

## ÇOKTAN SEÇMELİ TESTLERİN KLASİK TEST TEORİSİ VE ÖRTÜK ÖZELLİKLER TEORİSİNE GÖRE HESAPLANAN PSİKOMETRİK ÖZELLİKLERİNİN İKİ KATEGORİLİ VE AĞIRLIKLANDIRILMIŞ PUANLANMASI YÖNÜNDEN KARŞILAŞTIRILMASI\*

### A COMPARISON OF PSYCHOMETRIC CHARACTERISTICS OF MULTIPLE CHOICE TESTS BASED ON THE BINARY AND WEIGHTED SCORING IN RESPECT TO CLASSICAL TEST THEORY AND LATENT TRAIT THEORY

Devrim ÖZDEMİR\*\*

**ÖZET:** Bu araştırmada, çoktan seçmeli test maddelerini iki kategorili (1,0) ve ağırlıklı (1,2,3,4) puanlama yöntemlerinin testin güvenilirlik ve geçerliğine etkisi, klasik test teorisi ve örtük özellikler teorisine göre incelenmiştir.

Araştırma verileri, 2001-2002 öğretim yılında çeşitli ilköğretim okullarının 4., 5., 6. ve 7. sınıflarında okuyan 1608 öğrenciye uygulanan çoktan seçmeli bir testle elde edilmiştir.

Araştırma bulgularına göre, test geliştirme çalışmalarında örtük özellikler teorisinin ve (1,0) puanlamadan yararlanılmasının daha uygun olacağı düşünülmektedir. Klasik test teorisiyle yapılacak çalışmalarda ise ağırlıklı puanlamadan yararlanılması önerilebilir.

**Anahtar sözcükler:** Puanlama yöntemleri, iki kategorili puanlama, ağırlıklı puanlama, klasik test teorisi, örtük özellikler teorisi, derecelendirilmiş cevap modeli.

**ABSTRACT:** In this research, the effects of binary scoring (1,0) and weighted scoring (1,2,3,4) methods to the reliability and validity of the test have been analysed regarding classical test theory and latent trait theory.

The data were collected through the administration of a multiple choice test to 1608 students of 4., 5., 6. and 7. grades of various primary schools in 2001-2002.

Regarding the results of the study, it has been concluded that the use of latent trait theory and binary scoring for the test development studies can be more suitable. It is also recommended that the use of weighted scoring is suitable for the test development studies made through classical test theory.

**Keywords:** Scoring methods, binary scoring, weighted scoring, classical test theory, latent trait theory, graded response model.

## 1. GİRİŞ

Test geliştirmede kullanılan yöntemlerin hepsinde temel amaç, uygun maddeleri seçerek istenilen nitelikte testler meydana getirmektir. Bunu sağlayabilmek amacıyla, madde analizi çalışmalarından yararlanılır (Baykul, 1979). Madde analizi çalışmalarında en çok kullanılan teori, klasik test teorisidir. Klasik test teorisinde, bireyin bir testteki gözlenen puanı, gerçek puanın ve hata puanlarının toplamı olarak gösterilir. Teori, ölçme hatalarının tesadüfi olduğu varsayımına dayalıdır. Bu hatalar, gerçek puanlarla ve farklı ölçmelerle korelasyon vermezler. Klasik test teorisinde sorunların çoğu, elde edilen madde ve test istatistiklerinin gruba bağımlı olmasından kaynaklanmaktadır. Bu sorunu ortadan kaldırmanın bir yolu olarak “örtük özellikler teorisi” adı altında bir başka teori geliştirilmiştir. Örtük özellikler teorisinin güçlü yanları üç maddeye sıralanabilir; (1) bir madde, o maddeyi cevaplandırabilecek herhangi bir cevaplayıcı grubundan bağımsızdır; (2) bir bireyin yeteneği, o bireye uygulanan herhangi bir madde grubundan bağımsızdır; (3) bir tes-

\* Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Eğitimde Ölçme ve Değerlendirme Bilim Dalı Yayınlanmamış Doktora Tezi, 2002

\*\* Yrd. Doç. Dr., Mersin Üniversitesi Eğitim Fakültesi, devrimozdemir@mersin.edu.tr

tin özelliklerini, testi uygulamadan önce kestirmek mümkündür (Keeves, 1997). Teorinin bir sınırlılığı, parametrelerin kestirilmesi için büyük bir örneklem kullanılmasını gerektirmesidir. Her iki teoride de maddeye verilen cevapların puanlanması birbirine benzemektedir. Bu teorilerde kullanılan puanlama yöntemleri, iki kategorili ve çok kategorili yöntemler olmak üzere iki grupta incelenebilir (Linden ve Hambleton, 1997).

Bu araştırmada, iki kategorili puanlama yöntemiyle birlikte, kısmi bilgiyi de dikkate alan ve seçeneklerin farklı ağırlıklandırıldığı çok kategorili puanlama yöntemlerinden, klasik test teorisi için önsel ağırlıklandırma yöntemi ve örtük özellikler teorisi için derecelendirilmiş cevap modeli ile çalışılmıştır. Önsel ağırlıklandırmada, her bir maddenin seçeneklerinin ağırlıklandırılmasında uzman görüşlerine başvurulur ve uzmanların seçeneklere verdikleri ağırlıklara göre puanlama cetveli geliştirilir (Echternacht, 1976). Ağırlıklı seçenekler için IRT modelleri ilk kez 1969 yılında Samejima tarafından geliştirilmiştir. Samejima'nın geliştirdiği bu modeller, derecelendirilmiş cevap modelleri olarak adlandırılır ve sadece doğruluk derecesine göre sıralanabilen seçenekler için kullanılabilir (Frery, 1989).

### 1.1. Araştırmanın Önemi

Klasik test teorisinde farklı gruplara ait sonuçların karşılaştırılması söz konusu olmazken, örtük özellikler teorisi buna imkân vermektedir, örtük özellikler teorisinin gücü, sayıtların yerine getirilmesine bağlıdır. Bu araştırmada, yapılan karşılaştırmalar sonucunda, sonuçlar arasında farklılık olmadığı belirlenirse, çoktan seçmeli testlerin, örtük özellikler teorisinin gerektirdiği güçlü sayıtlara gitmeksizin, klasik test teorisi yardımıyla puanlanması ve test istatistiklerinin bu teoriye göre hesaplanması yoluna gidilebileceği şeklinde bir yargı elde edilmiş olacaktır.

### 1.2. Araştırma Problemi

Bu araştırmada şu alt sorulara cevap aranmıştır:

1. Testin iki kategorili (1,0) olarak puanlanması durumunda, klasik test teorisi ve örtük özellikler teorisine göre elde edilen güvenilirlikleri nasıldır?
2. Testin ağırlıklı (1,2,3,4) olarak puanlanması durumunda, klasik test teorisi ve örtük özellikler teorisine göre elde edilen güvenilirlikleri nasıldır?
3. Testin klasik test teorisine göre iki kategorili (1,0) ve ağırlıklı (1,2,3,4) olarak puanlanması durumunda elde edilen güvenilirlikleri nasıldır?
4. Testin örtük özellikler teorisine göre iki kategorili (1,0) ve ağırlıklı (1,2,3,4) olarak puanlanması durumunda elde edilen güvenilirlikleri nasıldır?
5. Testin örtük özellikler teorisine göre iki kategorili (1,0) ve ağırlıklı (1,2,3,4) olarak puanlanması durumunda elde edilen geçerlikleri nasıldır?
6. Testin klasik test teorisine göre iki kategorili (1,0) ve ağırlıklı (1,2,3,4) olarak puanlanması durumunda elde edilen geçerlikleri nasıldır?

