



<http://kefad.ahievran.edu.tr>

# Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi

ISSN: 2147 - 1037

## Review of Psychometric Properties of Minnesota Multiphasic Personality Inventory According to Item Response Theory

Şeyma Erbay Mermer  
Selahattin Gelbal

### Article Information



CrossMark

DOI: 10.29299/kefad.989587

Received: 21.09.2021

Revised: 02.02.2022

Accepted: 01.03.2022

### Keywords:

MMPI,

Item Response Theory,

2 Parameter Logistic Model

### Abstract

In the current research, the psychometric properties of the subtests of the MMPI inventory were examined within the scope of item response theory. In this context, item fit indexes were compared under 1PLM, 2PLM and 3PLM, item fit and discrimination indexes were examined, test information functions and total informations were calculated for all subtests. The research data consists of 730 people over the age of 18 who were chosen randomly. All the results of the research were analyzed in the R studio program. Unidimensionality and local independence, which are the assumptions of IRT, were checked. Parallel analysis method confirmed that each subtest was unidimensional. The unidimensionality assumption can be presented as evidence for the local independence assumption, the assumptions thought to be provided. In terms of model-data fit, it was found that 2PLM was the best fitting model of all. In test information functions at medium range of  $-1 < \theta < +1$ , were calculated that the most informative subtest was Sc(34.97), and the least information subtest was L(3.5). Considering the test information, it was calculated that the highest information belonged to the Sc subtest with 84.61, followed by the Hy subtest with 45.76.

## Minnesota Çok Yönlü Kişilik Envanteri'nin Madde Tepki Kuramına Göre Psikometrik Özelliklerinin İncelenmesi

### Makale Bilgileri



CrossMark

DOI: 10.29299/kefad.989587

Yükleme: 21.09.2021

Düzeltilme: 02.02.2022

Kabul: 01.03.2022

### Anahtar Kelimeler:

MMPI,

Madde Tepki Kuramı,

2 Parametrelili Lojistik Model

### Öz

Mevcut araştırma kapsamında MMPI envanterinin alt testlerinin psikometrik özellikleri madde tepki kuramı kapsamında incelenmiştir. Bu bağlamda 1PLM, 2PLM ve 3PLM altında madde uyum indeksleri karşılaştırılmış, madde uyumları ve ayırt edicilik indeksleri incelenmiş, test bilgi fonksiyonları toplam bilgi miktarları tüm alt testler için hesaplanmıştır. Araştırma verileri 18 yaş üstü seçkisiz olarak seçilen 730 kişiden oluşmaktadır. Araştırmaya ait tüm analizler R studio programında yapılmıştır. Araştırma sorularına geçilmeden önce MTK varsayımları sınanmış, tek boyutluluk ve yerel bağımsızlık durumları test edilmiştir. Tek boyutluluk için paralel analiz yöntemi uygulanmış ve her alt testin tek boyutlu olduğu doğrulanmıştır. Tek boyutluluk varsayımının karşılanmasının yerel bağımsızlık varsayımı için kanıt olarak sunulabileceğinden varsayımların karşılandığı görülmüştür. Model-veri uyumu kapsamında en uygun modelin 2PLM olduğu görülmüş ve madde uyumlarının yine en iyi 2PLM için sağlandığı hesaplanmıştır. Test bilgi fonksiyonlarında orta düzeyde yeteneğe karşı gelen  $-1 < \theta < +1$  aralığında en fazla bilgi sağlayan testin Sc (34.97) olduğu, en az bilginin ise L (3.5) alt testine ait olduğu, toplam bilgi miktarına bakıldığında en yüksek bilginin 84.61 ile Sc alt testine ait olduğu, onu 45.76 ile Hy alt testinin takip ettiği görülmüştür.

**Sorumlu Yazar :** Şeyma Erbay Mermer, Dr, Bilecik Şeyh Edebali Üniversitesi, Türkiye, sey.erbay@gmail.com, ORCID ID: 0000-0002-7747-9545

Yazar2 Selahattin Gelbal, Prof, Hacettepe Üniversitesi, Türkiye, gelbal@hacettepe.edu.tr, ORCID ID: 0000-0001-5181-7262

**Atf için:** Erbay Mermer, Ş. & Gelbal, S. (2024). Minnesota çok yönlü kişilik envanterinin madde tepki kuramına göre psikometrik özelliklerinin incelenmesi. *Kırşehir Eğitim Fakültesi Dergisi*, 25(1), 1-28.

## GİRİŞ

Minnesota Çok Yönlü Kişilik Envanteri (Minnesota Multiphasic Personality Inventory/MMPI) (Hathaway ve McKinley 1940), kişilik özelliklerini ve bireyin kişisel ve toplumsal uyumunu objektif olarak değerlendirmeyi amaçlayan bir ölçektir. Klinik alanda sıkça başvurulmuş insan kaynakları, eğitim gibi alanlarda da kullanılan bu ölçeğin Türkiye uyarlaması Savaşır (1981) tarafından gerçekleştirilmiştir. Yaygın olarak kullanılan MMPI, alt ölçekleri arasındaki korelasyonun yüksek olması; alt ölçeklerin adlandırıldıkları belirti kümesini (ör., şizofreni veya depresyon alt ölçekleri gibi) net bir şekilde temsil etmemeleri; bazı maddelerin ifadelerinin net olmayışı, cinsiyetçi olması, doğrusal puanlama olmayışından dolayı karşılaşılan güçlükler gibi birçok eleştirinin de hedefi olmuştur (Butcher ve Pope 1992).

MMPI, zihinsel sağlık bozukluklarını değerlendirmek, tanı ve teşhis koymak ya da tedavi sonrası tedavinin etkinliğini değerlendirmek için kullanılan bir kişilik testidir. Klinik alanda sıkça başvurulmuş bu testin incelenmesi, analiz edilmesi ve yorumlanmasını gerektirmektedir. Test kuramları ile dolaylı yollarla ölçülen psikometrik özelliklerin analizinde karşılaşılan problemlerin çözülmesi amaçlanmaktadır. Bu kuramlar Klasik Test Kuramı ve Madde Tepki Kuramı olmak üzere ikiye ayrılır.

Klasik Test Kuramı (KTK) gerçek puanın, gözlenen puan ve rastgele hatadan oluştuğunu varsayar. MTK ise madde parametrelerine ve bireyin özelliklerine göre yetenek kestirimi yapan bir kuramdır. KTK'nın yeterli olmadığı gözlenerek geliştirilen MTK, madde parametrelerinin değişmezliği, yetenek kestiriminin teste bağlı olmaması gibi özellikleri ile KTK'ya karşı daha güçlüdür. Kavramsal olarak MTK'nın, birey özellikleriyle madde etkileşiminin sonucunda bireyin performansını tanımlayan olasılık temelli bir teori olduğu söylenebilir (Faulkner-Bond ve Wells, 2016). MTK'da testten alınan toplam puan yerine her bir maddeye verilen cevaplarla ilgilenilir ve bu maddeler sürekli veya süreksiz, iki kategorili veya çok kategorili olarak puanlanabilir. Kuram kapsamında maddelere verilen cevapların her biri ile bireyin yeteneğini ortaya çıkarmaya yarayan birçok model tanımlanmıştır (Hambleton ve Jones, 1993).

MTK model geçerliliğinin sağlanması için verinin modele kabul edilebilir derecede uyum göstermesi gerekmektedir. Model-veri uyumsuzluğu varsayımlarının karşılanmaması, uygun modelin seçilmemesi veya örneklem büyüklüğünün yeteri büyüklükte olmaması gibi sebeplerden kaynaklanmaktadır (Hambleton ve Jones, 1993).

MTK'da maddelerin iki kategorili ve çok kategorili puanlanmasına göre 2 model vardır. İki kategorili modeller arasında en çok kullanılan tek boyutlu MTK modelleri bir parametrelili model (1PLM), iki parametrelili model (2PLM) ve üç parametrelili model (3PLM) olarak adlandırılır. MTK modelleri arasındaki temel ayırım maddeleri tanımlamada kullanılan parametre sayısıdır. 1PLM en temel model olup bir maddeye doğru cevap verme olasılığının bireyin yeteneğine ( $\theta$ ) ve madde

güçlük düzeyine (b) bağlı olduğu durumdur. 2PLM'nin 1PLM'den farkı, güçlük parametresine ek olarak madde ayırt edicilik (a) indeksi de analize dahil edilmektedir. 3PLM'de ise doğru yanıt verme olasılığına güçlük, ayırt edicilik ve şans parametrelerinin (c) etki etmektedir. Rasch modeli ise 1PLM'nin özelleşmiş hali olup, ayırt edicilik parametreleri sabittir. 1PLM, 2PLM ve 3PLM'nin her birinde, çeşitli yetenek düzeylerinde maddeyi doğru yanıtlama olasılığının tespiti için matematiksel eşitlikler kullanılır.

MTK temelde, tek boyutluluk ve yerel bağımsızlık olmak üzere iki varsayıma dayanmaktadır. Tek boyutluluk, tek bir örtük özelliğin ölçüldüğü varsayılan bir durumdur. Bu sebeple MTK kapsamlı bir analiz yapmak için ölçme aracının baskın bir faktörü ölçtüğünün gösterilmesi gerekmektedir (Hambleton ve Swaminogton, 1985; Embretson ve Reise, 2000). Yapılan analizler sonucunda tek boyutluluk varsayımı sağlanamadığı zaman çok boyutlu MTK modellerine başvurulabilir. Yerel bağımsızlık ise, bireyin bir maddeye doğru cevap verme olasılığının testte yer alan diğer maddelere vereceği cevaplardan etkilenmeyeceğini ifade eder.

