

ÇOKTAN SEÇMELİ TEST MADDELERİNİ PUANLAMADA, SEÇENEKLERİ FARKLI BİÇİMLERDE AĞIRLIKLANDIRMANIN MADDE VE TEST İSTATİSTİKLERİNE OLAN ETKİSİNİN İNCELENMESİ

AN INVESTIGATION ON THE EFFECTS OF DIFFERENT ITEM-OPTION SCORING METHODS ON ITEM AND TEST PARAMETERS

Oylum AKKUŞ *, Yaşar BAYKUL**

ÖZET: Bu çalışmada çoktan seçmeli test maddelerini puanlama yöntemlerinden; (1,0) metoduyla puanlama, uzman kanısına dayalı puanlama, Zinger Z_1 ve Z_2 puanlama yöntemlerinin madde ve test istatistiklerine etkisi araştırılmıştır. Araştırma, 1999-2000 Eğitim-Öğretim yılı Güz Dönemi'nde Özel Arı Lisesi 8. sınıf (30 öğrenci) ve Arı Fen Lisesi 9. (30 öğrenci), 10. (28 öğrenci) sınıf öğrencilerinden toplanan veriler üzerinde yapılmıştır. Uygulamada 4 seçenekli 20 maddeden oluşan sayılar arası ilişkileri yoklayan bir yetenek testi kullanılmıştır. Araştırmanın bulgularından; dört farklı puanlama yönteminden elde edilen madde ayırtıcılıklarının birbirinden farklı olmadığı görülmüştür. Elde edilen test güvenilirlik katsayıları yeterli düzeyde yüksek bulunmuş, ancak bu katsayılar arasında manidar bir fark çıkmamıştır. Diğer taraftan, kullanılan puanlama metodlarının test geçerliğine aynı etkiyi yaptığı sonucuna varılmıştır. Bu çalışma ve yapılan bu gibi çalışmalar göstermiştir ki; farklı madde puanlama metodları kullanıldığında elde edilen sonuç çoğu zaman, güvenilirliğin değişmemesi ya da çok az artması ve geçerliğin dikkate değer farklılık göstermemesidir.

ANAHTAR SÖZCÜKLER: (1,0) metoduyla puanlama, uzman kanısına dayalı puanlama, Zinger Z_1 ve Z_2 puanlamaları

ABSTRACT: The purpose of this research was to determine the effect of different multiple choice test item-option scoring methods of (a) conventional scoring –number right scoring -, (b) priori weighting, (c) Z_1 scoring, (d) Z_2 scoring on the item and test parameters. This study was carried out in Arı College (Grade 8, 30 students) and Arı Science High School (Grade 9, 30 student – Grade 10, 28 student) on 88 students in fall semester of the school year 1999-2000. A 20-item dealing with determining the relationship between numbers was constructed for this study. Each item has four options. The research findings can be summarized as follows; from all the item-option scoring methods used, approximately the same item discrimination

values were found. The indices of test reliabilities were found adequately high, but there was no significant relationship between these four reliability indices. It was also found that, the four item option scoring schemes did the same effect on the validity of test. Studies on validity and reliability showed that, when using numerous item scoring methods, one gets most of the time slight improvement in test reliability and no remarkable change in validity.

KEY WORDS: (1,0) scoring (number-right), priori weighting, Zinger Z_1 and Z_2 scoring.

1. GİRİŞ

Ölçme literatüründe çoktan seçmeli testlerin puanlanmasında yaygın olarak kullanılan metod, (1,0) yöntemiyle puanlamadır (doğru yanıtta 1; boş, yanlış ve birden fazla yanıt olduğu hallere 0 vererek puanlama). Bu metodun yaygın olarak kullanılmasının nedenleri arasında, binom teoremine dayalı sağlam bir istatistik temeli olması ve puanlama kolaylığı belirtilebilir.

