

Ölçek Madde Analizleri: Klasik ve Yeni Yöntemler

Doç. Dr. Tuncay Duruöz

Celal Bayar Üniversitesi Tıp Fakültesi

Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı

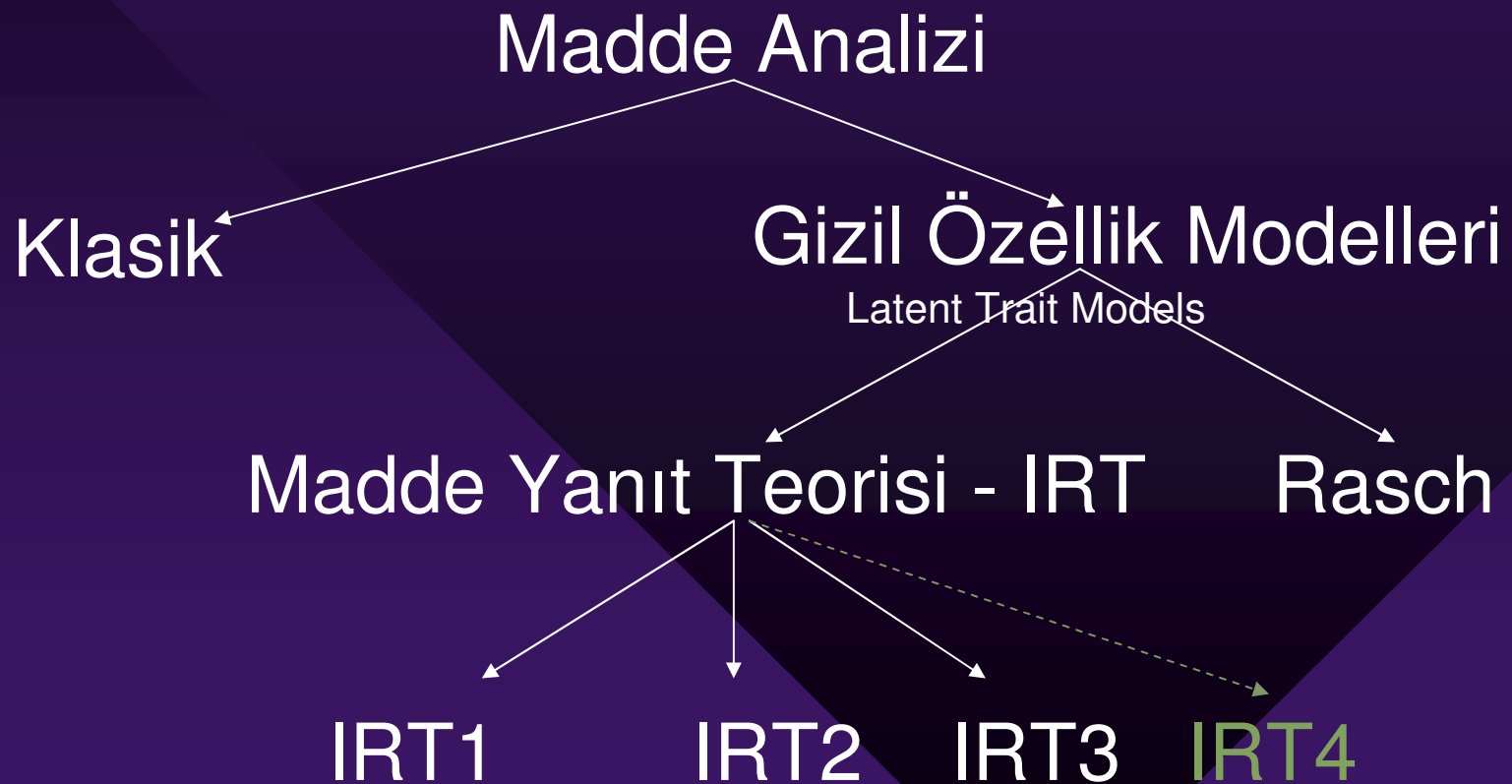
Manisa

Madde Analizi Niçin Gerekli ? (Item Analysis)

Madde analizi soruların kalitesini deęerlendirmeyi saęlayan bir yöntemdir. Soruların hedef populusyona özgülüęünü ve amaçlanan deęerlendirmeyi yapma performansını belirleyebilir.

