

Kişilik Testlerinde Klasik ve Modern Test Kuramları ile Madde Analizi

Oya Somer*
Ege Üniversitesi

Özet

Bu çalışmada, klasik test kuramının bir uzantısı sayılabilen, ancak test geliştirmede klasik kurama göre ilave bazı avantajlar sağlayan Madde - Cevap Kuramı'na dayalı madde analizleri, bir kişilik ölçeği örnek alınarak (Beck Umutsuzluk Ölçeği) incelenmiştir. Madde güçlüğü, ayırdedicilik, madde-model uyumu, bilgi (informasyon) kavramları örnek veriler üzerinde yorumlanmış, klasik ve modern kuramla madde analizlerinin özellikleri, avantaj ve dezavantajları karşılaştırılmıştır. Modern kuramdaki kavramlarla yürütülen analizlerin madde ile ölçülen psikolojik özellik arasındaki ilişkinin anlaşılmasında klasik kurama göre çok daha ayrıntılı incelemelere olanak sağladığı sonucuna varılmıştır. Yetenek ölçeklerinde olduğu gibi, kişilik ölçeklerinde de madde-cevap modellerinden daha yaygın biçimde yararlanılmasının, yapısal özellikleri daha güçlü ölçekler oluşturmaya katkıda bulunacağı görüşü vurgulanmıştır.

Anahtar sözcükler: *Kişilik ölçümü, madde analizi, madde-cevap kuramı, klasik test kuramı*

Abstract

Item analysis procedures based on Item-Response Theory, which can be seen as an extension of Classical Test Theory with additional advantages, were applied to a personality scale. Item difficulty, item discrimination, item-model fit and information concepts were tested on the Beck Hopelessness Scale items. Specifications, advantages and disadvantages of modern and classical item analysis procedures were compared. In comparing modern test theory with classical theory, it was shown that modern test theory allows for a more detailed analysis of the relations between items and latent traits. It was pointed out that, as with ability scales, using modern test theory more widely in personality measurement may contribute to the development of personality scales with higher construct validity.

Key words: *Personality measurement, item analysis, item-response theory, classical test theory*

Psikolojide ölçek geliştirmede klasik kurama dayalı madde analizlerinin, grup değişikliklerinde madde parametrelerinin değerini fazlasıyla etkilemesi (Gulliksen, 1958), eşit aralıklı ölçekler oluşturmaya imkan vermemesi (Wright ve Stone, 1979) gibi zayıflıkları, araştırmacıları yeni modeller aramaya yöneltmiştir. Bu arayışlar (benzer kavramlar üzerindeki çalışmaların 1930 ve 1940'lı yıllarda başlayıp çeşitli araştırmacıların katkılarıyla sürmüş olmasına rağmen) bugün bilinen haliyle Birnbaum'un etkisi ile Lord tarafından logistik bir tepki

modeli olarak geliştirilen ve Madde-Cevap Kuramı (Item Response Theory, IRT) ya da Örtük Özellikler Kuramı (Latent Trait Theory) adıyla anılan modern test geliştirme modellerinin gelişimiyle sonuçlanmıştır. Modern kuram klasik test kuramının bir uzantısı sayılmakla birlikte, test geliştirmede maddelerin psikometrik özelliklerini incelemeye klasik kuramın sağlayamadığı bazı avantajlar sağlamıştır.

Başlangıçta çoğunlukla yetenek ölçeklerinin geliştirilmesinde kullanılan modern test geliştirme modellerinin, Steinberg

ve Thissen'in (1995)'de belirttiği gibi ölçek geliştirme problemlerinin çok daha yoğun olduğu kişilik ölçümünde son yıllara kadar yaygın bir kullanımı olmadığı görülmektedir. Ancak klasik kuramların çözülmemeyen birçok ölçekleme problemine açıklık getirmesi, doğru-yanlış (dikotomik) olarak puanlanan ölçeklerin yanısıra, derecelenme ölçekleri için de madde analizi yapan programların geliştirilmesiyle, kişilik ve tutum ölçümlerinde de modern test geliştirme modellerine ilgi artmaktadır.

Madde-Cevap Kuramı, kişilerin gözlenen tepkilerinin psikolojik kuramlarda içerilen kuramsal yapılara, olasılıksal bir yolla bağlanmasını sağlayan modeller sunmaktadır. Diğer bir deyişle, kişilerin ve maddenin belirli özelliklerine göre, kişinin bir maddeye göstereceği belirli bir tepkinin olasılığını veren matematiksel fonksiyonları içermektedir. Kuram, kişinin belirli bir test maddesine belirli bir tepkiyi gösterme olasılığının, kişinin test maddelerinin altında yatan örtük özellik (latent trait, θ) üzerindeki konumu ile bağlantılı olduğunu varsaymakta ve bu doğrudan gözlenemeyen özelliklerle, gözlenen tepkiler arasında bağlantı kurmaktadır. Bu gözlenen ve gözlenemeyen büyüklükler arasındaki ilişki, matematiksel bir fonksiyonla kurulmaktadır. Bu fonksiyonlar normal ogive ve logistik fonksiyonlardır. Hambleton ve Swaminathan (1989), klasik test modellerinin basit varsayımları ve çok gelişmiş pratik kullanım uygulamaları varken, kuvvetli varsayımlara dayanan ve çok pratik olmayan madde-cevap modellerinin neden tercih edileceği sorusuna, bu modellerin aşağıdaki üç temel avantajı ile cevap vermektedirler. Bu modellerin başlıca avantajları; (1) madde parametreleri tahminlerinin örneklemden bağımsız olarak yapılabilmesi, (2) deneklerin yeteneklerinin, ölçeklenmiş (kalibrasyon) maddeler içinden seçilen madde grubundan bağımsız olarak tahmin edilebilmesi, (3) yetenek tahminlerinin doğruluğunun farklı yetenek düzeyleri için bilenebilmesi olarak sıralanmaktadır.

Madde-cevap modellerinin üzerine inşa edildiği temel kavram madde karakteristik eğrisidir (item characteristic curve, ICC). Model, kişilerin test maddelerinde

gösterdiği performansla, bu performans üzerinde etkili olan özellik arasındaki ilişkinin, matematiksel olarak artan bir fonksiyon olan madde karakteristik fonksiyonuyla tanımlanabileceğini varsaymaktadır. Madde karakteristik eğrisi, bireyin ölçülen özellik boyutundaki düzeyi ile, maddeye doğru cevap verme olasılığı arasındaki ilişkinin grafik gösterimidir (Lord ve Novick, 1968). Bu fonksiyon, yetenek ya da ölçülen özellik boyutu üzerinde daha yüksek puanlara sahip kişilerin, maddeyi doğru cevaplama olasılıklarının, düşük puanlı kişilerden daha yüksek olacağını belirtir. Madde karakteristik eğrisi yoluyla, yetenek boyutu üzerinde birbirinden farklı noktalarda bulunan deneklerin bir maddeye doğru cevap verme olasılıkları elde edilebilmektedir.

Bu olasılığın elde edilmesinde bir, iki ya da üç parametre kullanan farklı madde-cevap modelleri bulunmaktadır. Aşağıda, en mükemmel biçimi ile Frederic M. Lord tarafından formüle edilmiş üç parametrelilik modelin deneklerin maddeyi doğru cevaplama olasılıklarını temsil eden logistik fonksiyon verilmiştir.

$$P = P(\theta) = c + \frac{1 - c}{1 + e^{-1.7a(\theta - b)}} \quad (\text{Lord, 1980})$$

($e = 2.71828$)

Üç parametrelilik modelin parametreleri, (b_i) madde güçlüğü (location, item threshold), (a_i) ayırtetme (item discrimination, slope) ve (c_i) tahmin (guessing) parametreleridir.

Güçlük parametresi, maddenin örtük özellik (θ) boyutu üzerindeki konumunu, yerini belirtmektedir. Güçlük parametresinin değeri, i maddesini 0.50 oranında doğru cevaplayan deneklerin buldukları örtük özellik puanına karşılık gelmektedir. Ayırtetme parametresi ise, bir madde karakteristik eğrisinin eğimine ya da dikliğine karşılık gelmektedir. Farklı yetenek seviyelerini ayırtetmekte başarısız olan maddeler, nispeten düz bir madde karakteristik eğrisi göstermekte ve düşük bir a_i değerine sahip olmaktadır. Üçüncü parametre olan c_i parametresi ise, tahmin parametresi olarak

isimlendirilmekte ve madde karakteristik eğrisinin düşük asimptotuna karşılık gelmektedir. Yani θ üzerinde en düşük düzeylerde olan kişilerin maddeye doğru cevap verme oranıdır. Bu parametre şans ya da tahminle doğru cevaplanma olasılığı olan, özellikle çoktan seçmeli maddeler için önem taşımaktadır. İki parametrelili modelde, c_j parametresi 0 olarak sabit bir değerde tutulmakta, üç parametrelili modelin özel bir hali olan tek parametrelili Rasch modelinde ise, hem a_j hem de c_j parametreleri sabit kabul edilmekte ve yalnızca güçlük parametresi tahmin edilmektedir.

