

# TEST VE MADDE YANLILIĐI

Nuri DOĐAN\*

Tuncay ÖĐRETMEN\*\*

## ÖZET

Test geliştirme, çok sayıda işlem ve basamaktan oluşur. Testlerin geçerliğini belirlemek bu işlem ve basamakların en önemlisidir. Test ve madde yanlılıđını belirlemek ise testin geçerliğini belirleme teknikleri arasında yer alır. Test ve madde yanlılıđını belirlemek amacıyla yapılan çalışmalar, son yıllarda yurt dışında yaygınlaşmasına rağmen, ülkemizde, hem kuramsal hem de uygulama alanında henüz çok sınırlı sayıdadır. Bu eksikliđin giderilebilmesi için çok sayıda araştırma yapılması gerekmektedir. Bu amaçla, test ve madde yanlılıđı üzerine alanyazın incelemesine dayanan bu çalışma gerçekleştirilmiştir. Yapılan incelemeler sonucunda; test ve madde yanlılıđını belirlemenin önemi, test ve madde yanlılıđı arasındaki ilişki, test ve madde yanlılıđını belirleme teknikleri, bu tekniklerin hangi durumlarda nasıl kullanılacağı ve elde edilen sonuçların nasıl yorumlanacağı hakkında bilgiler sunulmuş ve tartışılmıştır.

**Anahtar Kelimeler:** Test yanlılıđı, madde yanlılıđı, geçerlik

### Abstract

Test developing is consist of several processes and stages. Identifying validity of tests is the most important step of these processes and stages. Therefore, identifying item and test bias is involved in determining of tests' validity. Although in recent years, studies of identifying test and item biases have been widespreaded in abroad, theoretical and practical studies in this field are very limited in our country. In order to satisfy this deficiency, many studies should be done. For that purpose, this study is based on literature review about test and item biases. At the end of the study, the information are presented and discussed about the following topics: the importance of identifying test and item biases, the relationship between test bias and item bias, techniques for identifying test and item biases, in which situations these techniques were used and how the obtained results were interpreted.

**Key words:** Test bias, item bias, validity

---

\* Dr. Öğretim Görevlisi, H. Ü. Eğt. Fak. Eğitimde Ölçme ve Deđerlendirme Anabilim dalı

\*\* Araştırma Görevlisi, H. Ü. Eğt. Fak. Eğitimde Ölçme ve Deđerlendirme Anabilim dalı

## GİRİŞ

Genellikle, eğitim ve psikoloji alanında kullanılan testlerin yansız ve yordama geçerliğine sahip olduğu varsayılır. Örneđin, üniversite giriş sınavlarında cinsiyet veya bölge faktörlerine bađlı olarak öğrenci grupları arasında test puanı ortalamaları yönünden farklılık çıktığında, grupların gerçek başarılarının böyle olduğu varsayılır. Ancak bu durum kesin değildir. Testin yanlı ölçümler yapması, farklı grupların ortalamaları arasında fark çıkmasına neden olabilir. Farklı grupların ortalamaları arasındaki farkın gerçek kaynađını ortaya koyabilmek için testin veya test maddelerinin yanlılık çalışması yapılmalıdır.

Yanlılık kelimesi, önyargı, taraf tutma, kayırma, özel bir eğilim veya yönelim olarak tanımlanabilir (MEB 1996; Köklü, 2002). Bir araç yanlı ise, bu araçla yapılan ölçümlerin geçerliğinden kuşkulunmak gerekir. Eğitim ve psikolojide kullanılan ölçme araçlarının yanlılık çalışmaları, araçların geçerliği ile ilgili bilgi verdikleri için önemlidir. Başka bir deyişle, eğitim ve psikolojide kullanılan araçların yanlılık çalışmalarını yapmak, araçların geçerliğiyle ilgili kanıtlar elde etmenin önemli tekniklerinden biridir.

Test yanlılığıyla ilgili ilk çalışmalar, 1900'lerde, farklı etnik gruplardan anadili İngilizce olmayan çocukların zeka ölçülerini belirlemede karşılaşılan sorunlar üzerine yapılmıştır. Birinci dünya savaşı sırasında, Ordu-Alfa testleri ilk kez geniş ölçekte kullanılmış, grup zeka testleri geliştirilmiş ve sonuçları geniş bir alanda kullanılmıştır. Aynı dönemde, zeka testi puanlarının belirlenmiş yorumlarına karşı çıkılmış ve bu testlerin kültürel yanlılığı vurgulanmıştır. 1930'larda dilin ve kültürün test puanlarına etkisi üzerine çalışmalar yapılmıştır. 1940'larda ve 1950'lerin ilk yıllarında zeka ölçümlerinde ırksal farklılıkların yanlılıkla ilgisini gösteren tekrarlı çalışmalar bulunmaktadır. 1960'larda Havighurst ve Davis (Akt. Berk 1982) test performansı ve sosyal sınıf ilişkisini araştırmıştır. Testin yanlılığının kültürle yakın ilgisinin olduğu; kültürler arası çalışmalarla ortaya konmuştur. Eells ve diğerleri (Akt. Berk 1982) ise kültürden bağımsız test geliştirmeye çalışmışlardır. Bu nedenle yanlılık çalışmalarının sosyal bir değeri – anlamı – olduğu belirtilebilir. 1970'lerde getirilen yasal düzenlemeler, eğitim kurumlarında kullanılan testlerin eşitliğe uygun olması için geçerlik kanıtlarının daha fazla gözlenmesi, test kuramının geliştirilmesini ve test geliştirme süreçlerinin yenilenmesini sağlamıştır (Tittle, 1988).

Test puanlarının geçerliği eğitimde çok önemlidir. Çünkü, kişilerin test puanları, onlar hakkında alacağımız kararları etkiler. Test yanlılığı sorununda geçerliğin, özellikle yapı ve yordama geçerliğinin temel bir rolü vardır. Test yanlı ise, istenmeyen bir başka boyuta işaret eder. Test yanlılığı, ölçme sonuçlarına, bir grup öğrencinin lehine veya aleyhine işleyen ve istenmeyen bir başka boyutun karışması anlamına gelir. Test yanlılığı çalışması, bir dış ölçüt yokluğunda, aracın yapısal olarak

geçerliliğini ortaya koyma çalışmasıdır (Tittle, 1988). Eğer bir test, bir grup için başka bir gruptan daha avantajlı ise, yanlıdır ve eşitlik ilkesini bozar.

