

ÇOKTAN SEÇMELİ TESTLERİ PUANLAMA YÖNTEMLERİNE BİR BAKIŞ

Devrim ÖZDEMİR*

Özet

Çoktan seçmeli testleri puanlamada sıklıkla kullanılan yöntem, iki kategorili (1,0) puanlama yöntemidir. Ancak bu yöntem, özellikle kısmi bilgiyi puanlamadaki yetersizliği nedeniyle eleştirilmiş ve bu zayıflığı ortadan kaldırma çabaları sonucunda farklı puanlama yöntemleri geliştirilmiştir. Bu çalışmada, çoktan seçmeli testleri puanlama yöntemlerinden bazıları incelenmiştir.

Anahtar Kelimeler: Çoktan seçmeli testler, iki kategorili puanlama, ağırlıklı puanlama, çok kategorili puanlama, seçenek ağırlıklandırma, madde ağırlıklandırma.

Summary

The method often used for scoring the multiple choice tests is binary scoring (1,0) method. But this method has been criticized particularly due to the its inefficiency in scoring the partial knowledge and owing to the efforts to prevent this weakness, different scoring methods have been developed. In this study, some of the scoring methods for the multiple choice tests have been studied.

Key Words: Multiple choice tests, scoring methods, dichotomous scoring, weighted scoring, polythomous scoring, option weighting, item weighting.

GİRİŞ

Çoktan seçmeli testler, eğitim ve psikolojide sıklıkla kullanılan ölçme araçlarıdır. Bu testler, kapsamı geniş bir şekilde örnekleyen, güvenilirlikleri yüksek, uygulaması ve puanlanması kolay ölçme araçları olmaları nedeniyle, hem test geliştirme uzmanları ve hem de eğitimciler tarafından tercih edilmektedir. Yine bu testler, bilgiyi hatırlama düzeyinden çok daha karmaşık olan düzeylere kadar, hedef davranışların ölçülmesinde oldukça büyük bir öneme sahiptir (Kurz 1999). Psikometrik özelliklerinin önceden kestirilebilmesi, güvenilirlik ve geçerliklerinin yüksek değerlerde olması ve büyük gruplara uygulanabilmesi gibi nedenlerle çoktan seçmeli testler, günümüzde sıklıkla kullanılan ölçme aracı durumundadır.

*Dr., Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü
devrimo@hacettepe.edu.tr

Çoktan seçmeli testleri kullanmanın önemli bir yararı, puanlama kolaylığıdır. Bu testler genellikle, doğru cevaplar için 1 ve yanlış, boş veya birden fazla olan cevaplar için 0 puanın verildiği, doğru cevabı puanlama yöntemi (number right scoring method) ile puanlanmaktadır. Bu yöntem, iki kategorili (1,0) puanlama yöntemi olarak da adlandırılır. Bu yöntemle tüm maddeler eşit olarak ağırlıklandırılır. Çoktan seçmeli testler için iki kategorili (1,0) puanlamada doğru cevabın seçilmesi durumunda verilen 1 puan tam bilmeyi (complete information), 0 puan ise yanlış bilmeyi diğer bir deyişle, tam bilgisizlik durumunu (misinformation) göstermektedir. Yöntem, tam bilgiye sahip bireylerin cevapları ile şans yoluyla doğru cevabı işaretleyen bireylerin cevaplarını “doğru cevap” kategorisine atarken, tamamıyla yanlış veya kısmen yanlış bilgiye sahip, veya şansla yanlış cevabı işaretleyen bireylerin cevaplarını da “yanlış cevap” kategorisine almaktadır. Bu yöntemin kullanımı basit olsa da, belirli zayıflıkları nedeniyle eleştirilmektedir. En temel eleştirilerden biri, cevaplayıcıların bilgi düzeylerinin “tam bilme”den “tam yanlış bilme”ye düşmesi ve şansla tahminde bulunmaları kontrol etmedeki zayıflığına yöneliktir (Jaradat ve Tollefson 1988).