### 1.3. Sayıtlar

1. Uygulamanın yapıldığı tüm sınıflarda test, yönergeye uygun bir biçimde uygulanmıştır.
2. Öğrenciler test maddelerini cevaplarırken gerçek bilgilerini yansıtmışlardır.

### 1.4. Sınırlılıklar

1. Araştırma, uygulamanın yapıldığı okullarla sınırlıdır.
2. Araştırma, uygulamanın yapıldığı ilköğretim 4, 5, 6 ve 7. sınıflarıyla sınırlıdır.
3. Araştırma, Türkçe okuduğunu anlama yeteneği ile sınırlıdır.

## 2. YÖNTEM

Bu bölümde, araştırmanın türü, verilerin elde edildiği grup, ölçme aracının hazırlanması, verilerin elde edilmesi ve analizi ile ilgili açıklamalara yer verilmiştir.

### 2.1. Araştırmanın Türü

Bu çalışmada, örneklem bilgilerinin evrene genellenmesi amacı güdülmeyip, iki teoriye göre puanlama yöntemleri hakkındaki temel özellikler üzerinde durulmuştur. Bu nedenle araştırmanın “temel araştırma” niteliğinde olduğu söylenebilir.

### 2.2. Verilerin Elde Edildiği Grup

Araştırmanın verileri, 2001-2002 öğretim yılı güz döneminde çeşitli ilköğretim okullarında 4., 5., 6. ve 7. sınıflarda okuyan 1608 öğrenciye uygulanan ve Türkçe okuduğunu anlama yeteneğini ölçmeye yönelik olarak hazırlanan, her biri 4 seçenekli 20 maddelik bir çoktan seçmeli test yardımı ile elde edilmiştir.

### 2.3. Ölçme Aracının Hazırlanması

Testin deneme formunun geliştirilmesi çalışmaları sonrasında, 55 maddelik bir test formu elde edilmiştir. Hazırlanan form, 9 Türkçe uzmanına verilmiş ve bu uzmanlardan, her bir madde için seçenekleri ağırlıklandırmaları istenmiştir. Uzmanlardan alınan ağırlıklı puanların uyumlu olup olmadığı Kendall'ın W uyum istatistiği yardımıyla incelenmiştir. Elde edilen sonuçlara göre, 5 maddede uzmanlar arasında uyum olmadığı gözlenmiştir ( $\alpha=0,05$ ). Bu analizler sonucunda oluşturulan 50 maddelik deneme formu, iki ayrı ilköğretim okulunda okuyan toplam 237 öğrenciye uygulanmıştır. Deneme uygulamasına ilişkin madde analizi çalışmaları hem klasik test teorisi hem de örtük özellikler teorisi için yapılmıştır. Bu çalışmalar sonucunda, 20 maddenin nihai forma alınmasına karar verilmiştir.

### 2.4. Verilerin Çözümlemesi

Verilerin analizinden önce, örtük özellikler teorisinin sayıtlarının (normallik, yerel bağımsızlık, tek boyutluluk) karşılanıp karşılanmadığı ve model-data uyumu incelenmiştir.

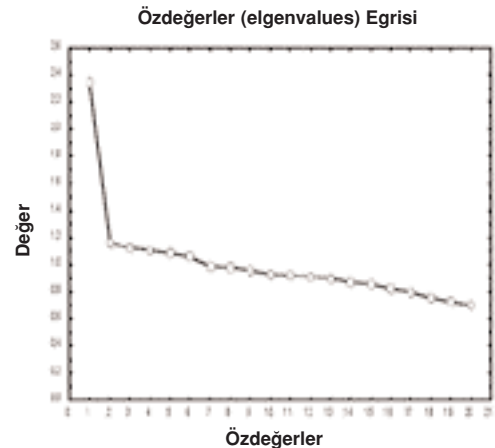
#### Normallik Testi:

Örtük özellikler teorisinde bireyin ölçülen özelliğinin evrende normal dağıldığı varsayımı yer almaktadır. Dağılımın normal olup olmadığı hakkında karar verebilmek için Kolmogorov Smirnov tek örneklem testinden yararlanılmış, ve test istatistiği  $d=0,039$  bulunmuştur. Elde edilen bu sonuca göre, deneysel dağılım ile kuramsal dağılım arasında manidar bir farklılık olmadığı ( $\alpha=0,05$ ); deneme testine ilişkin veri kümesinin normal dağıldığı söylenebilir.