Tahmin edilen parametre değerleri ayırdetme indeksi için yaklaşık 0.30 ile 2.0 arasında olmakta, güçlük parametresi genelde -3 ve +3 arasında değerler almakta, tahmin parametresi de 0 ile 1.0 arasında değişen değerler almaktadır (Hulin, Drasgow ve Parsons, 1983).

Parametre tahminlerinde kullanılan yaygın bir istatistiksel yöntem gözlenen verilerin olasılıklarını maksimize eden model parametrelerini parametre değeri olarak almaktır. Bu şekilde elde edilen parametre tahminleri Maksimum Olabilirlik tahminleri olarak bilinmektedir (Mislevy ve Stocking, 1989). Tahmin etme (estimation) süreci ile, maddeler ve denekler, gerçek performans olasılıkları ile beklenen olasılık parametreleri arasındaki ilişkinin en yakın olmasını sağlayacak biçimde, yetenek ölçeği üzerinde yerleştirilirler. Madde parametresi tahminleri ve deneklerin yetenek tahminleri, gerçek test verileri ile madde ve yetenek tahminleri arasında en yüksek uyum sağlanana kadar sürdürülür.

Modelin beklentilerine uymayan maddeler, test ile uyum sağlamadığı düşünülerek, ya testten çıkarılır, ya da üzerinde yeni çalışmalar yapılır. Uyum analizleri sonucu elenen bir maddenin (1) yazımında teknik bir hata olduğu, (2) ayıricılığının düşük ya da testteki diğer maddelerden farklı olduğu, (3) test ile ölçülen yetenek dışında bir boyutu ölçtüğü, (4) şansla cevaplanabildiği, (5) testteki diğer maddelerden güç olduğu ya da yetenek dağılımı çok geniş bir kitleye uygulandığı düşünülmektedir (Özçelik ve Berberoğlu, 1989).

Madde-cevap modellerinin çoğu bir-takım varsayımlar üzerine kurulmuştur. Bunlardan birincisi, maddelerin yansıtıkları düşünülen özellik üzerinde derece farklarına göre sıralanabileceği, ikincisi kişilerin bu özellik üzerinde aynı şekilde sıralanabileceği, üçüncüsü ise bu özelliğe sahip olma derecelerinin kişinin bir grup maddeye göstereceği tepkiyi belirleyeceğidir. Helmes ve Jackson (1989), bu varsayımların temel olan ancak nadiren sözü edilen varsayımlar olduğunu, diğer varsayımların, cevaplayıcıların özellikleri ile maddeye verilen tepkiler arasındaki ilişkiyi veren matematiksel fonksiyona ilişkin terimlerle ifade edilebileceğini belirtmektedirler. Madde-cevap kuramının temel varsayımlarından birisi, belirli bir maddeye doğru cevap verme olasılığının, kişinin tek bir karakteristiği tarafından belirlendiğidir. Bu temel yetenek ya da özellik boyutu (θ) ile gösterilmektedir. Bazı çok boyutlu Madde-cevap modelleri olmakla birlikte, bunlar pratik kullanım açısından tek boyutlu modeller kadar yaygın değildir. Andrich (1988), tek boyutlu ölçümler söz konusu olduğunda, ölçümler arasındaki farklılıkları kullanarak karşılaştırmalar yapmanın mümkün olduğunu ve bunun bir derece farklılığı olduğunu belirtmektedir.

Madde-cevap kuramının bir diğer varsayımı lokal bağımsızlıktır. Bu durum, örtük özellik uzayının tek boyutluluğunun bir sonucu olarak ortaya çıkmaktadır. Lokal bağımsızlık, θ üzerinde belirli bir yetenek düzeyinde olan alt grupların, herhangi bir maddeye tepkilerinin bağımsızlığına işaret etmektedir. Ancak bu bağımsızlık, genel anlamda bir bağımsızlık değildir. Farklı θ değerlerine sahip genel bir örneklemede, maddelerin birbiriyle korelasyon göstermesi beklenen bir durumdur. Andrich (1988), maddelerin her birinin ölçümlere eşit olarak katkıda bulunmaları için birbirinden bağımsız olarak tekrarlanabilmeleri gerektiğini, ancak bir özelliğe ilişkin değeri tahmin edebilmek için aynı zamanda birbirleriyle ilişkili olmaları gerektiğini belirtmektedir.

Farklı yetenek düzeylerini içeren bir grubun, iki maddeye gösterdikleri tepkilerin korelasyon göstermesi, bu korelasyonun o

yetenek boyutundan kaynaklandığına işaret ederken, yetenek bakımından aynı düzeyde olan kişilerin iki maddeye gösterdikleri tepkiler arasında bir ilişki olması, ölçülen yetenekten başka özelliklerin de maddedeki performansını etkilediğine işaret etmektedir.

Çalışmamızda, yukarıda ana hatlarıyla ele alınan Madde-cevap kuramına dayalı madde analizlerinin, kişilik boyutlarının ölçeklenmesinde sağladığı katkılar, örnek olarak ele alınan bir kişilik ölçeği üzerinde tartışılmış, klasik kurama dayalı madde analizi süreçleri ile arasındaki benzerlik ve farklılık ile avantaj ve dezavantajlar üzerinde durulmuştur.

Yöntem

Örnekleme

Çalışmanın örnekleme*, 1994-1995 öğretim yılında Ege Üniversitesinin çeşitli fakültelerinde okumakta olan, 700 öğrenciden oluşmaktadır. Bu öğrencilerin 378'i kız, 322'si erkektir; 156'sı sağlık bilimleri, 111'i güzel sanatlar ve 284'ü Fen ve Uygulamalı Bilimlerin çeşitli fakültelerine devam etmektedirler.

Araçlar

Çalışmada bir kişilik ölçeği örneği olarak Beck, Weisman, Lester ve Trexler (1974) tarafından geliştirilen, Beck Umutsuzluk Ölçeği kullanılmıştır. Ölçek 20 maddeden oluşan, kendini değerlendirme türünde bir ölçektir. Ölçeğin maddeleri geleceğe ilişkin düşüncelerle ilgili ifadelerden oluşmuştur. Bu ifadelerin doğru ya da yanlış bulunmasına göre, maddeler doğru yanlış (1 ya da 0) olarak puanlanmaktadır. Ölçekten elde edilen puanlarla kişilerin geleceğe yönelik karamsarlık derecesinin, belirlenmesi amaçlanmaktadır (Derebaşı, 1996).

Bulgular

Beck umutsuzluk ölçeği maddelerinin analizi, Madde-Cevap Kuramını temel

alan, BILOG (Mislevy ve Bock, 1986) madde analizi programı ile yapılmıştır.

Tek boyutluluk varsayımının test edilmesi için faktör analizi yapılmıştır. Ancak Hambleton ve Swaminathan (1989) ve Hulin, Drasgow ve Parsons (1983), bu varsayımın pratikte tam olarak karşılanmasının mümkün olmadığını, genellikle kişilik, test alma becerileri ve ölçülen temel boyuttan başka pek çok diğer faktörün test performansını etkilediğini belirtmektedirler. Ackerman (1989) ve Traub'un (1983) test görevleri ile ilgili bu bilişsel faktörlerin sayısının kişiden kişiye değişebileceği gibi, maddeye doğru cevap vermeyi etkileyen özelliklerin de maddeden maddeye değişiklik göstereceğini ileri sürdüğüne değinmektedir. Hambleton ve Swaminathan (1989) ve Hulin, Drasgow ve Parsons (1983), pratikte bu varsayımın karşılanması için, test performansını etkileyen 'baskın' bir özellik ya da faktörün bulunmasının yeterli olduğunu belirtmektedirler. Bu baskın faktör, test ile ölçülen yetenek olarak tanımlanmaktadır.