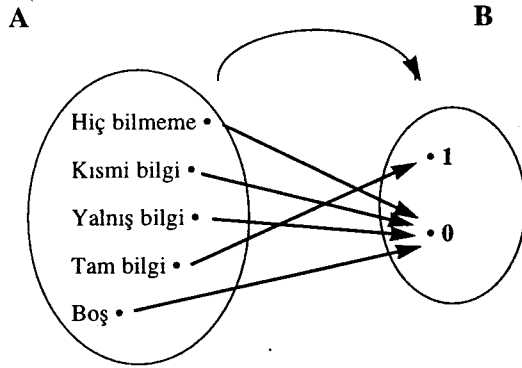
Bu puanlama yöntemine zaman zaman bazı eleştiriler getirilmiştir. Bunlardan biri, elde edilen puanların bireylerin buldukları grup içindeki sıralanışlarıyla ilgili bilgi vermesi ama bireyler arasındaki farklılıkları ortaya çıkarmada karşılaştırmaların grup dışında geçersiz kalması [1]; diğeri bu puanlama yönteminin test maddeleri hakkında fazla bilgi vermediği için, madde geçerliğini yükseltmemesidir [2].

Diğer taraftan, (1,0) ile puanlama yöntemi, sadece doğru cevap verip vermemeyi puanladığından, bireyin soru kökünde yoklanan davranış hakkında yanlış, kısmi bilgi, ya da hiç bilmeme

*Ar. Gör., Hacettepe Üniversitesi, İlköğretim Öğretmenliği Blm., İlköğretim Matematik Eğitimi ABD

** Prof. Dr., Başkent Üniversitesi, Eğitim Fakültesi, ANKARA

gibi durumları dikkate almamakta; bu durum da puanlamanın yetersizliği olarak ileri sürülmektedir. Bu puanlama öğrencilerin cevaplama durumlarına göre aşağıdaki gibi bir şekilde ifade edilebilir.



Şekil 1: (1,0) Yöntemiyle Puanlama

f fonksiyonunun (1,0) yöntemiyle puanlama metoduna göre, birebir olmadığı gözlenir. Bu puanlamanın bireyin o maddeyi tam olarak bilmesiyle hiç bilmemesi arasındaki dereceleme yer vermediği görülür. Başka bir deyişle, maddeyi hiç bilmeyen, madde hakkında kısmi bilgiye sahip olan, maddeye yanlış yanıt veren ve maddeyi boş bırakan bireylerin aynı puanı (yani 0) aldıkları gözlenir. Bu durum, bu puanlama yöntemiyle birey hakkındaki değerlendirmenin duyarlı olarak yapıldığı hakkında kuşku doğurur. Ayrıca, bu puanlama yöntemiyle eğitim ve psikolojide ölçme konusu yapılan pek çok özellik sürekli olduğu halde; ölçme sonuçları 1 ve 0 ile ifade edildiği için iki kategorili hale getirilmektedir. Bu da yukarıda da belirtildiği gibi; bilgi kaybına sebep olmaktadır.

Frery'nin [3] belirttiğine göre; Womer, Şekil 1'deki A kümesinin elemanlarını aşağıdaki gibi tanımlamıştır:

Hiç bilmeme (complete ignorance): Yanıtlayıcı ona yöneltilen madde hakkında en ufak bir bilgiye sahip değil.

Kısmi bilgiye sahip olma (partial information): Yanıtlayıcı ona yöneltilen maddenin doğru yanıtını bilmiyor ama seçeneklerden birkaçını

eleyecek kadar bir şeyler biliyor.

Yanlış bilgiye sahip olma (misinformation): Yanıtlayıcı ona yöneltilen maddenin çeldiricilerinden birini doğru yanıt sanıyor.

Tam bilgiye sahip olma (complete information): Yanıtlayıcı ona yöneltilen maddenin doğru yanıtını ve yanlış olan tüm seçenekleri biliyor.

Boş bırakma (nonmarked): Yanıtlayıcı maddeyi işaretlemiyor.