#### Tek Boyutluluk Testi:

Testin tek boyutlu olup olmadığının belirlenmesinde, maddelerarası tetrakorik korelasyon katsayısı hesaplanmıştır. Faktör analizinde tetrakorik korelasyon matrisinden yararlanılmıştır. Şekil 2.4.1.'de 20 maddelik nihai forma ilişkin özdeğerler eğrisi yer almaktadır.

Ölçülen özelliğin tek boyutlu olmasının bir göstergesi, birinci faktöre ait özdeğerin ikinci faktöre ait özdeğerden büyük olması ve diğer faktörlerin büyük farklılıklar gösterme-



Şekil 2.4.1. Testin 20 maddelik nihai formuna ilişkin özdeğerler eğrisi

yen özdeğerlere sahip olmasıdır (Lord, 1980). Özdeğerler eğrisine göre, veri kümesinin tek bir boyutta ele alınamayacağı görülmektedir.

#### *Yerel Bağımsızlık Testi:*

Tek boyutluluk varsayımının gerçekleşmesi, yerel bağımsızlık varsayımının sağlandığının da göstergesidir (Hambleton ve Swaminathan, 1992). Nihai forma ilişkin verilerin tek boyutlu olup olmadığının belirlenmesinde elde edilen faktör analizi sonuçları yerel bağımsızlığın test edilmesi için kullanıldığında, yerel bağımsızlık şartının yerine getirilmediği söylenebilir.

#### *Model-Data Uyumunun Belirlenmesi:*

Model ve verilerin uyumunun test edilmesinde yararlanılan istatistiklerden biri,  $c^2$  dağılımı gösteren  $-2 \log l$  istatistiğidir (Tatlıdıl, 1996; Keeves, 1997). İki parametrelili modele göre  $-2 \log l$  istatistiği, 15806,9 olarak elde edilmiştir. Elde edilen bu sonuç, verinin iki parametrelili modele uyumlu olduğunu göstermektedir ( $\alpha=0,05$ ).

Her bir alt soruya ilişkin güvenilirlik araştırmasının yapılabilmesi amacıyla klasik test teorisinde (1,0) puanlama için KR-20 ve ağırlıklı puanlama için Cronbach  $\alpha$  güvenilirlik katsayıları hesaplanmıştır. Örtük özellikler teorisinde (1,0) puanlama için Lord'un önerdiği güvenilirlik katsayısı, ağırlıklı puanlama içinse marjinal güvenilirlik katsayıları elde edilmiştir. İki puanlama yönteminin test geçerliğine etkisini ortaya koyabilmek amacıyla, 4., 6. ve 7. sınıflara uygulanan testlerden elde edilen puanlar ölçüt olarak kabul edilerek, klasik test teorisi için bu testlerden elde edilen puanların ortalamaları 5. sınıflara uygulanan testten elde edilen puanların ortalaması; örtük özellikler teorisi için farklı sınıflardaki cevaplayıcıların  $q$  yetenekleri karşılaştırılmıştır. Ortalamalar arasında ve  $q$  yetenekleri arasında manidar bir farklılık olup olmadığı ilişkisiz örneklem için  $t$  testi yardımıyla test edilmiştir.

### **3. BULGULAR VE YORUMLAR**

Bu bölümde, araştırma sonucunda elde edilen verilerin bulgularına yer verilmiş ve yorumlar yapılmıştır.