Lord ve Novick (1968), okuma, sözcük dağarcığı, telaffuz ya da uzaysal yetenek gibi özelliklerin tek boyutlu olarak kabul edilebileceğini, ancak bir matematik testinin, hem matematiksel akıl yürütmeyi, hem de düzlem geometrisi ile ilgili maddeleri içermesi durumunda, en azından iki temel boyuttan söz edilebileceğini belirtmektedirler. Hulin, Drasgow ve Parsons (1983), bu varsayımın test edilmesinde, faktör analizlerinden yararlanılabileceğine işaret etmektedirler. Bu temel varsayımın karşılanıp, karşılanmadığını incelemek üzere, Beck Umutsuzluk Ölçeği maddelerine faktör analizi uygulanmıştır. Ana bileşenler analizi ile elde edilen bulgular Tablo 1'de verilmiştir. Analiz sonucunda, test puanlarındaki varyasyonun %22'sini tek başına açıklayan bir temel faktörün bulunduğu görülmüştür.

* Umutsuzluk Ölçeği ham puanları Derebaşı'nın (1996) yüksek lisans tezi için yaptığı uygulamadan alınmıştır.

Tablo 1
Umutsuzluk Ölçeği Faktör Analizi Sonuçları

Faktör İstatistikleri						
Faktör	1'den Büyük Özdeğerler	Açıklanan Varyans Yüzdesi			Toplamlı Yüzde	
1	4.41284	22.1			22.1	
2	1.26540	6.3			28.4	
3	1.10050	5.5			33.9	
4	1.06412	5.3			39.2	
5	1.04558	5.2			44.4	
Döndürülmemiş Faktör Yükleri						
Maddeler	Faktör 1	Faktör 2	Faktör 3	Faktör 4	Faktör 5	Kommunaliti
Madde 15	.62635	.01137	.00200	-.11755	-.07871	.54307
Madde 18	.60842	.16339	.15088	-.18743	-.08718	.71664
Madde 7	.59491	-.20427	.17491	-.08434	.07214	.30409
Madde 11	.59125	-.10248	.21280	-.13237	.01587	.32276
Madde 1	.58390	.10473	-.29851	-.29687	-.11801	.50704
Madde 6	.55736	-.16996	.18842	-.04658	.10973	.38925
Madde 19	.54732	-.31920	.01250	.02292	.10797	.43855
Madde 20	.49599	-.09281	-.15802	.47408	.08455	.38708
Madde 17	.49558	.26345	.04851	.27007	-.21496	.28621
Madde 14	.48564	.32224	.17378	.08288	-.14102	.66415
Madde 16	.47755	.00850	-.18223	.39484	.05687	.42313
Madde 13	.46994	-.36655	.09530	-.09719	-.15268	.45572
Madde 9	.43944	.10804	-.12655	.20733	.14976	.39704
Madde 3	.38395	-.26735	.10141	.20802	-.17786	.39663
Madde 12	-.12648	.48223	.22471	.39262	-.05023	.41246
Madde 4	.23198	.44618	.14776	-.06405	-.11182	.42047
Madde 5	.30108	.43499	-.20187	-.41659	.11347	.43650
Madde 2	.32989	.07468	-.60862	.04490	.47938	.46236
Madde 10	.25494	.13546	.43291	-.06976	.62332	.41378
Madde 8	.33633	.02019	-.28296	-.06975	-.43431	.51149

Tablonun incelenmesinden de görüleceği gibi, oldukça güçlü olarak ortaya çıkan 1. Faktörden, 10 ve 12. Maddeler dışında test maddelerinin tamamı 0.30'dan yüksek yük almaktadırlar. Analiz sonucunda, özdeğeri 1'den büyük beş faktörün ortaya çıkması, bu maddeler içinde bazı özellikler gösteren alt gruplanmaların olabilece-

ğine işaret etmekle birlikte, ilk faktörün diğerlerine göre oldukça güçlü bir açıklama yüzdesine (%22) sahip olması (diğer faktörlerin açıklama oranları %5 civarındadır) ve hemen hemen maddelerin tümünün bu faktörden yük alması, maddelerdeki tepkileri etkileyen temel bir boyutun varlığına işaret etmektedir.

Bu deęerlendirmeler çerçevesinde, Beck Umutsuzluk Ölçeęi maddelerinin boyutsal özelliklerinin, madde-cevap kuramı ile madde analizine uygun olduęu sonucuna varılmıř ve maddelerin analizinde iki parametrelili logistik model temel alınmıřtır. Bu parametrelerden birincisi, maddenin örtük özellik üzerindeki yerini, konumunu

tayin eden 'madde-güçlük' parametresi, dięeri ise maddenin örtük özellik üzerinde farklı konumlarda bulunan kişileri ayırdedip edemediğini gösteren, yani madde karakteristik eğrisinin eğimini veren 'ayırdetme' parametresidir. Maddeler için bu parametre deęerleri Tablo 2'de verilmiştir.

Tablo 2

Umutsuzluk Ölçeęi Maddelerinin İki Parametrelili Modelle Madde Analizi Sonuçları

Madde	Eđitim (a) S.H.	Güçlük (b) S.H.	Daęılım (Hata) S.H.	Asimtot S.H.	Ki-Kare Uyum Testi	S.D.	Anlamlılık Düzeyi
1	1.055 .118*	1.140 .087*	.948 .106*	.000 .000*	15.2	7.0	.0336
2	.701 .121*	2.827 .374*	1.426 .245*	.000 .000*	5.8	6.0	.4478
3	.567 .079*	2.090 .254*	1.763 .245*	.000 .000*	7.2	7.0	.4053
4	.422 .059*	-.013 .112*	2.370 .330*	.000 .000*	10.7	8.0	.2206
5	.421 .060*	.822 .151*	2.378 .338*	.000 .000*	3.9	8.0	.8634
6	1.066 .136*	1.777 .141*	.938 .119*	.000 .000*	4.7	6.0	.5868
7	1.389 .173*	1.871 .127*	.720 .089*	.000 .000*	7.2	4.0	.1230
8	.436 .060*	.871 .151*	2.291 .316*	.000 .000*	7.4	8.0	.4962
9	.649 .078*	1.604 .164*	1.540 .185*	.000 .000*	12.0	8.0	.1517
10	.366 .058*	1.516 .249*	2.730 .432*	.000 .000*	7.3	8.0	.5073
11	1.124 .128*	1.405 .103*	.890 .101*	.000 .000*	4.2	6.0	.6468
12	.117 .028*	-3.477 .927*	8.578 2.036*	.000 .000*	24.0	8.0	.0024
13	.644 .080*	1.403 .161*	1.552 .193*	.000 .000*	7.1	8.0	.5239
14	.800 .089*	.169 .068*	1.249 .139*	.000 .000*	12.0	7.0	.0984
15	1.203 .134*	1.128 .083*	.831 .093*	.000 .000*	2.7	6.0	.8452
16	.906 .118*	2.044 .185*	1.103 .143*	.000 .000*	4.0	6.0	.6722
17	.815 .096*	1.351 .123*	1.227 .144*	.000 .000*	7.0	7.0	.4344
18	1.183 .138*	.765 .065*	.845 .098*	.000 .000*	9.1	6.0	.1652
19	1.010 .123*	1.687 .130*	.990 .120*	.000 .000*	3.2	6.0	.7880
20	1.099 .164*	2.189 .194*	.910 .136*	.000 .000*	3.5	5.0	.6212

*Standart Hata

Tablo 2'de yer alan değerlerden bir diğeri de, maddenin standart hatasını veren, 'dağılım' (dispersion) parametresidir. Bu değerler madde karakteristik eğrisinin eğimi ile ters orantılıdır (reciprocal). Eğimin azaldığı durumda bu hata değerleri artış göstermektedir. Yani örtük özellik üzerinde aynı değere sahip olan kişilerin, maddenin 'doğru' seçeneğini işaretleme oranlarının farklılaşma derecesini yani dağılımını göstermektedir. Bu değerlerin büyük olması, kişilerin test performansında eşit olmakla birlikte, maddeyi çok farklı düzeylerde seçtiklerini göstermektedir ki bu da maddenin ayırdedici konumunun yetersizliğine, hataya işaret etmektedir.