Madde analizi aynı zamanda her bir sorunun performansını belirledięi için, bu sorunun dięer testlerde tekrar kullanılmasını da saęlamış olur.

Madde Analiz Yöntemleri Nelerdir ?



Madde (item) Klasik Analizi

- Zorluk
(Difficulty) madde düzeyinde
- Ayırt edicilik
(Discrimination) madde düzeyinde
- Güvenilirlik
(Reliability) test düzeyinde

Klasik Analiz Yöntemi Zorluk

Klasik analizde soruya yanlış cevap verenlerin oranı basit olarak bize ZORLUK derecesini gösterir. Çoklu sorularda yanlış cevap sayısının ortalaması zorluk derecesidir.

Klasik Analiz Yöntemi Ayırt Edicilik

Bir sorunun ortalama puanının, toplam test ortalama puanıyla korelasyonu (Pearson) ayırt ediciliđi gösterir.

+1 , - 1 arasında deđişen deđerı vardır.
Yüksek deđerler yüksek ayırt ediciliđi gösterir.

Klasik Analiz Yöntemi Güvenilirlik

İç tutarlılık (internal consistency) testteki soruların birbiriyle ilişkisini değerlendirir.

İç tutarlılık değeri 0-1 arasında değişir.
Büyük değer, yüksek güvenilirliği gösterir.

Gizil Özellik Modelleri (Latent Trait Models)

- Gizil Özellik Modelleri test performansına sebep olan özelliklerle ilgilenir. Bunlar madde (item) düzeyinde incelenir ve başka etkilerden bağımsız ölçüm olmasını sağlar.

Madde Yanıt Teorisi (Item Response Theory – IRT)

Madde Yanıt Teorisi (IRT), klasik test teorisi (CTT) diye de adlandırılan klasik ölçüm teorisinin (CMT) alternatifidir.

IRT çeşitli modellerin genel ismidir. En fazla “**lojistik madde yanıt modeli**” kullanılır.

IRT – Genel Model

$$P_g(\theta) = c_g + (1 - c_g) \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}}$$

a_g = ICC nin θ noktasında derecelendirilmesi
(ayırt edicilik - item discrimination)

b_g = a_g maksimum olduğunda yeti düzeyi
(zorluk - item difficulty)

c_g = g sorusuna kişilerin düşük düzeyde doğru cevap verme oranı

Klasik ölçüm teorisinin (CMT) temel fikri:
cevaplayanın doğru skoru + hata payı
olarak bilinir.

Hata payının, zamanlama, soru kurgusu ve soru maddesi gibi birçok sebebi olabilir.

IRT hata payını daha detaylı olarak ve özellikle soruların karakteristiklerine göre irdeleyebilir.

Klasik ölçüm teorisi öncelikli olarak toplam skorlarla ilgilenirken, IRT daha çok soru maddelerine ve onların alt özelliklerine odaklanmıştır.

Her maddenin ilgilenilen (ölçümü yapılan) değişkenle (attribute) ilişkisi değerlendirilir.

- IRT nin amacı, arařtırmacılara soruları cevaplayan kiřilerin etkilerinden bağımsız soru elde etmelerine yardımcı olmaktır.
- Bunun fiziksel ölçülerden örnekle açıklamak gerekirse, boy veya ağırlığı neyle ölçersek ölçelim deęişmemesidir.

- Son yıllarda IRT uygulamalarında yaygın olarak maddelerin (items) performans değerlendirmelerine 3 farklı boyutta yaklaşılmaktadır :

- Maddenin zorluk düzeyi (**difficulty**)
- Maddenin ayırt edicilik (**discrimination**) kapasitesi
- Yanlış pozitiflik eğilimi (**false positive**)

Bu madde parametrelerinin her biri ölçüm hata payıyla belirgin ilişki gösterir.

Eğer :

- a. Maddenin zorluğu amaçlanan parametreye uygun değilse
- b. “Doğru-Yanlış” arasındaki belirsizlik büyükse
- c. Madde, “olmayan” bir özelliği “var” diye gösteriyorsa

madde **HATALI** (performansı zayıf)

kabul edilir.