### **Araştırmanın Önemi**

Childs, Dahlstrom, Kemp ve Panter (2000), MMPI-2 Depresyon Ölçeğinin MTK kapsamında incelemiş, depresyon alt testinin ölçekte her ne kadara yarı bir boyut olarak gösterilse de parametrelerin çok boyutluluğa eğiliminin daha fazla olduğu, kadın ve erkeklerde farklı yetenek düzeyleri gösterdiğini bulmuşlardır. Thomas ve Locke (2010), MMPI-2-RF ölçeğinin somatik şikayetler alt boyutunun psikometrik özelliklerini incelemiş, 2PLM'in 1PLM ve 3PLM modele göre daha iyi uyum sağladığını bulmuşlardır. Tatar (2016), beş faktör ve on yedi alt boyutu olan Beş Faktör Kişilik Ölçeğine yönelik kısa formunun geliştirilmeyi amaçlanmış, MTK kapsamında yapılan analizlerle kısa ve uzun formların beş faktörünün birbirine çok yakın değerler gösterdiklerini hesaplamışlardır. Demir, Uluğ, Batur ve Mercan (2002), alkol bağımlılarında MMPI-2 bağımlılık alt ölçeklerinin psikometrik özelliklerini KTK kapsamında incelemiş, belirlenen ölçeklerin geçerli ve güvenilir oldukları, kendi içlerinde birbirleriyle tutarlı ve aynı özelliği yoklayan maddelerden oluştukları sonucunu bulmuşlardır.

İlgili alan yazında kişilik ölçeklerinin psikometrik özelliklerine ilişkin çok sayıda çalışma mevcuttur. Fakat Türkçeye uyarlanmış MMPI testi ile ilgili herhangi bir çalışmaya rastlanmamıştır. Bu bağlamda MMPI testinin madde parametrelerinin ve model uyumlarının belirlenmesinin sonraki çalışmalara da bir temel oluşturacağı düşünülmektedir.

## **YÖNTEM**

### **Araştırmanın modeli**

Bu çalışmada MMPI ölçeği kapsamında toplanan 730 kişiden elde edilmiş veri seti üzerinden alt testlere ait 1PLM, 2PLM ve 3PLM modele dayalı olarak model-veri uyumları, madde

parametrelerinin kestirimi ve test bilgi fonksiyonları incelenmiştir. Bu amaç doğrultusunda mevcut çalışma, MMPI alt testlerinin test edilerek model uyumlarını belirlemeyi ve karşılaştırmayı amaçladığı için, temel bir araştırma niteliğindedir.

### **Veri Toplama Araçları**

MMPI, “doğru”, “yanlış” ve “bilmiyorum” şeklinde cevaplandırılan ve toplamda 566 maddeden oluşan bir ölçektir. Ölçeğin puanlaması 1-0 şeklindedir. Bu ölçek yalan(L), sıklık ya da nadirlik(F) ve düzeltme(K) olmak üzere 3 farklı geçerlik, hipokondriazis(Hs), depresyon(D), histeri(Hy), psikopatik sapma(Pd), kadınlık-erkeklik(Mf), paranoya(Pa), psikasteni(Pt), şizofreni(Sc), hipomani(Ma) ve sosyal içedönüklük(Si) olmak üzere 10 klinik alt testten oluşmaktadır. L, F, K olarak adlandırılan bu üç geçerlik ölçeği, bireyin test alma tutumlarını değerlendirmeyi amaçlamaktadır. Örneğin, histeri alt testinde yüksek puan alan bireye histerik tanısı konmamaktadır. Bu gibi yanlış anlaşılmalara yol açmaması için alt ölçekler kısaltmalar veya sayılarla anılmaktadır.

Mevcut ölçek 13 ayrı alt testten oluşmakta, ve ortak bir toplam puan elde edilmemektedir. Her bir alt testten elde edilen puanlara ilişkin bir kişilik profili oluşturulmaktadır. Envanterin kadın ve erkekler için ayrı normları söz konusudur. Özellikle Kadınlık-erkeklik alt testi olan Mf’de aynı maddeler için farklı puanlaması göz önüne alındığında, mevcut çalışma için Mf alt testinin Mfe (erkek) ve Mfk (kadın) olmak üzere iki ayrı alt test olarak ele alınması uygun görülmüştür.

### **Çalışma Grubu**

Örneklem basit seçkisiz örnekleme metodu ile oluşturulmuştur. 18 yaş üstü sağlıklı bireylerden oluşan ve envanteri eksiksiz dolduran 730 kişi mevcut araştırmanın çalışma grubunu oluşturmaktadır.

### **Araştırmanın Etik İzinleri**

Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

### **Etik kurul izin bilgileri:**

Etik değerlendirmeyi yapan kurul adı = Hacettepe Üniversitesi Etik Komisyonu

Etik değerlendirme kararının tarihi= 26 Ekim 2021

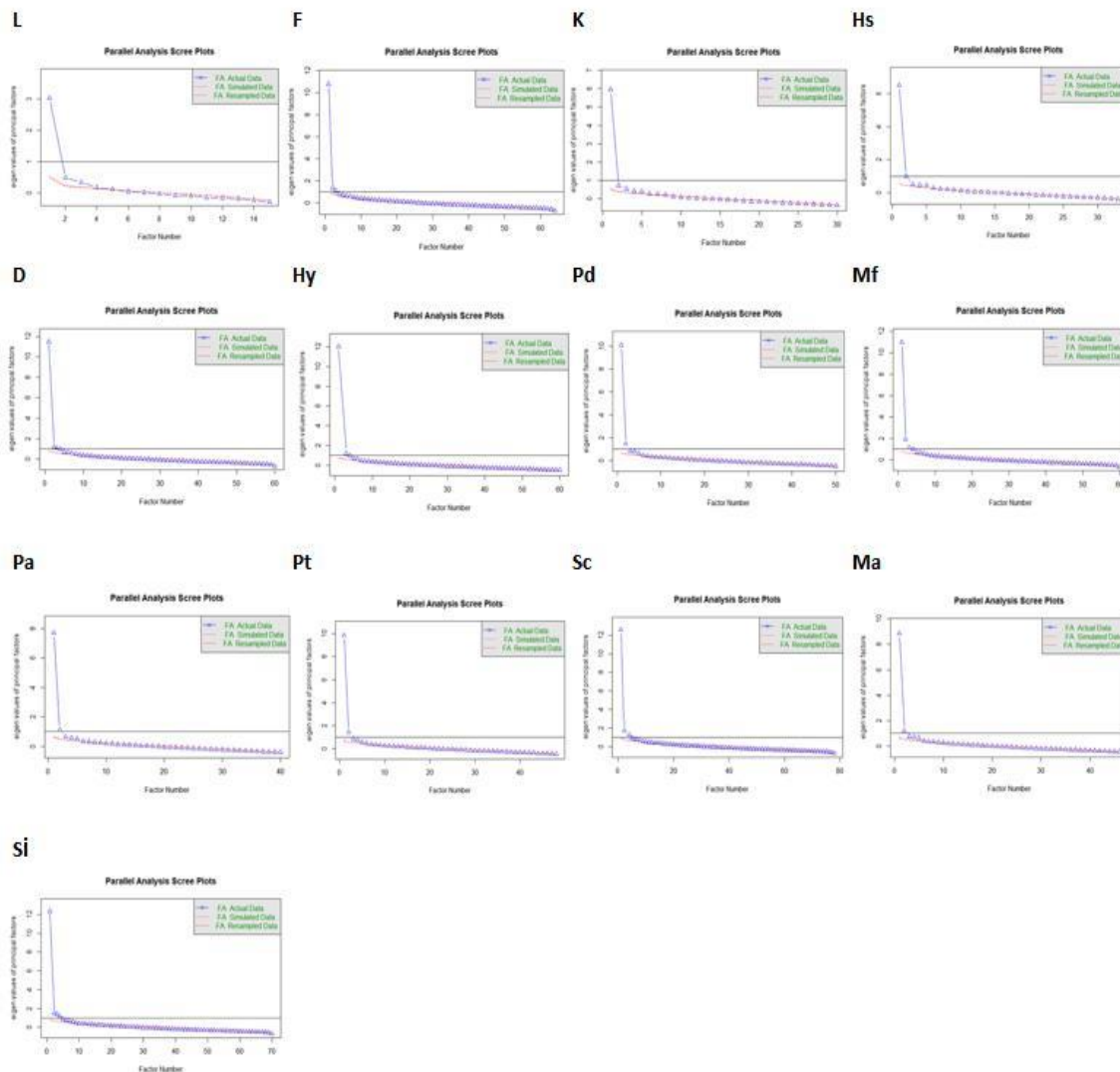
Etik değerlendirme belgesi sayı numarası= E-35853172-300-00001845927

## Verilerin Analizi

### Varsayımların test edilmesi

#### Tek Boyutluluk

Mevcut çalışmada tek boyutluluğun test edilmesi için özdeğerler ve paralel analiz yöntemleri kullanılmıştır. Paralel Analiz, mevcut veri setine ilişkin özdeğerler ile bu veriye paralel olarak üretilmiş verinin özdeğerlerinin karşılaştırılmasına dayalı bir yöntemdir (Horn, 1965). Paralel olarak üretilen veriye ait özdeğer ile gerçek veri setinde kestirilen özdeğerler karşılaştırılır ve paralel veriye ait özdeğerin gerçek veri setindeki özdeğerden büyük olduğu nokta anlamlı faktör sayısı olarak belirlenir (Ledesma ve Mora, 2007, Koçak, Çokluk ve Kayrı, 2020). R Studio programı kullanılarak yapılan paralel analizde “psych” paketinden yararlanılmıştır. MMPI alt testleri için yapılan paralel analiz şekilde verilmiştir.



Şekil 1. MMPI alt testlerine ait paralel analiz grafikleri

Şekil incelendiğinde L, K, Hs, D, Hy, Ma alt testlerinde tek boyutluluğun sağlandığı görülmektedir. Fakat F, Pd, Mf, Pa, Pt, Sc ve Si alt testlerinde paralel analiz sonuçları için tek boyutluluğun

sağlanamadığı gözlenmiş, bu sebeple her bir alt test için hesaplanan özdeğerler incelenmiştir. Özdeğer, başat faktörün varlığını ortaya çıkarmada etkili bir yöntemdir (Hambleton & Swaminathan, 1985). Veri setindeki başat boyut olduğuna, ilk faktörün açıkladığı varyansın %20 ve üzerinde olduğu durumda karar verilebilir (Lee, 1995). Hesaplanan özdeğer oranlarının 5 ve üzerinde olması tek boyutluluk varsayımını olumlu yönde desteklemektedir (Köse, 2015). MMPI'a ait alt testler için hesaplanan özdeğer oranları aşağıdaki tabloda verilmiştir.