Klasik test teorisi ve örtük özellikler teorisinde (1,0) puanlamaya dayalı olarak elde edilen güvenilirliklere ilişkin sonuçlar Tablo 3.1.'de verilmiştir. Tablo 3.1.'de, klasik test teorisi ve örtük özellikler teorisinde (1,0) puanlamaya göre elde edilen güvenilirlikler incelendiğinde, klasik test teorisinde grubun geneline ilişkin elde edilen güvenirliliğin (0,49) örtük özellikler teorisinden elde edilen güvenirliliğe (0,78) göre düşük olduğu görülmektedir. Yine, sınıflar düzeyinde güvenilirlikler incelendiğinde, örtük özellikler teorisi ile elde edilen güvenilirliklerin klasik test teorisinde elde edilen güvenilirliklere göre yüksek olduğu görülmektedir.

**Tablo 3.1.** KTT ve IRT'de (1,0) Puanlamaya Göre Elde Edilen Güvenirlik Sonuçları

Sınıf	N	KR-20 Güv.	Lord'un Güv. Kats.
4.	222	0,44	0,76
5.	887	0,45	0,75
6.	337	0,50	0,80
7.	162	0,57	0,84
Genel	1608	0,49	0,78

Klasik test teorisi ve örtük özellikler teorisinde ağırlıklı puanlamaya dayalı olarak elde edilen güvenilirliklere ilişkin sonuçlar Tablo 3.2.'de verilmiştir. Tablo 3.2.'de klasik test teorisi ve örtük özellikler teorisinde ağırlıklı puanlamaya göre elde edilen güvenilirlikler incelendiğinde, örtük özellikler teorisinden grubun geneline ilişkin olarak elde edilen güvenirliliğin (0,68) klasik test teorisinde elde edilen güvenirliliğe (0,54) göre yüksek olduğu görülmektedir.

Tablo 3.2.'de, sınıflar düzeyinde güvenilirlikler incelendiğinde de, örtük özellikler teorisi ile elde edilen güvenilirliklerin klasik test teorisinde elde edilen güvenilirliklere göre yüksek olduğu görülebilir. Her iki teoride de en yüksek güvenilirliklerin 7. sınıflara uygulanan test yardımıyla elde edildiği söylenebilir (0,59 ve 0,73).

Klasik test teorisinde (1,0) ve ağırlıklı puanlamaya dayalı olarak elde edilen güvenilirliklere ilişkin bilgiler Tablo 3.3.'de verilmiştir.

Klasik test teorisinde (1,0) ve ağırlıklı puanlamaya göre elde edilen güvenilirlikler karşılıklı olarak incelendiğinde, (1,0) puanlamada grubun geneline ilişkin olarak elde edilen güvenirliliğin (0,49) ağırlıklı puanlamadan elde edilen güvenirliliğe (0,54) göre yüksek olduğu görülmektedir. Sınıflar düzeyinde güvenilirlikler incelendiğinde de, ağırlıklı puanlanan testin (1,0) puanlamaya göre daha güvenilir sonuçlar verdiği gözlenmektedir.

Örtük özellikler teorisinde (1,0) ve ağırlıklı puanlamaya dayalı olarak elde edilen güvenilirliklere ilişkin bilgiler Tablo 3.4.'de verilmiştir.

**Tablo 3.3.** IRT'de (1,0) ve (1,2,3,4) Puanlamaya Göre Elde Edilen Güvenirlik Sonuçları

Sınıf	N	(1,0) Lord'un Güv. Kats.	(1,2,3,4) Marjinal Güvenirlik
4.	222	0,76	0,69
5.	887	0,75	0,65
6.	337	0,80	0,72
7.	162	0,84	0,73
Genel	1608	0,78	0,68