Yanlılık, iç ve dış yanlılık olarak iki alt başlık altında incelenebilir. Osterlid'e göre (1983, s.9), "dış yanlılık testin bağımsız değişkenlerle korelasyonel bir ilişki içinde olması anlamında test puanlarının uyumudur." Dış yanlılık, seçme kurallarında olduğu kadar testin eşitliğe uygun kullanımı konusunda sosyal anlaşmayı da kapsar. Dış yanlılık, "test yanlılığı" başlığı altında test maddelerinden çok bir testin yapı ve daha çok yordama geçerliliğine odaklanmıştır. İç yanlılık, test puanlarıyla madde puanları arasındaki ilişki anlamına gelen yapı geçerliliğini ilgi odağı yaparak test maddelerinin psikometrik özelliklerine öncelik verir. İç yanlılık "madde yanlılığı" olarak da bilinir.

### TEST YANLILIĞI

Tittle'a göre (1988) test yanlılığı, farklı gruplardan cevaplayıcıların test ortalamalarında ortaya çıkan sistematik bir fark (hata) olarak tanımlanabilir. Diğer bir deyişle, ilgilenilen bir özelliği kestirmede kullanılan testin, gruba bağlı olarak sistematik bir biçimde ortalamanın altında veya üstünde kestirimde bulunmasıdır. Bu gruplar genellikle etnik, cinsiyet, anadil farklılığı veya bir engele sahip olma özellikleriyle birbirinden ayrılır. Eğer bir test yanlı ise, testi alan farklı gruplardaki bireylere adil davranılmamasına neden olur. Testin yanlı olması, testlerin kullanımında, grupların ortalama performanslarında ve herhangi bir maddeyi doğru cevaplama üzerinde etkili olabilir (Tittle, 1988).

#### Test Yanlılığını Belirlemede Kullanılan İstatistiksel Teknikler

Test Yanlılığını belirlemek için bazı istatistiksel teknikler önerilmiştir. Bunlardan ikisi yaygındır.

##### *Ortalamalar Arası Fark*

Test yanlılığını belirlemenin en basit yolu ortalamalar arası farka bakmaktır. Örneğin, üniversite sınavı sözel testi sonuçlarına göre, A bölgesindeki öğrencilerin ortalama sözel puanları B bölgesindeki öğrencilerin ortalama sözel puanlarından 10 puan fazla olsun. Bu durumda testin A bölgesindeki öğrencilerin lehine, yanlı olduğunu düşünebiliriz. Ancak ortalama farklar tekniği, yanlılığı belirlemede kullanılacak yetersiz bir tekniktir. Çünkü, A bölgesindeki öğrencilerin ve B bölgesindeki öğrencilerin sözel puanları ortalamaları arasındaki 10 puan fark, yanlılığı gösterebileceği gibi, gruplar arasındaki gerçek farkı da gösterebilir. Gruplar arasında bulunabilecek ortalama farkların hepsi yanlılık olarak yorumlanamaz. Örneğin, Türkiye'deki kadın ve erkekleri temsil eden bir örneklem üzerinden erkek ve kadınların boylarını ölçmüş olsaydık, erkek ve kadın boy uzunluğu ortalamaları arasında erkekler lehine fark çıkabilirdi. Bu ölçümlere ait ortalamalar arasındaki farkın anlamı yanlılık olmayabilir. Bu fark, gerçek durumu yansıtan bir sonuç olabilir. Diğer yan-

dan, grup ortalamaları arasında fark bulunmaması da aracın yansız olduđu anlamına gelmeyebilir. Yine kadın ve erkeklerin sosyal konumlarını (örneğin, aylık gelirlerini) araştırma amacıyla geliştirilen bir ölçeğin uygulaması sonucunda, kadın ve erkeklerin statülerini eşit bulduğumuzu varsayalım. Toplumsal yapı düşünüldüğünde, kadın ve erkeklerin sosyal konumlarının ortalamaları arasında fark bulunmamışsa, araç veya seçilen örneklemin yanlı olması olasıdır. Verilen örneklerle de açıkça görüleceđi üzere, ortalamalar arası farklar tekniđi yanlılıđı belirlemede yetersizdir. Bu teknikle elde edilen sonuçlara bakarak karar vermek yanıltıcı olabilir (Tittle 1988).

### ***Regresyon Tekniđi***

Test yanlılıđını belirlemede daha fazla kabul gören bir tekniktir. Tekniđe göre, eđer farklı grupların test puanları aynı regresyon çizgisinde yer alıyorsa, grupların ortalamaları arasında fark olsa da, test yansızdır. Eđer farklı grupların test puanları farklı regresyon çizgisinde yer alırsa test yanlıdır. Çünkü, test farklı gruplar için başka şey ölçüyor demektir. Tekniđe göre, aynı test puanına sahip kişiler bir dış ölçüte göre eşit olmalıdır. Örneğin, üniversite giriş sınavı sayısal testi yansızsa, aynı sayısal puana sahip kız ve erkeklerin lise sayısal derslerindeki başarıları birbirine yakın olmalıdır. Aksine üniversite giriş sınavı sayısal testi kızlar aleyhine yanlıysa, kızlarla aynı sayısal puana sahip erkeklerin lise sayısal derslerindeki başarıları daha yüksek olacaktır (Tittle 1988).

Test puanlarının geçerliđinin düşmesinin veya test yanlılıđının yükselmesinin asıl nedeni, uygulamaların çoğunda bir test maddesine bir grubun diđerinden daha fazla doğru yanıt vermesi anlamında, gruplardan biri lehine adaletsizce işleyen yanlı maddelerdir. Yanlı maddelerin sayısı fazla ise testin yanlılıđı artacaktır. Testin yanlı çıkmasının nedeni birkaç madde olabilir. Eđer bu maddeler belirlenebilirse, testten çıkarılarak test yansız hale getirilebilir. Bu nedenle yanlılık çalışmalarının büyük çoğunluđu test maddelerinin yanlılıđı üzerine odaklanmıştır.

