

# Likert Tipi Ölçeklere Madde Seçmede Geleneksel Madde Analizi Tekniklerinin Karşılaştırılması

A. Ata Tezbaşaran\*  
Hacettepe Üniversitesi

## Özet

Tutumların ölçülmesi amacıyla, çeşitli konularda iki farklı ölçüte göre beş ayrı Likert tipi tutum ölçeği geliştirme çalışması yapılmıştır. Birinci ölçüt madde seçmeye esas olan madde istatistiklerinin manidarlık düzeylerinin sabit tutulmasıdır. Bu çalışmada manidarlık düzeyi  $p \leq 0,01$  olarak alınmış ve bu ölçütü karşılayan maddeler seçilmiştir. İkinci ölçüt ise seçilecek madde sayısının sabit tutulmasıdır. Bu çalışmada seçilecek madde sayısı 20 olarak belirlenmiştir. Bu iki ölçüte dayalı olarak üç ayrı teknikle madde puanları ve ölçek puanlarına ait istatistikler belirlenmiştir. Ele alınan tekniklerden birincisi madde puanları dizisi ile ölçek puanları dizisi arasında Pearson Momentler Çarpımı korelasyon katsayılarının hesaplanmasıdır. İkinci teknik, hesaplama konu olan madde puanları dışta tutularak hesaplanan ölçek puanları ile madde puanları arasındaki Pearson Momentler Çarpımı korelasyon katsayılarının hesaplanmasıdır. Üçüncü teknik, her bir madde için, ölçek puanları ölçüt alınarak belirlenen alt ve üst gruptaki madde puanları dizilerinin ortalamaları arasındaki farkın manidarlığının belirlenmesidir. Likert Tipi ölçeklerin geliştirilmesi için bu iki ölçüte dayalı olarak üç ayrı teknikle elde edilen ölçekler, güvenilirlik katsayıları bakımından incelenmiştir. Güvenirlik katsayıları Cronbach Alpha (a) formülüyle hesaplanmıştır. İkinci ölçüte dayalı olarak daha az sayıda madde ile birinci ölçütte elde edilenlere denk ve yüksek güvenilirlik katsayıları elde edilmiş; üç ayrı madde seçme tekniğiyle elde edilen ölçeklerin güvenilirlik katsayıları arasında manidar fark bulunmamıştır. Madde istatistiklerinin manidarlığı ölçütüne dayalı olarak, madde-ölçek korelasyonları ile daha çok madde seçilmiş ve bu yolla oluşturulan ölçeklerin güvenilirlik katsayıları ile diğer tekniklerle elde edilen ölçeklerin güvenilirlik katsayıları arasında manidar fark bulunmamıştır.

**Anahtar kelimeler:** Likert Tipi Ölçekler, madde seçme işlemleri, iç tutarlık, tutumların ölçülmesi, tutum ölçeklerinin geliştirilmesi, korelasyon

## Abstract

*In this study two criteria and three techniques related with item selection procedures for construction of Likert Type scales were investigated. One of these criteria was item selection via fixed significance level of item statistics ( $p \leq 0.01$ ), and the other criterion was selection via fixed number of item ( $K = 20$  items). Three techniques for item selection procedures were examined under these criteria. The first technique depends on calculation of correlation coefficients between item scores and scale scores. The second technique depends on calculation of correlation coefficients between item scores and scale scores obtained with the related item score excluded. The third technique depends on significance of difference between averages of upper and lower group item scores. Five separate attitude scales were prepared for investigative purposes. Results indicated that using fixed number of item criterion provided equivalent reliability with less amount of items. Reliability coefficients of obtained scales via three item selection techniques were not significantly different.*

**Key words:** Likert Type Scales, item selection procedures, internal consistency, measurement of attitudes, construction of attitude scales, correlation

\* Yazışma Adresi: Dr. A. Ata Tezbaşaran, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Beytepe, Ankara.

E-posta: ata@hacettepe.edu.tr

Yayın Kurulu Notu: Bu araştırma, 1996 yılındaki yayın kurulunca yazara verilen kabul mektubuna istinaden, ancak yeniden değerlendirme sürecine alınarak yayımlanmıştır.

Bu araştırma, genel olarak "dereceleme toplamalarıyla ölçekleme" (summated ratings) yöntemine dayalı olan, yaygın adıyla "Likert tipi" ölçek geliştirme işlemlerinin bir bölümünü oluşturan madde analizi tekniklerinden başlıcalarının karşılaştırılmasını kapsamaktadır.

Likert tipi ölçekler, geliştirilmesinde yapılan işlemler diğer dereceleme ölçeklerinin geliştirilmesine göre daha kullanışlı olduğu için, davranış bilimlerinde birçok kişilik özelliğinin (trait) ve özellikle tutumların ölçülmesinde diğer ölçeklere göre en sık ve en yaygın olarak kullanılan ölçeklerdir (Oppenheim, 1979; Judd, Eliot ve Kidder, 1991; Fraenkel ve Wallen, 2003; Sommer ve Sommer, 2002). Likert tipi ölçekler, cevaplayıcı tepkilerine dayalı olarak geliştirilir. Likert'in ölçek geliştirme işlemleri "cevaplayıcı merkezli"dir (Torgerson, 1958). Bu bakımdan da tutum maddelerinin değil cevapların ölçeklenmesi amacını güder. Likert'in ölçekleme yönteminde, uyarıcılara gösterilen bütün sistematik değişimler, cevaplayıcılar arasındaki farklılıklara atfedilir. Uyarıcılardan her biri, diğerlerinin tekrarı gibi düşünülür (Likert, 1932).