IRT, maddenin performansının bu üç boyutunu **nicel** olarak değerlendirir ve bu seçilen maddelerin belirlenen içerikte güçlü (**iyi performanslı**) olduğunu gösterir.

Maddenin Zorluk Derecesi



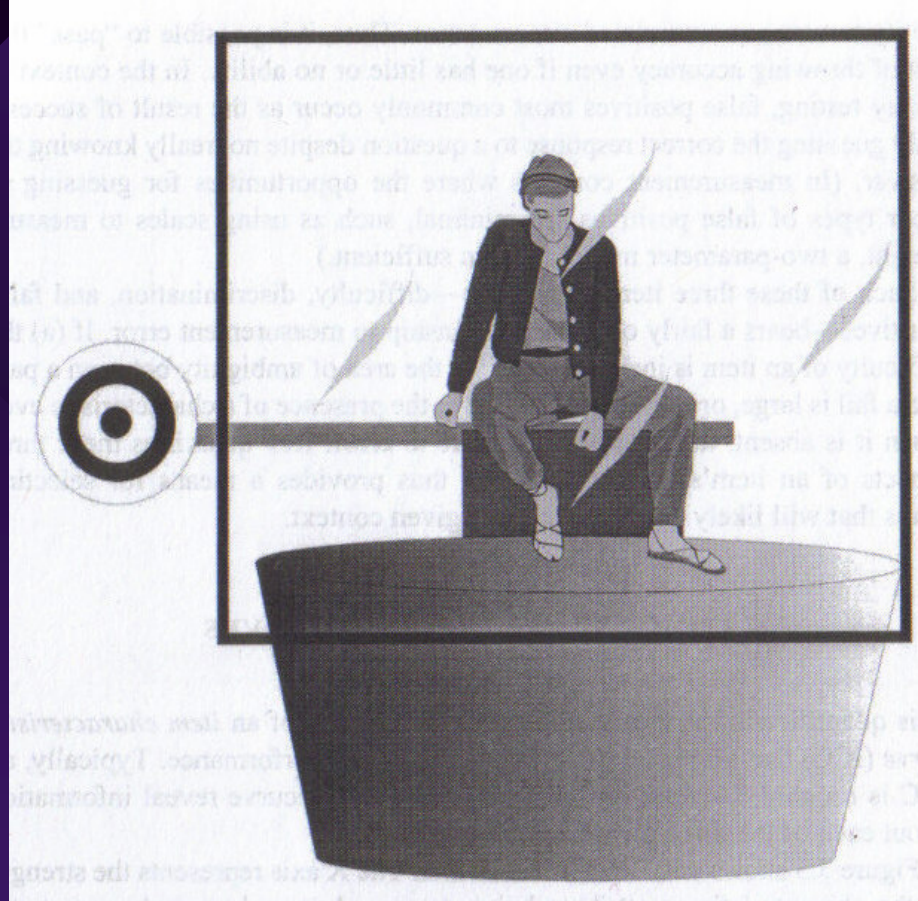
Maddenin Zorluk Derecesi

- Maddenin zorluk derecesini belirlemede amaç “ölçülmek istenen deęişkenin (attribute)” varlığının belirlenmesi için gerekli madde zorluk derecesidir.
- ÖR: Depresyonla ilgili soruda, bu sağlanırsa kişinin maddeye cevabı **depresyon düzeyine baęlı** olacaktır, kişinin **depresyon düzeyinin ortalamasının neresinde olduğundan baęımsız** olacaktır.

Maddenin Ayırt Edicilik Özelliđi

IRT modeli maddenin (sorunun) ne kadar keskinlikle “dođru” ve “yanlıř”ı sınıflandırabileceđini deđerlendirir.

