan madde türlerindedir. Puanlamasındaki nesnellik ve kolaylık, tekniğine uygun yazılmaları durumunda bilişsel alanın pek çok düzeyindeki davranışları ölçmek için kullanılabilmesi, makine ile puanlamaya olanak tanınması, çok sayıda öğrenciye aynı anda uygulanabilmesi gibi nedenlerle sıkça tercih edilmektedirler. Bununla birlikte bu tür maddelerin bazı dezavantajları da vardır. Bunlardan en önemlisi, doğru yanıtların tahmine giderek şansla bulunma olasılığıdır. Madde ile ölçülen davranışa sahip olmayan veya kısmen sahip olan yanıtlayıcıların, doğru yanıtı tahmine giderek bulmaları durumunda puanlara şans başarısı karışmakta ve bu da testin geçerliliğine ve güvenilirliğine olumsuz katkıda bulunmaktadır (Baykul, 2000).

Şans başarısına ve tahmin davranışına karşı ne yapılacağı sorusu, seçmeli test yönteminin başından beri test geliştiricilerini ve uygulayıcılarını uğraştırmıştır. Uzun bir süredir devam eden çoktan seçmeli test deneyimine rağmen, konuyla ilgili test uzmanları arasında tam bir uzlaşma sağlanamamıştır. Tahminde bulunmayı ilgilenilmesi gereken bir problem olarak görenler, öğrencileri hem test açıklamasında verilen yönergelerle hem de yanlış tahminde bulunanları cezalandırma yoluyla tahminde bulunmaktan vazgeçirmeyi bir çözüm olarak düşünmüşlerdir (Mehrens ve Leh-

* Ankara Üniversitesi Sağlık Eğitim Fakültesi Sağlık Eğitimi Bölümü, umitcelen@yahoo.com

** Yrd.Doc.Dr, Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı, email: mukhet@yahoo.com

mann, 1984, s.: 339). Tahmini önlemek için test yönergesine, “Yanıtını kesin olarak bilmediğiniz maddeleri yanıtsız bırakınız. Yanlış yanıtlarınızın bir kısmı doğru yanıtlarınızdan çıkartılacaktır.” gibi bir açıklama konulması önerilmiştir (Tekin, 1996, s.: 104). Thurstone (1919) ve Holzinger (1924) (aktaran Budescu ve Bar-Hillel, 1993) tarafından ortaya konulan ve düzeltme formülü olarak bilinen puanlama yöntemi de test puanlarına karışan şans hatasını yok etmeyi amaçlamıştır. Bu formül, düzeltilmiş test puanını (DTP); D, doğru yanıtlanan soru sayısını; Y, yanlış yanıtlanan soru sayısını ve k, seçenek sayısını göstermek üzere aşağıdaki

şekildedir: $DTP = D - \frac{Y}{k-1}$. Alternatif bir puanlama formülü de Traub, Hambleton ve Singh (1969) tarafından geliştirilmiştir. Boş bırakılan maddeler için fazladan puan verilerek rastgele tahminde bulunmanın engellenmesini amaçlayan bu yöntemde formül, B, boş bırakılan maddeleri göstermek üzere, $DTP = D + \frac{B}{k}$ şeklindedir (Budescu ve Bar-Hillel, 1993).

Rastgele tahminin güvenilirlik üzerinde olumsuz bir etkisi olduğu genel bir kabuldür (Frery, 1989). Düzeltme formülü uygulanması yönteminin, puanların geçerliliği ve güvenilirliğini yükselttiği, çünkü düzeltilmiş puanların düzeltilmemiş puanlara göre, bireyin ilgilenilen özelliğini ölçmede daha iyi tahmin ettiği öngörülmüştür. Kuramsal çalışmalar, düzeltilmiş puan sonuçlarının test puanlarının geçerliliğini (Lord, 1963) ve güvenilirliğini (Mattson, 1965) artırdığını göstermiştir (Muijtjens ve diğerleri, 1999). Ancak deneye dayalı çalışmalar bu kanıyı desteklememektedir. Bu durumun nedenleri arasında, testi alanın bir test maddesinin “k” sayıdaki seçeneğinden birini rastgele tahmin ettiğini kabul eden “rastgele tahmin modeli”nin fazla basit olması, testi alanların düzeltme formülü uygulanacağını anlamaması veya test açıklamasındaki yönergeleri yerine getirmemesi ya da test alma davranışlarının başka özelliklerden etkilenmesi sayılabilir (Crocker, 1986). Yanlışlardan bir kısmını test puanlarından çıkarmak ya da boş bırakılan maddelerden bir kısmını test puanına eklemek suretiyle yapılan düzeltme girişimlerinin, test puanını şans hatasından arıttığını ya da yanıtlayıcılar arasındaki şans başarısından ileri gelen farkları giderdiğini savunmak olanaksızdır (Tekin, 1996, s.: 104).

Yanlış yanıtlanan soruların bir kısmının doğru yanıtlardan çıkarılması ilkesine dayanan düzeltme formülleri, bir öğrenci bir sorunun yanıtını bilmediğinde, verilen seçenekler arasından rastgele olarak tahminde bulunup birini seçtiğini varsayar. Bu yüzden eğer bir birey, bir doğru-yanlış testinde 60 doğru, 15 yanlış yanıt vermişse, bireyin testteki 30 soruyu rastgele bir şekilde tahmin ettiği varsayılır. Bu varsayım yerine geldiğinde, düzeltme formülü uygun bir düzeltme sağlar. Ancak bu varsayım her zaman gerçekleşmemektedir. Bazı doğru tahminler, kısmi bilgiye dayanan tahminlerdir. Bazı yanlış yanıtlar ise, yanlış bilgiden veya makul görünen ve iyi çalışan bir çeldiriciden dolaydır. Bu iki durumda, düzeltme formülü, testi alanın puanını gereğinden fazla veya eksik düzeltir. Testi alanın, sorunun yanıtını bildiği, kısmen bildiği veya tamamen bilmediği kararını vermek olanaksız olduğundan, düzeltme girişimleri asla amacına ulaşamaz (Gronlund, 1976; Remmers, Gage ve Rummel, 1965).