(1,0) ile puanlama metodunun olumsuz yönleri nedeniyle, psikometristler çoktan seçmeli test maddelerini puanlamanın farklı yollarını araştırmaya başlamışlardır. İlk olarak 1953 yılında Coombs "eleme puanlaması" (elimination scoring) adıyla yeni bir metot ortaya atmıştır [6]. Bu metotta, tamamıyla yanlış bilgiden başlayıp kısmi bilginin çeşitli derecelerini de içine alan ve tam bilgiyle sonlanan seçenekleri kapsayan test maddelerinden oluşturulan bir ölçek puanları değerlendirilmesi yer almaktadır. Bu puanlama yönteminde, yanıtlayıcılar testi cevaplararken yanlış olduğunu düşündükleri tüm seçenekleri, kağıt üzerinde eleyerek gösterirler [4]. Bu yöntemle puanlamada, madde seçeneklerine verilen ağırlıklar $[-(n-1)$ ile $(n-1)$] arasında değişmektedir (n =seçenek sayısı) [5]. Örneğin 4 seçenekli madde için verilecek puanlar -3 ile $+3$ aralığında yer alır, -3 tamamıyla yanlış bilgiyi temsil ederken, $+3$ ise tamamıyla doğru bilgi için verilir; arada kalan diğer tamsayılar da seçeneklerin temsil ettikleri madde bilgisi düzeylerine göre ağırlık olarak verilir.

Coombs'un çalışmalarını, çeşitli psikometristler izlemiş ve çoktan seçmeli test maddelerini farklı puanlama metodlarına ilişkin çeşitli yöntemler ortaya atılmıştır. Frery [6] bu yöntemleri genel olarak aşağıdaki gibi sınıflandırmıştır:

A) Doğrudan Yanıtlama Metodu (Direct Response Method): Bu metotta, yanıtlayıcı seçenekler arasından hangisinin doğru olduğunu düşünüyorsa onu işaretler.

1) Uzmanlar ve testi geliştiren araştırmacı tarafından yapılan önsel ağırlıklandırma (priori/ logical weighting).

2) Görgül ağırlıklandırma (empirical weighting).

3) Doğru yanıtı bulana dek işaretleme (answer-until correct method).

4) Örtük özellikler kuramına göre yanıtları derecelendirerek yapılan puanlama (graded response).

5) Maddeden maddeye göre değişen doğru seçenek sayısına göre yapılan puanlama (multiple correct options).

B) Yanıtlayıcı Yargısına Dayalı Puanlama Metodu (Examinee Judgment Method): Bu metotta; yanıtlayıcı seçenekler arasından ya yanlış olduğunu bildiği seçeneği ya da doğru olduğunu sezdiği seçeneği ya da seçenekleri işaretler.

1) Güven derecesine dayalı puanlama metodu (confidence testing).

2) Alt kümelere bölme metodu (subset-selection methods).

3) Yanıtlama olasılığına göre puanlama metodu (probabilistic response scoring methods).

1950-80 yılları arasındaki ölçme literatüründe, farklı madde puanlama yöntemlerinin test istatistikleri üzerine olan etkisini araştırmak için yapılmış çalışmalara sıkça rastlanmaktadır. Bu çalışmada da; madde puanlama türlerinden (1,0) ile puanlama, uzman kanısına dayalı puanlama ve Zinger Z_1 ve Z_2 puanlama yöntemlerinin madde varyansına, madde ayırıcılık gücüne, test güvenilirliği ve geçerliğine olan etkisi araştırılmıştır.

2. YÖNTEM

2.1. Araştırmanın Niteliği: Çalışma kuramsal bilgi alanına yenilerini katma amacı taşıdığından, temel araştırma niteliğindedir.