İki kategorili puanlama öğrencilerin sahip olduğu bilginin miktarını doğrudan tahmin etmedeki başarısızlığı yönünden özellikle eleştirilmektedir. Çoğu çoktan seçmeli test, sadece sıralama bilgisi verir. Bu durum, test puanlarını bir programa girmeyi veya lisans belgesi almayı bekleyen adaylar arasında seçim yapma amacıyla kullanılan testlerde oldukça ciddi bir sorun ortaya çıkartmaktadır. Yöntemin bu zayıflıklarına yönelik tepkiler, bu sorunların bazılarında kaçınılmaz alternatif puanlama algoritmaları sağlamıştır. Bu alternatif puanlama stratejileri geleneksel puanlamanın zayıflıklarının üstesinden gelmeye çalışmakta ve cevaplayıcıların yeteneklerinin en iyi tahminini veren bilgileri elde etmeyi hedeflemektedir. İki kategorili puanlama yöntemine alternatif olarak geliştirilen bu puanlama yöntemleri tahminle cevap vermeyi önleyen ve kısmi bilgi için puan sağlayan yöntemlerdir. Bu yöntemler, kuramsal olarak test puanlarının güvenilirliğini ve geçerliğini yükseltmekte ve test tecrübesi daha az olan veya riske girmedikleri için daha düşük puan almak durumunda olan cevaplayıcılara yararlı olmaktadır (Kurz 1999).

Gerek klasik test teorisinde gerekse örtük özellikler teorisinde maddeye verilen cevaplar, cevaplayıcıların muhtemel cevap diye verilen seçeneklerden birini seçmeleri suretiyle elde edilmektedir. Maddeye verilen cevapların puanlanması da birbirine benzemektedir. Bu teorilerde kullanılan başlıca puanlama yöntemleri iki grupta incelenebilir: iki kategorili puanlama (dichotomous scoring) ve çok kategorili puanlama (polytomous scoring) (Hambleton ve Van der Linden 1996:4). Çoktan seçmeli testleri iki kategorili puanlama ile ilgili olarak yukarıda açıklanan zayıflıkları ortadan kaldırma çabalarında psikometristler, çok kategorili

puanlama algoritmaları geliştirmişlerdir. Çok kategorili puanlamada, maddelerin veya madde seçeneklerinin farklı ağırlıklandırıldığı yöntemler söz konusudur.

Klasik test teorisinde maddelerin veya madde seçeneklerinin farklı ağırlıklandırıldığı yöntemlerle ilgili olarak literatürde bir çok araştırma bulunmaktadır (Davis 1959; De Finetti 1965; Stanley ve Wang 1970; Waters 1974; Echternacht 1973; Echternacht 1976; Frary 1982; Kurz 1999). Cevaplayıcıların çoktan seçmeli test maddelerine ilişkin bilgi düzeyleri ile ilgili bir inceleme, Coombs, Milholland ve Womer (1953) tarafından yapılmıştır. Bu incelemeye ilişkin bilgiler Tablo 1’de verilmiştir.

Tablo 1.
CEVAPLAYICILARIN ÇOKTAN SEÇMELİ BİR TEST
MADDESİNE İLİŞKİN BİLGİ DÜZEYLERİ

Kısmi Bilme (<i>Partial Information</i>)	Birey, maddenin doğru cevabını bilmemesine karşın, yanlış olan bazı seçenekleri bilir.
Tam Bilme (<i>Complete Information</i>)	Birey, doğru cevabı ve diğer seçeneklerin yanlış olduğunu bilir.
Yanlış Bilme (<i>Misinformation</i>)	Birey, doğru olan seçeneğin yanlış olduğunu düşünmektedir.
Kısmi Yanlış Bilme (<i>Partial Misinformation</i>)	Birey, doğru olan seçeneğin yanlış olduğunu ve ek olarak bazı çeldiricilerin de yanlış olduğunu düşünmektedir.