Doğru yanıtından emin olmadan bir soruyu tahminle yanıtlamayı, gerçekte bilmediği bir şeyi bildiğini düşündürerek test uygulayıcısını aldatmak ve testten geçer notu almak için bir girişim olarak görenler, tahmin davranışına olumsuz olarak bakmaktadır. Eğer öğrenme “ya hep-ya hiç” tarzı bir iş olsaydı, o zaman tahminde bulunmaya yapılan bu itirazların önemli bir ağırlığı olacaktı. Fakat öğrenmenin genellikle bir “daha az-daha çok” meselesi olması gerçeğinden dolayı, bir bireyden bildiği şeyleri bilmediklerinden açıkça ayırmasını ve yalnızca kesinlikle emin olduğu soruları yanıtlamasını istemek çok mantıklı ve makul görünmemektedir. Hayatın karmaşık problemleri hakkındaki kararlar nadiren şüphelerden uzak ve kesin emin olarak alınır. Hayatta en etkin kişilerin, kanıtlara dayanan tahminlerde bulunan ve bu tahminleri en sık işe yarayanlar olduğu görülmektedir (Ebel, 1965, s.: 230). Problem çözme daima, belirli tahminleri içerir (Gronlund, 1976, s.: 258). Bu nedenle bazı bilgilerden yararlanılarak yapılan tahmin, sanıldığı kadar kötü değildir.

Şans başarısı için yapılan düzeltme, sadece soruları boş bırakmak yerine kör bir şekilde tahminde bulunan öğrencinin avantajını ortadan kaldırma eğilimindedir. Test alma konusunda deneyimli olan bir birey, soruları yanıtlarken, her bilgi kırıntısını kullanarak tahminde bulunmaktan kaçınılmazdır. Tecrübesiz veya şansını denemekten kaçınmaya eğilimli olan kişi ise, düzeltme yönergesinden etkilenerek ve doğru yanıt olma olasılığı şans başarısı düzeyinden oldukça yüksek olduğu birçok soruyu da boş bırakacaktır. Düzeltmiş puanların, cesur ve testler konusunda deneyimli kişilere özel bir avantaj sağlaması ölçüsünde, başarı ölçümleri olarak geçerliliği tartışılabilir (Ebel, 1965). Cronbach (1984) tarafından belirtildiği gibi, bir soruda şüpheye düşüldüğünde tahminde bulunma ya da bulunmama eğilimi, kişilik faktörleriyle belirlenir ve tahminde bulunmaya karşı verilen yönergelerle tamamıyla ortadan kaldırılamaz. Düzeltme formülü uygulanması da, riski göze alan bireylerin avantajını ortadan kaldıramaz. Bu faktörleri ortadan kaldırmanın tek yolu, öğrencilere her soruyu yanıtlamalarını söylemektir. Bu yapıldığında, hiçbir öğrenciye özel bir avantaj verilmemiş olur ve puanlamada tahmin düzeltmesi yapmak gereksizdir (Gronlund, 1976, s: 256).

Şans başarısı için düzeltme uygulanmasına karşı çıkanların dayandıkları bir diğer nokta da, şansla elde edilebilecek başarının çok küçük olması ve bir testten şansla en yüksek puanı elde etme olasılığının neredeyse yok denecek kadar düşük olmasıdır. Bir kişi, hiçbir şey bilmediği bir konuda 40 soruluk bir doğru-yanlış testi aldığında, her soruda rastgele bir şekilde tahminde bulunursa, 40 soruda 20 puan alması beklenebilir. Ancak şansla 40 puan alma olasılığı $1.099.511.627.776$ 'da 1'dir. Dört seçenekli, 40 sorudan oluşan bir testte, hiçbir şey bilmeden ve sadece rastgele tahminle en yüksek puanı alma olasılığı yaklaşık olarak 1.200×10^{21} 'de birer iner (Mehrens ve Lehmann, 1984).

Düzeltmiş ve düzeltilmemiş test puanları ile ilgili karşılaştırmalar, bu iki puan arasındaki korelasyonun genellikle çok yüksek olduğunu göstermektedir (Ebel, 1965). Yani, pratikte düzeltilmiş puanlar testi alanlar hakkında düzeltilmemiş puanlarla benzer (sırada) bilgi verir (Lord, 1975). Eğer testi alan bireylerin tümü soruların tamamını yanıtlamaya ikna edilirse bu korelasyon değeri 1.00 'a ulaşacaktır (Mehrens ve Lehman, 1984). O halde, test puanlarında düzeltme uygulamak, testi alan bireylerin düzeltme uygulanmaması durumundaki sıralamasını neredeyse hiç değiştirmemektedir.

Özetle, seçmeli testlerin, "şans hatası" gibi önemli bir sorunu olduğu ve bu sorunu gidermek veya etkisini en aza indirmek için bir takım yöntemler geliştirildiği ancak bu yöntemlerin de amacına ulaşmada ne derece yeterli olduğunun tartışmalı olduğu ileri sürülebilir. Bu çalışmada da, test puanlarını şans hatasından arındırmak için test yönergesine konulan şans başarısı için düzeltme uygulanacağına ilişkin bir açıklamanın, testin psikometrik özelliklerini (madde ve test puanlarının geçerliliği ve güvenilirliğini) nasıl etkilediği araştırılmıştır. Bu amaçla, a) Test yönergesinde, şans başarısı için düzeltme formülü kullanılacağına işaret eden bir belirlemenin yer aldığı ve yer almadığı koşullarda uygulanan testin, test istatistikleri arasında anlamlı farklar var mıdır? b) Test yönergesinde, şans başarısı için düzeltme formülü kullanılacağına işaret eden bir belirlemenin yer aldığı ve yer almadığı koşullarda uygulanan bir testin, madde istatistikleri arasında anlamlı farklar var mıdır? sorularına yanıt aranmıştır.