Tablo 1. *MMPI alt testlerine ait özdeğer oranları*

Alt test	$\lambda_1$	$\lambda_2$	$\lambda_1 / \lambda_2$
L	1.342	0.362	3.707
F	8.671	2.121	4.088
K	3.680	1.083	3.397
Hs	5.179	1.412	3.667
D	6.165	2.021	3.050
Hy	6.351	2.041	3.111
Pd	5.834	1.860	3.136
MF	3.645	1.156	3.153
Pa	5.974	1.638	3.647
Pt	8.916	1.300	6.858
Sc	12.624	2.509	5.031
Ma	4.561	1.519	3.002
Si	6.745	1.580	4.268

MMPI alt testleri için özdeğerler hesaplanmış ve ilk özdeğerin ( $\lambda_1$ ) ikinci özdeğere ( $\lambda_2$ ) oranlarının 3.002 ile 6.858 arasında değiştiği gözlemlenmiştir. Hambleton ve diğ. (1991) ilk özdeğerin ikinci özdeğere oranının büyük olması durumunda tek boyutluluk varsayımının karşılanacağını belirtmiştir (Şimşek, 2017). Hesaplanan özdeğer oranlarının 3'ün üzerinde olması tek boyutluluk varsayımını destekleyecek niceliktedir.

### **Yerel Bağımsızlık**

Yerel bağımsızlık, aynı yetenek düzeyindeki bireylerin farklı maddelere verdiği tepkilerin istatistiksel olarak bağımsız olması olarak tanımlanır (Çakıcı Eser ve Gelbal, 2015). Alan yazın incelendiğinde tek boyutluluk varsayımının karşılanmasının yerel bağımsızlık varsayımı için kanıt olarak sunulabileceği belirtilmektedir (Embretson ve Reise, 2000, s. 48; Hambleton vd., 1991, s.11). Bireyin maddelere verdiği yanıtlar üzerinde tek bir faktör etkili olacağından yerel bağımsızlık varsayımının karşılandığı sonucuna varılabilir (Şimşek, 2017).

Güvenirlik, ölçme araçlarında bulunması gereken bir özelliktir. Güvenirlik katsayısının yüksek olması, yapılan ölçmenin hatalardan o derece arınık olduğu anlamına gelir. Mevcut çalışmada MMPI alt testleri için "0-1" puanlanan veri setlerinde başvuru KR-20 güvenilirlik katsayıları hesaplanmıştır.

Tablo 2. *MMPI alt testlerine ait güvenilirlik katsayıları*

Alt Test	Madde Sayısı	Güvenirlilik
L	15	0.73
F	64	0.89
K	30	0.76
Hs	33	0.86
D	60	0.76
Hy	60	0.74
Pd	50	0.78
Mf	60	0.71
Pa	40	0.74
Pt	48	0.90
Sc	78	0.92
Ma	46	0.76
Si	70	0.81

Yukarıdaki tablo incelendiğinde MMPI alt testlerinin iç tutarlık katsayılarının .71 ile .92 arasında değiştiği gözlenmektedir. En düşük iç tutarlık katsayısının Erkeklik/Kadınlık alt testine ( $r = .71$ ), en yüksek katsayısının Şizofreni alt testine ( $r = .92$ ) ait olduğu hesaplanmıştır. Genel olarak tüm alt testlerin kabul edilebilir sınır olan .70 üzerinde olması ölçekten elde edilen puanların güvenirliliğini doğrulamaktadır.

## BULGULAR

### Birinci Araştırma Sorusuna İlişkin Bulgular

“1PLM, 2PLM ve 3PLM modellerinden hangisi araştırma verisine daha fazla uyum sağlamaktadır?” sorusu için her bir alt testin modellere uyum düzeyleri test edilmiştir. Her alt test için yapılan model uyum analizleri için R Studio programının “ltm” paketinden yararlanılmıştır.

Tablo 3. Tüm alt testler için 1PLM, 2PLM ve 3PLM model veri uyumlarının ikili karşılaştırılması

Alt Test	Model	-2LL	sd	p
L	1PLM-2PLM	146,29	14	< .001
	2PLM-3PLM	24,16	15	>.001
F	1PLM-2PLM	1213,39	63	< .001
	2PLM-3PLM	95,13	64	>.001
K	1PLM-2PLM	632,12	29	< .001
	2PLM-3PLM	55,62	30	>.001
Hs	1PLM-2PLM	607,44	32	< .001
	2PLM-3PLM	29,34	33	>.001
D	1PLM-2PLM	2520,87	59	< .001
	2PLM-3PLM	37,04	60	>.001
Hy	1PLM-2PLM	2685,51	59	< .001
	2PLM-3PLM	74,02	60	>.001
Pd	1PLM-2PLM	2098,9	49	< .001
	2PLM-3PLM	40,33	50	>.001
Mfe	1PLM-2PLM	813,32	59	< .001
	2PLM-3PLM	65,72	60	>.001
Mfk	1PLM-2PLM	1183,36	59	< .001
	2PLM-3PLM	97,41	60	>.001
Pa	1PLM-2PLM	1818,35	39	< .001
	2PLM-3PLM	68,5	40	>.001
Pt	1PLM-2PLM	879,04	47	< .001
	2PLM-3PLM	21,74	48	>.001
Sc	1PLM-2PLM	1274,53	77	< .001
	2PLM-3PLM	86,35	78	>.001
Ma	1PLM-2PLM	1332,42	45	< .001
	2PLM-3PLM	72,14	46	>.001
Si	1PLM-2PLM	1708,70	69	< .001
	2PLM-3PLM	110,35	70	>.001

Her alt test için oluşturulan 1PLM, 2PLM ve 3PLM uyumları için 2 loglikelihood(-2LL) ve ki-kare indekslerine bakılmıştır. Belirlenen iki model arasındaki -2LL değerinin, serbestlik derecesi ve .001 anlamlılık düzeyinde anlamlı olup olmadığına bakılır. Tüm alt testler için 1PLM-2PLM karşılaştırılmasında aradaki fark anlamlı olduğu için çok parametrelili olan 2PLM'nin seçilmesinin daha uygun olacağı düşünülmüştür. 2PLM-3PLM karşılaştırılmasında aradaki fark anlamlı çıkmamış bu sebeple daha az parametrelili olan 2PLM seçilmesinin uygun olacağına karar verilmiştir.

Model uyumu değerlendirmelerinde de kullanılan madde uyumu istatistikleri için Bock'un (1972)  $\chi^2$  indeksinden yararlanılmıştır. 1PLM, 2PLM ve 3PLM altında her alt testte yer alan maddelerin uyum değerlerinde  $\alpha = 0,001$  manidarlık düzeyi esas alınmıştır. Hesaplamalar sonucu elde edilen  $\chi^2$  değerinin manidar olması, maddenin modele iyi uyum sergilemediği anlamına gelmektedir.



Tablo 4. Tüm alt testler için 1PLM, 2PLM ve 3PLM uyumlu madde sayısı

Alt Test	Uyumlu Madde Sayısı		
	1PLM	2PLM	3PLM
L	11	12	9
F	47	63	62
K	25	29	26
Hs	28	33	32
D	40	60	59
Hy	45	60	60
Pd	31	50	50
Mfe	60	60	57
Mfk	59	60	57
Pa	21	39	38
Pt	38	48	47
Sc	65	78	77
Ma	34	46	43
Si	55	70	69

Yukarıdaki tablo incelendiğinde madde uyumlarının parametre sayısına bağlı olarak değiştiği, her alt test için en fazla uyumlu madde sayısının 2PLM ait olduğu görülmektedir. L testinden 3, F, K ve Pa testlerinden 1'er maddenin madde-model uyumunun sağlanmadığı görülmüştür. Fakat MMPI'nin madde sayısı göz önüne alındığında uyumsuz madde sayısı %1.01'e denk gelmektedir.

### İkinci Araştırma Sorusuna İlişkin Bulgular

“MMPI alt testleri için kestirilen ayırt edicilik indeksleri ne düzeydedir?” sorusu için her alt madde için ayırt edicilik indeksleri ve betimsel istatistikleri hesaplanmıştır. Her alt test için yapılan analizler için R Studio programının “ltm” paketinden yararlanılmıştır.

Tablo 5. MMPI alt testlerine ait betimsel istatistikler

Alt Test	n	Min.	Maks.	Ort.	ss
L	15	0.14	1.80	0.74	.51
F	64	-0.51	2.71	1.03	.74
K	30	-0.57	1.91	0.56	.61
Hs	33	0.22	2.28	1.00	.58
D	60	-0.77	1.31	0.43	.82
Hy	60	-0.87	2.39	0.35	.84
Pd	50	-0.88	1.78	0.57	.73
Mfe	60	-1.36	1.70	0.20	.73
Mfk	60	-1.40	1.26	0.19	.67
Pa	40	-0.99	3.70	0.87	1.12
Pt	48	-0.49	2.22	1.13	.60
Sc	78	-0.48	2.24	1.08	.56
Ma	46	-0.62	1.73	0.57	.65
Si	70	-0.95	1.95	0.57	.58

Ayırt edicilik indeksi (a), bir tutuma ait pozitif ve negatif tutumların ayırt edilmesine yardımcı olur. Teorik olarak a parametresi değeri  $-\infty$ ,  $+\infty$  arasında değer almaktadır. Yüksek ayırt edicilik istenen bir durumdur. İndeksin pozitif olması, maddenin doğru cevaplanma oranının üst yetenek düzeyine ait grupta yüksek olduğunu, negatif olması ise maddenin doğru cevaplanma oranının alt yetenek düzeyine ait grupta yüksek olduğunu gösterir. Yukarıdaki tabloya bakıldığında L ve Hs alt testleri hariç diğer boyutların tümünde negatif bir a değeri görülmektedir. Özellikle Mfk alt testi için ortalamanın negatif değerde olması alt grupta yer alan bireylerin doğru cevaplama durumlarının daha fazla olduğu anlamına gelir. Baker (2001, s.21) madde ayırt edicilik indekslerinin 0,64 altındaki maddelerin “Düşük”, 0,65-1,34 aralığındaki maddelerin “Orta”, 1,35 ve üzerindeki maddelerin ise “Yüksek” ayırt ediciliğe sahip olduğunu belirtmiştir (Şimşek, 2017). Bu değerler doğrultusunda MMPI alt testlerinde yer alan maddelerin ayırt edicilik düzeyleri aşağıdaki tabloda verilmiştir.