Bir cevaplayıcı, çoktan seçmeli bir test maddesinde doğru seçeneği dikkate almadığında “yanlış bilgiye sahip” (misinformed) olarak tanımlanmaktadır. Cevaplayıcı, (1) yanlış seçeneklerden birinin doğru olduğuna inanabilir; (2) doğru seçeneği bir ya da daha fazla yanlış seçenekle sınıflandırabilir ve kalan iki veya daha fazla seçenek arasından tahminde bulunabilir. Bu durumlar, cevaplayıcı bilgisinde bazı değişiklikler olduğunu gösterir. Dahası, farklı puanlama/ cevaplama yöntemleri, aynı yanlış bilme koşulu için farklı madde puanları verebilirler. Bazı puanlama/ cevaplama yöntemleri, yanlış bilmenin tüm koşullarını eşit olarak cezalandırmaktadır; diğerleri yanlış bilgiye sahip cevaplayıcının doğru seçeneği de aynı kategoriye soktuğu yanlış seçeneklerin sayısına göre cezalandırır. Bazı yöntemler tek bir yanlış seçeneğin doğru olduğuna inanıldığı durumlarda en büyük cezayı gerektirir (doğru seçenek diğer tüm yanlış seçeneklerle gruplandırılmıştır) (Frary 1980).

Kısmi bilginin ölçülmesi, geleneksel puanlama sistemlerinde de varsayıldığı gibi, bir cevaplayıcının doğru cevabı bilememesi durumunda, maddeyi boş bırakmak veya şansa cevaplamak yerine, çeldiriciler arasında akla uygun gelen bir seçim yaptığını varsayar. Bir cevaplayıcının elediği doğru olmayan seçenek sayısı ne kadar çoksa, maddeyle test edilmek istenen bilgiye ilişkin bilgisi de o kadar fazladır. Bu nedenle cevaplayıcının kısmi bilgisinin

değerlendirilmesindeki puanlama sistemlerinin elde edilen ölçümlerin güvenilirlik ve geçerliklerini artıracakı düşünölmektedir (Waters 1976).

Frary (1989), kısmi bilgiyi ortaya koyma amacını taşıyan çeşitli puanlama/cevaplama yöntemlerine ilişkin bir sınıflandırma önermiştir. Tablo 2’de bu sınıflandırma verilmiştir.

Tablo 2.
ÇOKTAN SEÇMELİ TEST MADDELERİNİ
PUANLAMA YÖNTEMLERİ

Doğrudan Cevaplama Yöntemi (Direct Response Method)	Doğruyu Bulana Dek Cevaplama Yöntemi (<i>Answer Untill Correct-AUC</i>) Seçenek Ağırlıklandırma Yöntemi (<i>Option Weighting</i>) Çoklu Doğru Cevap Yöntemi (<i>Multiple Correct Response</i>) IRT yöntemleri (<i>Item Response Theory Methods</i>)
Cevaplayıcı Kararları Yöntemi (Examinee Judgements Method)	Güven Testi (<i>Confidence Testing</i>) Altgrup Seçme Yöntemleri (<i>Subset Selecting Methods</i>)

Frary (1980) tarafından önerilen sınıflama, puanlama yöntemlerini iki temel kategoride ele almaktadır. Doğrudan cevaplama yöntemi olarak adlandırılan birinci yöntemde cevaplayıcının basit olarak cevaplayıcıların en doğru olduğuna inandıkları cevabı işaretlemelerini gerektirmektedir. İkinci yöntem, cevaplayıcı kararlarına dayalı yöntem olarak adlandırılmaktadır. Bu yöntem genel olarak, cevaplayıcıların bir maddeyle ilgili bilgilerinin derecesine karar vermelerini gerektirmektedir. İki yöntem aşağıda ayrıntılı olarak ele alınmaktadır.