2.2. Ölçme Aracının Hazırlanması: Veri toplama amacıyla sayılar arası ilişkileri konu alan 20 maddelik her biri 4 seçenekli bir yetenek testi geliştirilmiştir. Önce bu amaca uygun 75 madde yazılmıştır. Bu maddelerin yazımında, önce x, y, z değişkenleri belirlenmiş, bu değişkenler arasında ikili ilişkiler tespit edilmiş ancak

öğrencilere bu ilişkiler yerine, ilişkileri bulmalarına yardımcı olacak sayı grupları verilmiştir. Madde kökünde x, y, z değerlerinden hangisinin sorulacağı baz alınarak, 75 madde 15'er maddelik 5 guruba ayrılmıştır. I. grup maddelerde, x değeri verilmiş buna bağlı y değeri ve y değeri belliyken buna bağlı z değeri sorulmuştur. II. grup maddelerde, y değeri verilmiş buna bağlı x ve z değerleri; III. grup maddelerde, y değeri verilmiş buna bağlı x değeri ve z değeri belli iken buna bağlı y değeri sorulmuştur. IV. grup maddelerde, x değeri verilmiş buna bağlı y ve z değerleri; V. grup maddelerde ise z değeri verilmiş, buna bağlı y ve x değerleri sorulmuştur.

Her maddede bir tamamiyle doğru, bir tamamiyle yanlış, 2 tane de kısmen doğru seçenek bulunmaktadır. Kısmen doğru olarak adlandırılan seçenekler; soruda bulunması istenen iki sayısal bağıntının birbirlerine göre zorluk düzeyleri temel alınarak hazırlanmıştır. Hazırlanan her bir test maddesi, seçeneklerin 3, 2, 1, 0 olarak ağırlıklandırılması için 16 uzmana verilmiştir. Uzmanlara yazılan 75 maddenin A seçeneğinde tamamiyle doğru, D seçeneğinde tamamiyle yanlış yanıtın olduğu söylenmiş ve onlardan öncelikle maddeleri yanıtlamaları, dil ve ifade yanlışları yapılmışsa düzeltmeleri, kısmen doğru olduğu düşünülen B ve C seçeneklerine verilmesi düşünülen 2 ve 1 puanlarının uygun olup olmadığının tespiti istenmiştir. Uzmanlardan alınan sonuçlara göre 75 maddeden 15 tanesi, maddelerin çok zor olması ve B ve C seçeneklerinin (kısmen doğru/yanlış) eşit olarak ağırlıklandırmaya uygun olması nedenleriyle elenmiştir. Deneme uygulamasına geriye kalan 60 madde alınmıştır.

Uzman kanısına başvurularak puanlanmış olan 60 madde, her bir madde tipi kendi içinde seçkisiz olarak 20'şer maddelik 3 gruba ayrılmış; bu gruplara uygulamada kolaylık sağlanması açısından A, B, C grupları adları verilmiştir.

2.3. Deneme Uygulaması: Deneme formları, 1999-2000 Eğitim-Öğretim Yılı Güz Dönemi'nde; Özel Bilim Lisesi (8. Sınıf-30 kişi), Fatih Sultan Mehmet Lisesi (9. Sınıf-27 kişi), 10. Sınıf-32 kişi) ve Ankara Fen Lisesi (11. Sınıf-

32 kişi) olmak üzere toplam 112 kişiye uygulanmıştır.

Ön uygulamadan elde edilen veriler, (1,0) puanlama metoduyla puanlanmış, 100-150 kişilik gruplara uygulanan "Henrysson Metodu" (Baykul 1999:335) kullanılarak analiz edilmiş, esas uygulama için madde güçlüğü 0,24 ile 0,79 ve madde ayıricılığı 0,30 ile 0,59 arasında değişen 20 madde seçilmiştir.

2.4. Verilerin Elde Edilmesi: 20 maddeden oluşan son test, araştırmanın verilerini elde etmek amacıyla kullanılmıştır. Veriler, 1999-2000 Eğitim ve Öğretim yılı Güz Dönemi'nde Özel Arı Lisesi (8. Sınıf-30 kişi) ve Arı Fen Lisesi (9. sınıf-30 kişi, 10. Sınıf- 28 kişi) 'nde okuyan toplam 88 öğrenciden testin tek oturumda uygulanmasıyla elde edilmiştir.