Ölçeğin bütünüyle ölçülmek istenen tutumu ölçmede, her bir maddenin ölçme gücünü belirlemek için Likert tarafından özgün olarak iki çeşit madde analizi önerilmiştir: 1) Korelasyonlara dayalı analiz, 2) "İç tutarlık ölçütüne" (alt ve üst grup ortalamaları farkına) dayalı analiz (McIver ve Carmines, 1982).

**Korelasyonlara Dayalı Analiz.** Her bir maddeye ait madde puanları dizisi ile ölçek puanları dizisi arasındaki korelasyonların hesaplanması Likert tarafından önerilen ilk nesnel denetimdir. Bir madde için bu yolla hesaplanan korelasyon katsayısı sıfır veya çok düşük ise bu korelasyon katsayısı, maddenin diğer maddelerle ölçülmek istenen tutumu ölçmede yetersiz kaldığını gösterir. Ölçek puanı (diğer maddelerin oluşturduğu toplam) ile ilişkisi düşük maddeler, nihai ölçekle ölçülmek istenen tutumun ölçülmesine pek az katkıda bulunabilir.

Aynı durum bir maddenin diğer maddelerle ilişkisi bakımından da söz konusudur. Birbirleri ile düşük ilişki gösteren veya ilişkisiz olan maddelerin bir araya getirilmesiyle oluşturulan bir ölçeğin güvenilirliği ve geçerliği düşük olur. Sonuç olarak, düşük korelasyonlara sahip maddeler nihai ölçeğe alınmamalıdır.

**Alt ve Üst Grup Ortalamaları Farkına Dayalı Analiz.** Likert ölçekleme tekniği, yüksek hızlı ve kullanışlı bilgisayarların ortaya çıkışından önce geliştirilmiş olduğundan, korelasyon katsayılarının hesaplanması, özellikle büyük veri kümeleri için zahmetli, emek ve zaman tüketici görülmüştür. Bu nedenle, bu tür ölçeklerin madde analizinin yapılmasında daha az zahmetli ikinci bir yol önermiş ve bu yola "iç tutarlık ölçütüne dayalı madde analizi" adını vermiştir. İç tutarlık ölçütü ile nihai ölçeğe madde seçme işlemini şöyle özetleyebiliriz: Deneme ölçeğindeki maddeler puanlanıp ölçek puanları hesaplandıktan sonra, cevaplayıcıların aldığı puanlar en yüksekte en düşüğe doğru sıralanır. Cevaplayıcılardan ölçek puanları dağılımının üst ucundaki %27'si üst grup, alt ucundaki %27'si alt grup olarak belirlenir. Üst gruptaki cevaplayıcılar ölçeğin tümüyle ölçülmek istenen tutuma olumlu yönde sahip olanlar, alt gruptaki cevaplayıcılar da olumsuz yönde sahip olanlardır. Her bir maddeye ait tepki dağılımında da üst gruptakilerin ölçülen tutuma olumlu derecede sahip olanları, alt gruptakilerin de olumsuz derecede sahip olanları göstermesi beklenir. Eğer bir madde bu iki gruptaki tepkileri birbirinden ayırt edemiyorsa, tutuma olumlu derecede sahip olanlar ile olumsuz derecede sahip olanları ayırt etmiyor demektir. Bu durum değişik şekillerde ortaya çıkabilir, örneğin üst gruptaki cevaplayıcılar alt gruptakilerden yeterince yüksek puan alamamış olabilir. Böylesi maddelerin nihai ölçeğin dışında kalması gerekir. Bir madde, ölçülen tutum boyunca birbirlerinden kesin bir şekilde ayrılabilen gruplarda özdeş tepkileri uyandırıyor ise bu madde ölçülmek istenen tutumun iyi bir göstergesi değildir. Diğer yandan bir madde aşırı uçtaki iki grubu birbirinden

manidar bir biçimde ayırt edebiliyorsa, bu madde nihai ölçüğe konulması uygun olan bir maddedir.

Likert'in tanımıyla "iç tutarlık ölçütü"ne dayalı madde seçme işlemi, denenen ölçek puanları dağılımının iki ucundaki % 27'lik alt ve üst gruptaki cevaplayıcılardan elde edilen veriler üzerinde gerçekleştirilir. Bu ölçütle her bir madde için madde puanları dağılımı incelenir. Her bir madde için, üst gruptaki cevaplayıcıların madde puanları ortalaması ile alt gruptaki cevaplayıcıların madde puanları ortalaması arasındaki farkın manidarlığı t testi ile sınanır. Buradaki t değeri, alt ve üst gruplar birbirinden bağımsız olduğu için, bağımsız gruplar için uygulanan t testidir. Hesaplanan t değerine "Ayırıcılık Gücü İndeksi" adı verilmektedir.