Doğrudan Cevaplama Yöntemleri

Bu sınıfta yer alan doğru cevabı bulana dek cevaplama yöntemi (AUC), doğru cevap olduğuna inandığı bir seçeneği işaretleyen bir cevaplayıcının, bu işlemi doğru cevabı buluncaya dek tekrarlamasını gerektirir. Doğru cevap seçildiğinde cevaplayıcı, bir sonraki maddeye geçmesi için yönlendirilir. Bugün bu yöntem bilgisayar kullanılarak rahatlıkla kullanılabilir. AUC testlerini puanlamada geleneksel yöntem, cevaplayıcıların verdiği toplam cevap sayısını olası tüm cevapların sayısından çıkarmaktır. Yöntem, hemen yapılan dönüt verme sayesinde öğrenmeyi ilerletebilmesi; cevaplayıcıya, verilen dönüt başarıya ulaşmaya değin cevaplama bulunma imkanı sağlaması; cevaplayıcılar doğru seçeneği buluncaya kadar soruları cevaplama sürdürdüğünde, olası puanların ranjının artması ve buradan da puanların geçerlik ve güvenilirliğini iyileştirmesi gibi avantajlara sahiptir. Özel araçlara ihtiyaç duyulması nedeniyle yüksek uygulama maliyetleri de bu yöntemin kullanılmasını güçleştirmektedir (Frary 1989; Ben Simon, Budescu ve Nevo 1997).

Doğrudan cevaplama yönteminin bir alt kategorisi olan seçenek ağırlıklandırma yaklaşımı kullanıldığında, cevaplayıcılara, her madde için bir seçeneği işaretlemeleri yönergesi verilmektedir. Her seçeneğe göre değişik puan değerleri atanmaktadır ve madde başına verilen maksimum puan genellikle 1 olmaktadır. Seçenek ağırlıklandırma, iki farklı yolla yapılabilir. Bu ağırlıklandırmalar, bir iç veya dış ölçüt yardımıyla yapılan önsel (uzman kanısına dayalı) ağırlıklandırma ve deneysel (görgül) ağırlıklandırma olarak adlandırılır. İç veya dış ölçüt kullanarak ağırlıklandırmanın geçmişi, Strong'un ilgi envanteri üzerine çalışmaya başladığı 1920'lere kadar gitmektedir (Frery 1989). Deneysel ağırlıklandırmada cevaplayıcıların maddeye verdikleri cevapların dağılımından yararlanılmakta ve her bir seçeneğin işaretlenme yüzdesi hesaplanarak seçeneklerin ağırlıklandırılması sağlanmaktadır. Önsel ağırlıklandırma yönteminde ise, her bir maddenin seçeneğinin ağırlıklandırılmasında genellikle uzman görüşlerine başvurulmakta ve genellikle uzmanların seçeneklere verdikleri ağırlıkların ortalamaları alınarak puanlama cetveli geliştirilmektedir (Echternacht 1976). Ancak, test maddelerini önsel ağırlıklandırma yaygın biçimde uygulanmamaktadır.

Önsel ağırlıklandırma yöntemi diğerlerine göre daha basit ve anlaşılabilir bir yöntemdir ve genellikle bir grup uzman tarafından yapılır. Bunun yanında, bazı durumlarda uzmanlar arasında tutarlık elde edilememesi ve fazla maliyete neden olması gibi nedenler, önsel ağırlıklandırmanın dezavantajı olarak düşünülmektedir. Bu nedenle, önsel ağırlıklandırmanın kullanılması durumunda, madde yazarının kendisinin seçenekleri daha önceden ağırlıklandırması önerilmektedir. Deneysel ağırlıklandırmanın olumsuz yönü ise, cevaplayıcının hangi seçeneği işaretlediğinde hangi puanı alacağını önceden bilememesidir; kullanılan puanlama sistemi hakkında cevaplayıcıya daha önceden bilgi verilmemektedir (Echternacht 1973).