2. YÖNTEM

Araştırma, yarı deneysel bir araştırmadır. Kullanılan ölçme aracı, rasgele olarak ikiye ayrılan ve birbirine benzeyen iki grup öğrenciye, sadece şans başarısı için düzeltme uygulanıp uygulanmayacağına ilişkin ifadesi farklı olan iki değişik yönerge ile verilmiş, bu iki koşulda uygulanan testten elde edilen sonuçlar karşılaştırılmıştır.

Araştırma Grubu: Araştırma grubunu, 14 sağlık meslek lisesinde 2000–2001 öğretim döneminde öğrenim gören 452 birinci sınıf öğrencisi oluşturmuştur. Öğrencilerin düzeltme yönergesi verilip verilmemesi durumuna ve cinsiyete göre dağılımları Tablo 1'de verilmiştir.

Tablo 1. Araştırma grubunun düzeltme yönergesi verilip verilmemesi durumuna ve cinsiyete göre dağılımı

Cinsiyet	Düzeltilme Yönergesi Verilen		Düzeltilme Yönergesi Verilmeyen		Toplam	
	Sayı	%	Sayı	%	Sayı	%
Kız	127	56.2	146	64.6	273	60.4
Erkek	99	43.8	80	35.4	179	39.6
Toplam	226	100.0	226	100.0	452	100.0

Ölçme Aracı: Araştırmada kullanılan ölçme aracı, "Lise 1. Sınıf Matematik Başarı Testi"dir (L1-MBT). Test araştırmacı tarafından geliştirilmiş olup çoktan seçmeli (beş seçenekli) 40 sorudan oluşmaktadır. Testin geliştirilmesinde, test geliştirme sürecindeki temel aşamalar izlenmiştir.

İşlem: L1-MBT uygulamaları her sınıfta iki oturumda yapılmıştır. Öğrenciler, sınıf listesindeki sıra numaralarına göre ikiye ayrılmış, tek sıra numaralı öğrencilere birinci ders saatinde düzeltme yönergeli L1-MBT, çift sıra numaralı öğrencilere ise ikinci ders saatinde düzeltme yönergesiz L1-MBT verilmiştir. Testi yanıtlamaya başlamadan önce öğrencilerin test yönergesini okumaları sağlanmış, ders öğretmeni tarafından, yanlış yanıtlanan soruların doğruları götürüp götürmeyeceğine ilişkin açıklama öğrencilere tekrar hatırlatılmıştır. Bir geçerlilik ölçütü olarak kullanılmak üzere öğrencilerin lise 1 matematik dersi yıl sonu geçme notları okul idarelerinden alınmıştır.

Verilerin Analizi: Verilerin analizi SPSS 10.0 ve İteman 3.6 programları aracılığı ile yapılmıştır. Testi farklı yönergelerle alan iki gruptaki öğrenciler, lise 1. sınıf matematik dersi yıl sonu geçme notlarına göre sıralanmış, öğrenci sayısının iki grup için eşit olmasını sağlamak amacıyla üst gruptaki 100 öğrenci başarılı; alt gruptaki son 100 öğrenci başarısız olarak tanımlanmışlardır. Araştırmada, bağımsız (ilişkisiz) iki grup ortalaması arasındaki farkın manidarlık kontrolü t-testi, bağımlı (ilişkili) iki grup ortalaması arasındaki farkın manidarlık kontrolü için de t-testi ve Wilcoxon İşaret Testi tekniklerinden yararlanılmıştır. L1-MBT'nin iki uygulama koşulunda hesaplanan test puanları ile lise 1 matematik dersi başarı notu arasındaki geçerlilik katsayısı olarak korelasyonları karşılaştırmak için Z testi kullanılmış; düzeltilmiş ve düzeltilmemiş test puanları ve bu iki uygulama koşulunda hesaplanan madde güçlük ve ayırıcılık gücü indeksleri aralarındaki ilişkinin miktarını belirlemek için Pearson Çarpım Momentler korelasyon tekniğinden yararlanılmıştır. Çift serili madde ayırıcılık gücü indeksleri, hem testten alınan ham puanlar ve hem de düzeltilmiş puanlar üzerinden hesaplanmıştır. Düzeltilmiş test puanlarının (DTP) hesaplanmasında, D, doğru yanıtlanan soru sayısını; Y, yanlış yanıtlanan soru sayısını ve k, seçenek sayısını göstermek üzere, $DTP = D - \frac{Y}{k-1}$ formülü kullanılmıştır.

3. BULGULAR

Araştırma grubundaki öğrencilerin testten aldıkları puanlara ait dağılım istatistikleri Tablo 2'de verilmiştir. Düzeltme yönergesi verilen ve verilmeyen koşullardaki öğrencilerin, düzeltilmemiş test puanları ortalamaları arasında istatistiksel olarak anlamlı bir farklılık yoktur. Bu durum, düzeltme yönergesi verilen ve verilmeyen koşullardaki başarılı ve başarısız öğrenciler için de değişmemektedir.