Bu iki çeşit madde analizinden hangisi kullanılmalıdır? 1938'de Likert ile Murphy'nin birlikte yaptıkları bir çalışmada iki çeşit madde analizi karşılaştırılmıştır (akt., Edwards, 1957). Bu çalışmada deneme ölçeğindeki 15 madde, alt ve üst grup ortalamaları arasındaki farkın büyüklüğüne göre ve ayrıca, madde puanları ile ölçek puanları arasındaki korelasyonların büyüklüğüne göre sıralanmış, bu iki sıra arasındaki korelasyon 0,91 olarak bulunmuştur. Bu yüzden Edwards (1957), korelasyonlarla analiz yerine, alt ve üst grup ortalamaları farkının test edilmesine dayalı olarak madde analizi yapılmasının daha kolay ve pratik olduğu sonucuna varmıştır. Likert'in öne sürdüğü, korelasyon hesaplamasının çok zaman ve emek alacağı gerekçesi bugün için geçersizdir. Çünkü günümüzde bilgisayarların yaygın olarak kullanılması bu gerekçeyi ortadan kaldırmıştır. Üstelik korelasyon tekniğinin, t testine göre önemli avantajları olduğu, sadece alt ve üst gruptan alınan veriler yerine bütün gruptaki verileri dikkate aldığı vurgulanmaktadır (McIver ve Carmines, 1982). Diğer yandan her zaman bütün gruptan elde edilen verilerin kullanılması büyük avantajlar getirebilir. McIver ve Carmines (1982) aynı verilere iki madde analizi çeşidinin de uygulanabileceğini, iki çeşit madde analizinden ayrı ayrı

yararlanılabileceğini ve bunların aynı sonuca götürmesi gerektiğini de öne sürmektedirler.

Gamache (1983) tarafından çeşitli geleneksel modellere dayalı ölçütlere göre ve örtük özellikler modeline göre geliştirilen ölçekler (Guttman-Scales; Likert-Scales; Thurstone-Scales; Birnbaum-Models ve Rasch-Model), ölçeklerin taşınması gereken nitelikler (güvenirlik ve geçerlik) bakımından incelenmiştir. Ölçekler ortak madde havuzundan madde seçilmesi yoluyla Likert'in geleneksel yöntemiyle ve bir parametrelili Rasch modeline dayalı olarak oluşturulmuştur. Ölçekler, madde-ölçek puanları korelasyonları, Guttman, iki parametrelili Birnbaum ve tek parametrelili Rasch modeli için analiz edilmiştir. Bulgular, madde-ölçek korelasyonlarına dayalı geleneksel yöntemle az bir farkla daha güvenilir bir ölçek elde edildiğini göstermiştir. Bununla birlikte beş yöntemeye dayalı ölçeklerin tamamı, ölçek nitelikleri bakımından birbirlerine benzer bulunmuş, aralarında manidar fark görülmemiştir.

Tutumların ölçülmesi ile ilgili bir başka çalışmada O'Neal ve Chissom (1993) üç ölçek modelini karşılaştırmışlardır. Bu modeller sıralama, ikili karşılaştırmalar ve Likert tipi ölçeklerdir. Bu çalışma için, üç tutum konusu seçilerek her biri beş madde içeren, incelenen modellere dayalı üç ayrı madde kümesi geliştirilmiştir. Yüksek öğretimde öğretmen niteliklerine ilişkin tutumları ölçmek için geliştirilen maddeler, çoğunluğu kızlardan oluşan ve eğitim kurslarına katılan üniversite öğrencilerine uygulanmıştır. İyi öğretmenin niteliklerine ilişkin madde kümeleri için, modeller arasında oldukça yüksek korelasyonlar bulunmuştur. Sıralama ve ikili karşılaştırma modellerine ait madde kümeleri arasındaki korelasyonlar daha yüksek; Likert tipi maddeler ile diğer iki modele dayalı madde kümeleri arasındaki korelasyonlar da tutarlı ve yüksek olmakla birlikte diğerleri kadar yüksek bulunmamıştır. Bu çalışma üç küçük madde kümesiyle sınırlı olmakla birlikte, belirli koşullar altında bu üç model arasında uyuma bulunduğunu gösteren kanıtları

sağlamıştır. Araştırmacılar, kullanılan maddelerin yeniden gözden geçirilmesi ve sayısının artırılarak ölçütleri karşılayan iyi maddelerin seçilmesiyle bu üç model arasındaki uyumun daha yüksek olabileceğini belirtmektedirler.