2.5. Puanlama: Esas uygulamadan elde edilen veriler, dört ayrı yöntem [(1,0) puanlama metodu, uzman kanısına dayalı puanlama metodu, Z_1 puanlaması ve Z_2 puanlaması] kullanılarak puanlanmıştır. (1,0) ile puanlama metodunda, sorulardan ikisini birden doğru cevaplayana 1, birinin doğru olması, hiçbirinin doğru olmaması ve sorunun boş bırakılması haline 0 (sıfır) puan verilmiştir.

Uzman kanısına dayalı puanlamada, sorulardan her ikisini de doğru cevaplayana 3, birini doğru cevaplayanlardan zor olanı doğru cevaplayana 2, kolay olanı doğru cevaplayana 1, her ikisini birden yanlış cevaplayan ve soruyu boş bırakanlara 0 (sıfır) puan verilmiştir.

Z_1 ve Z_2 puanlamalarında aşağıdaki formüller kullanılmıştır.

Zinger Z_1 Puanlaması: Bu puanlama yöntemine göre, doğru yanıt işaretleyenlere 1 puan, çeldiricilerden birini işaretleyenlere;

$$c = \frac{\sum_{i=1}^{a-1} n_i^2}{\left(\sum_{i=1}^{a-1} n_i \right)^2} \quad \begin{array}{l} a: \text{seçenek sayısı} \\ n_i: i \text{ çeldiricisini} \\ \text{işaretleyen birey sayısı.} \end{array}$$

olmak üzere (-c) puanı verilir.

Zinger Z_2 Puanlaması: Bu puanlama yönteminde,

$$b = \frac{\sum_{i=1}^{a-1} (n_i - n)^2}{n_T \cdot \sum_{i=1}^{a-1} n_i} \quad \begin{array}{l} n_T: i \text{ maddesine} \\ \text{doğru yanıt} \\ \text{verenlerin sayısı.} \\ n = \frac{\sum_{i=1}^{a-1} n_i}{a-1} \end{array}$$

olmak üzere, doğru yanıt işaretleyenlere 1+b puanı, çeldiricilerden birini işaretleyenlere (-c) puanı verilir (c puanı Z_1 puanlamasında tanımlandığı gibidir). Kullanılan formüller sonucu elde edilen sayılar, madde seçeneklerine puan olarak verilmiştir.

2.6. Verilerin Analizi: Geliştirilen çoktan seçmeli test maddelerinin dört ayrı puanlama yöntemine göre madde varyansı genel varyans formülü ve madde ayıricılığı için madde-test korelasyonu (Pearson momentler çarpım korelasyonu) kullanılarak hesaplanmıştır. Test güvenilirlikleri için Cronbacha katsayısı hesaplanmıştır. Testlerin geçerli olup olmadığını anlamak amacıyla 8., 9., ve 10. sınıflara ait araştırma grubu öğrencilerinin test ortalamaları arasında farkın anlamlı olup olmadığına bakılmıştır. Bu amaçla tek yönlü varyans analizi yapılmış, varyans analizi sonucunun anlamlı olması durumunda Scheffé testi kullanılarak ikili karşılaştırmalar yapılmıştır.

3. BULGULAR VE YORUMLAR

Doğru yanıt 1; boş, yanlış ve birden fazla yanıt olduğu hallere 0, uzman kanısına başvurup doğru yanıt yakınlıklarına göre çeldiricilere 3-2-1-0 değerlerini vererek puanlamanın, Zinger Z_1 Z_2 puanlamalarından madde varyansına etkisini incelemek için; genel varyans formülüyle madde varyansları hesaplanmış, elde edilen değerler aşağıdaki tabloda sunulmuştur. Ayrıca madde varyans değerlerinin ikişerli grafikleri de çizilmiştir.