Tablo 6. MMPI alt testleri ayırt edicilik düzeyleri

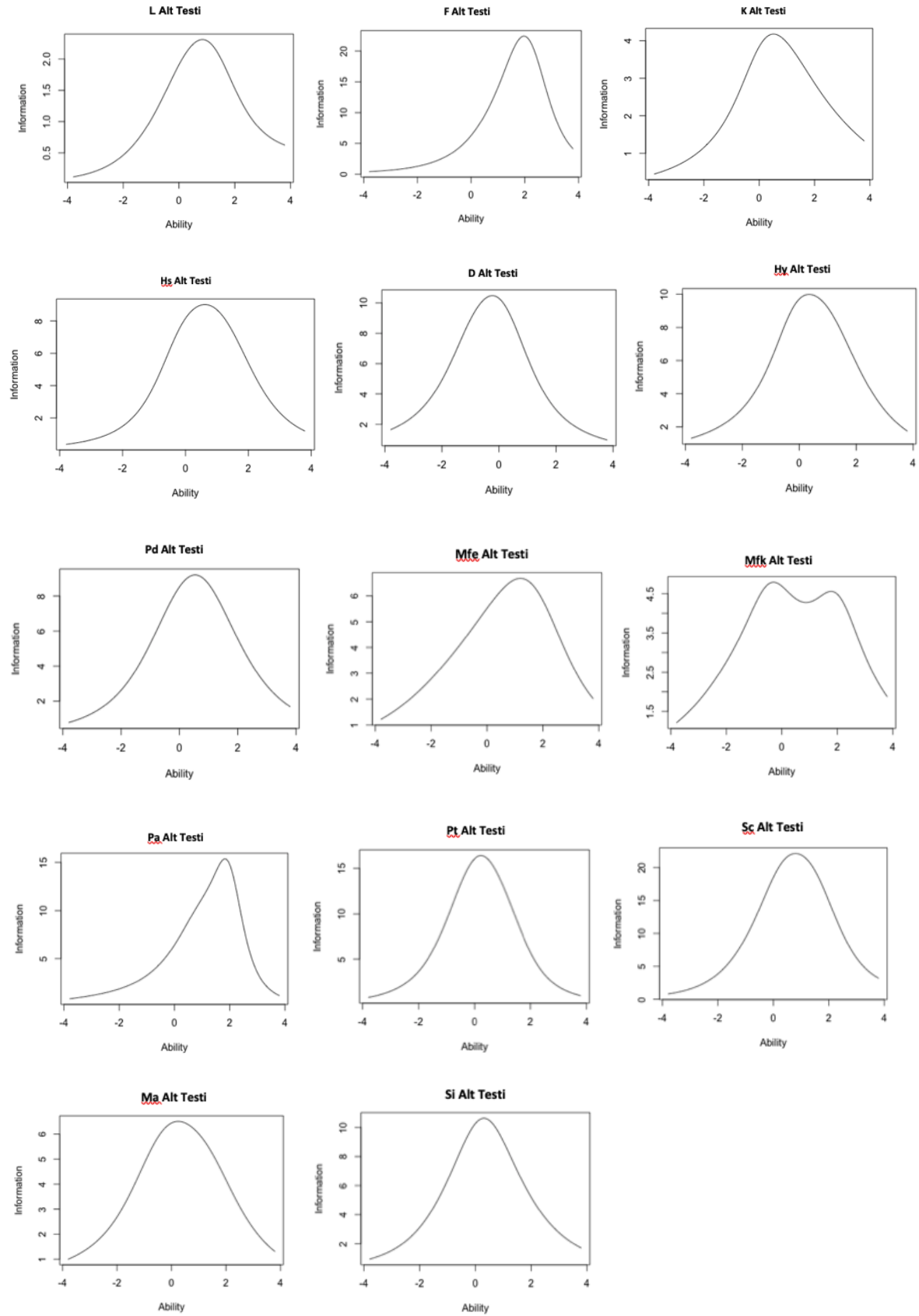
Alt Test	Soru Sayısı	Ayırt Edicilik Düzeyi		
		Düşük	Orta	Yüksek
L	15	10	3	2
F	64	23	18	23
K	30	15	13	2
Hs	33	10	11	12
D	60	50	8	2
Hy	60	36	18	7
Pd	50	24	18	8
Mfe	60	39	18	3
Mfk	60	45	13	2
Pa	40	17	9	14
Pt	48	11	9	20
Sc	78	20	31	27
Ma	46	24	15	7
Si	70	38	25	7

Tablo 6 incelendiğinde MMPI alt testlerinde yer alan maddelerin genel olarak ayırt edicilik düzeylerinin düşük olduğu görülmektedir. Pt ve Sc alt testlerinde diğer alt testlere nispeten madde ayırt ediciliklerinin daha yüksek olduğu, Mf alt testinde ise hem kadın hem erkek bireyler için ayırt ediciliğin madde bazında düşük seyrettiği görülmüştür. Madde ayırt ediciliği ile madde bilgi fonksiyonunu arasında pozitif yönlü bir ilişki bulunmaktadır (Hambleton ve diğ., 1991). Bu sebeple madde ayırt edicilik düzeyinin yüksek olduğunda madde bilgi düzeyi de yüksek olmaktadır. Bu sebeple madde bilgi düzeyi için madde ayırt ediciliği oldukça önemlidir.

### Üçüncü Araştırma Sorusuna İlişkin Bulgular

“Tanımlanan model altında test bilgi fonksiyonuna dayalı olarak MMPI alt testleri ne kadar bilgi vermektedir?” sorusu için her alt teste ait test bilgi fonksiyonları ve  $\theta$  aralıklarına göre test bilgi miktarı hesaplanmıştır. Test bilgi fonksiyonu, testin yapısını göstermesi açısından oldukça önemlidir. Temel olarak test bilgi fonksiyonu, madde bilgi fonksiyonlarının bileşkesidir ve testte yer alan tüm

maddelerin bilgi fonksiyonlarının toplamı test bilgi fonksiyonunu meydana getirmektedir(Şimşek, 2017). Test bilgi fonksiyonu, sürekli örtük özellik boyunca testin ayırt ediciliğinde meydana gelen değişimin görselleştirilmesini sağlamaktadır. MMPI alt testleri için oluşturulan test bilgi fonksiyonu grafikleri aşağıda verilmiştir.



## Şekil 2. MMPI test bilgi fonksiyonları

Tablo 7. 2PLM  $\theta$  aralıklarına göre test bilgi miktarı

Alt test	$\theta (-1,1)$	$\theta(-4,4)$ Toplam
L	3.5	10.8
F	13.96	66.32
K	7.11	20.78
Hs	14.98	33.62
D	18.67	44.17
Hy	17.85	45.76
Pd	15.97	39.98
Mfe	10.94	37.15
Mfk	9.04	32.36
Pa	13.5	45.72
Pt	28.59	55.12
Sc	34.97	84.61
Ma	12.17	33.12
Si	18.66	44.62

Test bilgi fonksiyonlarında orta düzeyde yeteneğe karşı gelen  $-1 < \theta < +1$  aralığında en fazla bilgi sağlayan testin Sc(34.97) olduğu, en az bilginin ise L(3.5) alt testinde olduğu hesaplanmıştır. Toplam bilgi miktarına bakıldığında en yüksek bilginin 84.61 ile Sc alt testine ait olduğu, onu 45.76 ile Hy alt testinin takip ettiği görülmektedir. MMPI madde seti genel anlamda düşük-orta seviyede ayırt ediciliğe sahip maddelerden oluşmaktadır. Bu durum test bilgi fonksiyonlarından da görülmektedir. Sc alt testinin bu denli yüksek olmasının sebebi, madde sayısının fazlalığına ve diğer alt testlere oranla yüksek ayırt ediciliği olan maddelere sahip olmasına bağlanmaktadır.

## SONUÇ VE TARTIŞMA

Mevcut araştırmada MMPI envanteri ile elde edilen verinin 1PLM, 2PLM ve 3PLM'den hangisine daha iyi uyum sergilediği, uyumlu madde sayısı, madde ayırt edicilik düzeyleri, test bilgi fonksiyonları incelenmiştir. Bu kapsam doğrultusunda ilk önce tek boyutlu MTK varsayımlarının karşılanıp karşılanmadığı incelenmiştir. Tek boyutlu MTK için tek boyutluluk ve yerel bağımsızlık varsayımları ele alınmıştır. Tek boyutluluk için paralel analiz yapılmış, her alt test için özdeğerler hesaplanmış ve ilk özdeğerin ( $\lambda_1$ ) ikinci özdeğere ( $\lambda_2$ ) oranları gözlemlenmiştir. Hesaplanan özdeğer oranlarının 3'ün üzerinde olması tek boyutluluk varsayımını desteklemiştir. Tek boyutluluk varsayımının karşılanmasının yerel bağımsızlık varsayımı için kanıt olarak sunulabileceği belirtilmektedir (Embretson ve Reise, 2000, s. 48; Hambleton vd., 1991, s.11). Mevcut çalışma için yerel bağımsızlık da sağlanmıştır denilebilir. MMPI alt testleri için güvenilirlik katsayıları da hesaplanmış mevcut veriler kapsamında güvenilirlik düzeylerinin kabul edilir seviye olan 0.70'in üzerinde olduğu görülmüştür.