Söz konusu iki koşuldaki öğrencilerin düzeltilmiş test puanları ortalamalarına bakıldığında, bu iki test puanı ortalaması arasında da anlamlı fark olmadığı görülmektedir. Düzeltme yönergesi verilen ve verilmeyen koşullardaki başarısız öğrenciler için de benzer bir durum söz konusudur. Ancak, düzeltme yönergesi verilen grupta yer alan başarılı öğrencilerin düzeltilmiş test puanları ortalamasının, düzeltme yönergesi verilmeyen grupta yer alan başarılı öğrencilerin düzeltilmiş test puanları ortalamasından istatistiksel olarak anlamlı şekilde yüksek olduğu görülmektedir. Test açıklamasında düzeltme yönergesi verilmediği halde, puanlarda şans başarısı için düzeltme yapmak, sadece başarılı gruptaki öğrencilerin puanlarının test açıklamasında düzeltme yönergesi verilen diğer başarılı grubun test ortalamasından anlamlı ölçüde daha düşük olmasına yol açmıştır.

Düzeltilmiş ve düzeltilmemiş puanlar arasında 0.97 ile 0.99 arasında değişen oldukça yüksek korelasyonlar gözlenmiştir (Tablo 2).

Düzeltilmemiş test puanlarının standart sapmaları, düzeltme yönergesi verilmeyen grupta diğer gruba göre daha düşük bulunmuştur. Test puanlarında yapılan düzeltme, test puanı ortalamalarını bütün durumlarda istatistiksel olarak anlamlı şekilde azaltmış, ancak standart sapmalarının büyümesine neden olmuştur. Düzeltme yönergesi verilmesi durumunda test puanları dağılımının çarpıklığı, yönerge verilmemesi durumuna göre, daha büyük olmakta ve bu durum başarısız öğrenciler için daha da belirginleşmektedir. Düzeltme yönergesi verilen grupta dağılımın sivrililiği diğer gruba göre normale daha yakın ve başarılı grupta normalden daha basık olmakla birlikte, düzeltme yönergesi verilen başarısız öğrencilerin puan dağılımı oldukça sivri hale gelmektedir (Tablo 2).

Tablo 2. Düzeltme yönergesi verilen ve verilmeyen koşullarda hesaplanan test puanlarına ait dağılım istatistikleri ve korelasyon değerleri

Koşul	Düzeltilme Yönergesi Verilen						Düzeltilme Yönergesi Verilmeyen						t _a
	\bar{x}	Mod	Med	S	Çarp.	Bas.	\bar{x}	Mod.	Med	S	Çarp.	Bas.	
Genel (n=226+226)													
Düzeltilmemiş Test P.	15.74	13.00	14.00	6.39	0.89	0.25	15.51	15.00	15.00	5.06	0.49	0.39	0.42
Düzeltilmiş Test P.	10.90	6.25	8.88	7.86	0.84	0.15	9.83	8.75	8.75	6.28	0.52	0.36	1.60
t _b =39.58*, r=.99*						t _b =56.07*, r=.99*							
Başarılı (n=100+100)													
Düzeltilmemiş Test P.	20.28	19.00	19.00	5.90	0.46	-0.42	19.46	17.00	19.00	4.16	0.50	1.12	1.14
Düzeltilmiş Test P.	16.63	14.50	15.00	7.07	0.46	-0.41	14.81	15.00	13.88	5.02	0.76	0.74	2.09*
t _b =23.69*, r=.99*						t _b =38.19*, r=.98*							
Başarısız (n=100+100)													
Düzeltilmemiş Test P.	11.77	11.00	11.00	4.45	1.96	5.72	11.84	11.00	11.50	3.31	0.40	1.35	-0.13
Düzeltilmiş Test P.	5.94	3.75	5.25	5.44	1.82	5.17	5.24	3.75	5.00	4.13	0.45	1.01	1.02
t _b =38.10*, r=.97*						t _b =55.50*, r=.97*							

* p<0.05

r : Düzeltilmemiş ve düzeltilmiş test puanları arasındaki ilişki.

t_a : Düzeltme yönergesi verilen ve verilmeyen grupların test puanları ortalamaları arasındaki farka ait t değeri.

t_b : Düzeltilmemiş ve düzeltilmiş test puanları ortalamaları arasındaki farka ait t değeri.

Düzeltilme yönergesi verilen ve verilmeyen koşullarda hesaplanan KR-20 iç tutarlılık katsayılarının 0.34 ile 0.89 arasında değerler aldığı ve en düşük değere düzeltme yönergesi verilmeyen başarısız öğrencilerin düzeltilmemiş test puanlarında; en yüksek değere de düzeltme yönergesi verilen öğrencilerin düzeltilmiş test puanlarında ulaştığı görülmektedir (Tablo 3).

Tablo 3. Düzeltme yönergesi verilen ve verilmeyen koşullarda ve başarıları farklı gruplardan hesaplanan test puanlarının kr-20 iç tutarlılık katsayıları

Grup	Hesaplama Esas Alınan Puan Türü	Düzeltilme Yönergesi Verilen		Düzeltilme Yönergesi Verilmeyen	
		n	KR-20	n	KR-20
Genel	Düzeltilmemiş	226	.81	226	.69
	Düzeltilmiş		.89		.81
Başarılı	Düzeltilmemiş	100	.77	100	.51
	Düzeltilmiş		.85		.67
Başarısız	Düzeltilmemiş	100	.65	100	.34
	Düzeltilmiş		.77		.58

Buna göre düzeltme yönergesi verilen grupta yer alan öğrencilerin düzeltilmiş ve düzeltilmemiş test puanlarından hesaplanan KR-20 iç tutarlılık katsayılarının, düzeltme yönergesi verilmeyen grupta yer alan öğrencilerden hesaplanan katsayılardan yüksek olduğu görülmektedir. Bu durum her iki koşul altında çalışmaya alınan başarılı ve başarısız öğrenciler için de aynı şekilde gerçekleşmiş; düzeltme yönergesi verilen grupta daha yüksek bulunmuştur.