Tablo 2'deki son sütun ise, testteki her bir maddenin analizlerin yapıldığı iki parametrelili logistik modele uygunluğunu test eden χ^2 değerlerini içermektedir. Büyük bir χ^2 değeri (alttaki parantezlerde anlamlılık düzeyleri verilmiştir), maddenin modele uyumsuzluğuna işaret etmektedir. Madde-model uyumunda beklenen durum, kişilerin örtük özellik üzerindeki konumları ilerledikçe, maddeleri seçme oranlarında da düzenli bir artış olmasıdır, yani madde karakteristik eğrisinin monotonik bir şekilde

yükselmesidir. Burada madde karakteristik eğrisinin teorik konumu ile, gözlenen, deneysel eğri arasındaki fark arttıkça maddenin modelle uyum göstermediği anlaşılmaktadır. Örneğin, umutsuzluk boyutunda yüksek puanlara sahip olan kişilerin, umutsuzluğun göstergesi olan bir maddeyi olumlu olarak seçmeleri beklenirken, düşük oranda seçmeleri ya da düşük umutsuzluk puanına sahip kişilerin maddeyi seçmeleri gibi bir durum düşünülebilir. Yani madde, istenen kuramsal yapıyı ölçmeye gerektiği şekilde katkıda bulunmamaktadır. Bu durum maddenin yazımındaki teknik hatalardan kaynaklanabileceği gibi, kuramsal açıdan o maddenin ölçülen psikolojik yapının iyi bir göstergesi olmasından da kaynaklanabilmektedir.

Tablo 3'de ise, Umutsuzluk Ölçeği maddelerinin klasik kurama göre madde analizi sonuçları verilmiştir. Son iki sütundaki madde-toplam korelasyonları [Pearson ve çiftserili (biserial) korelasyonlar] klasik kuramda maddenin "ayırdetme" indeksi olarak ifade edilmektedir. Bu değerler, madde puanı ile, testten alınan toplam puan arasındaki ilişkinin derecesini göstermektedir.

Tablo 3
Umutsuzluk Ölçeği Maddelerinin Klasik Kurama Göre Madde Analizi Sonuçları

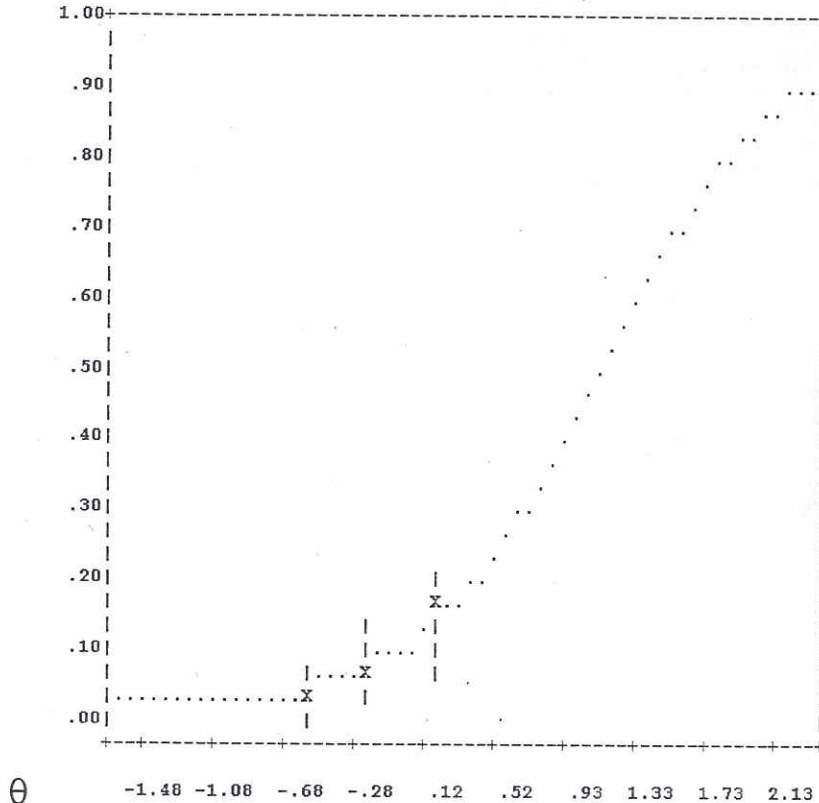
Madde	İsim	Deneme Sayısı	Doğru Sayısı	Yüzdesi	Madde Test Korelasyonu		Mad. Çık. Alfa Değeri
					Pearson	Çiftserili	
1	0001	700.0	141.0	.201	.471	.671	.7421
2	0002	700.0	40.0	.057	.249	.504	.7582
3	0003	700.0	105.0	.150	.281	.430	.7559
4	0004	700.0	351.0	.501	.240	.301	.7615
5	0005	700.0	259.0	.370	.247	.316	.7604
6	0006	700.0	69.0	.099	.434	.746	.7477
7	0007	700.0	46.0	.066	.459	.888	.7486
8	0008	700.0	251.0	.359	.254	.326	.7597
9	0009	700.0	132.0	.189	.350	.506	.7512
10	0010	700.0	206.0	.294	.198	.262	.7636
11	0011	700.0	102.0	.146	.461	.712	.7442
12	0012	700.0	465.0	.664	-.080	-.104	.7865
13	0013	700.0	154.0	.220	.335	.468	.7522
14	0014	700.0	319.0	.456	.408	.513	.7462
15	0015	700.0	133.0	.190	.506	.732	.7397
16	0016	700.0	61.0	.087	.364	.648	.7520
17	0017	700.0	136.0	.194	.413	.594	.7465
18	0018	700.0	193.0	.276	.505	.676	.7381
19	0019	700.0	81.0	.116	.408	.670	.7485
20	0020	700.0	39.0	.056	.378	.771	.7529

Klasik Kuramda 'madde güçlüğü' indeksi ise, maddeye doğru cevap verenlerin, toplam örneklem sayısına oranıdır. Bu oran yetenek testlerinde maddenin gruba göre ne kadar güç ya da kolay olduğunu bir göstergesi olarak ele alınmaktadır. Kişilik testlerinde ise bu değerler, tam olarak yetenek testlerindeki gibi fonksiyonel bir kullanıma sahip olmasa da, maddelerin o kişilik boyutu üzerindeki yerine ilişkin bir fikir verebilir. Ancak, madde-güçlük değerlerinin yorumlanmasında önemli bir problem, bu değerlerin büyük ölçüde örneklem özelliklerinden etkilenmesidir. Dolayısıyla madde-güçlüğü parametresinin değerleri, örneklem yetenek düzeyine göre önemli farklılıklar gösterebilmektedir. Bu açıdan klasik modele göre geliştirilen testlerde, örneklem büyüklüğü ve temsil ediciliği, maddelerin güçlük derecelerine göre ölçeklenmesinde fazlasıyla önem taşımaktadır.

Wright ve Stone (1979), üç parametrelili logistik modelin özel bir hali olan Rasch modeline göre yapılan analizlerde, örneklem yetenek dağılımından bağımsız madde ölçeklenmesinin ve maddelerin dağılımından bağımsız olarak kişilerin yetenek tahminlerinin yapılmasının mümkün olduğunu belirtmektedirler.

Madde-cevap kuramına göre yapılan madde ölçeklenmesinde, maddelerin güçlük değerleri yetenek boyutu ya da ölçülen kişilik boyutu üzerindeki konumlarına göre ölçeklenmektedir. Bu durum test geliştiricilere büyük bir avantaj sağlamaktadır. Böylece kişilerin yetenek düzeyleri ya da kişilik boyutu üzerindeki konumları ile, maddenin aynı boyut üzerindeki konumlarını karşılaştırmak mümkün olmakta ve böylece ölçmeyi hedeflediğimiz grubu merkez alacak maddeleri seçmekte, klasik kuramdan çok daha etkin bir biçimde yol gösterici olmaktadır.