### **MADDE YANLILIĐI**

Madde yanlılıđı, bir test maddesine aynı yetenek düzeyinde fakat farklı alt gruplardan gelen bireylerin doğru cevap verme olasılıklarının aynı olmaması olarak tanımlanmaktadır. Yanlı bir madde farklılaşan madde fonksiyonu (Differential Item Function - DIF) verir. DIF eşit bilgiye sahip (aynı yetenek düzeyinde) farklı gruplardan cevaplayıcıların, bir maddedeki başarı olasılıklarının farklılaşması olarak ortaya çıkar. Vurgulanmalıdır ki; DIF kendi başına madde yanlılıđının bir kanıtı değildir. Bir maddeye gelen cevaplardaki farklılık, cevaplayıcı grupların bilgilerindeki (başarılarındaki) farklılık olduđu zaman beklenen bir durumdur. Genellikle farklı yetenek düzeyindeki grupların madde cevaplarındaki (performanslarındaki, başarılarındaki) farklılık madde yanlılıđı değildir; tam tersine maddenin geçerliđi ve

yansızlığıdır (Osterlind, 1983; Raju, 1988; Adams ve Rowe, 1988; Mellenberg, 1989; Hambleton, Swaminathan ve Rogers, 1991).

Tanımdan da anlaşılacağı üzere, madde yanlılık analizleri her bir test maddesinin aynı evrenden elde edilen farklı alt gruplara, benzer şekilde davranıp davranmadığı sorunu üzerine odaklanmaktadır. Diğer bir deyişle, bu analizler eğitimsel ve psikolojik yapı ölçümlerinin gruplara göre farklı olup olmadığını araştırmak amacıyla yapılmaktadır (Mellenberg, 1983). Eğer bir test çok heterojen bir evren için kullanılacaksa, yanlılık analizleri madde seçimi sürecinin en önemli parçası durumuna gelir. Çünkü, bu ölçme işleminde test uzmanı için en önemli kriter farklı alt gruplardan gelen ve testi alan bireyler için en adil ve en doğru sonuçları elde etmek olacaktır.

Madde yanlılığı analizlerinin iki temel amacı vardır. Bunlardan birincisi, test puanlarının değişik gruplardan gelen farklı değişkenlik kaynaklarından etkilenip etkilenmediğinin araştırılması; ikincisi ise bu ilişkisiz kaynakların bazı alt gruplar için adil olmayan bir avantaj sağlayıp sağlamadığıdır (Crocker ve Algina, 1986). Eğer test, ilgilenilen farklı alt gruplar için adil ise testteki maddelerin farklı alt gruplardaki aynı değişkenlik kaynağından etkilendiği söylenebilir. Buna ek olarak cevaplayıcılar, aynı test tarafından ölçülen psikolojik yapı açısından benzer düzeyde olacaklardır. "İlişkisiz değişkenlik kaynağının dağılımı bütün alt gruplar için benzer olacaktır" (Crocker ve Algina, 1986, pp.376-377).

Madde yanlılığını belirleme işlemlerinin çoğu, aynı evrenden çekilen farklı örneklem için, bir test maddesinin, dış bir kriteri referans almaksızın ölçülen özellik yönünden nasıl işlediğini araştırmayı içerir. Bu bağlamda kontrol edilen gruplar arasında madde performansı bakımından fark olduğu zaman test maddesinin yanlı olduğu söylenir. Yordama yanlılığı gösteren testlerin iç yanlılık (madde yanlılığı) olasılığı arttığından, seçme testleri gibi çoğunlukla yordama geçerliğine bağımlı testler için madde yanlılığını belirlemek önemlidir.

Testlerdeki yanlı maddeleri araştırmak için önerilen yöntemler Klasik Test Kuramı veya Örtük Özellikler Kuramına dayanır. Çoğu teknik bu iki kuramdan birinin uzantısıdır. Bu nedenle de teknik yeterlikleri yanında kuramlardan gelen avantaj ve dezavantajları da bünyelerinde barındırır. Klasik Test Kuramına dayanan teknikler genellikle, madde indekslerinin (madde güçlük indeksi ve madde ayırıcılık gücü indeksi) alt gruplara göre karşılaştırılması üzerine kuruludur. Klasik test teorisi kapsamında başvurulan diğer bir teknik ise chi-square yaklaşımıdır. Bu yöntem cevaplayıcı farklarının, ilgilenilen gruplarda bir maddeye doğru cevap verme oranlarının toplam test puanları kategorilerine göre hesaplanmasını içermektedir (Adams ve Rowe, 1988). Bu teknikler, klasik teorinin getirdiği bazı dezavantajlar göz önünde bulundurularak bir çok araştırmacı tarafından eleştirilmektedir (Shepard, Camilli ve Williams, 1984; Rodney ve Drasgow, 1990). Klasik yöntemlerle hesaplanan madde