L1-MBT'nin ölçüt dayanaklı geçerliliğine ilişkin kanıt bulmak amacıyla araştırma grubundaki öğrencilerin testten aldıkları puanlar ile lise 1. sınıf matematik dersi yıl sonu başarı notları arasındaki korelasyonlar hesaplanmış ve bu değerler Tablo 4'te verilmiştir. Grubun genelinde bu korelasyon katsayıları 0.62 ile 0.73 arasında değerler almış, en düşük değere, düzeltme yönergesi verilen grupta yer alan öğrencilerin düzeltilmemiş test puanları ile ders notları arasında; en yüksek değere ise, düzeltme yönergesi verilmeyen öğrencilerin düzeltilmiş test puanları ile ders notları arasında hesaplanan korelasyonlarda ulaşmıştır. Her iki grupta yer alan başarısız öğrencilerin puanlarından hesaplanan geçerlilik katsayıları dışındaki tüm katsayılar 0.05 düzeyinde istatistiksel olarak anlamlıdır.

Düzeltilme yönergesi verilen gruptaki öğrencilerin puanlarından hesaplanan geçerlilik katsayıları, düzeltme yönergesi verilmeyen gruptaki öğrencilerin puanlarından hesaplanarlardan daha düşük bulunmuştur. Bu durum başarılı ve başarısız öğrenciler için de aynı şekilde gözlenmektedir. Ancak korelasyon katsayıları arasındaki farklar istatistiksel olarak anlamlı bulunmamıştır.

Düzeltilmiş puanlardan hesaplanan geçerlilik katsayıları, düzeltilmemiş puanlardan hesaplanan değerlerden yüksek veya düzeltilmemiş puanlardan hesaplanarlara eşit olarak bulunmuştur. Ancak bu korelasyon katsayıları arasındaki farklar da istatistiksel olarak anlamlı bulunmamıştır.

Tablo 4. Düzeltme yönergesi verilen ve verilmeyen koşullarda ve başarısı farklı gruptardan hesaplanan test puanlarının geçerlilik katsayıları

		Düzeltilme Yönergesi Verilen	Düzeltilme Yönergesi Verilmeyen	
Grup	Puan Türü	$r_{\text{test-ysbn}}$	$r_{\text{test-ysbn}}$	Z
Genel	Düzeltilmemiş	.62*	.72*	-1.90
	Düzeltilmiş	.64*	.73*	-1.80
	z	-0.32	-0.21	
Başarılı	Düzeltilmemiş	.36*	.51*	-1.77
	Düzeltilmiş	.39*	.53*	-1.77
	z	-0.30	-0.10	
Başarısız	Düzeltilmemiş	-.09	.01	-0.98
	Düzeltilmiş	-.09	.02	-1.08
	z	0.00	-0.10	

* $p < 0.05$

Test açıklamasında düzeltme yönergesi verilen ve verilmeyen koşullarda hesaplanan madde güçlük indekslerine ait ortalama değerlerin t testi ile karşılaştırılmasına ilişkin sonuçlar ve düzeltme yönergesi verilen ve verilmeyen gruptardan elde edilen madde güçlük indeksleri değerleri arasındaki korelasyonlar Tablo 5'te verilmiştir.

Tablo 5. Düzeltme yönergesi verilen ve verilmeyen gruplardan hesaplanan madde güçlük indeksleri ortalamalarının ilişkili t testi ile karşılaştırılmasına ilişkin sonuçlar

Grup	Düzeltme Yönergesi Verilen		Düzeltme Yönergesi Verilmeyen		r	t
	\bar{P}_j	S	\bar{P}_j	S		
Genel	.39	.17	.39	.17	.97*	1.15
Başarılı	.51	.19	.49	.19	.90*	1.57
Başarısız	.29	.17	.30	.16	.90*	-0.15

* p<0.05

Yapılan ilişkili t testi sonuçlarına göre, düzeltme yönergesi verilen ve verilmeyen gruptan elde edilen madde güçlük indeksleri ortalamaları arasında istatistiksel olarak anlamlı bir farklılık bulunmamıştır. Her iki koşuldaki grubun başarılı ve başarısız öğrencilerinden elde edilen madde güçlük indeksleri ortalamaları arasında da istatistiksel olarak anlamlı bir farklılık yoktur. İki gruptan elde edilen madde güçlük indeksleri arasında anlamlı korelasyonlar bulunmuştur. Bu korelasyonlar genel için 0.97; başarılı ve başarısız öğrenciler için de 0.90 olarak gerçekleşmiştir.

Düzeltilme yönergesi verilen ve verilmeyen koşullarda ve başarıları farklı gruplardan düzeltilmiş ve düzeltilmemiş puanlar üzerinden hesaplanan çift serili madde ayırıcılık gücü indeksleri (r_{ij}) değerlerinin Wilcoxon İşaret Testi ile karşılaştırılmasına ilişkin sonuçlar Tablo 6'da verilmiştir. Maddeler için ayırıcılık gücünün en üst düzeye çıktığı durumlar daha çok düzeltme yönergesi verilen grupta, en düşük seviyesine indiği durumlar ise daha çok düzeltme yönergesi verilmeyen grupta gözlenmektedir.

Bütün durumlarda, düzeltme yönergesi verilen grupta yer alan öğrenciler üzerinden hesaplanan madde ayırıcılık gücü değerleri ortalamaları ve medyanları, düzeltme yönergesi verilmeyen grupta yer alan öğrenciler üzerinden hesaplanarlardan daha yüksektir ve aralarındaki fark istatistiksel olarak anlamlıdır. Buna göre, testin düzeltme yönergesi verilerek uygulandığı durumlarda hesaplanan ayırıcılık gücü indeksleri, düzeltme yönergesi verilmeyen durumlarda uygulandığında hesaplanarlardan daha yüksek olarak gözlenmiştir. Grubun geneli için düzeltme yönergesi verilen ve verilmeyen gruplardan hesaplanan madde ayırıcılık gücü değerleri arasında istatistiksel olarak anlamlı korelasyonlar gözlenmiş ve bu değerler düzeltilmemiş puanlar için hesaplandığında, 0.39; düzeltilmiş puanlar için ise 0.42 olarak bulunmuştur. Ancak başarılı ve başarısız öğrenciler için bu korelasyon değerleri düşmüş ve istatistiksel olarak anlamlı bulunmamıştır.