Madde 15 Olasılık < .8452



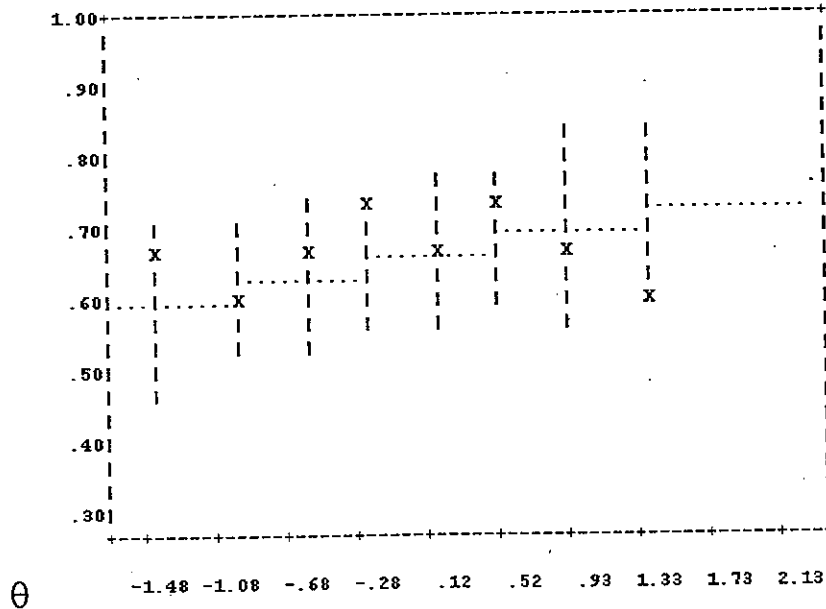
Şekil 1. Modelle uyumlu ve ayırıcı bir maddenin (Madde 15) karakteristik eğrisi

Maddelerin kalitesinin incelenmesinde yol gösterici bir diğer bilgi kaynağı Madde Karakteristik Eğrileridir. Ayırdetme (a_i) ve güçlük -tutum ve kişilik ölçeklerinde popülerlik- (b_i) parametreleri madde karakteristik eğrisini tanımlayan parametrelerdir. Ancak bu parametreler bir dereceye kadar özet bilgi sağlamaktadır. Örneğin ayırdetme indeksi, maddenin sahip olduğu güçlük düzeyinde eğimi vermektedir. Bu güçlük düzeyinde ayırdedici olan yani dik bir eğim gösteren madde karakteristik eğrisinin eğimi diğer bölgelerde oldukça düşük olabilir. Ya da Hulin, Drasgow ve Parsons'un (1983) değindiği gibi, orta ve daha düşük düzeylerde ayırdedici olan bir madde, hedef grubun özellikleri iyi tanınmadığında, daha çok kişiyi ayırdetme gibi bir fonksiyonla testin oluşumuna katkıda bulunabilir. Bu nedenle madde karakteristik eğrilerini içeren grafiklerin incelenmesi, araştırmacıya, farklı θ düzeylerinde maddenin sağladığı farklı ayırdetme düzeyleri ve θ düzeylerine göre maddelerin doğru cevaplanma olasılıkları hakkında görsel bir bilgi sağlayacak ve problemlili maddelerin yorumuna katkıda bulunacaktır. Şekil 2 ve 3'te Umutsuzluk Ölçeği içerisinde problemlili iki madde, Şekil

1'de ise problemsiz bir madde için madde karakteristik eğrisi örnekleri verilmiştir.

Şekil 2'de yer alan 12. Madde ("Gerçekten özlediğim şeylere kavuşabileceğimi ummuyorum."), analiz sonuçlarına göre Umutsuzluk Ölçeği içinde görülen en problemlili maddedir. Tablo 2'nin son sütunu incelendiğinde, bu maddenin modelle uyum sağlamadığı yani beklenen madde karakteristik eğrisi ile, gözlenen eğrinin anlamlı biçimde birbirinden farklılaştığı, elde edilen yüksek χ^2 değerinden (24) ve anlamlılık düzeyinden (0.0024) anlaşılmaktadır. Yine Tablo 2'den maddenin ayırdetme değerine bakıldığında, 0.117 olduğu görülmektedir. Bu değer sıfıra oldukça yakın ve çok düşük bir ayırdetme değeridir. Yine maddenin güçlük düzeyinin -3.447 olması (madde ve yetenek düzeylerinin ölçeklemesinde kullanılan ölçeğin başlangıç noktası keyfidir, ancak ölçekleme genellikle ortalaması sıfır olan ve birim varyansa sahip bir ölçek üzerinde yapılmaktadır), maddenin umutsuzluk boyutunda çok uzak bir konumda olduğunun yani umutsuzluk düzeyi çok düşük kişilerin yüzde ellisinin bu maddeyi doğru olarak işaretlediklerinin bir göstergesidir.

Madde 12 Olasılık < .0024



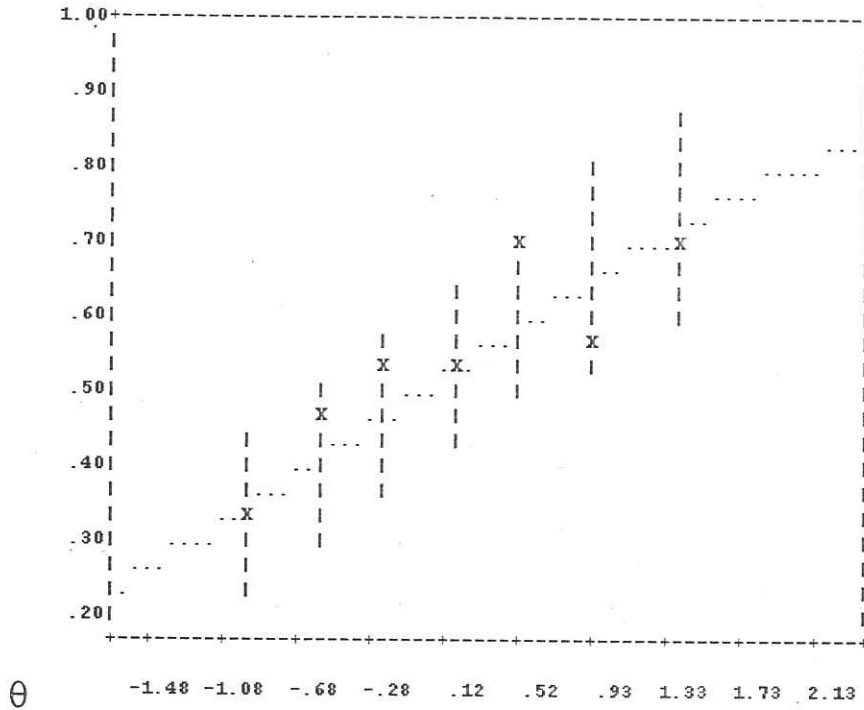
Şekil 2. Modele uyumsuz ve ayırdediciliği düşük bir maddenin (Madde12) karakteristik eğrisi

Ayrıca Şekil 2'deki madde karakteristik eğrisinin incelenmesi de, maddenin ayırdedicilik düzeyinin çok düşük olduğunu net bir şekilde görselleştirmektedir. Örtük özellik boyutu (θ) üzerinde en düşük umutsuzluk düzeyinden, en yüksek umutsuzluk düzeyine doğru gidildikçe, maddeyi "doğru" olarak cevaplayanların oranında hemen hemen hiç bir artış görülmemekte, eğri θ boyutuna yaklaşık olarak paralel bir konum sergilemektedir.

Ayırdedicilik özelliği yüksek bir maddenin karakteristik eğrisi ise, Şekil 1'de görülmektedir (Madde 15: "Ailemdeki kişiler nasıl davranacakları ya da bazı şeyleri nasıl yapacakları konusunda benim fikirlerimden

yararlanırlar."). Şekil 1'den görülebileceği gibi, eğri oldukça dik bir şekilde yükselmekte, kişilerin umutsuzluk boyutu üzerindeki konumları ilerledikçe maddeye "yanlış" cevabını verme oranları düzenli bir şekilde artış göstermekte ve aynı düzeydeki kişilerin, cevap oranlarında fazla bir dağılıma görülmemektedir. Tablo 2'den bakılabileceği gibi, maddenin ayırdetme indeksi ($\alpha = 1.203$) de oldukça yüksek bir değere sahiptir. Yine Tablo 2'nin son sütununa bakıldığında, görülen düşük χ^2 değeri de maddenin modelle uyumunun iyi olduğuna işaret etmektedir. Yani gözlenen eğri, teorik eğriden anlamlı bir şekilde farklılaşmamaktadır.

Madde 4 Olasılık < .2206



Şekil 3. Modelle uyumlu, ayırdediciliği nispeten düşük bir maddenin (Madde 4) karakteristik eğrisi.

Tablo 3'te görülebileceği gibi, 15. maddenin klasik kuramla yapılan madde analizlerinden elde edilen ayırdetme indeksi (madde toplam korelasyonu) 0.732'dir. Bu değer de madde-cevap kuramıyla yürütülen madde analizi sonuçları ile uygunluk göstermekte ve maddenin kaliteli, bu tür öl-

çeklerde istenen türde bir madde olduğuna işaret etmektedir.