Tablo 1: Dört Ayrı Puanlama Yöntemi Kullanılarak Hesaplanan Madde Varyansları

Madde N.	(1,0)P.	Uzman P.	Z ₁ P.	Z ₂ P.
1	0,12	0,82	0,12	0,12
2	0,21	1,34	0,21	0,23
3	0,23	1,60	0,27	0,28
4	0,23	1,20	0,28	0,29
5	0,23	1,17	0,28	0,29
6	0,22	1,18	0,27	0,28
7	0,25	1,38	0,29	0,30
8	0,24	1,06	0,30	0,32
9	0,23	1,15	0,27	0,29
10	0,25	1,60	0,30	0,32
11	0,22	1,41	0,27	0,28
12	0,22	1,24	0,26	0,27
13	0,25	1,21	0,34	0,40
14	0,24	1,38	0,28	0,29
15	0,24	1,38	0,35	0,42
16	0,25	1,46	0,30	0,30
17	0,25	1,37	0,33	0,37
18	0,25	1,30	0,33	0,37
19	0,21	1,51	0,27	0,28
20	0,23	1,68	0,29	0,30

Tablo ve grafiklerden (1,0) ile puanlama metodu, Z₁ ve Z₂ puanlamalarından elde edilen madde varyanslarının, uzman kanısına dayalı puanlama metodundan elde edilen değerlerden daha düşük olduğu ve birbirlerine yakın değerler taşıdığı görülmektedir. Uzman kanısına dayalı puanlamadan elde edilen varyans değerlerinin diğerlerine göre daha yüksek çıkması zaten beklenen bir sonuçtur. Çünkü bu puanlamada 3-2-1-0 değerleri puan olarak kullanılmış ve bu da madde varyans değerlerini arttırmıştır.

Dört ayrı puanlama yönteminin madde ayrıncılığını nasıl etkilediğini bulmak için, madde ayrıncılıkları hesaplanmış, elde edilen değerler aşağıdaki tabloda sunulmuştur.

Tablo 2: Dört Ayrı Puanlama Yönteminden Elde Edilen Madde Ayrıncılık Güçleri

Madde N.	(1,0)P.	Uzman P.	Z ₁ P.	Z ₂ P.
1	0,27	-0,06	-0,03	-0,03
2	0,3	0,01	0,01	0,01
3	0,25	0	0,11	0,11
4	0,54	0,5	0,54	0,54
5	0,58	0,51	0,56	0,56
6	0,58	0,4	0,48	0,48
7	0,54	0,55	0,48	0,48
8	0,72	0,66	0,67	0,67
9	0,82	0,66	0,78	0,78
10	0,54	0,45	0,56	0,56
11	0,68	0,62	0,62	0,63
12	0,61	0,49	0,6	0,6
13	0,7	0,65	0,69	0,69
14	0,69	0,55	0,65	0,64
15	0,54	0,58	0,58	0,59
16	0,62	0,58	0,62	0,62
17	0,73	0,67	0,7	0,71
18	0,62	0,42	0,59	0,59
19	0,57	0,62	0,6	0,6
20	0,64	0,73	0,75	0,75

Çizilen tablolar incelendiğinde, dört ayrı madde puanlama metodundan elde edilen ayrıncılık katsayılarının birbirlerine yakın değerler taşıdığı görülmektedir.

Dört ayrı puanlama yönteminin, test güvenilirliğine olan etkisini bulmak için Cronbacha güvenirlik katsayıları hesaplanmış, bu katsayılar aşağıdaki tabloda sunulmuştur.