Düzeltilmiş ve düzeltilmemiş puanlar üzerinden hesaplanan madde ayırıcılık gücü değerleri ortalamaları arasında, düzeltilmemiş puanlar üzerinden hesaplanarlardan lehine küçük farklılıklar bulunmuş ve bu farkların düzeltme yönergesi verilmeyen grubun genelinde ve düzeltme yönergesi verilen başarısız öğrencilerde istatistiksel olarak anlamlı olduğu görülmüştür. Diğer bir deyişle, düzeltilmiş puanlar üzerinden hesaplanan ayırıcılık gücü değerleri daha düşük olmakta, bazı durumlarda bu fark istatistiksel olarak da anlamlı hale gelmektedir. Ancak ortalamalar arasındaki fark, anlamlı olanlarda dahi, 0.005 ve 0.011 gibi çok küçük farklardır. Düzeltilmemiş ve düzeltilmiş puanlar üzerinden hesaplanan madde ayırıcılık gücü indeksi değerleri arasında 0.97 ile 0.99 arasında değişen istatistiksel olarak anlamlı korelasyonlar bulunmuştur (Tablo 6).

Tablo 6. Düzeltme yönergesi verilen ve verilmeyen gruplardan hesaplanan madde ayırıcılık gücü indeksleri [çift serili korelasyon ($r_{j\phi}$)] değerlerinin karşılaştırılmasına ilişkin sonuçlar

Grup	Puan Türü	Düzeltme Yönergesi Verilen			Düzeltme Yönergesi Verilmeyen			r_a	Z_a
		n	$\bar{r}_{j\phi}$	Medyan	N	$\bar{r}_{j\phi}$	Medyan		
Genel	Düzeltilmemiş Puan	226	.449	.460	226	.362	.375	.39*	3.11*
	Düzeltilmiş Puan		.442	.450		.357	.365	.42*	3.04*
	r_b	.99*			.99*				
	Z_b	1.53			2.14*				
Başarılı	Düzeltilmemiş Puan	100	.410	.430	100	.294	.290	.07	3.54*
	Düzeltilmiş Puan		.405	.420		.290	.280	.16	3.41*
	r_b	.98*			.98*				
	Z_b	0.92			0.94				
Başarısız	Düzeltilmemiş Puan	100	.357	.340	100	.255	.235	.21	2.86*
	Düzeltilmiş Puan		.346	.320		.247	.245	.23	2.75*
	r_b	.97*			.98*				
	Z_b	2.07*			1.56				

* $p < 0.05$ r_a : Düzeltme yönergesi verilen ve verilmeyen gruplardan hesaplanan madde ayırıcılık indeksleri ($r_{j\phi}$) değerleri arasındaki ilişki. r_b : Düzeltilmemiş ve düzeltilmiş puanlardan hesaplanan madde ayırıcılık indeksleri ($r_{j\phi}$) değerleri arasındaki ilişki. Z_a : Düzeltme yönergesi verilen ve verilmeyen koşullarda ve başarıları farklı gruplardan elde edilen madde ayırıcılık indeksleri ($r_{j\phi}$) ortalamaları arasındaki farka ait Z değeri. Z_b : Düzeltilmemiş ve düzeltilmiş puanlardan hesaplanan madde ayırıcılık indeksleri ($r_{j\phi}$) değerleri ortalamaları arasındaki farka ait Z değeri.

4. TARTIŞMA

Düzeltme yönergesi verilen ve verilmeyen koşullarda L1-MBT'yi alan öğrencilerin, düzeltilmemiş test puanları ortalamaları arasında istatistiksel olarak anlamlı bir farklılık yoktur. İstatistiksel olarak anlamlı olmamakla birlikte, düzeltme yönergesi verilen gruptaki öğrencilerin düzeltilmemiş test puanları ortalaması, düzeltme yönergesi verilmeyen gruptaki öğrencilerin düzeltilmemiş test puanı ortalamasından daha yüksektir ve aralarındaki fark, test puanlarında yapılan düzeltme ile daha belirginleşmektedir. L1-MBT'nin açıklamasında düzeltme yönergesi verilmediği halde, puanlarda şans başarısı için düzeltme yapmak, sadece başarılı gruptaki öğrencilerin puanlarının, test açıklamasında düzeltme yönergesi verilen diğer başarılı grubun ortalamasından anlamlı ölçüde daha düşük olmasına yol açmıştır. Angoff (1989) tarafından ortaya konulan "eşit yetenek düzeyindeki iki gruba, bir test farklı yönergelerle verildiğinde, puanlarda yapılan düzeltme, iki grubun ortalamasını eşitlet" hipotezi bu çalışmada sadece başarılı gruptaki öğrenciler için doğrulanmıştır.

Grubun genelinde, düzeltilmiş ve düzeltilmemiş puanlar arasında oldukça yüksek korelasyonlar gözlenmiştir ($r=0.99$). Bu korelasyon değerleri, başarılı ve başarısız grupta yer alan öğrenciler için de benzer şekilde olup, en düşük değerler başarısız grupta yer alan öğrencilerin düzeltilmiş ve düzeltilmemiş test puanları arasında ($r=0.97$) saptanmıştır. Mehrens ve Lehman (1984), yapılan çalışmalarda bu korelasyon değerinin 0.95 ve üzerinde olduğunu belirtmiştir.