Yukarıda özellikleri tartışılan 12. maddenin ise çiftserili korelasyon katsayısı -0.104'tür. Bu çok düşük değer de klasik ayırdetme indeksinden yorumlanabileceği gibi, bu maddenin test açısından istenme-

yen özelliklere sahip, umutsuzluk boyutu üzerinde kişilerin farklı konumlarını ayırdetme gücü düşük bir madde olduğunu göstermektedir. Klasik kuramla elde edilen bu yorumlar, yukarıdaki paragraflarda tartıştığımız modern kuramdan elde edilen sonuçlarla uygunluk göstermektedir.

Kuramsal temellerinde fazla bir farklılık olmaması nedeniyle, klasik ve modern modellere dayalı olarak yürütülen madde analizi sonuçları arasında genel bir uygunluk bulunması beklenen bir sonuçtur. Ancak, Wright ve Stone (1979), klasik kuramda ayırdetme indeksi olarak kullanılan nokta-çiftserili (point-biserial) korelasyonun, maddenin uyum durumunun ve ayırdediciğinin incelenmesinde olumsuz bazı özellikleri olduğuna dikkat çekmektedirler. Yazarlar, çiftserili korelasyon değerlerinin puanların varyansından fazla etkilenmesi nedeniyle, örneğin dar dağılımı olan bir grupta ya da maddenin hedef gruba uzaklığına bağlı olarak, nokta-çiftserili korelasyon değerinin, maddenin kalitesine duyarız kalarak sifıra yaklaşabileceğini belirtmektedirler. Berberoğlu (1989) ve Somer (1994) de bulgularında, klasik ve modern madde analizi sonuçlarının genelde bir uyum göstermekle birlikte, bazı maddelerde ve özellikle madde-model uyumu ve ayırdetme düzeyi açısından farklılıklar gösterebildiğine işaret etmektedirler.

Çalışmamız sonuçlarında, 4. Madde ("Gelecek on yıl içinde hayatımın nasıl olacağını hayal bile edemiyorum.") bu tür bir uyumsuzluğa örnek olarak verilebilir. Dördüncü maddenin çiftserili korelasyon değeri 0.30'dur. Klasik kuramla yapılan madde analizlerinde, genellikle 0.30 ve üzerindeki madde toplam korelasyon değerleri maddenin ayırdetme kalitesinin yeterli olduğuna işaret etmektedir. Oysa iki-parametrelili modelle elde edilen ayırdetme indeksi 0.422'dir ve 0 ile 2 arasında elde edilecek ayırdetme değerleri göz önüne alındığında oldukça düşük bir değer olarak görülmektedir. Şekil 3'te maddenin karakteristik eğrisi verilmiştir. Bu eğrinin incelenmesi duruma açıklık getirmektedir. Şekilden görülebileceği gibi, kişilerin umutsuzluk boyutundaki düzeyi arttıkça, maddeye olumlu cevap verme oranlarında da bir artış görülmekte, yani

eğri monotonik bir şekilde fonksiyon göstermektedir. Ancak ayırdetme indeksinin de işaret ettiği gibi, eğrinin yeteri kadar dik olmaması maddenin standart hatasını (dispersion) yükseltmektedir. Bu durum, örtük özellik boyutu üzerinde aynı konumlarda olan kişilerin, maddeyi olumlu olarak cevaplama oranlarının geniş bir dağılım gösterdiğine işaret etmektedir. Yani farklı düzeylerde, olumlu cevap oranlarındaki örtüşmeler fazla olmaktadır. Maddenin içeriği incelendiğinde, kültürel etkinin ve yaşam koşullarının maddenin ayırdediciği üzerinde olumsuz bir şekilde etkili olduğu düşünülebilir. Klasik kuramla elde edilen sonuçların işaret ettiği gibi maddenin bir dereceye kadar ayırdetme özelliği vardır. Yani genelde umutsuzluk düzeyi yüksek kişiler bu maddeyi daha çok olumlu olarak yanıtlamaktadırlar. Ancak yine Şekil 1'den, umutsuzluk özelliği farklı düzeylerde olan kişilerin de bu maddeyi sıklıkla doğru cevaplayabildiği görülmektedir. Ülkemizdeki sosyal, ekonomik ve politik koşulların istikrarı düşünüldüğünde maddenin farklı kültürler ve yaşam koşulları içerisinde farklı biçimde fonksiyon göstermesinin anlamlı olabileceği düşünülmektedir.

Beşinci maddede de, benzer bir durum görülmekte, ayrıca 1. Madde de yüksek bir ayırdetme düzeyine sahip olmakla birlikte modelle uyumsuzluğu anlamlı düzeyde (0.03) görülmektedir. Diğer maddeler genelde her iki modelde uyumlu sonuçlar vermektedir.

Madde-cevap kuramıyla madde analizlerinde maddelerin seçiminde bir diğer önemli bilgi kaynağı, madde ve test bilgi eğrilerinin incelenmesidir. Hambleton ve Swaminathan (1989), madde-cevap kuramlarında, test bilgi fonksiyonlarının klasik kuramlardaki güvenilirlik ve ölçmenin standart hatası kavramlarına karşılık geldiğini belirtmektedir. Test bilgi fonksiyonlarının, testle yapılan yetenek tahminlerinin doğruluğunun bir ölçüsü olabileceğini gösteren iki neden vardır. Bunlardan birincisi, formunun tamamıyla testte içerilen maddelere dayanması, ikincisi ise, yetenek düzeylerinin her biri için ayrı ayrı ölçme hatası vermesidir.

Herhangi bir test geliştirildiğinde ölçmeye yöneldiği bir hedef grup vardır. Test maddelerinin bu hedef grubun göstermesi beklenen dağılımı ölçmeye uygun maddelerden oluşması ya da bu dağılımdan uzakta kalan maddeleri içermesi, test sonuçlarının bu grup için bize sağlayacağı bilginin kalitesini etkiler. Örneğin, geniş bir madde dağılımı olan bir başarı testinde, kendilerine güç gelen bir çok maddeyi cevaplamak durumunda olan düşük yetenekli kişiler söz konusu olduğunda, bu kişiler kendi yeteneklerine uygun az sayıda madde ile uğraşmanın yanı sıra, kendi yeteneklerinin çok üzerinde olan maddeleri cevaplamakla uğraşacaklar ve muhtemelen bu maddelerde tahmine gideceklerdir. Bu durumda, testin bu kişiler açısından sağlayacağı bilginin kalitesi düşük olacaktır.

Test maddelerinin seçimindeki bu kaygılar, teknik olarak bilgi terimi ile ifade edilmektedir. Madde-cevap kuramında, ölçmenin doğruluğu, maddelerin sayısı ile her maddenin belirli bir kişiye göre olan durumuna bağlıdır. Madde ve kişi birbirine yakın olduklarında, örneğin tam hedefte olduğunda, madde kişinin ölçümüne uzak olduğu durumdan daha fazla katkıda bulunur. Fark arttıkça maddenin etkinliği azalır ve aynı doğrulukta ölçüm yapmak için daha fazla sayıda madde gerekir.

Bir test, belirli bir özelliği ölçen ölçek üzerindeki farklı bölgelerde bulunan kişiler için farklı düzeylerde bilgi sağlamaktadır. Daha hassas ayırımların yapılabilirdiği bölgelerdeki kişiler için test daha fazla bilgi sağlamakta ve bir testin sağladığı bilgi maddelerin karakteristik eğrilerine dayanmaktadır (Crocker ve Algina, 1986). Her madde, örtük özellik boyutundaki farklı puanlar için, farklı düzeylerde bilgi sağlamaktadır. Madde karakteristik eğrisinin eğiminin en dik olduğu konuma karşılık gelen puan aralıkları için, madde en fazla bilgiyi sağlamaktadır. Dolayısıyla her mad-