Bir çoktan seçmeli test maddesi, birden fazla doğru cevabı içerebilir. Doğru seçenek sayısı bir maddeden diğerine farklı ise, cevaplayıcı her seçeneğin doğru olup olmadığını incelemek durumundadır. Bu durumda her bir seçenek bir doğru-yanlış maddesi haline gelir. Çoklu doğru cevaplama yöntemi olarak adlandırılan bu tip testleri puanlamanın bir yolu, doğru olarak sınıflandırılan seçenekleri toplamaktır. Bu test yaklaşımının cevaplayıcının kısmi bilgisini (1,0) puanlamaya göre daha iyi değerlendireceği düşünülmektedir (Frery 1989). Çok cevaplı formatta cevaplayıcılardan doğru olabilecek seçeneklerin hepsini işaretlemesi istenir. Bu formatta her bir madde, işaretlenen doğru seçeneklerin sayısından işaretlenen yanlış seçeneklerin sayısı çıkartılarak puanlanır. Çok cevaplı formatın tek yararı, ortalama ve ortalamanın üstünde olan cevaplayıcıları test etmede kısmen doğru cevaplar için kısmi puan vermesi olarak görülmektedir (Kurz 1999).

Frery (1989)'nin belirttiği gibi, ağırlıklı seçenekler için IRT modelleri ilk kez Samejima (1969) tarafından geliştirilmiştir. Bu modeller, derecelendirilmiş cevap modelleri olarak adlandırılmakta ve sadece doğruluk derecesine göre sıralanabilen seçenekler için kullanılabilir. Benzer bir sınırlılık, Masters (1982)'in önerdiği kısmi puan modeli için de geçerlidir. Bunların yanında, Bock (1972) seçeneklerin sıralanmasında sınırlama olmaksızın kullanılabilir bir IRT modeli (nominal cevap modeli) önermiştir.

Cevaplayıcı Kararlarına Dayalı Yöntemler

Bu yöntemler, cevaplayıcıların kesinlikle yanlış olduğuna inandıkları seçeneği işaretlemeleri (alt grup seçme yöntemi) veya cevaplayıcının işaretlediği seçeneğin doğruluğuna olan güvenini ifade etmesi (güven testi) biçiminde açıklanabilir.

Alt grup seçme yöntemleri, eleme puanlaması ve kapsayarak puanlama yöntemleri olarak iki grupta incelenebilir. Eleme puanlamasında (ET-elimination testing), cevaplayıcılara yanlış olduğuna kesin olarak inandıkları tüm seçenekleri işaretlemeleri yönergesi verilir ve puanlama sürecinde şans nedeniyle kazanılan puanlar çıkarılacağından, bilgileri yoksa tahminde bulunmamaları gerektiği şeklinde uyarıda bulunulur. Eleme puanlaması Coombs, Milholland ve Womer (1953) tarafından ortaya konmuş ve yoğun ilgi nedeniyle pek çok deneysel ve teorik çalışmaya yol açmıştır. Bu yöntem, bilginin mümkün tüm düzeylerini ayırmaktadır. Cevaplayıcı tüm çeldiricileri elediğinde “tam bilgi”ye (full knowledge) sahip olduğu kabul edilir. Kısmi bilgi (partial knowlege), çeldiricilerin alt grubunun elenmesiyle tanımlanır. Kısmi yanlış bilme (partial misinformation) cevaplayıcının, çeldiricilerin alt grubuyla birlikte doğru seçeneği de elemesi durumudur. Tam yanlış bilme (full misinformation), sadece doğru cevabın elenmesini ifade eder. Cevaplayıcının maddeyi boş bırakması veya tüm seçenekleri elemesi, bilgi yokluğu (absence of knowledge) olarak adlandırılır. Bu yaklaşımın bir üstünlüğü, özel maddeler yazmayı gerektirmemesidir (Coombs, Milholland ve Womer 1953).