Tablo 3: Dört Puanlama Yönteminden Elde Edilen Güvenirlik Katsayıları ve Bu Katsayılara Ait Z Değerleri

	(1,0) P.	Uzman P.	Z ₁ P.	Z ₂ P.
Cronbach α	0,90	0,86	0,88	0,88
Fischer z	0,64	0,56	0,60	0,60

Bu katsayılardan (1,0) puanlamasına ait 0,90 ile en düşük olan uzman kanısına ait 0,86 arasındaki fark, katsayılar Fischer'in z istatistiğine dönüştürülerek test edilmiş, aradaki fark manidar

bulunmamıştır ($\alpha = 0,01$). Ancak bu dört puanlama yöntemiyle elde edilen güvenilirlik katsayılarının 0,86 ve daha yüksek değerler olduğu dikate alındığında, bu dört puanlama yöntemiyle de güvenilirliği yüksek testler elde edilebileceği; ancak birinin diğerinden güvenilirlik yönüyle farklı olamayacağı sonucu çıkarılabilir.

Dört farklı puanlama yönteminin test geçerliliğine etkisini incelemek amacıyla, testin 8., 9., ve 10. sınıflara uygulanmasından elde edilen ölçme sonuçları, ölçüt puanlar olarak kullanılmış, 8., 9. ve 10. sınıflara ait test ortalamaları arasındaki farkın manidar olup olmadığına her puanlama yöntemi için ayrı ayrı varyans analizi ile bakılmıştır. Yapılan analizler sonucunda her puanlama yönteminden elde edilen F değerleri $\alpha = 0,01$ düzeyinde manidar bulunmuş, farklılıkların hangi sınıflar arasında olduğunu anlamak amacıyla Scheffe testine başvurulmuştur. Scheffe testi sonucunda 8. sınıfla 9. ve 10. sınıflara ait ortalamalar arasındaki farklar $\alpha = 0,01$ düzeyinde manidar bulunmuş, 9. ile 10. sınıflar arasındaki fark manidar bulunmamıştır.

Bu duruma göre, 8., 9. ve 10. sınıf başarıları ölçüt alındığında, puanlama yöntemlerinin hepsi grupları aynı şekilde birbirinden ayırmaktadır. Dolayısıyla, dört puanlama yönteminin de geçerli olduğu söylenebilir.

4. SONUÇLAR

Üçüncü bölümde belirtilen bulgular ışığında varılan sonuçların yanısıra, bu çalışma bitiminde elde edilen başka sonuçlar da vardır.

Araştırmada kullanılan ve ölçme literatüründe adı geçen puanlama metotlarının birey hakkında bilgi verme gücü bakımından oldukça yararlı ama aşağıda sayılan nedenler yüzünden kullanımının olumsuz yönleri olduğu düşünülmüştür.

- Test uygulaması için gereken zaman fazladır.
- Test yönergelerini bireylere açıklamak zordur.

- Bazı puanlama yöntemleri için gerekli olan, yanıtlayıcıların bire bir kontrolünü sağlama, uygulama anında güçtür.
- Puanlama yöntemleri karmaşıktır.
- Bireylere testten aldıkları puanları açıklamak ve bu puanlar doğrultusunda not vermek zordur.
- Bu puanlama metotlarına uygun test maddesi yazmak çok zordur.

Ayrıca literatüre bakıldığında kullanılan puanlama türlerinin genelde test güvenilirliğini biraz artırırken, geçerliği hiç etkilemediği saptanmıştır. Testin en önemli iki özelliğine bir katkı getirmemesi nedeniyle bu puanlama türlerinin, (1,0) ile puanlama metodu yerine kullanılmasının bir yarar sağlamayacağı sonucuna varılmıştır.

Araştırmadan elde edilen diğer bir sonuç da, farklı puanlama metotları için madde yazımıyla ilgilidir. Farklı madde puanlama yöntemlerinin olumsuz yönlerinden biri, bu puanlama türlerine uygun olacak şekilde madde yazmaktır. Çünkü bu yöntemler kullanıldığında, maddenin her bir seçeneğine de puan verilecek ve verilen bu puanlar o seçeneği işaretleyen bireyin hangi bilgi seviyesinde olduğunu göstermeye yönelik olacaktır. Günümüzde çoktan seçmeli testlerin en çok uygulandığı kurum olarak okullar düşünülürse, bu puanlamalar için uygun olacak maddeleri de konu alanı uzmanı olan öğretmenlerin yazması gerekecektir. Madde yazma işlemi konuya hakimiyet, emek, sabır ve yaratıcılık isteyen bir süreçtir, bu süreçte rol alacak öğretmenlerin konu alanında uzman, sezgi gücü yüksek ve yaratıcı olması gerekmektedir. Bu tip madde yazarlarının sağlanması da her zaman mümkün olmayabilir.