test geçerliğine etkisini ortaya koyabilmek amacıyla, 4., 6. ve 7. sınıflardaki öğrencilerin q yetenekleri ölçüt kabul edilmiş ve bu yeteneklerin ortalamaları ile 5. sınıftaki öğrencilerin q yeteneklerinin ortalaması arasındaki fark test edilmiş ve elde edilen sonuçlar Tablo 3.5.'de verilmiştir. Tablo 3.5.'den görülebileceği gibi, hem (1,0) puanlama hem de ağırlıklı puanlama için 4-5. ve 5-7. sınıflara uygulanan testlerin ortalamaları arasında, 4-5. sınıflara uygulanan testlerde 5. sınıf öğrencilerinin q yetenek ortalamaları lehinde; 5-7. sınıflara uygulanan testlerde ise 7. sınıf öğrencilerinin q yetenek ortalamaları lehinde 0,05 düzeyinde manidar bir farklılık elde edilmiştir. Buna göre, hem ağırlıklı hem de (1,0) puanlama yönteminin 6. sınıflar dışındaki grupları birbirinden ayırdığı, dolayısıyla, her iki puanlama yönteminin geçerli olduğu söylenebilir.

Klasik test teorisinde iki puanlama yönteminin test geçerliğine etkisini ortaya koyabil-

**Tablo 3.2.** KTT'de (1,0) ve (1,2,3,4) Puanlamaya Göre Elde Edilen Güvenirlik Sonuçları

Sınıf	N	(1,0) KR-20 Güv.	(1,2,3,4) $\alpha$ Güvenirliği
4.	222	0,44	0,50
5.	887	0,45	0,49
6.	337	0,50	0,57
7.	162	0,57	0,59
Genel	1608	0,49	0,54

Örtük özellikler teorisinde (1,0) ve ağırlıklı puanlamaya göre elde edilen güvenilirlikler karşılıklı olarak incelendiğinde, (1,0) puanlamada grubun geneline ilişkin elde edilen güvenirliliğin (0,78) ağırlıklı puanlamadan elde edilen güvenirliliğe (0,68) göre yüksek olduğu Tablo 3.4.'de görülmektedir. Sınıflar düzeyinde güvenilirlikler incelendiğinde de, (1,0) puanlama ile elde edilen güvenilirliklerin ağırlıklı puanlama ile elde edilen güvenilirliklere göre yüksek olduğu görülmektedir.

Örtük özellikler teorisinde iki puanlama yönteminin

**Tablo 3.4.** IRT'de (1,0) ve (1,2,3,4) Puanlamaya Göre q Yeteneklerinin Karşılaştırılması

Puan	Grp.	N	$\theta$ Ort.	sd	t	p
Puanlama	4.	222	-0,34	1107	4,92	0,00*
	5.	887	-0,12			
	5.	887	-0,12	1222	1,55	0,12
	6.	337	-0,06			
Ağırlıklı Puanlama	5.	887	-0,12	1047	5,05	0,00*
	7.	162	0,14			
	4.	222	0,73	1107	5,44	0,00*
	5.	887	0,97			
	5.	887	0,97	1222	1,29	0,19
	6.	337	1,02			
	5.	887	0,97	1047	4,91	0,00*
	7.	162	1,22			

\* p < 0,05 düzeyinde manidardır.

mek amacıyla, 4., 6. ve 7. sınıflara uygulanmasından edilen puanlar ölçüt kabul edilerek bu testlerin ortalamaları ile 5. sınıflara uygulanan testin ortalaması arasındaki fark test edilmiş ve elde edilen sonuçlar Tablo 3.6.'da verilmiştir.