Yanlış Pozitif



Bu platformu hedefe atışı değerlendiren bir madde (item) olarak düşünürsek

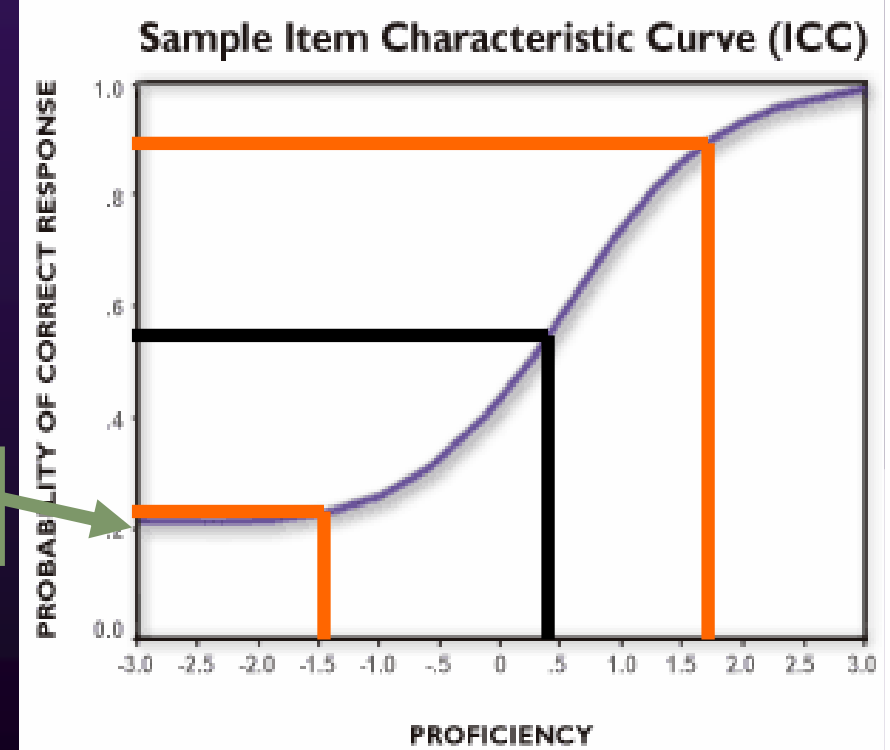
Yanlış Pozitiflik Eğilimi

- Maddeye verilen cevap hedeflenen parametre “olmadığı” halde “olduğunu” belirtiyorsa.
- Bu yeti testlerinde genellikle kişinin cevabı bilmediği halde doğru tahmin edebilmesiyle belirlenir.

Madde Karakteristik Eğrileri

Item Characteristic Curves (ICC)

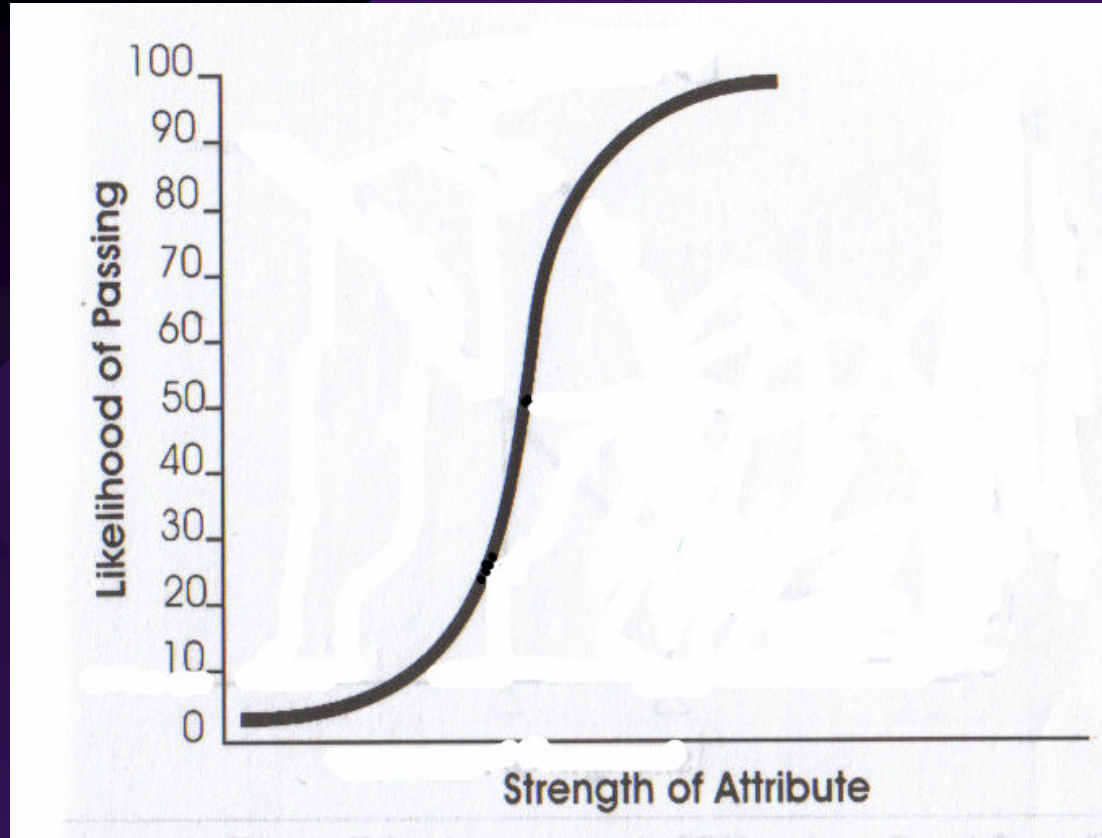
- ICC kişilerin soruya doğru cevap verme ihtimali yetisinin grafiğidir. Yüksek yeti ve yüksek derecede doğru cevap verilebilir.