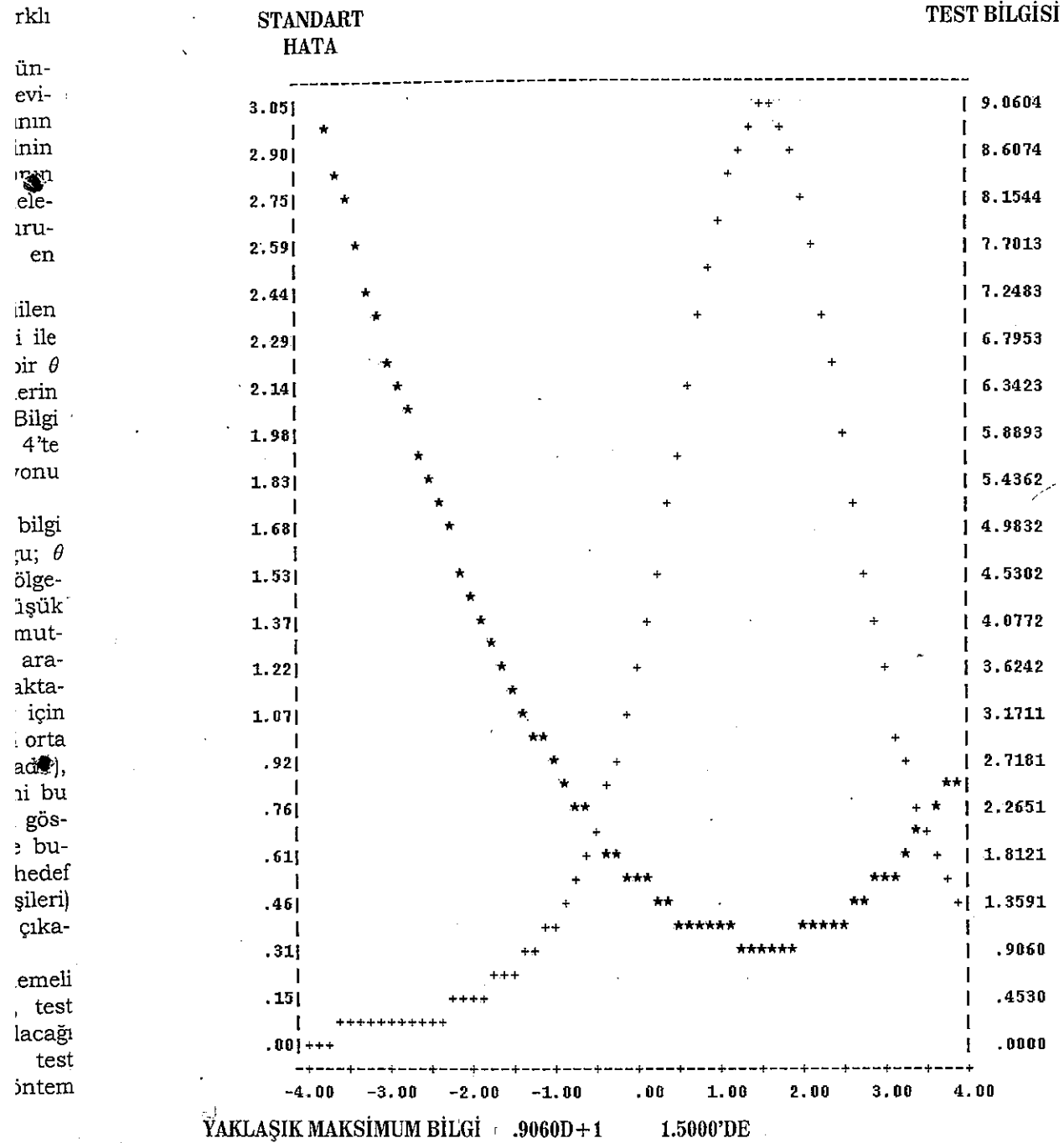
de, θ ölçeği üzerinde farklı noktalarda farklı bilgi vermektedir.

Burada standart hata kavramı gündeme gelmektedir. Belirli bir yetenek seviyesi için bilgi, ölçmenin standart hatasının karesi ile ters orantılıdır. Test ölçümlerinin doğruluğu, ölçmenin standart hatasının küçülmesi oranında artmaktadır. Maddelerin tam hedefe uygun dağılımları durumunda, standart hataların dağılımını en küçüğe indirmek olanaklıdır.

θ 'nın her bir değeri için elde edilen bilgilerin θ 'ya karşı grafiğinin çizilmesi ile Madde Bilgi Eğrileri elde edilir. Belirli bir θ için madde bilgilerinin, testteki maddelerin tümü için toplanması ile de, Test Bilgi Fonksiyonu elde edilmektedir. Şekil 4'te Umutsuzluk Ölçeğinin test bilgi fonksiyonu verilmiştir.

Şekil 4'te standart hata ve test bilgi eğrilerinin birbiriyle ters orantılı olduğu; θ boyutu üzerinde, test bilginin arttığı bölgelerde standart hata düzeylerinin düşük olduğu görülmektedir. Test bilgisi, umutsuzluk düzeyi tahminleri 0.50 ve 3.0 arasında olan kişiler için en yüksek olmaktadır. Bu durum, bu aralıktaki kişiler için (bu aralık daha çok umutsuzluk düzeyi orta ve yüksek olan kişileri kapsamaktadır), testin daha doğru bilgi sağladığını yani bu bölgedeki kişileri daha iyi ayırdettiğini göstermektedir. Testin amacı göz önünde bulundurulduğunda, test maddelerinin hedef grubu (umutsuzluk düzeyi yüksek kişileri) temsil etmeye uygun olduğu sonucu çıkarılabilir.

Lord (1980), maddelerin kademeli olarak teste sokulup çıkartılmasıyla, test bilgi fonksiyonunun hedef grup için alacağı en uygun konumun bulunmasının test maddelerinin seçiminde uygun bir yöntem olacağını belirtmektedir.



Şekil 4. Umutsuzluk Ölçeği test bilgi ve standart hata eğrileri grafiği

Tartışma

Bu makalede, bir kişilik ölçeğinin psikometrik özellikleri, madde faktör analizi, klasik ve modern kuramlarla madde a-

nalizi süreçleriyle incelenmiş, klasik ve modern kuramlarla elde edilen sonuçların benzer ve farklı yönleri tartışılmıştır.

Madde faktör analizleri kişilik testlerinde, boyut oluşturmada, yapı geçerliğinin

incelenmesinde sıklıkla kullanılan bir yöntemdir. Ancak madde faktör analizleri maddelerin ölçülmek istenen kişilik boyutuyla ilişkilerini incelemekte önemli bir adım olmakla beraber, maddelerin boyut üzerindeki konumları, ölçülen özellik bakımından farklı düzeylerdeki kişileri ayırdedebilmek gibi psikometrik özelliklerini inceleme imkanı vermemektedir.

Bu aşamadan sonra maddelerin kalitesinin incelenmesi, gerekli düzeltmelerin yapılabilmesi için madde analizleri gerekmektedir. Modern kuramla madde analizleri, yetenek ve başarı testlerinin geliştirilmesi ve düzeltilmesinde son yıllarda yaygın olarak kullanılmasına rağmen, Steinberg ve Thissen'in (1995) de belirttiği gibi kişilik testlerinde kullanımı pek yaygın değildir.

Oysa, yukarıdaki incelemeden de görülebileceği gibi, madde-cevap kuramıyla elde edilen madde analizi sonuçları, maddelerin kalitesi hakkında klasik kurama göre çok daha ayrıntılı inceleme yapma fırsatı sağlamaktadır. Steinberg ve Thissen (1995), bir testte birbiriyle az da olsa korelasyon gösteren çok sayıda madde bulunmasının, güvenilirliği yüksek bir ölçekle sonuçlanacağını, oysa maddeler arasındaki korelasyonların nedeninin aynı boyutu ölçmeleri olabileceği gibi, birbiriyle korelasyon gösteren farklı boyutları ölçmelerinden de kaynaklanabileceğini belirtmektedirler. Nunnally (1978) de, iç tutarlılık güvenilirlik katsayısının yüksek olmasının tek boyutluluk anlamına gelmeyeceğine dikkat çekmektedir. Steinberg ve Thissen (1995), madde-cevap kuramı ve madde faktör analizlerinin, klasik kuramla incelenemeyen bu durumu ayırdedebileceğine işaret etmektedirler.

Lord (1980), madde cevap kuramında hiçbir şeyin, klasik test kuramının varsayımları ve çıkardığı temel sonuçlar ile çelişmediğini belirtmekle birlikte, klasik kuramın cevaplayamadığı bazı soruları yanıtlayabilmek için ek bazı varsayımların kullanıldığına işaret etmektedir. Lord, madde cevap kuramının, klasik kuramla çelişmediğini aksine ileriye götürdüğünü belirtirken, yine de modern kuramda klasik kuramdan bilinen haliyle şartıcı biçimde

az yararlanıldığını ifade etmektedir. Berberoğlu'nun (1989) belirttiği üzere, Helsley, Suber ve Ryan (1977), Rasch modeli ile elenen maddelerin klasik kuram yöntemleri ile de eleneceğini öne sürmüşler ve Rasch modelini klasik test kuramının matematiksel bir uzantısı olduğunu göstermişlerdir.