### Amaç

Likert tipi ölçeklerin geliştirilmesinde bu araştırmada incelenecek tekniklerin dışında kalan birçok madde analizi tekniğinden söz edilebilir. Bu araştırmada incelenen madde analizi teknikleri, en sık ve en yaygın olarak kullanılan tekniklerdir. Bunlardan birincisi ölçek puanlarına dayalı olarak cevaplayıcılar en yüksek puanlı olandan en düşük puanlı olana doğru sıralandıktan sonra, madde puanları dağılımının iki ucunda bulunan alt ve üst gruba ait ortalamalarının birbirinden ayırt edilmesi esasına dayalıdır. Bu teknikle ilgili olarak Edwards (1957) ile Turgut ve Baykul (1992) bilgi vermektedir. İkinci teknik ise, tek boyutlu ölçekleme sayıtlarına dayalı iç tutarlığı artırmaya yönelik, madde puanları ile ölçek puanları arasındaki korelasyonlara dayalı olarak madde seçme tekniğidir. Bu teknikte madde puanları dizisi ile ölçek puanları dizisi arasında Pearson Momentler Çarpımı korelasyon katsayısı hesaplanır (Judd, Eliot ve Kidder, 1991). Bu araştırmada korelasyonlar iki ayrı şekilde hesaplanmıştır. Esasen madde-ölçek korelasyonlarının hesaplanmasında söz konusu madde puanları hariç ölçek puanları hesaplandıktan sonra madde-ölçek korelasyonunun hesaplanması gerekir. Çünkü madde puanları ölçek puanlarından çıkarılmadığı takdirde, madde puanları ile ölçek puanları arasındaki kovaryans büyüyeceği için, hesaplanacak korelasyon katsayısı gerçekte olduğundan daha yüksek çıkacaktır (Horst, 1966). Durum böyle olmasına rağmen birçok uygulamacının bunu göz ardı ettiği düşünülerek, madde puanları dahil olarak korelasyon hesaplama tekniği de araştırma kapsamına alınmıştır.

Bu araştırmada Likert tipi ölçeklerin madde analizinde, hesaplanan madde istatistiklerine ait

manidarlık düzeyini sabit tutma ölçütüne dayalı olarak, alt ve üst grup ayırıcılığına göre ve iki ayrı yoldan hesaplanan madde - ölçek korelasyonlarıyla nihai ölçeğe seçilen madde sayıları değişip değişmediği; hem bu ölçüte hem de seçilecek madde sayısını sabit tutma ölçütüne dayalı olarak bu madde seçme teknikleriyle oluşturulan nihai ölçeklerin güvenilirlik katsayılarının arasında manidar fark olup olmadığına bakılacaktır.

### Araştırma Soruları

1. Hesaplanan istatistiklere ait manidarlık düzeyini ( $p \leq 0,01$ 'de) sabit tutma ölçütüne (Ölçüt I) göre;

a) "Söz konusu madde puanları çıkarılarak hesaplanan ölçek puanları ile bu madde puanları arasındaki korelasyonlara dayalı olarak madde seçme" (R mp.H) tekniğiyle seçilecek madde sayısı kaçtır?

b) "Söz konusu madde puanları çıkarılmadan hesaplanan ölçek puanları ile bu madde puanları arasındaki korelasyonlara dayalı madde seçme" (R mp.D) yoluyla seçilecek madde sayısı kaçtır?

c) "Ölçek puanlarından tanımlanan alt - üst gruplardaki madde puanları ortalamaları arasındaki farkın manidarlığının t testiyle sınanmasına dayalı madde seçme" (AÜGOF) yoluyla seçilecek madde sayısı kaçtır?

d) Yukarıda sözü edilen R mp.H, R mp.D ve AÜGOF teknikleriyle oluşturulan ölçeklerin güvenilirlik katsayıları arasında manidar farklar var mıdır?

2. Seçilecek madde sayısı sabit ( $K = 20$ ) tutma ölçütüne (Ölçüt II) göre deneme ölçeğinden madde seçmede, R mp.H, R mp.D ve AÜGOF teknikleriyle oluşturulan ölçeklerin güvenilirlik katsayıları arasında manidar farklar var mıdır?

3. Ölçüt I ve Ölçüt II'ye göre aynı teknikle oluşturulan ölçeklerin güvenilirlik katsayıları arasında manidar fark var mıdır?

### Yöntem

Üzerinde madde seçme tekniklerini incelemek amacıyla, araştırmacının denetiminde bir öğrenci grubu tarafından beş ayrı denemelik tutum ölçeği hazırlanmıştır. Bu ölçekler şunlardır: Fala İlişkin Tutum Ölçeği (Ölçek 1), Görücü Usulüyle Evliliğe İlişkin Tutum Ölçeği (Ölçek 2), Ölçme Dersine İlişkin Tutum Ölçeği (Ölçek 3), Öğretmen Dayağına İlişkin Tutum Ölçeği (Ölçek 4) ve Sigara İçmeye İlişkin Tutum Ölçeği (Ölçek 5).