c - kesişme

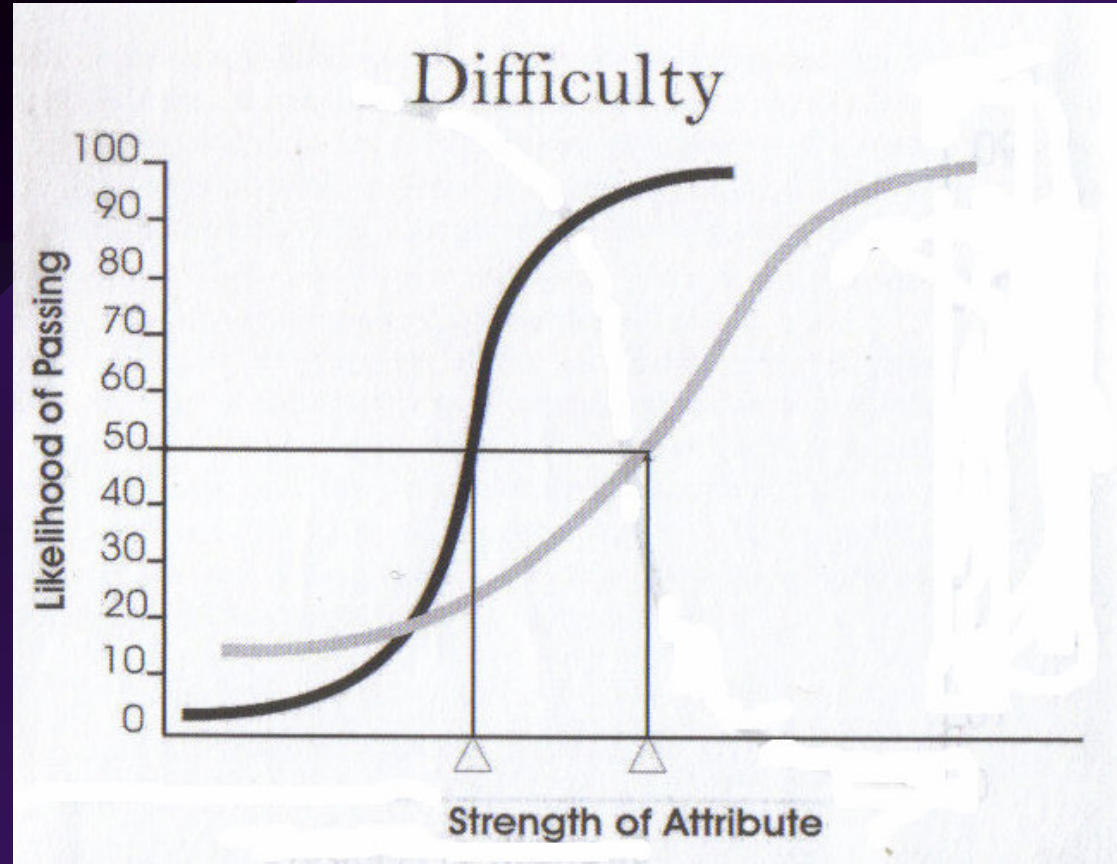
b - maksimum yeti (a)