Model uyumlarında her alt test için oluşturulan 1PLM, 2PLM ve 3PLM için 2 loglikelihood(-2LL) ve ki-kare indekslerine bakılmıştır. Belirlenen iki model arasındaki -2LL değerinin, serbestlik derecesi ve .001 anlamlılık düzeyinde anlamlı olup olmadığına bakılmıştır. 1PLM-2PLM, 2PLM-3PLM karşılaştırmalarında 2PLM'in veri setine daha uygun olduğuna karar verilmiştir.

Model uyumu değerlendirmelerinde de kullanılan madde uyumu istatistikleri için Bock'un (1972)  $\chi^2$  indeksinden yararlanılmıştır. 1PLM, 2PLM ve 3PLM altında her alt testte yer alan maddelerin uyum değerlerinde  $\alpha = 0,001$  manidarlık düzeyi esas alınmıştır. Hesaplamalar sonucu her alt test için en fazla uyumlu madde sayısının 2PLM ait olduğu görülmüştür. L testinden 3, F, K ve Pa testlerinden 1'er maddenin madde-model uyumunun sağlanmadığı bu değer MMPI'nin madde sayısı göz önüne alındığında uyumsuz madde sayısının %1.01'e denk geldiği hesaplanmıştır.

Ayırt edicilik indeksi (a), bir tutuma ait pozitif ve negatif tutumların ayırt edilmesine yardımcı olur. Mevcut araştırmada L ve Hs alt testleri hariç diğer boyutların tümünde negatif bir a değeri görülmektedir. Özellikle Mfk alt testi için ortalamanın negatif değerinde olması alt grupta yer alan bireylerin doğru cevaplama durumlarının daha fazla olduğu anlamına gelir. Baker (2001, s.21) madde ayırt edicilik indekslerinin düşük, orta, yüksek düzey aralığı göz önüne alınarak MMPI alt testlerinde yer alan maddelerin ayırt edicilik düzeylerine bakılmış, genel olarak ayırt edicilik düzeylerinin düşük olduğu görülmüştür. Pt ve Sc alt testlerinde diğer alt testlere nispeten madde ayırt ediciliklerinin daha yüksek olduğu, Mf alt testinde ise hem kadın hem erkek bireyler için ayırt ediciliğin madde bazında düşük seyrettiği görülmüştür.

Test bilgi fonksiyonlarında orta düzeyde yeteneğe karşı gelen  $-1 < \theta < +1$  aralığında en fazla bilgi sağlayan testin Sc (34.97) olduğu, en az bilginin ise L (3.5) alt testinde olduğu hesaplanmıştır. Toplam bilgi miktarına bakıldığında en yüksek bilginin 84.61 ile Sc alt testine, en düşük bilginin 10.8 ile L testine ait olduğu görülmektedir. MMPI madde seti genel anlamda düşük-orta seviyede ayırt ediciliğe sahip maddelerden oluşmaktadır. Bu durum test bilgi fonksiyonlarından da görülmektedir. Sc alt testinin bu denli yüksek olmasının sebebi, madde sayısının fazlalığına ve diğer alt testlere oranla yüksek ayırt ediciliği olan maddelere sahip olmasından kaynaklanmaktadır. L testinde bilgi miktarının düşük olmasının sebebi ise düşük madde sayısı, uyumsuz madde miktarının toplam madde oranına bakılarak yüksek olmasına bağlanmaktadır. Bilgi miktarının azalması standart hatayı da artıracığından L testi en yüksek hata ile en az bilgi veren alt test olarak karşımıza çıkmaktadır.

MMPI testinin doğrusal bir puanlama örüntüsünün olmayışı MTK için uygun olmayan bir durumdur. Özellikle kadın-erkek profillerinin farklı puanlanmasından kaynaklanan farklılık, Mf alt testinin farklı boyutlarda değerlendirilmesine, bu da örneklem sayısını o alt test için düşürerek bireylerdeki yetenek kestirim hatasının artmasına sebep olmuştur. Elde edilen sonuçlara göre MMPI alt testlerinde soru sayısının artması hem yeteneklerin doğru kestirilmesine, hem de daha fazla bilgi sağlanmasına yardımcı olmuştur. Mevcut çalışmanın sınırlılığı olan 18 yaş üstü sağlıklı bireylerden farklı olarak tanı konulmuş bireylerin cevap örüntülerinin incelenmesi gelecek çalışmalar için önerilmektedir.



<http://kefad.ahievran.edu.tr>

## Ahi Evran University Journal of Kırşehir Education Faculty

ISSN: 2147 - 1037

### ENGLISH VERSION

#### INTRODUCTION

The Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway and McKinley 1940) is a scale that aims to objectively assess personality traits and an individual's personal and social adjustment. The Turkish adaptation of this scale, which is frequently used in the clinical field and also in fields such as human resources and education, was conducted by Savaşır (1981). The MMPI, which is widely used, has been the target of many criticisms such as the high correlation between its subscales; the fact that the subscales do not clearly represent the symptom cluster they are named for (e.g., schizophrenia or depression subscales); the unclear wording of some items, sexism, and difficulties encountered due to the lack of linear scoring (Butcher and Pope 1992).

The MMPI is a personality test used to assess, diagnose and diagnose mental health disorders or to evaluate the effectiveness of treatment after treatment. This test, which is frequently used in the clinical field, needs to be examined, analyzed and interpreted. Test theories aim to solve the problems encountered in the analysis of psychometric properties measured indirectly.

These theories are divided into two as Classical Test Theory and Item Response Theory. Classical Test Theory (CTT) assumes that the true score consists of the observed score and random error.

Classical Test Theory (CTT) assumes that the true score consists of the observed score and random error. IRT, on the other hand, is a theory that predicts ability according to item parameters and individual characteristics. The CTT, which was developed after observing that the CTT was not sufficient, is stronger than the CTT with its features such as the invariance of item parameters and the fact that ability estimation does not depend on the test. Conceptually, it can be said that IRT is a probability-based theory that describes an individual's performance as a result of the interaction of individual characteristics and items (Faulkner-Bond & Wells, 2016). In IRT, instead of the total score obtained from the test, the responses to each item are considered and these items can be scored as continuous or discontinuous, dichotomous or multicategorical. Within the scope of the theory, many

models have been defined to reveal the ability of the individual with each of the answers given to the items (Hambleton & Jones, 1993).

In order to ensure IRT model validity, the data should fit the model at an acceptable level. Model-data mismatch arises due to reasons such as not meeting the model-data mismatch assumptions, not selecting the appropriate model or not having a large enough sample size (Hambleton & Jones, 1993).

There are 2 models in IRT according to the bi-categorical and multi-categorical scoring of items. Among the two-category models, the most commonly used unidimensional ITS models are called one-parameter model (1PLM), two-parameter model (2PLM) and three-parameter model (3PLM). The main distinction between IRT models is the number of parameters used to describe the items. 1PLM is the most basic model in which the probability of answering an item correctly depends on the individual's ability and (⊕) the item difficulty level (b). The difference between 2PLM and 1PLM is that in addition to the difficulty parameter, the item discrimination index (a) is also included in the analysis. In 3PLM, difficulty, discrimination and chance parameters (c) affect the probability of giving the correct answer. The Rasch model is a specialized version of the 1PLM and the discrimination parameters are fixed. In each of the 1PLM, 2PLM and 3PLM, mathematical equations are used to determine the probability of answering the item correctly at various ability levels.

The IRT is basically based on two assumptions: unidimensionality and local independence. Unidimensionality is a situation in which a single latent trait is assumed to be measured. Therefore, it is necessary to show that the measurement tool measures a dominant factor in order to conduct a comprehensive analysis (Hambleton & Swaminogton, 1985; Embretson & Reise, 2000). When the unidimensionality assumption cannot be met as a result of the analyses, multidimensional IRT models can be used. Local independence states that the probability of an individual's correct response to an item will not be affected by their responses to other items in the test.

### **Purpose of the Research**

In this study, it is aimed to examine the 1PLM, 2PLM and 3PLM model-data fit levels of MMPI subtests comparatively, to examine item fit and discrimination indices under the appropriate model, and to calculate test information functions and total amount of information in subtests.

In the research conducted for this purpose, answers to the following questions were sought:

1. Which of the 1PLM, 2PLM and 3PLM models fits the research data better?
2. What are the discrimination indices estimated for the MMPI subtests?
3. How informative are the MMPI subtests based on the test information function under the defined model?

### **Importance of the Research**

Childs, Dahlstrom, Kemp, and Panter (2000) examined the psychometric properties of the MMPI-2 Depression Scale within the scope of IRT, and found that although the depression subtest is shown as a semi-dimension in the scale, the tendency of the parameters to multidimensionality is higher and it shows different levels of ability in men and women. Thomas and Locke (2010) examined the psychometric properties of the somatic complaints sub-dimension of the MMPI-2-RF scale and found that the 2PLM provided a better fit than the 1PLM and 3PLM models. Tatar (2016) aimed to develop a short form of the Five Factor Personality Scale, which has five factors and seventeen sub-dimensions, and calculated that the five factors of the short and long forms showed very close values to each other with the analyzes conducted within the scope of IRT. Demir, Uluğ, Batur, and Mercan (2002) examined the psychometric properties of MMPI-2 addiction subscales in alcohol addicts within the scope of the CTT and found that the scales were valid and reliable, consistent with each other and consisted of items that tested the same feature.

There are many studies on the psychometric properties of personality scales in the related literature. However, there is no study on the MMPI test adapted into Turkish. In this context, it is thought that determining the item parameters and model fits of the MMPI test will provide a basis for future studies.