5. ÖNERİLER

Araştırma sonucunda aşağıdakiler önerilmektedir.

- a) Bu çalışmada veri toplamak için bir sayısal yetenek testi geliştirilmiş sayılar arası ilişkileri yoklayan test maddeleri yazılmıştır. Bu tip

bir çalışmanın, daha farklı davranışları yoklamak için farklı derslerde yapılması önerilir. Örneğin; Türkçe dersinin, okuduğunu anlama konusunun farklı puanlama yöntemleri ile ilgili bir çalışma yapmak için, uygun olduğu düşünülmektedir.

b) Ölçme literatüründeki; bu çalışmada kullanılmayan diğer madde puanlama türleri ve (1,0) ile puanlama metodunun karşılaştırmalı olarak incelenmesi önerilebilir.

c) Çoktan seçmeli test maddelerini puanlamada her ne kadar (1,0) metoduyla ile yapılan puanlama kolay ve güvenilir olarak düşünülse de seçeneklere doğru yanıtla yakınlıklarına göre puan vererek puanlama öğrenciler hakkında daha fazla bilgi vermektedir. Bu yüzden ders öğretmenlerine uygun madde yazabilmeleri halinde bu puanlama türlerini, öğrenciyi tanımak için kullanmaları önerilir.

ç) Farklı puanlama türlerinin test geçerliğini nasıl etkilediğini bulmak için, ölçüt grupları dışında farklı metotlar kullanılabilir.

d) Farklı puanlama yöntemlerinin diğer test teorilerinde de örneğin; örtük özellikler (latent traits)'de çalışılması, ölçme alanına katkı sağlayıcı olabilir.

e) Eldeki çalışmada ve başka pek çok çalışmada ortaya çıktığı gibi, klasik test teorisinde farklı puanlama yöntemlerinin test geçerlik ve güvenilirliğine önemli katkılar sağlamadığı gözlenmektedir. Bunun nedeni test geçerliği için,

yanıtlayıcıların maddeleri seçerek yanıtlamalarından ileri gelebilir. Bu düşünceyle, farklı puanlama yöntemlerinin kısa yanıtlı sınavlar üzerinde ve seçme gerektiren testlerle karşılaştırmalı olarak çalışılması test geçerliği açısından yararlı olabilir.

KAYNAKLAR

- [1] Anastasia, A. "Principles of Psychological Testing", USA: The Macmillan Company (1957)
- [2] Merwin, Jack. "Rational and mathematical Relationships of Six Scoring Procedures Applicable to Three-Choice Items" in *Journal of Educational Psychology* vol.50, no:4, (1959).
- [3] Frary, Robert. "A Simulation Study of Reliability and Validity of Multiple Choice Test Scores under Six Response-Scoring Modes." In *Journal of Educational Statistics* vol:7, no:4 (1982).
- [4] Jaradat, Derar; N. Tollefson, "The Impact of Alternative Scoring Procedures for Multiple Choice Items on test Reliability, Validity and Grading." in *Educational and Psychological Measurement* vol: 48. (1988)
- [5] Coombs, Cylde, "The Assessment of
- [6] Partial Knowledge" in *Educational and Psychological Measurement* vol: 13. (1953)
- [7] Frary, Robert, "Partial Credit Scoring Methods for Multiple Choice Tests" in *Applied Measurement in Education* vol: 2(1) (1989).
- [8] Baykul, Y. "Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulamaları", ANKARA: ÖSYM Yayınları (1999)