### Verilerin Toplanması ve Analizi

Birbirinden bağımsız olarak, deneme ölçeklerinden dördü 100'er kişiye, biri 120 kişiye uygulanmıştır. Seçilen maddelerden oluşturulan nihai ölçeklerin güvenilirliklerini hesaplamada deneme örneklemindeki verilerden yararlanılmıştır. Her bir ölçeğe ilişkin olarak toplanan veriler için, bilgisayarda ClarisWorks-Spreadsheet (1991) programıyla madde cevaplayıcı matrisleri oluşturulmuş ve Cronbach  $\alpha$  güvenilirlik katsayıları bu programla hesaplanmıştır. Madde puanı dahil ve hariç olmak üzere, madde puanları ile ölçek puanları arasındaki Pearson Momentler Çarpımı Korelasyon katsayıları, alt ve üst grup ortalamaları arasındaki farkın manidarlığı StatView paket programıyla (Abacus Concepts, Inc. 1992) hesaplanmış ve katsayıların manidarlık düzeyleri bu programla belirlenmiştir. Güvenirlik katsayıları arasındaki farkın manidarlığını sınamak için Microsoft Excel programı (1993) ile iki korelasyon katsayısı arasındaki farkın manidarlığını test eden (Kutsal ve Muluk, 1972; Akhun 1991) bir program araştırmacı tarafından yazılmıştır.

### Bulgular

#### 1. Ölçüt I: Manidarlık Düzeyi ( $p \leq 0,01$ 'de) Sabit Tutularak Madde Seçilmesiyle İlgili Bulgular

1.1. Ölçek 1'in deneme formunda 60 madde 100 cevaplayıcıya uygulanmıştır. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,78$ 'dir. Birinci ölçüte dayalı olarak, birinci teknikle (RmpH) 60 maddeden 26'sı seçilmiş (%43); elde

edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,55'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,80$ 'dir. Birinci ölçüte dayalı olarak, ikinci teknikle (RmpD) 60 maddeden 39'u seçilmiş (%65); elde edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,58'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,83$ 'tür. Birinci ölçüte dayalı olarak, üçüncü teknikle (AÜGOF) 60 maddeden 29'u seçilmiş (%48); elde edilen en düşük t değeri 2,788, en yüksek t değeri 6,086'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,81$ 'dir. Ölçüt I'e göre üç ayrı teknikle deneme ölçeğinden seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farkların hiçbirisi manidar değildir ( $p \leq 0,01$ ). Diğer yandan deneme ölçeğinden elde edilen güvenilirlik katsayısı ile üç ayrı tekniğe dayalı olarak oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasında da manidar fark yoktur. Bununla birlikte, seçilen madde sayısı birbirinden farklıdır. Buna göre, madde puanları hariç tutularak hesaplanan korelasyonlara dayalı olarak oluşturulan ölçek için, diğerleriyle denk güvenilirlikte, daha az sayıda madde seçilmiştir.

1.2. Ölçek 2'nin deneme formunda 60 madde 100 cevaplayıcıya uygulanmıştır. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,75$ 'tir. Birinci ölçüte dayalı olarak, birinci teknikle (RmpH) 60 maddeden 26'sı seçilmiş (%43); elde edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,56'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dir. Birinci ölçüte dayalı olarak, ikinci teknikle (RmpD) 60 maddeden 30'u seçilmiş (%50); elde edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,61'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dir. Birinci ölçüte dayalı olarak, üçüncü teknikle (AÜGOF) 60 maddeden 28'i seçilmiş (%47); elde edilen en düşük t değeri 2,743, en yüksek t değeri 7,038'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,91$ 'dir. Ölçüt I'e göre deneme ölçeğinden elde edilen güvenilirlik katsayısı ile incelenen tekniklerle oluşturulan ölçeklerden

elde edilen güvenilirlik katsayıları arasındaki farklar manidardır ( $p \leq 0,001$ ). Diğer yandan üç ayrı teknikte seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Bununla birlikte, madde-ölçek korelasyonları ve t değerleri için Ölçüt I'e göre seçilen maddelerin sayısı aynı değildir.

Buna göre, madde puanları hariç tutularak hesaplanan madde-ölçek korelasyonlarına dayalı olarak oluşturulan ölçek, diğer tekniklerle oluşturulan ölçeklerden daha az sayıda maddeden oluşmaktadır.

1.3. Ölçek 3'ün deneme formunda 60 madde 100 cevaplayıcıya uygulanmıştır. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,88$ 'dir. Birinci ölçüte dayalı olarak, birinci teknikle (RmpH) 60 maddeden 41'i seçilmiş (%68); elde edilen en düşük korelasyon katsayısı 0,27, en yüksek korelasyon katsayısı 0,65'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,91$ 'dir. Birinci ölçüte dayalı olarak, ikinci teknikle (RmpD) 60 maddeden 47'si seçilmiş (%78); elde edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,68'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,91$ 'dir. Birinci ölçüte dayalı olarak, üçüncü teknikle (AÜGOF) 60 maddeden 42'si seçilmiş (%70); elde edilen en düşük t değeri 2,794, en yüksek t değeri 10,997'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,91$ 'dir. Ölçüt I'e göre üç ayrı madde seçme tekniğinde seçilen maddeler aynı sayıda değildir. Buna göre, madde puanları hariç tutularak hesaplanan madde-ölçek korelasyonlarına dayalı olarak oluşturulan ölçek için diğerlerinden daha az sayıda madde seçilmiştir. Ölçüt I'e göre üç ayrı teknikte seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Aynı zamanda deneme ölçeğinden elde edilen güvenilirlik katsayısı ile üç ayrı tekniğe dayalı olarak oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasında da manidar fark yoktur.