Düzeltilme yönergesi verilmesi ve bu koşuldaki test puanlarının düzeltilmesiyle güvenilirlikte, düzeltme yönergesi verilmemesi ve düzeltme uygulanmaması durumuna göre 0.20 artış meydana gelmiştir. Testin düzeltme yönergesi verilerek uygulanması durumunda, puanlara düzeltme işlemi

uygulanırsa da uygulanmasa da, KR-20 iç tutarlılık katsayıları, testin düzeltme yönergesi verilmeden uygulanması durumuna göre daha yüksektir.

Tüm koşullarda, düzeltilmiş test puanları üzerinden hesaplanan KR-20 iç tutarlılık katsayıları, düzeltilmemiş test puanları üzerinden hesaplanan katsayılardan yüksektir. Bu durum, düzeltilmiş test puanları varyansının, düzeltilmemiş test puanları varyansından yüksek olmasının testin iç tutarlılık anlamındaki güvenilirliğini de yükselttiği biçiminde yorumlanabileceği gibi, bu bulgu pek çok araştırma sonucu ile de (Burton, 2001a; Muijtijens ve diğerleri, 1999; Burton ve Miller, 1999; Harris, Changas ve Palmore, 1996; Telli, 1993; Abu-Sayf ve Diamond, 1976) paralellik göstermekte ve rasgele tahminin güvenilirlik üzerine olumsuz etkisinin olduğu genel kabulünü doğrulamaktadır.

Testin düzeltme yönergesi ile uygulanıp uygulanmaması veya test puanlarında düzeltme yapıp yapılmaması, test puanlarının geçerliliğini istatistiksel olarak anlamlı şekilde değiştirmemektedir. Literatürde de, şans hatasını önleme çabalarının, testin geçerliliğini istatistiksel olarak anlamlı şekilde artırdığına ilişkin bir bulguya rastlanmamaktadır. Matematik dersi yılsonu başarı notları ile test puanları arasındaki ilişkinin düzeltme yönergesi verilen grupta azalması, Abu-Sayf ve Diamond'ın (1976) çalışmasındaki bulguyla paralellik göstermektedir.

Testin düzeltme yönergesi verilerek veya verilmeden uygulanmış olması, madde güçlük indekslerinin değişmesine yol açmamıştır. Bu durum, başarılı ve başarısız gruptaki öğrenciler için de geçerlidir. Bu sonuç, düzeltme yönergesi verilmeyen grupta, soruların şansla doğru yanıtlanma olasılığının maddeleri olduğundan daha kolay göstereceği kanısını desteklememektedir.

Testin düzeltme yönergesi verilmeden uygulanması durumunda, maddelerin rasgele tahminle yanıtlanma oranının daha fazla olması nedeniyle, maddeler olduğundan daha az ayırdedici görünmektedirler. Burton (2001b), Harris ve diğerleri (1996) ve Yurdabakan (1991) da çalışmalarında benzer sonuçlara ulaşmışlardır. Ayırıcılık gücü indekslerinin düzeltilmiş test puanları üzerinden hesaplanması ise maddelerin ayırıcılık güçlerini değiştirmemektedir.

5. SONUÇ ve ÖNERİLER

Araştırmada, başarı testinde şans başarısı için düzeltme yönergesi bulunmasının, testin test ve madde istatistiklerini nasıl etkilediği test edilmiştir.

1) Düzeltme yönergesi verilen ve verilmeyen iki grubun düzeltilmemiş test puanları ortalamaları arasında fark bulunmamıştır. Ancak, test puanlarını şans başarısından arındırmak için yapılan düzeltme sonucunda, düzeltilmiş test puanları ortalamaları, düzeltilmemiş test puanları ortalamasından istatistiksel olarak anlamlı şekilde düşük çıkmıştır. Bununla birlikte, iki grubun düzeltilmiş puan ortalamaları arasında da anlamlı bir fark meydana gelmemiştir. Bunun yanında, sadece düzeltme yönergesi verilen başarılı öğrencilerin düzeltilmiş test puanları ortalaması, düzeltme yönergesi verilmeyen başarılı öğrencilerin düzeltilmiş test puanları ortalamasından istatistiksel olarak anlamlı biçimde yüksektir. Bu durum, düzeltme yönergesi verilerek test puanlarına şans başarısının karışmasının engellenmesinin tamamen mümkün olmadığını, sadece başarılı bir öğrenci grubunda bir miktar etkili olabildiğini göstermektedir.

2) Düzeltmiş ve düzeltilmemiş test puanları arasındaki ilişki her iki test yönergesi koşulunda da 0.99 gibi oldukça yüksek bir düzeyde bulunmuştur. İlişki miktarının bu derece yüksek bulunması, düzeltme yönergesi verilse de verilmese de, test puanlarında şans başarısı için yapılan düzeltmenin, testi alanların sıralamasını neredeyse hiç değiştirmedeğini ortaya koymaktadır.

3) Test puanlarında şans başarısı için düzeltme yapmak, test puanı ortalamalarının küçülmesine neden olurken, standart sapmalarının ve varyanslarının büyümesine neden olmuştur. Düzeltme yönergesi verilmesi durumu, verilmemesi durumuna göre test puanları dağılımının çarpıklığı artmış ve bu durum başarısız öğrenciler için daha da belirginleşmiştir. Düzeltme yönergesi verilen grupta düzeltilmemiş ve düzeltilmiş test puanlarının dağılımının sivriliği, yönerge verilmeyen koşula göre normale daha yakın ve başarılı grupta normalden daha basık bulunmakla birlikte, düzeltme yönergesi vermek başarısız öğrencilerin test puanları dağılımını oldukça sivri hale getirmiştir.

4) Testin düzeltme yönergesi ile uygulandığı durumda, testin iç tutarlılık güvenilirliği düzeltme

yönergesiz uygulandığı duruma göre daha yüksek bulunmuştur. Test puanlarında yapılan düzeltme de testin güvenilirliğini artırmaktadır. En büyük artış, düzeltme yönergesi verilmeyen başarısız öğrencilerde meydana gelmiştir.