parametrelerinin bir gruptan diđerine deđişkenlik göstermesi yani sabit parametreler olmaması ve örnekleme bağımlı olması nedeniyle yüksek oranlarda hatalı ölçme sonuçları doğuracağı; bu nedenle, yapılan madde yanlılıđı analizlerinin hatalı yorumlara yol açacağı ileri sürülmektedir. Diđer taraftan, madde yanlılıđı analizi ile bağlantılı olarak Örtük Özellikler Kuramı modellerinin kullanılmasının Klasik Test Kuramı ve bu kurama dayanan tekniklere göre üstün tarafları olduđu iddia edilmektedir. Bu modellerle elde edilen madde karakteristik eğrisi (Item Characteristic Curve) ve madde parametrelerinin sabit olduđu, yani bir gruptan diđerine deđişmeyen deđerler aldıđı, grupların kıyaslanmasında arařtırmacıya maddenin yanlı olup olmadıđı konusunda daha güvenilir kanıt sağladıđı ileri sürülmektedir. Bu anlamda madde yanlılıđı, bir tek test maddesine aynı yetenek düzeyinde fakat farklı gruplardan gelen iki bireyin dođru cevap verme olasılıđının aynı olmaması olarak tanımlanmaktadır (Adams ve Rowe, 1988; Mellenberg, 1989; Raju, 1990; Hambleton, Swaminathan ve Rogers, 1991). Bir maddeye ait parametrelerin belirli bir yetenek düzeyinde farklı örneklemler için aynı olması, yani aynı madde için farklı gruplardan elde edilen madde karakteristik eğrilerinin benzer olması gerekir (Devine ve Raju, 1982; Rodney ve Drasgow, 1990). Buna dayanarak, Örtük Özellikler kuramı modellerinin, madde yanlılıđı analizleri için gerekli çerçeveyi sağlayabileceđi görüşü yaygındır. Çünkü bu teoride, madde karakteristik eğrileri, kapsanan yetenek düzeyinde dođru bir tepkide bulunma olasılıđı ile madde özelliklerini ilişkilendirmektedir. Bu anlamda, madde yanlılıđı analizini Madde Tepki Kuramı çerçevesinde yapmak, ancak bir test maddesinin iki ayrı gruptan elde edilen madde karakteristik eğrilerinin karşılaştırılması ile olasıdır (örneğin, kız ve erkek cinsiyet gruplarına göre). Bu karşılařtırmalar çeřitli şekillerde yapılabilir. Madde parametreleri karşılaştırılarak, aynı parametrelere sahip madde karakteristik eğrilerinin de aynı olduđu kabul edilebilir veya madde karakteristik eğrileri arasında kalan alan madde yanlılıđı açısından delil olarak kabul edilebilir (Lord, 1980; Rudner ve diđerleri, 1980). Hesaplanan alan deđeri sıfır olduđunda, o maddenin alt gruplara göre yanlı davranmadıđına karar verilir (Rudner ve diđerleri, 1980; Raju, 1988).

Adams ve Rowe (1988), bu teknikleri iki sınıfta incelemektedir. Birincisi, Klasik Test Kuramı'na dayanan kolay, hızlı ama yetersiz (eksik) olan teknikler; ikincisi, karmařık, zor ama güçlü olan yeni teknikler.

Genel olarak, madde yanlılıđı analiz tekniklerini ařađıdaki kategorilerde sınıflamak mümkün görünmektedir.

1. Madde ayırcılık gücü, madde güçlüđu ve kısmi korelasyon teknikleri
2. Varyans analizi tekniđi
3. Korelasyonel teknikler (Güvenirliklerin veya  $P_{ort}$  Karşılaştırılması)
4. Faktör analizi teknikleri

5. Chi-square tekniğı
6. Foil tekniğı (chi-square istatistiğı kullanılıyor)
7. Madde güçlük dönüşümü (MGD) tekniğı
8. Örtük Özellikler Kuramı teknikleri
9. Deneysel kontrol teknikleri
10. Kapsam geçerliğı ve güvenilir olan testler oluşturma teknikleri (Hills, 1984; Adams ve Rowe, 1988).

### **Basit Teknikler**

Bu tekniklerin her biri, cevaplayıcı örneklemeden elde edilen yetenek dağılımına bağılı istatistiklerin ürünü olan geleneksel madde analizlerine dayanır. Klasik Test Kuramı'nda, madde ve test istatistiklerinin örnekleme bağılı olarak değışmesi söz konusudur. Bu nedenle, araştırılan gruplarda yetenek dağılımı değıştiğinde, yanlılık göstergelerinde kestirilemeyen (belirlenemeyen) etkiler oluştüğundan madde yanlılığını belirlemek zorlaşır (Crocker ve Algina, 1986; Hambleton, Swaminathan and Rogers, 1991). Bu tekniklerden çok kullanılanları aşağıda açıklanmıştır.

#### ***Madde Güçlük İndeksi Tekniğı***

Öncelikle her bir grup için madde güçlüğü (p) hesaplamak gerekir. Maddeler gruplar için aldığı p değerlerine göre sıralanır ve her bir gruptaki sıralama karşılaştırılır. Karşılaştırma sırasında sapan madde, yanlılık belirtisi gösterir. Bununla birlikte sıralama karşılaştırmasına dayanan teknikler yanlıcı olabilir. Örneğın, eğer gruplar farklı yetenek dağılımına sahipse, maddeler aynı ayırcılığa sahip olmadıkça aynı p değerleri almayacaktır.

#### ***Madde Ayırcılık Gücü İndeksi Tekniğı***

Her bir grup için madde ayırcılık güçleri karşılaştırılır. Bunlar genellikle madde ve test puanları arasındaki nokta çift serili veya çift serili korelasyonların karşılaştırılmasıdır. Yanlılığın göstergesi olarak, gruplara ait ayırcılık değerleri arasındaki farklar, her bir grup için puan dağılımlarındaki farklılara duyarlı olduğundan kullanışlı olabilir.

#### ***Faktör Analizi Tekniğı***

Bir testin faktör yapısı her bir grup için hesaplanır. Testin faktör sayısı gruplar için değışirse teste veya gruplara bağılı olarak farklı faktörde yükü olan maddelere yanlı gözüyle bakılabilir. Özellikle çift serili korelasyonlar matrisinden elde edilen faktör analizi sonuçlarına bakmak kuramsal olarak kabul görmektedir. Ancak, yanlı olduğunu bildiğimiz bir test için faktör analizi tekniğı sık sık farklı gruplar için aynı yapıyı vermektedir. Ayrıca, normal olarak eğitim ve psikolojideki yetenek testlerin-

de, ikili (1-0) data tipi için faktör analizinin uygun olmadığı belirtilmektedir (Adams ve Rowe, 1988 ).

### ***Varyans Analizi Tekniği***

Genel kabul gören ve sık uygulanan bir tekniktir. Aynı test verileri için, iki ya da daha fazla grup arasında madde x grup etkileşimi manidar çıkarsa ilgili maddelerin yanlı olduğu yorumu yapılır. Ancak, gerçekte maddeler aynı ortalamada olmayabilir. Bu nedenle, tüm grup performansındaki farkların veya test alan gruplar arasında yetenek farklılıklarının bir fonksiyonu olarak sık sık gözlenen madde x grup etkileşiminin önemliliği hatalı yorumlanabilir (Hills 1984; Crocker ve Algina, 1986; Tittle, 1988; Adams ve Rowe 1988; Hambleton, Swaminathan and Rogers, 1991).