1.4. Ölçek 4'ün deneme formunda 50 madde 100 cevaplayıcıya uygulanmıştır. 50 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,97$ 'dir.

Birinci ölçüte dayalı olarak, birinci teknikle (RmpH) 50 maddeden 50'si seçilmiş (%100); elde edilen en düşük korelasyon katsayısı 0,27, en yüksek korelasyon katsayısı 0,83'tür ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,97$ 'dir. Birinci ölçüte dayalı olarak, ikinci teknikle (RmpD) 50 maddeden 50'si seçilmiş (%100); elde edilen en düşük korelasyon katsayısı 0,31, en yüksek korelasyon katsayısı 0,84'tür ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,97$ 'dir. Birinci ölçüte dayalı olarak, üçüncü teknikle (AÜGOF) 50 maddeden 49'u seçilmiş (%98); elde edilen en düşük t değeri 2,905, en yüksek t değeri 13,618'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,96$ 'dir. Ölçüt I'e göre üç ayrı teknikte seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farkların hiçbirisi manidar değildir. Diğer yandan deneme ölçeğinden elde edilen güvenilirlik katsayısı ile üç ayrı tekniğe dayalı olarak oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasında da manidar fark yoktur. Bununla birlikte, Ölçüt I'e göre, alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı olarak oluşturulan ölçek için diğerlerinden bir madde daha az seçilmiştir.

1.5. Ölçek 5'in deneme formunda 60 madde 120 cevaplayıcıya uygulanmıştır. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,93$ 'tür. Birinci ölçüte dayalı olarak, birinci teknikle (RmpH) 60 maddeden 53'ü seçilmiş (%88); elde edilen en düşük korelasyon katsayısı 0,26, en yüksek korelasyon katsayısı 0,56'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dir. Birinci ölçüte dayalı olarak, ikinci teknikle (RmpD) 60 maddeden 30'u seçilmiş (%50); elde edilen en düşük korelasyon katsayısı 0,25, en yüksek korelasyon katsayısı 0,66'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,93$ 'tür. Birinci ölçüte dayalı olarak, üçüncü teknikle (AÜGOF) 60 maddeden 50'si seçilmiş (%83); elde edilen en düşük t değeri 2,846, en yüksek t değeri 10,005'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,93$ 'tür. Ölçüt I'e göre üç ayrı teknikte seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasında fark yoktur. Diğer yandan deneme ölçeğinden elde edilen güvenilirlik katsayısı ile üç ayrı tekniğe dayalı

olarak oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasında da fark yoktur. Bununla birlikte, Ölçüt I'e göre, alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı olarak oluşturulan ölçek için diğerlerinden daha az madde seçilmiştir.

Tablo 1'de Ölçüt I'e göre deneme ölçeklerinden üç ayrı teknikle nihai ölçekler için seçilen madde sayılarının yüzdeleri verilmiştir.

**Tablo 1.**

*İstatistiklerin Manidarlık Düzeyi ( $p \leq 0.01$ ) Ölçütüne Dayalı Olarak Üç Ayrı Teknikle Deneme Ölçeklerinden Seçilen Madde Sayısı Yüzdeleri*

Madde Seçme Tekniği	Ölçekler				
	1	2	3	4	5
R mp.H	43	43	68	100	88
R mp.D	65	50	78	100	92
AÜGOF	48	47	70	98	83

Tablo 1'de gösterilen beş ölçekten son ikisinde seçilme ölçütünü ( $p \leq 0.01$ ) karşılayan madde sayısı denenen madde sayısının %80'inden fazladır. Tablo 1'de gösterilen ilk üç ölçek dikkate alındığında, nihai ölçek için madde seçme oranı %80'in altındadır. Bu ölçüte dayalı olarak üç teknikle seçilecek madde sayıları arasında şöyle bir sistematik ilişki görünmektedir: R mp.H < AÜGOF < R mp.D.

## **2. Ölçüt II: Madde Sayısı Sabit ( $K = 20$ ) Tutularak Madde Seçmeyle İlgili Bulgular**

2.1. Ölçek 1 için madde puanları hariç ve dahil tutularak hesaplanan korelasyonlara dayalı olarak yapılan 20 maddelik seçme işlemi aynı maddeler seçilmiş, buna karşılık alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı 20 maddelik seçme işlemi, önceki işlemlerden farklı dört madde seçilmiştir. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,78$ 'dir. İkinci ölçüte dayalı olarak, birinci teknikle (RmpH) elde edilen en düşük korelasyon katsayısı 0,29, en yüksek korelasyon katsayısı 0,55'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,79$ 'dur. İkinci teknikle

(RmpD) elde edilen en düşük korelasyon katsayısı 0,34, en yüksek korelasyon katsayısı 0,58'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,79$ 'dur. Üçüncü teknikle (AÜGOF) elde edilen en düşük t değeri 3,376, en yüksek t değeri 6,086'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,77$ 'dir. Üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki fark manidar değildir. Diğer yandan deneme ölçeğinden elde edilen güvenilirlik katsayısı ile üç ayrı tekniğe dayalı olarak oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasında da manidar fark yoktur.