a - dereceleme

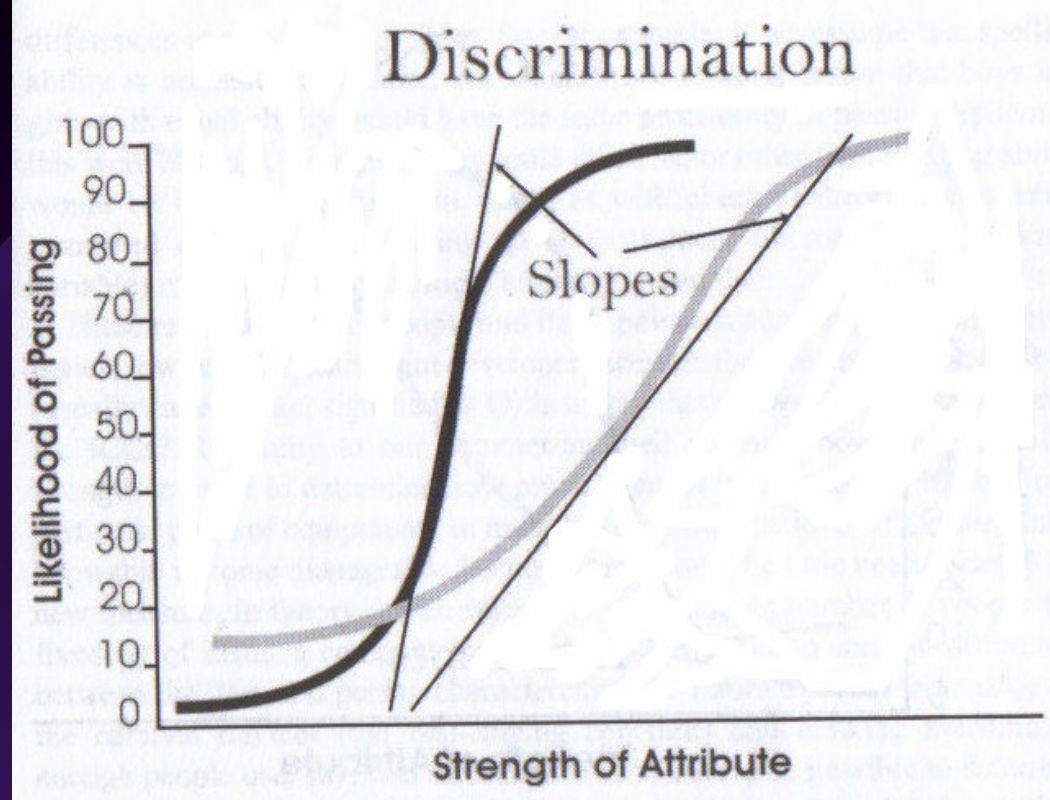


“X” aksı ölçülmek istenen parametrenin (bilgi düzeyi, kuvvet, depresyon vb) karakteristiklerinin gücünü temsil etmekte.