Kapsayarak puanlama (IS-including scoring) eleyerek puanlama yöntemiyle aynı ölçeği kullanır. Bununla birlikte cevaplayıcılardan, her bir madde için, doğru seçeneği veya doğru olduğuna inandığı seçenekler grubunu işaretlemeleri istenir (Kurz 1999). Jaradat ve Tollefson (1988)'un çalışmasında şans başarısını düzeltme formülleriyle eleyerek ve kapsayarak puanlama karşılaştırılmış, eleyerek puanlamanın kısmi bilgi için en fazla puanı verdiği ve bu yöntemin cevaplayıcıların bir maddeyle ilgili gerçek bilgi durumlarını ifade etmelerine imkan verdiği bulunmuştur. Hem eleyerek puanlama hem de kapsayarak puanlamanın iki kategorili puanlamaya göre bir miktar daha güvenilir ve geçerli puanlar verdiği gözlenmiştir. Ancak, test yönergelerinin bazı cevaplayıcıların kafasını karıştırdığı görülmüştür.

Güven testi, cevaplayıcıların işaretledikleri seçeneğin veya seçeneklerin doğruluğuna inançlarını yansıtan madde cevaplarına doğrudan ya da dolaylı olarak ağırlıkların atandığı bir test yöntemidir. Bu ağırlıklandırma yöntemi kullanıldığında, cevaplayıcıya, doğruluğuna inandığı cevabın ne olduğu ve bu cevabın doğruluğundan nasıl emin olduğu sorulur. Tereddüt etmeden verilen bir doğru cevaba, emin olmadan verilen yanlış bir cevaptan daha fazla puan verilir. Aynı cevabı seçen cevaplayıcılar, emin olma derecelerinin göstergeleri olması nedeniyle, farklı puanlar alabilirler. Güvenirlik ve geçerlikte artış olmaması, kişilik değişkeni, test alma ve test puanlama tekniklerindeki karmaşıklık ve gerekli olan test uygulama süresi, bu yöntemin geniş bir şekilde kullanılmamasına ilişkin nedenler olarak sayılabilir (Kurz 1999).

Kısmi bilgiyi dikkate alan başka puanlama yöntemleri de vardır. Bu yöntemlerden bazıları, olasılığa dayalı puanlama (PT), tam sıralı puanlama (CO), ve kısmi sıralı puanlama (PO) şeklinde sıralanabilir. Bu yöntemler, tam yanlış bilgiden tam doğru bilgiye kısmi bilginin çeşitli dereceleri için bir ölçek sağlamaktadır.

Olasılığa dayalı puanlama (PT- probability testing) yönteminde cevaplayıcıdan her bir seçeneğin doğru cevap olma olasılığını belirtmesi istenir. Bu şekliyle yöntem, cevaplayıcıdan kısmi bilgisi hakkında daha ayrıntılı ve belirgin bir bilgi alınmasına olanak sağlamaktadır. Tam sıralı puanlama (CO- complete ordering) yöntemi, olasılığa dayalı puanlama yönteminin özel bir durumudur. Bu yöntemde cevaplayıcıdan her bir seçeneğin doğru cevap olması olasılığını belirtmesi yerine, bu olasılıklara göre seçenekleri doğruluk sırasına koyması istenmektedir (Ben Simon, Budescu ve Nevo 1997). Kısmi sıralı puanlama (PO- partial ordering), eleme puanlaması ve tam sıralı puanlama yöntemlerinin karışımı olan bir yöntemdir. Bu yöntemde cevaplayıcıdan, eleme puanlamasında olduğu gibi yanlış olduğunu düşündüğü tüm seçenekleri elemesi ve sonra da kalan seçenekleri, tam sıralı puanlamada olduğu gibi, doğru cevap olma olasılığına göre sıraya koyması istenir (De Finetti 1965).