### ***Karmaşık Teknikler***

Özellikle Örtük Özellikler kuramına benzeyen veya ona alternatif olabilecek çok sayıda teknik geliştirilmiştir. Ancak bunların çoğu henüz kabul görmemiş veya kuramsal yetersizlikleri giderilmemiştir. Bu tekniklerden aşağıda açıklananlar, çok kullanılan ve kabul gören teknikler olarak bilinmektedir.

### ***Madde Güçlük Dönüşümü (MGD) Tekniği***

Bu tekniğe Angoff Delta grafik tekniği de denir. Angoff ve Ford (1973) tarafından önerilen bir tekniktir. Madde güçlüklerinin bir grafikte gösterilmesini içeren yöntemdir. MGD tekniği, maddeler ve grup etkileşimine her bir maddenin katkısını belirlemeye çalışan varyans analizi yaklaşımının bir uzantısıdır.

Madde güçlük indeksi p (maddeyi doğru cevaplayanların gruba oranı) her bir madde için gruplarda ayrı ayrı hesaplanır. Bulunan p değerleri için standart normal dağılımda 1-p'ye karşılık gelen z değerleri elde edilir. Bu dönüştürmeden sonra z değerlerinin grafiği doğrusal olacaktır. Gruplar için bulunan z değerlerinden delta değerleri hesaplanır. Burada amaç, doğrusal dönüştürme ile negatif z değerlerinin elimine edilmesidir. Delta dönüştürmesi  $\Delta=4z+13$  eşitliği kullanılarak yapılmaktadır. Büyük olan bir  $\Delta$  değeri zor olan bir maddeye işaret eder. Ölçme hatasından dolayı, bu iki gruba ait dönüştürülmüş madde güçlükleri arasında 0,98 veya daha fazla bir korelasyon beklenir (Angoff,1975). İki Grup için hesaplanan her bir maddeye ait  $\Delta$  değerleri saçılım grafiğinde gösterilir. MGD yaklaşımında, bir maddenin delta değerlerine grafikte karşılık gelen nokta regresyon eğrisine manidar bir uzaklıkta ise bu maddenin yanlı olduğuna karar verilir. Böyle bir durumda madde-grup etkileşimi var demektir. Bu madde, bir gruptaki cevaplayıcılar için diğer gruptaki cevaplayıcılara göre daha zor bir madde olmaktadır. Angoff bu sapma farklarının, maddelerin gruplara göre farklı psikolojik anlamlar taşımasından kaynaklandığını ileri sürmektedir. Buradaki esas sorun, yanlılığı belirleme de kullanılacak uzaklığın büyüklüğünün görece olmasıdır (Devine ve Raju, 1982).



Çoğu durumda madde yanlılığını belirlemek için MGD yeterlidir. Basittir, hesaplaması kolaydır ve yanlılık miktarı için görsel bir sonuç verir. Madde yanlılığını belirleme tekniklerinin karşılaştırıldığı çalışmalar göstermiştir ki, MGD yaklaşımı geleneksel madde analizlerine dayanan eksik metotlara benzemekle birlikte, özellikle küçük örneklem ve kısa testler olduğunda çoğu teknik kadar iyi işlemektedir. Ne yazık ki, MGD kullanıldığı zaman, gruplar arasında yetenek farkının istatistiksel kontrolü mümkün olmadığı için grup puan dağılımlarındaki farklar maddelerin yanlılığını söylemede yanıltıcı olabilir. Seong ve Subkoviak (1987), MGD tekniğinin kuramsal sınırlamalarına dikkat çekmektedirler. Eğer iki grubun ortalama yetenekleri arasında büyük bir fark var ise, MGD tekniği gerçekte yanlılık olmasa bile yanlılığa işaret edebilir. Bu nedenle Angoff, z değerlerinin klasik madde ayıricılığını veren madde-test korelasyonuna bölünmesini ve  $\bar{z}$  değerlerinin elde edilmesini, bu yapıldıktan sonra yukarıda verilen işlemlere aynı sıra ile devam edilmesini önermektedir.

### ***Chi-Square Tekniği:***

Scheuneman (1979) ve Camilli (1979) tarafından geliştirilmiştir. Örtük Özellikler kuramı tekniklerine benzemektedir. Bu tekniğin avantajı, Örtük Özellikler kuramına göre uygulanmasının daha kolay olmasıdır. Bu tekniğe göre bir maddenin yansızlığı, aynı test puanı kategorisinde fakat farklı alt gruplarda yer alan cevaplayıcıların, maddeyi doğru cevaplama oranlarının aynı olması olarak tanımlanmaktadır. Bu teknik, örtük özellikler kuramının bir parçası olarak görülebilir. Çünkü, burada gözlenen test puanları örtük özellik yerine yanlı bir maddenin tanımlanmasında kullanılmakta ve gözlenen bu test puanları, örtük özelliğin hatalı bir ölçümü olarak göz önünde bulundurulmaktadır.

Chi-square tekniğinde, gözlenen puanlar ölçeği birkaç kategoriye ayrılır. Her bir kategori içinde, alt gruplar bir maddeyi doğru cevaplama oranları açısından karşılaştırılır. Eğer oranlar gruplar arasında farklılık gösteriyorsa bu durum madde yanlılığı için bir kanıttır. Grupların kategorilerde verdiği cevaplar oranından Chi-square değeri elde edilir. Elde edilen Chi-square değeri, J serbestlik derecesinde,  $100(1-\alpha)$  düzeyinde chi-square dağılımı ile karşılaştırarak manidarlık testi yapılmaktadır. Burada J toplam aralık (kategori) sayısıdır. Hesaplanan  $\chi^2$  nin büyüklüğü maddenin yanlılık indeksinin miktarı olarak değerlendirilebilir. Chi-square yaklaşımı kolay hesaplanmasına rağmen, yetenek düzeylerinin seçimindeki teknik kadar, gruplar için madde ayıricılıklarına duyarlı olmanın dezavantajına sahiptir.