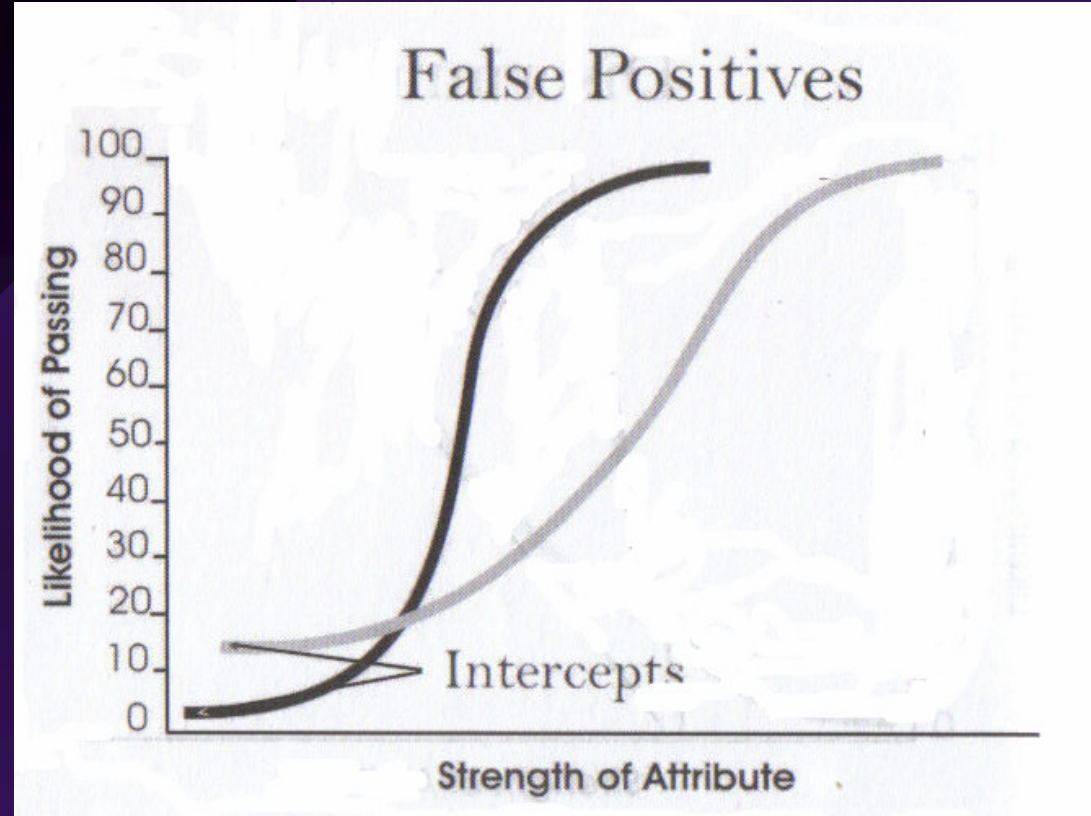
“Y” aksı gözlenen yanlış/doğru skor oranı esas alınarak amaçlanan parametrenin olabilirlik ihtimalini göstermektedir.



Her iki eğride ölçülen parametrenin %50 olma ihtimali farklı görülmektedir. Açık renkli çizginin temsil ettiği sorunun zorluğu koyu renkliden daha fazladır. Bu durumda zorluk subjektif bir değerlendirme değil gerçek değerlendirmedir.



Ayırt ediciliği göstermekte. Koyu eğrinin %50 noktasına paralel eğri açık renkli olandan daha dik. Koyu çizginin temsil ettiği soru açık renkliye göre olumsuz ve olumlu cevaplar arası geçiş daha kesin hatlıdır. Belirsiz alan koyuda daha azdır. Koyu çizginin sorusu diğerine göre ayırt edicilik özelliği daha fazladır.



Koyu çizginin Y aksını kesme noktası hemen hemen “sıfır” noktasındadır. Bu soruda kişinin yanlış pozitif cevap verme ihtimali 0 veya çok düşüktür.

Açık renkli çizginin sorusunda %15 kişi doğru cevabı bilmeden yanlış pozitif cevap verebilir.

SONUÇ:

- IRT maddenin hata payını daha detaylı olarak ve özellikle soruların karakteristiklerine göre objektif olarak irdeleyebilmektedir.
- Her kişide hatayla ilişkiyi irdeleyebilir.
- Her bir maddenin performansı bilindiği için maddelerden farklı ölçüt setleri yapılabilir.
- Ölçütlerin boyutlarını detaylı değerlendirebilir.
- *IRT parametrelerinin iyi değerlendirilmesi için fazla kişide denenmesi uygundur.*
- *IRT tek boyutluluğu değerlendirebilmektedir.*

Kaynaklar - IRT

Bilgisayar Programı

- BILOG (available at www.assess.com)
- Xcalibre available at www.assess.com

Okuma

- Lord (1980) Applications of Item Response Theory to Practical Testing Problems
- Baker, Frank (2001). *The Basics of Item Response Theory* - available at <http://ericcae.net/irt/baker/>

TEŞEKKÜRLER...