### **Method**

#### **Model of the Research**

In this study, model-data fits, estimation of item parameters and test information functions based on 1PLM, 2PLM and 3PLM models of the subtests were examined over the data set obtained from 730 individuals collected within the scope of the MMPI scale. In line with this purpose, the present study is a basic research as it aims to determine and compare the model fits by testing the MMPI subtests.

#### **Data Collection Tools**

The MMPI is a scale consisting of 566 items that are answered as "true", "false" and "don't know". The scale is scored 1-0. The scale consists of 10 clinical subtests including 3 different validity subscales: lie (L), frequency or rarity (F) and correction (K), hypochondriasis (Hs), depression (D), hysteria (Hy), psychopathic deviation (Pd), femininity-masculinity (Mf), paranoia (Pa), psychasthenia (Pt), schizophrenia (Sc), hypomania (Ma) and social introversion (Si). These three validity scales, labeled L, F, K, aim to assess an individual's test-taking attitudes. For example, an individual who scores high on the hysteria subtest is not diagnosed as hysterical. The subscales are referred to by abbreviations or numbers to avoid such misunderstandings.



The current scale consists of 13 separate subtests, and there is no common total score. A personality profile is created based on the scores obtained from each subtest. The inventory has separate norms for men and women. Especially considering the different scoring for the same items in the Mf subtest of femininity-masculinity, it was deemed appropriate to consider the Mf subtest as two separate subtests, Mfe (male) and Mfk (female), for the present study.

### **Study Group**

The sample was formed by simple random sampling method. 730 healthy individuals over 18 years of age who completed the inventory completely constituted the study group of the present study.

### **Data Analysis**

#### **Testing Assumptions**

##### *Unidimensionality*

In the present study, eigenvalues and parallel analysis methods were used to test unidimensionality. Parallel Analysis is a method based on comparing the eigenvalues of the current data set with the eigenvalues of the data generated in parallel (Horn, 1965). The eigenvalues of the parallel data are compared with the eigenvalues estimated in the real data set and the point where the eigenvalue of the parallel data is greater than the eigenvalue in the real data set is determined as the number of significant factors (Ledesma & Mora, 2007, Koçak, Çokluk, & Kayri, 2020). The "psych" package was used in the parallel analysis using the R Studio program. The parallel analysis for the MMPI subtests is given in the figure.

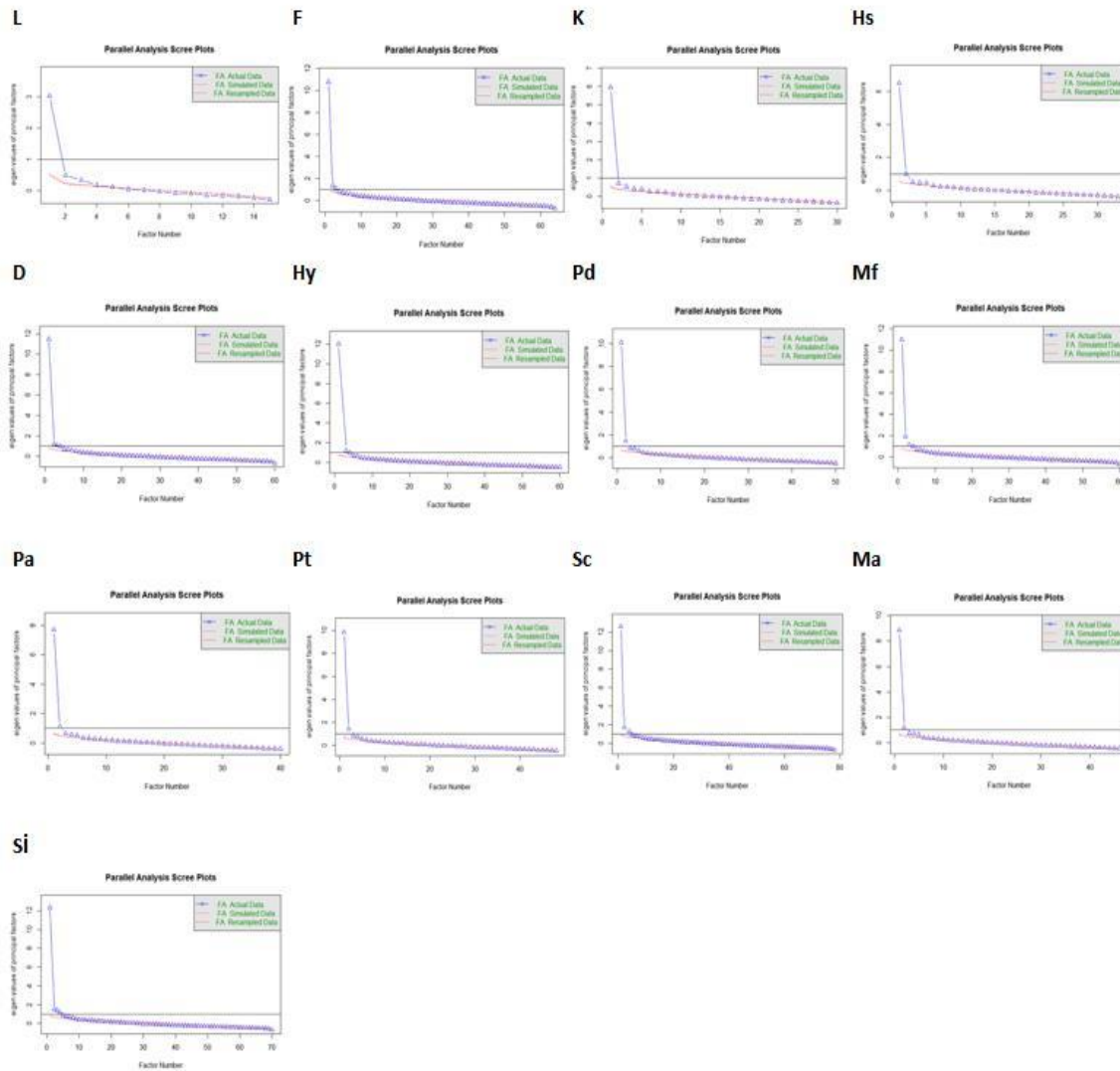


Figure 1. Parallel analysis graphs of MMPI subtests

When the figure is analyzed, it is seen that unidimensionality is achieved in L, K, Hs, D, Hy, Ma subtests. However, it was observed that unidimensionality was not achieved for parallel analysis results for the F, Pd, Mf, Pa, Pt, Sc and Si subtests, so the eigenvalues calculated for each subtest were examined. Eigenvalue is an effective method to reveal the presence of a dominant factor (Hambleton & Swaminathan, 1985). It can be decided that there is a dominant dimension in the data set when the variance explained by the first factor is 20% or more (Lee, 1995). The calculated eigenvalue ratios of 5 and above positively support the unidimensionality assumption (Köse, 2015). The eigenvalue ratios calculated for the subtests of the MMPI are given in the table below.

Table 1. Eigenvalue ratios of MMPI subtests

Sub-test	$\lambda_1$	$\lambda_2$	$\lambda_1 / \lambda_2$
L	1.342	0.362	3.707
F	8.671	2.121	4.088
K	3.680	1.083	3.397
Hs	5.179	1.412	3.667
D	6.165	2.021	3.050
Hy	6.351	2.041	3.111
Pd	5.834	1.860	3.136
MF	3.645	1.156	3.153
Pa	5.974	1.638	3.647
Pt	8.916	1.300	6.858
Sc	12.624	2.509	5.031
Ma	4.561	1.519	3.002
Si	6.745	1.580	4.268

Eigenvalues were calculated for the MMPI subtests and it was observed that the ratio of the first eigenvalue ( $\lambda_1$ ) to the second eigenvalue ( $\lambda_2$ ) ranged between 3.002 and 6.858. Hambleton et al. (1991) stated that if the ratio of the first eigenvalue to the second eigenvalue is larger, the unidimensionality assumption is met (Şimşek, 2017). The calculated eigenvalue ratios above 3 are quantitative enough to support the unidimensionality assumption.

#### *Local Independence*

Local independence is defined as the statistical independence of the responses of individuals at the same ability level to different items (Çakıcı Eser & Gelbal, 2015). When the literature is examined, it is stated that meeting the unidimensionality assumption can be presented as evidence for the local independence assumption (Embretson & Reise, 2000, p. 48; Hambleton et al., 1991, p.11). Since a single factor will be effective on the individual's responses to the items, it can be concluded that the local independence assumption is met (Şimşek, 2017).

Reliability is a feature that should be present in measurement tools. A high reliability coefficient means that the measurement is free from errors. In the current study, KR-20 reliability coefficients were calculated for the MMPI subtests for data sets scored "0-1".

Table 2. Reliability coefficients of MMPI subtests

Sub-Test	Number of Items	Güvenirlilik
L	15	0.73
F	64	0.89
K	30	0.76
Hs	33	0.86
D	60	0.76
Hy	60	0.74
Pd	50	0.78
Mf	60	0.71
Pa	40	0.74
Pt	48	0.90
Sc	78	0.92
Ma	46	0.76
Si	70	0.81

When the table above is analyzed, it is observed that the internal consistency coefficients of the MMPI subtests vary between .71 and .92. The lowest internal consistency coefficient belongs to the Masculinity/Femininity subtest ( $r = .71$ ) and the highest coefficient belongs to the Schizophrenia subtest ( $r = .92$ ). In general, all subtests were above the acceptable limit of .70, which confirms the reliability of the scores obtained from the scale.

### Findings

#### Findings related to the first research question

"Which of the 1PLM, 2PLM and 3PLM models fits the research data better?" For this question, the fit levels of each subtest to the models were tested. The "lrm" package of the R Studio program was used for model fit analyses for each subtest.