Chi-square tekniği, kategori grupları oluşturmada kesme puanlarının seçimi yönünden eleştirilmektedir. Kesme puanlarının seçimi, elde edilecek  $\chi^2$  değerlerinin büyüklüğünü etkileyecektir. Diğer taraftan, her grup ve aralık için, doğru cevap sayısının en az 5 olması gerektiği vurgulanmaktadır (Ironson ve Craig, 1982; Crocker ve Algina, 1986, pp. 384-386).

### **Örtük Özellikler Kuramı Teknikleri**

Örtük Özellik kuramı modelleri altında, bir madde, eđer, farklı gruptaki aynı yetenek düzeyine sahip bireylerin maddeye doğru cevap verme olasılıkları eşitse yansızdır. Başka bir deyişle, grupların madde karakteristik eğrisindeki farklar önemsiz olmalıdır. Madde yanlılığını belirlemek için yapılan iş, farklı gruplar için modellere göre elde edilen madde karakteristik eğrilerini karşılaştırmaktır. Yetenekler eşit olduğu zaman iki grubun madde karakteristik eğrilerinde farklılık gözleniyorsa madde için yanlılık belirlenir. Bu, iki grubun başarı olasılıklarının farklılığının kanıtıdır. Diğer taraftan yanlılık, sabit (uniform) yanlılık ve deđişken (non-uniform) yanlılık olarak gözlemlenebilir. İki gruba ait madde karakteristik eğrileri temelde b parametrelerinde farklı olmak üzere aşıđı yukarı paralel ise bu çeşit yanlılığa sabit yanlılık denir. Sabit yanlılıkta maddenin doğru cevaplandırılma olasılığındaki farklılık, tüm yetenek düzeylerinde iki grup için de deđişmemektedir. Bu tür yanlılıkta, madde tüm yetenek düzeylerinde aynı grubun lehine işlemektedir. Eđer madde karakteristik eğrileri kesişiyorsa, bunun anlamı, maddenin doğru cevaplandırılma olasılığının alt yetenek düzeyinde bir grup için yüksek, üst yetenek düzeyinde ise diğer grup için yüksek olduğudur. Olasılıklardaki farklılık, yetenek düzeylerinde gruplara göre deđişkenlik gösterdiği için (sadece bir grubun lehine devamlılık göstermediği için) deđişken yanlılık olarak adlandırılmaktadır (Yu 2002; Brannick 2004).

Madde yanlılığını Örtük Özellikler kuramı ile belirlemenin avantajlarından belki de en önemlisi, test ve madde parametrelerinin örneklem deđişmelerinden etkilenmemesi, aynı kalmasıdır. Bunun anlamı; madde parametreleri, örneklemelerinden dağılımından bağımsızdır. Yani, herhangi bir Örtük Özellik modeli ile farklı örneklemelerden elde edilen herhangi bir maddeye ait madde parametreleri eşit olmalıdır. Model varsayımları çerçevesinde gruplar için madde parametreleri eşit çıkmıyorsa o maddenin yanlı olduğu söylenir. Çünkü her bir grup için farklı bir şey ölçülüyor olabilir.

Karşılaştırmalı çalışmalarda, madde yanlılığın belirlemek için üç parametrelilik model genellikle daha dikkat çekici sonuçlar vermektedir. Madde karakteristik eğrisi farklarına dayanan çalışmalar için çok geniş sayıda teknik vardır. Literatür incelendiğinde, sadece 3 parametrelilik lojistik model kullanarak madde yanlılığını belirleyen karşılaştırmalı çalışmalarda sekiz tekniğin tanımlandığı görülmektedir. Yine literatürdeki araştırma sonuçlarına bađlı olarak ağırlıklandırılmış kareler toplamının kullanılması önerilmektedir (Crocker ve Algina, 1986; Hambleton, Swaminathan ve Rogers, 1991; Camilli ve Shepard, 1994).

Madde yanlılığını belirleme çalışmaları Örtük Özellikler kuramı çerçevesinde aşıđıdaki aşamaları içermektedir (Mellenberg, 1989):

1. Madde karakteristik eğrilerinin elde edileceđi modele karar vermek,
2. Ölçeklemeyi, karşılaştırmaların yapılacağı gruplarda gerçekleştirmek,

3. Gruplar arasındaki yetenek düzeyleri farkını ortadan kaldırmak için tekrar ölçekleme yapmak,
4. a- Gruplar için hesaplanan madde parametrelerini karşılaştırmak. b-İki farklı gruptan elde edilen madde karakteristik eğrileri arasındaki alan indekslerini hesaplamak.

Örtük Özellikler kuramı, çok kabul görmesine rağmen bazı temel güçlükleri vardır. Örneğin, a ve c parametreleri değişmezlik özelliğini gereğince sağlayamamaktadır (Fan, 1998; Doğan, 2002). Benzer şekilde, madde karakteristik eğrileri yetenek dağılımından bağımsız olmadığından olası yanlı maddeler yetenek kestirimiyle çıkarılamamaktadır. Analizler karmaşık, pahalı ve büyük örneklem gerektirmektedir (n>1000). Ayrıca, Örtük Özellik modelleri, maddeler arasında oluşan sabit yanlılığı tespit edemez. Diğer önemli nokta, her hangi bir test uygulaması durumunda, Örtük Özellik kuramının uygulanabilmesi için bazı varsayımların kanıtlanması gerekir. Bu varsayımlar, tek boyutluluk, normallik, yerel bağımsızlıktır. Bunlara model-veri uyumunun sağlanmasını da eklemek gerekir. Çoğu durumda, özellikle gerçek verilerin kullanıldığı çalışmalarda, varsayımları ve model-veri uyumunu sağlamak çok zordur (Lord, 1980; Crocker ve Algina, 1986; Hambleton, Swaminathan ve Rogers, 1991).

#### 3.2.4 Mantel – Haenszel Tekniği

Mantel-Haenszel istatistiği madde yanlılığında kullanılması önerilen göreceli olarak yeni bir metottur. Mantel-Haenszel yaklaşımı belli avantajlar sağlayan Chi-square tekniğidir. Bir serbestlik derecesindeki Chi-square dağılımıdır. Diğer Chi-square tekniklerinden daha güçlü olan yanı maddedeki yanlılık miktarının ölçüsünü hesaplamayı sağlamasıdır. Yanlılık ölçümü Mantel-Haenszel oranı olarak verilir.