Tablo 3.6.'dan görülebileceği gibi, hem (1,0) puanlama hem de ağırlıklı puanlama için, 4.-5. ve 5.-7. sınıflara uygulanan testlerin ortalamaları arasında, 4.-5. sınıflara uygulanan testlerde 5. sınıflara uygulanan testin ortalaması lehinde; 5.-7. sınıflara uygulanan testlerde ise 7. sınıflara uygulanan testin ortalaması lehinde 0,05 anlamlılık düzeyinde manidar bir farklılık elde edilmiştir. Buna göre, hem ağırlıklı hem de (1,0) puanlama yönteminin 6. sınıflar dışındaki grupları birbirinden ayırdığı, dolayısıyla, her iki puanlama yönteminin geçerli olduğu söylenebilir. 4. ve 7. sınıf düzeyindeki öğrencilerin okuduğunu anlama yeteneği bakımından 5. sınıf öğrencilerinden farklı olması, buna karşılık 6. sınıf öğrencilerinin okuduğunu anlama yeteneği ile 5. sınıf öğrencilerinin okuduğunu anlama yeteneği arasında manidar bir farklılık gözlenmemesi beklenen bir durumdur. 5.-6. sınıfların okuduğunu anlama yeteneğini ölçme amacıyla geliştirilen testten aldıkları puanların ortalamaları arasında farklılık bulunmamasının nedeni, yaş ve gelişim düzeyi bakımından bu grupların birbirine yakın olması nedeniyle açıklanabilir.

#### 4. SONUÇLAR VE ÖNERİLER

Araştırma sorularının istatistiksel bulguları ışığında aşağıdaki sonuçlar elde edilmiştir.

Klasik test teorisinde a güvenilirlik katsayısının, KR-20 güvenilirlik katsayısından yüksek sonuçlar verdiği; örtük özellikler teorisinde Lord'un güvenilirlik katsayısının, marjinal güvenilirlikten yüksek olduğu gözlenmiştir. Bu sonucun, Lord'un katsayısının hesaplanmasında güvenilirliğin klasik tanımından hareketle elde edilmesi, marjinal güvenilirliğin ise değişik yetenek düzeylerinden elde edilen güvenilirliklerin ortalamasına dayanılarak hesaplanması nedeniyle gözlendiği düşünülmektedir. (1,0) ve ağırlıklı puanlama için yapılan güvenilirlik çalışmaları incelendiğinde, iki kategorili puanlamada Lord'un katsayısının KR-20'den yüksek olduğu gözlenmiştir.

Klasik test teorisi ve örtük özellikler teorisinde ağırlıklı puanlama ile elde edilen güvenilirlikler incelendiğinde, örtük özellikler teorisinde marjinal güvenilirlik hesaplamalarına dayalı olarak elde edilen sonuçların klasik test teorisinde a katsayısı yardımıyla elde edilen sonuçlardan yüksek olduğu gözlenmiştir. Klasik test teorisinde ağırlıklı puanlama ve örtük özellikler teorisinde (1,0) puanlama ile elde edilen güvenilirlikler incelendiğinde, örtük özellikler teorisinde Lord'un güvenilirlik katsayısı yardımıyla elde edilen güvenilirliğin, klasik test teorisinde a katsayısı yardımıyla elde edilen güvenilirlikten yüksek sonuçlar verdiği belirlenmiştir. Klasik test teorisinde (1,0) puanlama ve örtük özellikler teorisinde ağırlıklı puanlama ile elde edilen güvenilirlikler incelendiğinde, örtük özellikler teorisinde marjinal güvenilirlik katsayısı yardımıyla elde edilen güvenilirliğin klasik test teorisinde KR-20 güvenilirlik katsayısı yardımıyla elde edilen güvenilirlikten yüksek sonuçlar verdiği görülmektedir.

**Tablo 3.5.** KTT'de (1,0) ve Ağırlıklı Puanlamaya Göre Aritmetik Ortalamaların Karşılaştırılması

Puan	Grp.	N	Ort.	sd	t	p
(1,0) Puanlama	4.	222	8,41	1107	4,99	0,00*
	5.	887	9,44			
	5.	887	9,44	1222	-1,58	0,11
	6.	337	9,72			
Ağırlıklı Puanlama	5.	887	9,44	1047	-4,91	0,00*
	7.	162	10,61			
	4.	222	58,43	1107	-5,48	0,00*
	5.	887	61,02			
Ağırlıklı Puanlama	5.	887	61,02	1222	-1,01	0,31
	6.	337	61,43			
	5.	887	61,02	1047	-4,26	0,00*
	7.	162	63,31			

\*p < 0,05 düzeyinde manidardır.