Table 3. Pairwise comparison of 1PLM, 2PLM and 3PLM model data fits for all subtests

Sub-Test	Model	-2LL	sd	p
L	1PLM-2PLM	146,29	14	< .001
	2PLM-3PLM	24,16	15	>.001
F	1PLM-2PLM	1213,39	63	< .001
	2PLM-3PLM	95,13	64	>.001
K	1PLM-2PLM	632,12	29	< .001
	2PLM-3PLM	55,62	30	>.001
Hs	1PLM-2PLM	607,44	32	< .001
	2PLM-3PLM	29,34	33	>.001
D	1PLM-2PLM	2520,87	59	< .001
	2PLM-3PLM	37,04	60	>.001
Hy	1PLM-2PLM	2685,51	59	< .001
	2PLM-3PLM	74,02	60	>.001
Pd	1PLM-2PLM	2098,9	49	< .001
	2PLM-3PLM	40,33	50	>.001
Mfe	1PLM-2PLM	813,32	59	< .001
	2PLM-3PLM	65,72	60	>.001
Mfk	1PLM-2PLM	1183,36	59	< .001

Pa	2PLM-3PLM	97,41	60	>.001
	1PLM-2PLM	1818,35	39	< .001
Pt	2PLM-3PLM	68,5	40	>.001
	1PLM-2PLM	879,04	47	< .001
Sc	2PLM-3PLM	21,74	48	>.001
	1PLM-2PLM	1274,53	77	< .001
Ma	2PLM-3PLM	86,35	78	>.001
	1PLM-2PLM	1332,42	45	< .001
Si	2PLM-3PLM	72,14	46	>.001
	1PLM-2PLM	1708,70	69	< .001
	2PLM-3PLM	110,35	70	>.001

For the 1PLM, 2PLM and 3PLM fits for each subtest, 2 loglikelihood (-2LL) and chi-square indices were examined. It is checked whether the -2LL value between the two models is significant at the degree of freedom and .001 significance level. Since the difference between 1PLM and 2PLM was significant for all sub-tests, it was thought that it would be more appropriate to choose the multi-parameter 2PLM. In the comparison of 2PLM-3PLM, the difference was not significant, so it was decided that it would be more appropriate to choose 2PLM with fewer parameters.

Bock's (1972)  $\chi^2$  index was used for item fit statistics, which are also used in model fit evaluations. The fit values of the items in each subtest under 1PLM, 2PLM and 3PLM were based on  $\alpha = 0.001$  significance level. If the  $\chi^2$  value obtained as a result of the calculations is significant, it means that the item does not fit the model well.

Table 4. Number of 1PLM, 2PLM and 3PLM compatible items for all subtests

Sub-Test	Number of Compatible Items		
	1PLM	2PLM	3PLM
L	11	12	9
F	47	63	62
K	25	29	26
Hs	28	33	32
D	40	60	59
Hy	45	60	60
Pd	31	50	50
Mfe	60	60	57
Mfk	59	60	57
Pa	21	39	38
Pt	38	48	47
Sc	65	78	77
Ma	34	46	43
Si	55	70	69

When the table above is examined, it is seen that the item fit varies depending on the number of parameters, and the highest number of compatible items for each subtest belongs to 2PLM. It was observed that 3 items from the L test and 1 item each from the F, K and Pa tests did not provide item-model fit. However, considering the number of items in the MMPI, the number of incompatible items corresponds to 1.01%.

### Findings related to the second research question

"What are the estimated discrimination indices for the MMPI subtests?" For this question, discrimination indices and descriptive statistics were calculated for each sub-item. The "lrm" package of the R Studio program was used for the analyses for each subtest.

Table 5. *Descriptives statistics for MMPI subtests*

Sub-Test	n	Min.	Maks.	Ort.	ss
L	15	0.14	1.80	0.74	.51
F	64	-0.51	2.71	1.03	.74
K	30	-0.57	1.91	0.56	.61
Hs	33	0.22	2.28	1.00	.58
D	60	-0.77	1.31	0.43	.82
Hy	60	-0.87	2.39	0.35	.84
Pd	50	-0.88	1.78	0.57	.73
Mfe	60	-1.36	1.70	0.20	.73
Mfk	60	-1.40	1.26	0.19	.67
Pa	40	-0.99	3.70	0.87	1.12
Pt	48	-0.49	2.22	1.13	.60
Sc	78	-0.48	2.24	1.08	.56
Ma	46	-0.62	1.73	0.57	.65
Si	70	-0.95	1.95	0.57	.58

The discrimination index ( $a$ ) helps to distinguish between positive and negative attitudes of an attitude. Theoretically, the parameter  $a$  takes values between  $-\infty$  and  $+\infty$ . High discrimination is desirable. A positive index indicates that the correct answer rate of the item is high in the group belonging to the higher ability level, while a negative index indicates that the correct answer rate of the item is high in the group belonging to the lower ability level. The table above shows a negative  $a$  value in all dimensions except for the L and Hs subtests. Especially for the Mfk subtest, the fact that the mean is negative means that the individuals in the subgroup are more likely to answer correctly. Baker (2001, p.21) stated that items with item discrimination indices below 0.64 have "Low" discrimination, items in the range of 0.65-1.34 have "Medium" discrimination, and items above 1.35 have "High" discrimination (Şimşek, 2017). In line with these values, the discrimination levels of the items in the MMPI subtests are given in the table below.

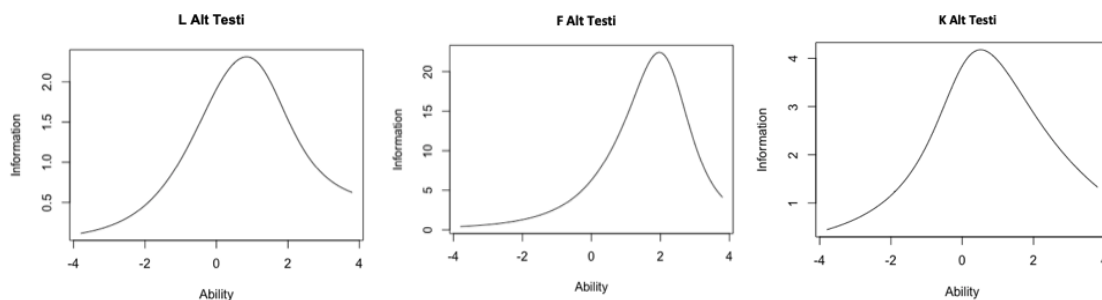
Table 6. MMPI subtests discrimination levels

Sub-Test	Number of Questions	Distinctiveness Level		
		Low	Medium	High
L	15	10	3	2
F	64	23	18	23
K	30	15	13	2
Hs	33	10	11	12
D	60	50	8	2
Hy	60	36	18	7
Pd	50	24	18	8
Mfe	60	39	18	3
Mfk	60	45	13	2
Pa	40	17	9	14
Pt	48	11	9	20
Sc	78	20	31	27
Ma	46	24	15	7
Si	70	38	25	7

When Table 6 is analyzed, it is seen that the discrimination levels of the items in the MMPI subtests are generally low. In the Pt and Sc subtests, item discrimination levels were higher compared to the other subtests, while in the Mf subtest, discrimination levels were low for both male and female individuals. There is a positive relationship between item discrimination and item information function (Hambleton et al., 1991). Therefore, when the item discrimination level is high, the item knowledge level is also high. Therefore, item discrimination is very important for item knowledge level.

### Findings related to the third research question

How informative are the MMPI subtests based on the test information function under the defined model? test information functions for each subtest and the amount of test information according to  $\theta$  intervals were calculated. The test information function is very important in terms of showing the structure of the test. Basically, the test information function is the combination of item information functions and the sum of the information functions of all items in the test constitutes the test information function (Şimşek, 2017). The test information function enables the visualization of the change in the discrimination of the test along the continuous latent trait. Test information function graphs created for MMPI subtests are given below.