Madde yanlılığını belirlemek için Mantel-Haenszel yaklaşımının ilk çalışmaları henüz gelişmektedir ve sonuçları henüz yeterince açık değildir. Yine de ucuz, hızlı, kolay ve daha karmaşık Örtük Özellik modellerine güçlü bir alternatiftir.

Son zamanlarda yukarıdaki modellere, Rash modelin varsayımlarını karşılamayan durumlarda kullanılmak üzere geliştirilen Log-Linear ve Mantel-Haenszel yaklaşımına benzeyen Logit-Model teknikleri de eklenmiştir. Bu yaklaşımların kuramsal formülasyonları umut vermeye birlikte henüz eksikler bulunmaktadır (Hills 1984; Crocker ve Algina, 1986; Adams ve Rowe, 1988; Hambleton, Swaminathan ve Rogers, 1991; Yu, 2002).

Yanlılığı tespit etmek için geliştirilmiş tekniklerden Mantel-Haenszel tekniği popülerdir. Ancak, lojistik regrasyon yaklaşımı daha çok tutulur. Lojistik regrasyon yaklaşımı karakteristik değişkenleri (cinsiyet, ırk, yaş vb.) ve/veya iki gruba uygulanan maddelerin parametrelerini karşılaştırır. Mantel-Haenszel ve lojistik regresyon yaklaşımları açık, belirgin avantajlar sunar. Örneğin Mantel-Haenszel tekniği küçük

örneklerde kullanılabilir. IRT metodu düzenli (simetrik) ve düzensiz (asimetrik) yanlılık durumlarında daha sağlam çözümler sunar. Olasılık tablolarında ise lojistik regresyon tekniđini kullanmak daha anlamlıdır.

### SONUÇLAR VE ÖNERİLER

Yanlılık çalışmaları genellikle maddeler üzerinde yürütölür, ama yanlılıđın kendisi testlerin kullanılması üzerine odaklanır. Eğitimde kullanılan başarı, yetenek ve ilgi testlerinin büyük çođunluđu için, yanlı veya adil olmadığı yorumunu yapmamızı sağlayacak bir dıř standart yoktur. Yanlılık çalışmalarının sosyal deđeri eřitlik kavramıyla açıklanabilir. Eřitlik, testin sonuçlarının kullanımının nasıl olduđuyla ilgilidir. Eřitlik ve yanlılık aynı řey deđildir. Eřitlik düřüncesi, testin içeriđi veya sonuçların kullanımı ile ilgili anlaşmalara ve deđerlere dayanır. Örneđin, ÖSS sonucunda öğrencilerin seçimiyle ilgili olarak iki farklı uzmandan iki seçme kuralı önerilebilir. Bunlar: 1-Testin eřitliđe uygun kullanılması için, bir gruptan diđer gruplara göre çok sayıda öğrenci almaya neden olsa bile, tüm uygulamalarda ırkı, cinsiyeti ve diđer grup özelliđi ne olursa olsun yüksek puanı olan öğrenciler alınmalıdır; 2-Uygulama yapılan öğrenci evrenindeki her bir alt gruptan oranlarına bađlı olarak öğrenci alınmalıdır. Böylece, farklı grupların oranları paralelinde, yüksek puana sahip öğrenci olarak test eřitliđe uygun kullanılmış olur. Eđer testimiz erkek veya herhangi bir bölge öğrencileri lehine yanlı ise, test sonuçları da 1. uzmanın önerdiđi řekilde kullanılırsa eřitlik zedelenecektir. Böyle yanlı bir test, eřitliđe uygun olarak kullanılabilir de. Yani yanlılıđı tespit edilen ölçme araçlarına ait sonuçları düzeltilebilir. Örneđin, test erkeklere kızlardan ortalama 10 puan fazla verecek řekilde yanlı ölçümler yapmış olsun. Eđer, kızların puanlarına 10 puan eklenirse ve elde edilen bu puanları karar almada kullanılırsa, test yanlı olmasına rađmen kararlar eřitliđe uygun olacaktır. Elbetteki, yansız araçlar kullanmak daha dođrudan ve emin bir eřitlik sağlayacaktır.

Testlerin ve test maddelerinin yansızlıđını belirlemek amacıyla kullanılacak çok sayıda istatistiksel teknik bulunmaktadır. Bu istatistiksel tekniklerin çođu, testteki maddelerin her biri ne kadar yansız ise testin de yansız olacađı varsayımıyla iře başlar. Böylece test, aynı yapıyı ölçen maddelerden oluşturulur. Maddelerin çođunu uymayan bir madde olduđunda bu maddenin yanlı olduđuna karar verilir. Bu tekniklerle deđerlendirilen yanlılıđın bir anlamı da, bir maddenin testteki diđer maddelerden farklılıđını bulmaktır.

İstatistik tekniklerin her birinin dayandıkları varsayımlara ve kuramlara bađlı olarak avantajı ve dezavantajı vardır. Varsayımlar ve diđer uygulama şartları sağlandığı takdirde, örtük özellik kuramına dayanan tekniklerin ve Mantel-Haensze tekniđinin genel kabul görüp, sıklıkla kullanıldıđı söylenebilir.

Yanlılığın belirlenmesinde istatistiksel teknikler yanında, test içeriğinin farklı perspektiflerden eleştirilmesi, testin belli bir kültüre dayanıp dayanmadığının veya eğitimsel, mesleki ve ırksal rollerin betimlemesi olup olmadığının incelenmesi de gereklidir. İstatistiksel olarak yansız olan testlerin veya testteki maddelerin, uzmanlarca incelenmesi sonucunda yanlı çıkma olasılığı vardır. Bu nedenle uzman kanısına başvurmak yanlılık çalışmalarının bir türü kabul edilebilir. Test yanlılığının etkisini iyi anlamak için, testi planlama, madde yazma ve gözden geçirme, maddeleri deneme ve madde seçme, ölçek ve normların geliştirilmesi ile diğer tüm aşamalarda uzman kanısına başvurulmalıdır.