5) Düzeltilmemiş test puanlarının öğretmen notlarıyla gösterdiği ilişki, istatistiksel olarak anlamlı olmamakla birlikte, düzeltme yönergesi verilmeyen grupta daha yüksektir. Puanlarda yapılan düzeltme, testin geçerliliğini çok küçük bir miktarda artırmıştır ancak bu fark da istatistiksel olarak anlamlı değildir.

6) Düzeltme yönergesi verilen ve verilmeyen gruptan hesaplanan madde güçlük indeksleri arasında istatistiksel olarak anlamlı bir fark bulunmamıştır. Yani, düzeltme yönergesi vermek, soruları daha güç hale getirmemiştir. Bu durum başarılı ve başarısız öğrenci gruplarından yapılan hesaplamalar için de değişmemektedir.

7) Düzeltme yönergesi vermek, test maddelerinin çift serili korelasyon katsayısı olarak hesaplanan ayırıcılık gücü indekslerinin daha yüksek olmasına neden olmuştur. Bu durum başarılı ve başarısız öğrenciler için de aynı şekilde gerçekleşmektedir. Düzeltilmiş puanlar üzerinden hesaplanan ayırıcılık güçleri, düzeltilmemiş puanlardan hesaplanandan istatistiksel olarak anlamlı şekilde düşük olmakla birlikte, aralarındaki farklar 0.01 veya daha az düzeydedir.

Araştırma bulgularının ışığında, seçmeli testlerde düzeltme uygulamanın testi alanların sıralamasını hiç değiştirmemesi nedeniyle gereksiz bir çaba olduğu söylenebilir ancak test uygulayıcıları, testlerin düzeltme yönergesi verilmeden uygulanması durumunda, madde ayırıcılık güçlerinin ve iç tutarlılığın daha düşük olacağını göz önünde bulundurmalıdır. Şansla kazanılabilecek başarıyı herkes için eşitlemenin yolu, herhangi bir düzeltme uygulamadan, öğrencilerin bütün soruları yanıtlamalarını sağlamakla mümkün olabilecektir.

Bu çalışma, farklı testler ve farklı örneklem gruplarıyla tekrarlanarak bu alandaki bilgi birikimine yeni katkılar getirilebilir. Ayrıca, yeni çalışmalarda kullanılacak testlerin Modern Test Kuramı'na (Madde Tepki Kuramı) göre geliştirildikten sonra bu tür araştırma soruları yeniden ele alınarak iki kurama göre elde edilen sonuçlarda farklılık olup olmadığı test edilebilir.

KAYNAKLAR

- Abu-Sayf, F. K., Diamond J. J. (1976). Effect of confidence level in multiple-choice test answers on reliability and validity of scores. *Journal of Educational Research*, 7(2), 62-63.
- Angoff, W. H. (1989). Does guessing really help? *Journal of Educational Measurement*, 26(4), 323-336.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulanması*, Ankara: ÖSYM Yayınları.
- 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.
- Burton, R. F. (2001a). Quantifying the effects of chance in multiple choice and true/false tests: Question selection and guessing of answers. *Assessment & Evaluation in Higher Education*, 26 (1), 41-50.
- Burton, R. F. (2001b). Do item-discrimination indices really help us to improve our tests?. *Assessment & Evaluation in Higher Education*, 26(3), 213-220.
- Burton, R. F., Miller, D. J. (1999). Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainty attributable to guessing. *Assessment & Evaluation in Higher Education*, 24 (4), 399-411.
- Crocker, L., Algina, J. (1986). *Introduction classical and modern test theory*. USA: CBS College Publishing Company.
- Cronbach, L. J. (1984). *Essentials of psychological testing*. New York: Harper & Row.
- Ebel, R. L. (1965). *Measuring educational achievement*. New Jersey: Prentice-Hall, Inc.
- Frary, R. B. (1989). The effect of inappropriate omissions on formula scores: a simulation study. *Journal of Educational Measurement*, 26 (1), 41-53.
- Gronlund, N. E. (1976). *Measurement and evaluation in teaching (Third edition)*. New York: Macmillan Publishing Co.
- Harris, D. K., Changas, P. S., Palmore, E. B. (1996). Palmore's first facts on aging quiz in a multiple-choice format. *Educational Gerontology*, 22(6), 575-589.
- Lord, F. M. (1963). Formula scoring and validity. *Educational and Psychological Measurement*, 23, 663-672.
- Lord, F. M. (1975). Formula scoring and number right scoring. *Journal of Educational Measurement*, 12, 7-11.
- Mattson, D. (1965). The effects of guessing on the standard error of measurement and reliability of test scores. *Educational and Psychological Measurement*, 25, 727-730.
- Mehrens, W. A., Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology (Third edition)*. New York: CBS College Publishing.
- Muijtjens, A. M. M., Mameren, H., Hoogenboom, R. J. I., Evers, J. L. H., Vleuten, C. P. M. (1999). The effect of a 'don't know' option on test scores: number-right and formula scoring compared. *Medical Education*, 33, 267-275.
- Remmers, H. H., Gage, N. L., Rummel, J. F. (1965). *A practical introduction to measurement and evaluation (Second edition)*. New York: Harper & Row, Publishers.
- Şahhüseyinoğlu, D. (1998). *Sayısal yetenek testlerinde seçenek sayısının test ve madde istatistikleri üzerindeki etkisinin şans başarısı ile birlikte incelenmesi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, yayımlanmamış yüksek lisans tezi.
- Telli, A. (1993). *Şans başarısının madde türlerindeki madde ve test istatistikleri üzerine etkisi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, yayımlanmamış yüksek lisans tezi.