İki kategorili puanlama yöntemi ile farklı puanlama yöntemlerinin karşılaştırıldığı güvenirlik ve geçerlik çalışmaları, genellikle Cronbach α güvenirliliği ve görünüş geçerliği bakımından iki kategorili puanlamanın biraz daha üstünde iyileşmeler olduğunu göstermekle birlikte, bazı dezavantajlar nedeniyle, bu yeni yöntemlerin kullanılmasını tam olarak haklı çıkarmamaktadır. Bu dezavantajlar, testlerin uygulama süresi ve maliyetlerindeki artışlar yanında test uygulama ve puanlamalarının karmaşıklığını ve test geliştirme sürecinde harcanan emek ve zamandaki artışı da kapsamaktadır. Bu nedenle, iki kategorili puanlamadan daha iyi bir yöntem bulununcaya kadar bu puanlama yönteminin kullanılmasının daha yararlı olacağı düşünülmektedir. Yapılacak diğer bir işlem, farklı yetenek düzeylerindeki cevaplayıcıları çekecek madde yazma çalışmalarının yapılmasıdır. Bunu yapmanın bir yolu, farklı doğruluk

düzeylerinde seçeneklere sahip maddeler yazmada madde yazarlarını eğitmektir. Diğer bir yaklaşım, yanlış olduğu düşünülen seçenekleri elemeye farklı cevaplama yöntemlerini gerektiren alışılmış çoktan seçmeli maddelerden yararlanmaktır. Bu test yöntemi, eleme puanlaması (elimination scoring) olarak açıklanan yöntemdir. Bu alanda yapılacak araştırmalar için, güvenilirliği arttırmanın ve kısmi bilgiyi değerlendirmenin bir yolu olarak, bu yöntem önerilebilir.

Ben-Simon, A., D. V. Budescu ve B. Nevo. "A Comparative Study of Measures of Partial Knowledge in Multiple-Choice Tests." *Applied Psychological Measurement*, No.21(1), 1997:65-88.

Bock, Darrell R. "Estimating Item Parameters and Latent Ability When Responses are Scored in Two or More Nominal Categories." *Psychometrika*, No.37(1), 1972:29-51.

Coombs, C. H., J. E. Milholland ve F. B. Womer. "The Assessment of Partial Knowledge." *Educational and Psychological Measurement*, vol.13, 1953:13-37.

Davis, Frederic B. "Estimation and Use of Weights for Each Choice in Multiple-Choice Test Items" *Educational and Psychological Measurement*, No.3, 1959:291-298.

De Finetti, B. "Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item." *British Journal of Mathematical and Statistical Psychology*, No.18, 1965:87-123.

Echternacht, Gary. "A Comparison of Various Item Option Weighting Schemes." *Educational Testing Service*, Report No: ETS-RB-73-6, 1973.

Echternacht, Gary. "Reliability and Validity of Item Option Weighting Schemes." *Educational and Psychological Measurement*, No:36, 1976:301-309.

Frary, Robert B. "The Effect of Misinformation, Partial Information, and Guessing on Expected Multiple Choice Test Item Scores." *Applied Psychological Measurement*, No.4(1), 1980:79-90.

Frary, Robert B. "Simulation Study of Reliability and Validity of Multiple Test Scores Under Six Response Scoring Models." *Journal of Educational Statistics*, No.7(4), 1982:333-351.

Frary, Robert B. "Partial Credit Scoring Methods For Multiple Choice Tests." *Applied Measurement in Education*, No.2(1), 1989:79-96.

Hambleton R.K. ve Van Der Linden. *Handbook of Modern Item Response Theory*, New York, Springer- Verlag Inc., 1997

Jaradat, D. ve N. Tollefson. "The Impact of Alternative Scoring Procedures for Multiple-Choice Items on Test Reliability, Validity and Grading." *Educational and Psychological Measurement*, No:48, 1988:627-635.

Kurz, Terri Barber. "A Review of Scoring Algorithms for Multiple-Choice Tests." *EDRS Publications*, Report No: ED 428 076, 1999.

Stanley, J. C. ve M. D. Wang. "A General Model For Free Response Data." *Educational and Psychological Measurement*, No:30, 1970:21-35.

Waters, Brian. "The Measurement of Partial Knowledge: A Comparison Between Two Empirical Option-Weighting Methods and Rights-Only Scoring." *The Journal of Educational Research*, No.70(6), 1977:256-60.