Yurt dışındaki test geliştirmeciler tarafından yıllardır kullanılmasına rağmen ülkemizde bu konu üzerine yeterli sayıda çalışma yapılmamıştır. Yurt dışında özellikle hayati öneme sahip sınavların çoğunun yanlılık çalışması yapılmaktadır. Oysa ülkemizde yapılan yanlılık çalışmalarının sayısı sınırlıdır ve hemen hepsi lisans üstü eğitimde tez çalışmaları olarak karşımıza çıkmaktadır. Ülkemizde uygulanan ÖSS, LGS, KPSS vb. gibi kişilerin gelecekleriyle ilgili kararlar verilen testlerin yanlılık çalışmaları genelde yapılmamaktadır. Kuramsal çalışmaların asıl amacının uygulama alanına katkı getirmek olduğu düşünülürse; çok sayıda insan hakkında hayati kararlar verilen bu tür sınavların yanlılık çalışmaları uygun tekniklerle yapılmalı ve sürekli hale getirilmelidir. Kuramsal çalışmalar birçok açıdan önemli olsa da, uygulama alanında kullanılmadığı sürece pratik katkı getirmeyecektir. Bu nedenle de akademik bir çaba olmanın dışına çıkmayacaktır.

## Kaynaklar

- Angoff, W. H. ve Ford, S. F., (1973). "Item-Race Interaction on a Test of Scholastic Aptitude." **Journal of Educational Measurement**, 10, 95-105.
- Angoff, W. H., (1975). **The investigation of test bias in the absence of an outside criterion**. Paper Presented at the NIE Conference on Test Bias.
- Adams, R. J. ve Rowe, K. J., (1988). **Item bias**.in Keeves, J.P.(ed.) Educational Research, Methodology, And Measurement:An International Handbook. Oxford: Pergamon Press. (pp 392-398)
- Berk, R., (1982). **Handbook of methods for detecting test bias**. Baltimore: Johns Hopkins University Pres.
- Brannick, Michael T., (2004, OCAK 21). *Item Response Theory*. <http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm>
- Camilli, G. (1979). **A Critique of The Chi-Square Method for Assessing Item Bias**. Unpublished Paper, Laboratory of Educational Research, University of Colorado.
- Camilli, G. Shepard, L. A., (1994). **Methods for Identifying Biased Test Items**. California: Sage Publications.
- Crocker, L. and Algina, J., (1986). **Introduction to Classical and Modern Test Theory**. Orlando:Rinehart and Winston, Inc.
- Devine, P. J. and Raju N. S., (1982). "Extent of Overlap Among Four Item Bias Methods." **Educational and Psychological Measurement** 42,pp. 1049-1066.
- Dođan, Nuri. (2002). "**Klasik Test Kuramı ve Örtük Özellikler Kuramının Örneklerle Bağlamında Karşılaştırılması**." (Yayımlanmamış Doktora Tezi). Ankara: Hacettepe Üniversitesi
- Fan, Xitao, (1998). "Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item-Person Statistics." **Educational and Psychological Measurement**, 58, 3, 357-381
- Hambleton, R K., Swaminathan, H. ve Rogers, H. J., (1991). **Fundamentals of Item Response Theory**. London: Sage Publication.
- Hills, J. R., (1984). "Quantitative Methods Used in the Study of Item Bias." **ERIC Document Reproduction Service No. ED 247 271**.
- Ironson, G. H. and Craig, R., (1982). "Item Bias Techniques When Amount Of Bias is Varied And Score Differences Groups Are Presented." University of South Florida, Tampa. Dept of Psychology. (**ERIC Document Reproduction Service No. ED 227 146**)

- Köklü, Nilgün, (2002). **Açıklamalı İstatistik Terimleri Sözlüğü**. Ankara: Nobel Yayınevi
- Lord, F. M., (1980). **Applications of Item Response Theory to Practical Testing Problems**. Hillsdale, NJ: Lawrence Erlbaum.
- Mellenberg, G. J., (1983). **Conditional Item Bias Methods**. In S. H. Irvine and W. J. Barry (Eds), *Human Assessment and Cultural Factors* (pp. 293-302), Newyork: Plenum Pres.
- Mellenberg, G. J., (1989). "Item Bias And Item Response Theory." **International Journal of Educational Research: Applications of Item Response Theory**.13,2,pp.123-144.
- Milli Eğitim Bakanlığı, (1996). **Örnekleriyle Türkçe Sözlük**. (4. Cilt) Ankara: Türk Tarih Kurumu Basımevi
- Osterlind, S., (1983). **Test Item Bias**. Newbury Park: Sage Publications.
- Raju, N. S., (1988). "The Area Between Two Item Characteristic Curves." **Psychometrika**, 53,495-502.
- Raju, N. S., (1990). "Determining The Significance Of Estimated Signed And Unsigned Areas Between Two Item Response Functions." **Applied Psychological Measurement**, 14, 197-207.
- Rodney. G. L. ve Drasgow, F., (1990). "Evaluation Of Two Methods For Estimating Item Response Theory Parameters When Assessing Differential Item Functioning." **Journal of Applied Psychology**. 75,2,pp.164-174.
- Rudner, L., Getson, P. R. ve Knight, D. L., (1980). Biased Item Detection Techniques. **Journal of Educational Statistics**. 5,pp. 213-233.
- Scheuneman, J. D. (1979). "A New Method of Assessing Bias in Test Items." **Journal of Educational Measurement**, 16, pp. 143-152..
- Seong, Tae-Je., ve Subkoviak, M. J., (1987). "A Comparative Study of Recently Proposed Item Bias Detection Methods." **Paper presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 157 942)**.
- Shepard, L. A., Camilli, G. ve Williams, D. M., (1984). "Validity of Approximation Techniques for Detecting Item Bias." **Journal of Educational Measurement**.22,2,pp77-105.
- Tittle, C. K., (1988). **Test Bias**. in Keeves, J.P.(ed.) *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Yu, Chong Ho. (2002, 1 Nisan). *True Score and Item Response Theory*. <[http://seamonkey.ed.asu.edu/~alex/computer/sas/math\\_reality.htm](http://seamonkey.ed.asu.edu/~alex/computer/sas/math_reality.htm)>