Ölçüt grup geçerliği sonuçları incelendiğinde, (1,0) puanlama ve ağırlıklı puanlama ile elde edilen ölçme sonuçlarına ilişkin olarak klasik test teorisi için 4.-5. ve 5-7. sınıfların test puanları ortalamaları; örtük özellikler teorisi için de 4.-5. ve 5-7. sınıfların q yetenekleri ortalamaları arasında manidar farklılık olduğu belirlenmiştir.

Araştırma sonuçlarına dayalı olarak aşağıda bazı önerilere yer verilmiştir.

Bu araştırmada, örtük özellikler teorisinde Lord'un güvenilirlik katsayısı yardımıyla elde edilen güvenilirliklerin en yüksek sonucu verdiği gözlenmiştir. Lord'un güvenilirlik katsayısını, örtük özellikler teorisinde ağırlıklı puanlama ile elde edilen marjinal güvenilirlik izlemektedir. Bu durumda, örtük özellikler teorisinin klasik test teorisine göre yüksek güvenilirlik verdiği; örtük özellikler teorisi kapsamında da (1,0) ile puanlama yönteminin güvenilirlik bakımından yüksek sonuçlar ortaya koyduğu söylenebilir. Bu bakımdan, daha sonra yapılabilecek test geliştirme çalışmalarında, örtük özellikler teorisi ve örtük özellikler teorisinde de (1,0) puanlamanın kullanılması önerilebilir. Yine, klasik test teorisi ile yapılması düşünülen test geliştirme çalışmalarında ağırlıklı puanlama yönteminden yararlanmanın uygun olabileceği söylenebilir.

Araştırma sürecinde yapılan literatür çalışmalarında, çoktan seçmeli testlerle ilgili olarak seçeneklerin ve test maddelerinin ağırlıklandırılmasına ilişkin çok sayıda puanlama yöntemi bulunduğu belirlenmiştir. Hem klasik test teorisi hem örtük özellikler teorisi için çeşitli seçenek ve madde ağırlıklandırma yöntemlerinin test ve madde istatistiklerine etkisi, karşılaştırmalı olarak incelenebilir. Yine, kullanılan puanlama yöntemlerinde madde yanlılığı olup olmadığı, cinsiyet, farklı okul türleri (devlet okulu, özel okulu gibi) ve diğer değişkenler bakımından incelenebilir.

## KAYNAKLAR

- Baykul, Y. (1979). Örtük özellikler ve klasik test kuramları üzerine bir karşılaştırma. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Ben-Simon, A., Budescu, D. V. ve Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. Applied Psychological Measurement, 21(1), 65-88.
- Echternacht, G. (1976). Reliability and validity of item option weighting schemes. Educational and Psychological Measurement, 36, 301-309.
- Frary, R. B. (1989). Partial credit scoring methods for multiple choice Tests. Applied Measurement in Education, 2(1), 79-96.
- Hambleton, R. K. ve Swaminathan, H. (1992). Item response theory: Principles and application. Boston: Kluwer Academic Publishers Group.
- Jaradat, D. ve Tollefson, N. (1988). The Impact of alternative scoring procedures for multiple-choice items on t test Reliability, validity and grading. Educational and Psychological Measurement, 48, 627-635.
- Keeves, J. P. (1997). Educational research methodology and measurement: An international handbook (Second Edition). Cambridge: Cambridge University Pres.
- Linden, V.D. ve Hambleton, R.K. (1997). Handbook of modern item response theory. New York: Springer- Verlag Inc.
- Lord, F. (1980). Applications of item response theory to practical testing problems. New Jersey: Lawrence Erlbaum Associates Publishers.
- Tatlıdil, H. (1996). Uygulamalı çok değişkenli istatistiksel yöntemler. Ankara.