2.2. Ölçek 2 için madde puanları hariç ve dahil tutularak hesaplanan korelasyonlara dayalı olarak yapılan 20 maddelik seçme işlemi aynı maddeler seçilmiş, buna karşılık alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı seçme işlemi üç farklı madde seçilmiştir. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,75$ 'tir. İkinci ölçüte dayalı olarak, birinci teknikle (RmpH) elde edilen en düşük korelasyon katsayısı 0,34, en yüksek korelasyon katsayısı 0,56'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,88$ 'dir. İkinci teknikle (RmpD) elde edilen en düşük korelasyon katsayısı 0,39, en yüksek korelasyon katsayısı 0,61'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,88$ 'dir. Üçüncü teknikle (AÜGOF) elde edilen en düşük t değeri 3,729, en yüksek t değeri 7,038'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,89$ 'dir. Üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasında manidar fark yoktur. Buna karşılık, üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları ile deneme ölçeğinin güvenilirlik katsayısı arasındaki farklar manidardır ( $p \leq 0,01$ ).

2.3. Ölçek 3 için madde puanları hariç ve dahil tutularak hesaplanan korelasyonlara dayalı olarak yapılan 20 maddelik seçme işlemi bir madde farklı seçilmiş, buna karşılık alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı seçme işlemi üç farklı madde seçilmiştir. 60

maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,88$ 'dir. İkinci ölçüte dayalı olarak, birinci teknikle (RmpH) elde edilen en düşük korelasyon katsayısı 0,42, en yüksek korelasyon katsayısı 0,65'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dır. İkinci teknikle (RmpD) elde edilen en düşük korelasyon katsayısı 0,46, en yüksek korelasyon katsayısı 0,68'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dır. Üçüncü teknikle (AÜGOF) elde edilen en düşük t değeri 4,002, en yüksek t değeri 10,997'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,89$ 'dur. Üç ayrı teknikle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Aynı şekilde, üç ayrı teknikle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları ile deneme ölçeğinin güvenilirlik katsayısı arasındaki farklar da manidar değildir.

2.4. Ölçek 4 için madde puanları hariç ve dahil tutularak hesaplanan korelasyonlara dayalı olarak yapılan 20 maddelik seçimde aynı maddeler seçilmiş, buna karşılık alt ve üst grup ortalamaları arasındaki farkın önemine dayalı seçme işlemi yedi farklı madde seçilmiştir. 50 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,97$ 'dir. İkinci ölçüte dayalı olarak, birinci teknikle (RmpH) elde edilen en düşük korelasyon katsayısı 0,66, en yüksek korelasyon katsayısı 0,83'tür ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,96$ 'dır. İkinci teknikle (RmpD) elde edilen en düşük korelasyon katsayısı 0,67, en yüksek korelasyon katsayısı 0,84'tür ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,96$ 'dır. Üçüncü teknikle (AÜGOF) elde edilen en düşük t değeri 6,325, en yüksek t değeri 13,618'dir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,95$ 'tir. Üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Aynı şekilde, üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları ile deneme ölçeğinin güvenilirlik katsayısı arasındaki farklar da manidar değildir.

2.5. Ölçek 5 için madde puanları hariç ve dahil tutularak hesaplanan korelasyonlara dayalı olarak yapılan 20 maddelik seçme işlemi aynı maddeler seçilmiş, buna karşılık alt ve üst grup ortalamaları arasındaki farkın manidarlığına dayalı seçme işlemi dört farklı madde seçilmiştir. 60 maddelik deneme ölçeğinin güvenilirlik katsayısı  $\alpha = 0,93$ 'tür. İkinci ölçüte dayalı olarak, birinci teknikle (RmpH) elde edilen en düşük korelasyon katsayısı 0,48, en yüksek korelasyon katsayısı 0,66'dır ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dır. İkinci teknikle (RmpD) elde edilen en düşük korelasyon katsayısı 0,51, en yüksek korelasyon katsayısı 0,69'dur ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dır. Üçüncü teknikle (AÜGOF) elde edilen en düşük t değeri 5,553, en yüksek t değeri 10,005'tir ve bu ölçeğin güvenilirlik katsayısı  $\alpha = 0,90$ 'dır. Üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Aynı şekilde, üç ayrı yöntemle seçilen maddelerden oluşturulan ölçeklerin güvenilirlik katsayıları ile deneme ölçeğinin güvenilirlik katsayısı arasındaki farklar da manidar değildir.