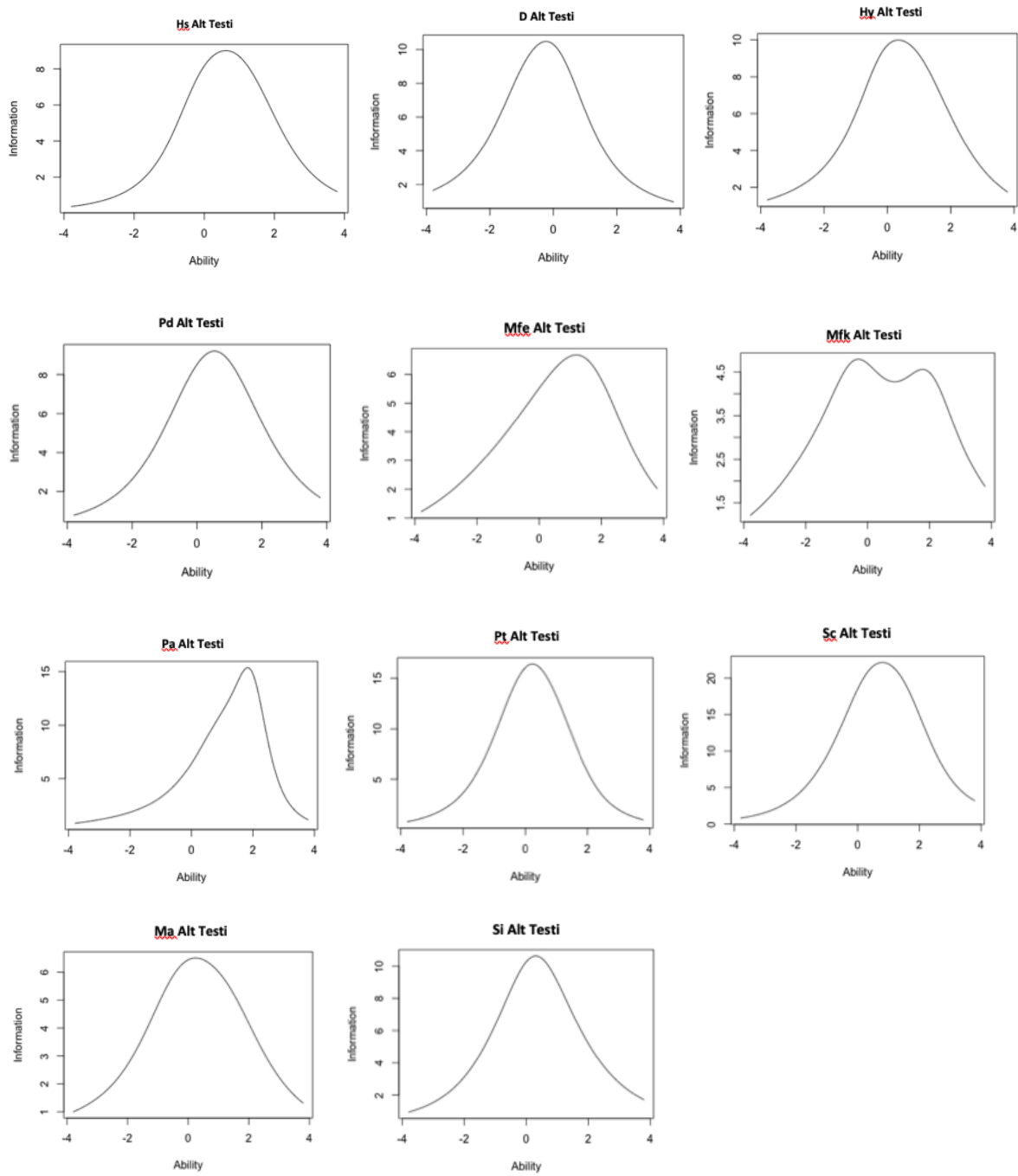


Figure 2. MMPI test information functions



Table 7. 2PLM test information amount by  $\theta$  intervals

Sub-Test	$\theta (-1,1)$	$\theta(-4,4)$ Total
L	3.5	10.8
F	13.96	66.32
K	7.11	20.78
Hs	14.98	33.62
D	18.67	44.17
Hy	17.85	45.76
Pd	15.97	39.98
Mfe	10.94	37.15
Mfk	9.04	32.36
Pa	13.5	45.72
Pt	28.59	55.12
Sc	34.97	84.61
Ma	12.17	33.12
Si	18.66	44.62

In the test information functions, it was calculated that Sc(34.97) was the test that provided the most information in the range of  $-1 < \theta < +1$ , which corresponds to moderate ability, while the least information was found in the L(3.5) subtest. When the total amount of knowledge is analyzed, it is seen that the highest amount of knowledge belongs to the Sc subtest with 84.61, followed by the Hy subtest with 45.76. In general, the MMPI item set consists of items with low-medium discrimination. This is also seen in the test information functions. The reason why the Sc subtest is so high is attributed to the high number of items and the fact that it has items with high discrimination compared to other subtests.

### Result and Discussion

In the current study, the data obtained with the MMPI inventory was examined to determine which of the 1PLM, 2PLM and 3PLM fits better, the number of compatible items, item discrimination levels, and test information functions. In line with this scope, it was first examined whether the one-dimensional IRT assumptions were met. For one-dimensional IRT, the assumptions of unidimensionality and local independence are discussed. Parallel analysis was performed for unidimensionality, eigenvalues were calculated for each subtest, and the ratios of the first eigenvalue ( $\lambda_1$ ) to the second eigenvalue ( $\lambda_2$ ) were observed. The fact that the calculated eigenvalue ratios were above 3 supported the unidimensionality assumption. It is stated that meeting the unidimensionality assumption can be presented as evidence for the local independence assumption (Embretson and Reise, 2000, p. 48; Hambleton et al., 1991, p. 11). It can be said that local independence was also achieved for the current study. Reliability coefficients for the MMPI subtests were also calculated and it was observed that the reliability levels were above 0.70, which is an acceptable level within the scope of the available data.

In model fit, 2 loglikelihood (-2LL) and chi-square indices for 1PLM, 2PLM and 3PLM created for each subtest were examined. It was checked whether the -2LL value between the two determined models was significant at the degree of freedom and .001 significance level. In 1PLM-2PLM, 2PLM-3PLM comparisons, it was decided that 2PLM was more suitable for the data set.

Bock's (1972)  $\chi^2$  index was used for item fit statistics, which were also used in model fit evaluations. The fit values of the items in each subtest under 1PLM, 2PLM and 3PLM were based on a significance level of  $\alpha = 0.001$ . As a result of the calculations, it was seen that the highest number of compatible items for each subtest belonged to 2PLM. Considering the number of items in the MMPI, 3 items from the L test and 1 item from the F, K and Pa tests were not compatible with the item-model, and it was calculated that the number of incompatible items corresponded to 1.01%.

The discrimination index ( $a$ ) helps distinguish positive and negative attitudes of an attitude. In the current study, a negative  $a$  value is observed in all dimensions except the L and Hs subtests. Especially for the Mfk subtest, the negative mean value means that individuals in the subgroup are more likely to answer correctly. Baker (2001, p.21) examined the discrimination levels of the items in the MMPI subtests, considering the low, medium and high level range of item discrimination indexes, and found that the discrimination levels were generally low. It was observed that item discrimination was higher in the Pt and Sc subtests compared to the other subtests, while in the Mf subtest, the discrimination was low on an item basis for both male and female individuals.

In test information functions, it was calculated that the test that provided the most information in the range of  $-1 < \theta < +1$ , corresponding to medium ability, was Sc (34.97), and the least information was in the L (3.5) subtest. When looking at the total amount of information, it is seen that the highest information belongs to the Sc subtest with 84.61, and the lowest information belongs to the L test with 10.8. The MMPI item set generally consists of items with low-medium level discrimination. This situation can also be seen from the test information functions. The reason why the Sc subtest is so high is due to the large number of items and the fact that it has items with high discrimination compared to other subtests. The reason for the low amount of information in the L test is due to the low number of items and the high amount of incompatible items considering the total item ratio. Since the decrease in the amount of information will increase the standard error, the L test appears as the subtest that gives the least information with the highest error.

The fact that the MMPI test does not have a linear scoring pattern is unsuitable for IRT. In particular, the difference resulting from different scoring of male and female profiles caused the Mf subtest to be evaluated in different dimensions, which reduced the number of samples for that subtest and increased the ability estimation error in individuals. According to the results obtained, increasing the number of questions in the MMPI subtests helped both accurately estimate abilities and provide more information. Examining the response patterns of diagnosed individuals, unlike healthy

individuals over the age of 18, which is a limitation of the current study, is recommended for future studies.

### Kaynakça

- Baker, F. (2001). The basics of item response theory (Second Ed). ERIC Clearinghouse on Assessment and Evaluation. <https://eric.ed.gov/?id=ED458219>
- Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, 16, 21–33. <https://doi.org/10.1111/j.1745-3992.1997.tb00605.x>
- Butcher, J. N. & Pope, K. S. (1992). The research base, psychometric properties, and clinical uses of the MMPI-2 and MMPI-A. *Canadian Psychol*, 33:61-78. <https://doi.org/10.1037/h0078693>
- Childs, R. A., Dahlstrom, W. G., Kemp, S. M., & Panter, A. T. (2000). Item response theory in personality assessment: A demonstration using the MMPI-2 Depression scale. *Assessment*, 7(1), 37–54. <https://doi.org/10.1177/107319110000700103>
- Demir, B., Uluğ, B., Batur, S. & Mercan, S. (2002). Alkol bağımlılarında MMPI-2 bağımlılık alt ölçeklerinin psikometrik özellikleri. *Türk Psikiyatri Dergisi*, 13:265-272. <https://www.psikolog.org.tr/tr/yayinlar/dergiler/1031828/tpd1300443320020000m000213.pdf>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Quality of Life Research. New Jersey: Lawrence Erlbaum Associates, Inc.
- Çakıcı Eser, D. & Gelbal, S. (2015). Farklı Boyutluluk Özelliklerindeki Basit ve Karmaşık Yapılı Testlerin Çok Boyutlu Madde Tepki Kuramına Dayalı Parametre Kestirimlerinin İncelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2).
- Faulkner-Bond, M., & Wells, C. S. (2016). A brief history of and introduction to Item Response Theory. In C. S. Wells ve M. Faulkner-Bond (Eds.), *Educational measurement: From foundations to future*. New York, NY: The Guilford Press.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.690.7561&rep=rep1&type=pdf>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. New York: Springer Science ve Business Media, LLC.
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of Item Response Theory*.
- Hathaway, S.R. & McKinley, J.C. (1940). A Multiphasic personality schedule (Minnesota): I. Construction of the Schedule. *J Psychol*, 10: 249-254.

- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Koçak, D., Çokluk, Ö., & Kayri, M. (2016). Faktör sayısının belirlenmesinde MAP testi, paralel analiz, K1 ve yamaç birikinti grafiği yöntemlerinin karşılaştırılması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 330-359.
- Köse, A. (2015). Aşamalı tepki modeli ve klasik test kuramı altında elde edilen test ve madde parametrelerinin karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 184-197.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 2–11.
- Lee, K.H. (1995). Application of the graded response model to the revised Tennessee self- concept scale: Unidimensionality, parameter invariance, and differential item functioning. Unpublished Doctoral Dissertation. University of Southern California.
- Savaşır, I. (1981). *Minnesota Çok Yönlü Kişilik Envanteri-Elkitabı (Türk Standardizasyonu)*. Sevinç Matbaası, Ankara.
- Tatar, A. (2016). Beş Faktör Kişilik Ölçeğinin Kısa Formunun geliştirilmesi. *Anadolu Psikiyatri Dergisi*, 17, 14-23. doi: 10.5455/apd.202977
- Şimşek, A. S. (2017). *Becerilere Güven Mesleki İlgi Envanterinin Uyarlanması Ve Bilgisayarlı Bireyselleştirilmiş Test Uygulamasının Geliştirilmesi*. Yayınlanmamış Doktora Tezi; Ankara Üniversitesi, Ankara.
- Thomas, M. L., & Locke, D. E. C. (2010). Psychometric properties of the MMPI-2-RF Somatic Complaints (RC1) Scale. *Psychological Assessment*, 22(3), 492–503. <https://doi.org/10.1037/a0019229>