Baykul (1980), yeteneği ölçmede örtük özellikler ve klasik kuramlarla geliştirilen testlerin geçerlik bağlamında farklı sonuçlar vermediğini ancak örtük özellikler kuramına dayalı modellerle, daha güvenilir ve örneklem farklılıklarından daha az etkilenen testler elde edilebileceğini belirtmektedir.

Wright ve Stone'da (1979), Rasch modelinin test geliştirmede eski istatistiklerin bir versiyonu olduğuna işaret etmekte ve bu yeni istatistiklerin geleneksel test geliştirmedeki ölçme problemlerinin birçoğunu çözdüğünü belirtmektedirler.

Hambleton ve Swaminathan (1989), madde ve test bilgi eğrilerinin incelenmesi yoluyla, farklı maddelerden oluşan örneklerin birbirlerine eşdeğerli olarak ölçeklenebilmesi gibi avantajlarla, madde cevap kuramının hedef grubu çok daha iyi temsil eden, daha az sayıda madde ile, çok daha güvenilir ölçekler oluşturmaya imkan sağladığına işaret etmektedirler. Hulin, Drasgow ve Parsons (1983) da, klasik kuramda düşük çiftserili korelasyona sahip oldukları için elenen bazı maddelerin, bilgi kavramının ışığında, hedef grup için yüksek bilgi sağlaması durumunda teste katkıları göz önüne alınarak seçilebileceklerini belirtmektedirler. Zira modern kuramlar maddenin ilgili boyuta katkısını incelememizde klasik kurama göre çok daha ayrıntılı veriler sağlamaktadır.

Yukarıda tartışılan noktalar göz önünde bulundurulduğunda, madde-cevap kuramının ortaya koyduğu madde analizi modellerinden yetenek testlerinde olduğu gibi, kişilik ölçeklerinin de geliştirilmesi ve adaptasyonunda daha yaygın olarak yararlanılmasının boyutsal özellikleri daha güçlü ölçekler oluşturmaya katkıda bulunacağı sonucu çıkmaktadır.

Kaynaklar

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement*, 13(2), 113-117.

Andrich, D. (1988). *Rasch models for measurement*. California: Sage Publications, Inc.

Baykul, Y. (1980). Örtük özellikler ve klasik test kuramları üzerine bir karşılaştırma. *Doğa*, 33-40.

Beck, A. T., Weisman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42, 861-865.

Berberoğlu, G. (1989). Erişi testlerine madde seçiminde Klasik Test Kuramı ve Rasch Modeli'nin karşılaştırılması. *Eğitim ve Bilim*, 13(74), 61-74.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College Publishing.

Derebaşı, I. (1996). *Beck Umutsuzluk Ölçeğinin Ege Üniversitesi öğrencileri üzerinde geliştirilmesi*. Yayınlanmamış yüksek lisans tezi, Ege Üniversitesi, İzmir.

Gulliksen, H. (1958). *Theory of mental tests*. New York: John Wiley.

Hambleton, R. K., & Swaminathan, H. (1989). *Item Response Theory: Principals and application*. Boston: Kluwer Nijhoff Publishing.

Helmes, E., & Jackson, D. N. (1989). Prediction models of personality item responding. *Multivariate Behavioral Research*, 24, 71-91.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item Response Theory: Application to psychological measurement*. Illinois: Dow Jones-Irwin.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum.

Lord, F. M., & Novick, M.R (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley Publishing.

Mislevy, R. J., & Bock, R.D. (1986). *BILOG. Item analysis and test scoring with binary logistic models*. Indiana: Scientific Software, Inc.

Mislevy, R. J., & Stocking, M. L. (1989). A Consumers guide to LOGIST and BILOG. *Applied Psychological Measurement*, 13(1), 57-75.

Noruris, M. J. (1990). *SPSS Base System User's guide*, Chicago: SPSS Inc.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Özçelik, D. A., & Berberoğlu, G. (1989). Ölçmede objektiflik ve Rasch modeli. *Eğitim ve Bilim*, 13(72), 34-44.

Somer, O. (1994). *İlkokul birinci sınıf okuma ve matematik ön becerilerinin değerlendirilmesinde bir Rasch Model Uygulaması*. Yayınlanmamış doktora tezi, Ege Üniversitesi, İzmir.

Steinberg, L., & Thissen, D. (1995). Item Response Theory in personality research. In P. E. Shrout, & T. Fiske, (Eds.), *Personality research, methods, and theory* (pp. 161-181). New Jersey: Lawrence Erlbaum.

Traub, R. E. (1983). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.), *Applications of Item Response Theory*. Vancouver, BC: Educational Research Institute of British Columbia.

Wright, B. D., & Stone, M. H. (1979). *Best test design, Rasch measurement*. Chicago: Mesa Press.

Summary

Item Analysis Procedures Based on Classical and Modern Test Theory in Personality Tests

Oya SOMER*
Aegean University

Limitations of Classical Test Theory, such as sample dependency of item parameters (Gulliksen, 1958) and inapplicability to interval scales (Wright and Stone, 1979), compelled researchers to search for new models. Beginning by the 1930's these efforts resulted in today's Latent Trait Theory or Item-Response Theory (IRT). Although it can be seen as an extension of the classical approach, IRT offers some new solutions and perspectives for the questions that the classical approach could not deal with effectively.

As pointed out by Steinberg and Thissen (1995), IRT has been used more extensively in ability testing than in personality testing, which seems to have more serious scaling problems. A growing interest has recently emerged for modern test developing models in the field of personality testing.

IRT offers new models which relate psychological constructs to the individual's observed responses in a probabilistic way. In other words, IRT contains mathematical functions for the probability of the individual's responses based on certain item and individual characteristics.

As Hambleton and Swaminathan (1989) indicated, estimating an examinee's ability independent of a particular sample of test items; estimating item parameters independent of a particular sample of examinees; and providing statistics indicating precision in estimating each examinee's ability are the primary advantages of item response models.

This study is an application of IRT to a personality scale, the Beck Hopelessness Scale. Item difficulty, item discrimination, item-model fit and information concepts were used to evaluate the Beck Hopelessness Scale items; and qualifications, advantages and disadvantages of modern and classical item analysis procedures were compared.

Method

Participants and Measures

Seven-hundred (378 women, 322 men) university students from several departments in Aegean University participated in this study. The Beck Hopelessness Scale is a self-report personality inventory and has 20 items scored dichotomously.

Analysis

Item analysis was performed by a computer program (BILOG) which is based on IRT. Item difficulties, item discriminations, item-model fit and test information function were obtained. Some of the items were taken as examples and were evaluated using both the modern and classical theories.

In order to examine the appropriateness of the scale for the IRT analysis, unidimensionality of items was checked by a factor analysis.

* Address for correspondence: Dr. Oya Somer, Ege Üniversitesi, Edebiyat Fakültesi, Psikoloji Bölümü, 35100, Bornova İzmir, TURKEY.

Results and Discussion

Hambleton and Swaminathan (1989) and Hulin, Drasgow, and Parsons (1983) point out that generally more than one factor emerge in responses of individuals, and that it is almost impossible to find a purely unidimensional response set in practical situations. However, they also assert that in practical situations it is sufficient to find a dominant factor which is influential on test performance. This dominant factor has generally been accepted as the basic ability that is being measured.

The factor analysis showed that, based on the argument presented above, the items of the Beck Hopelessness Scale can be accepted as unidimensional and one met. Factor loadings (unrotated) are presented in Table 1. Most of the items had loadings greater than .30 on the first factor.

Item analysis was conducted using a two-parameter logistic model. Classical item parameters were also computed. The item parameters obtained by the logistic model are presented in Table 2 and the parame-

ters obtained by the classical model are presented in Table 3. Examples of the item characteristic curves of some of the items which had good or poor fit to the model are presented in Figures 1, 2, and 3, and the test information curve is presented in Figure 4.

Results of item factor analysis, specifications of item parameters obtained through both the classical and modern procedures, and the contributions of the modern theory are discussed. Furthermore, the two item analysis procedures are compared and contrasted.

The results suggested that item analysis procedures based on IRT allow for a more detailed evaluation of the psychometric characteristics of the test items than do the procedures based on the classical model. It seems fair to conclude that the relations between the items and the latent trait can be analyzed more precisely by the modern test theory than the classical test theory. Extensive use of IRT in personality measurement might contribute to the development of personality scales with higher construct validity.