### 3. Ölçüt I ve Ölçüt II'ye Göre Elde Edilen Güvenirlik Katsayılarına İlişkin Bulgular

Hesaplanan istatistiklere ait manidarlık düzeyini sabit ( $p \leq 0,01$ 'de) tutma ölçütü (Ölçüt I) ile seçilecek madde sayısı sabit ( $K = 20$ 'de) tutma ölçütüne (Ölçüt II) göre oluşturulan ölçeklerden elde edilen güvenilirlik katsayıları arasındaki farkın manidarlığı, t testi ile sınımlanmış ve hesaplanan t değerleri Tablo 2'de gösterilmiştir.

Tablo 2'de görüleceği gibi, beş ayrı ölçekten dördünde iki ayrı ölçüte ve üç ayrı tekniğe dayalı olarak oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidar değildir. Sadece ikinci ölçeğin deneme formunun güvenilirlik katsayısı ile, bundan oluşturulan ölçeklerin güvenilirlik katsayıları arasındaki farklar manidardır ( $p \leq 0,01$ ).



**Tablo 2.**

*Değişik Ölçeklerin Geliştirilmesinde İki Ayrı Ölçüt (Manidarlık Düzeyi ve Madde Sayısı) ve Üç Ayrı Teknik İle Seçilen Maddelerden Oluşan Ölçeklerden Elde Edilen Güvenirlik Katsayıları ve t Değerleri*

Ölçek Adı	Madde Seçme Tekniği	Güvenirlik Katsayıları		t değerleri
		Ölçüt I p ≤ 0.01'de	Ölçüt II K = 20'de Sabit	
<b>Ölçek 1 Deneme (0.78)</b>				
	1. R mp.H	0.8	0.79	0.189
	2. R mp.D	0.83	0.79	0.813
	3. AÜGOF	0.81	0.77	0.743
<b>Ölçek 2 Deneme (0.75)*</b>				
	1. R mp.H	0.9	0.88	0.672
	2. R mp.D	0.9	0.88	0.672
	3. AÜGOF	0.91	0.89	1.057
<b>Ölçek 3 Deneme (0.88)</b>				
	1. R mp.H	0.91	0.9	0.385
	2. R mp.D	0.91	0.9	0.385
	3. AÜGOF	0.91	0.89	1.057
<b>Ölçek 4 Deneme (0.97)</b>				
	1. R mp.H	0.97	0.96	1.019
	2. R mp.D	0.97	0.96	1.019
	3. AÜGOF	0.96	0.95	0.795
<b>Ölçek 5 Deneme (0.93)</b>				
	1. R mp.H	0.93	0.9	1.297
	2. R mp.D	0.93	0.9	1.297
	3. AÜGOF	0.93	0.9	1.297

(p ≤ 0,01)

### Tartışma

Korelasyonun tanımı dikkate alındığında, bekleneneği gibi, ölçek puanlarından madde puanlarını çıkarmadan hesaplanan madde - ölçek puanları arasındaki korelasyon katsayıları, ölçek puanlarından madde puanları hariç tutularak hesaplanan madde - ölçek puanları arasındaki korelasyon katsayılarından daha yüksek çıkma eğilimindedir. Ne var ki bu durumun değişik tekniklerle oluşturulan ölçeklerin güvenirlilik katsayılarında önemli bir değişikliğe yol açmadığı da görülmektedir. Hesaplanan madde istatistiklerinin

manidarlık düzeyi ölçüt alınarak madde seçme işlemleri yapıldığında, ölçek puanlarından madde puanları çıkarılarak hesaplanan korelasyon katsayılarıyla, ölçek puanlarından madde puanları çıkarılmadığı duruma denk güvenirlilikte en az sayıda madde seçilmektedir. Bu durum deneme maddeleri arasından seçilecek madde sayısı %80'in altında olduğu zaman görülmektedir. Bu durumda, istatistiklerin manidarlık düzeyi ölçütüne dayalı olarak üç teknikle seçilen madde sayıları arasında ortaya çıkan ilişki şöyledir:

$$R_{mp.H} < AÜGOF < R_{mp.D}$$

İnceleme bulgularından, iki ölçüte dayalı her bir tekniğe ilişkin değişik açılardan ayrı ayrı sonuç çıkarmak ve yorumlamak olanağı vardır. Örneğin, seçilmeyen maddeler incelendiğinde, bu maddelerin Likert tipi madde yazma kurallarına uygun düşmediğine ilişkin ipuçları görülmektedir. Seçilen ve seçilmeyen maddelerin, madde yazım kurallarına uygunluğu ayrıca incelenmelidir. Bir ölçme aracının kullanışlı olması, istenen güvenirlilik ve geçerlik düzeyi ile bu aracın geliştirilmesi, uygulanması, sonuçlarının kullanılması gibi aşamalar-daki işlemlerin gerektirdiği unsurlar arasında bir "fayda - maliyet" dengesinin kurulmasını gerektirir. Bulgular kullanışlılık bakımından genellikle madde sayısının sabit tutulması ölçütünün tercih edilmesi gerektiğini ortaya koymaktadır. Çünkü bu ölçütle, beş ölçekten dördü daha az sayıda madde ile güvenirlilik düzeyini düşürmeden elde edilmiş; bir ölçekte ise deneme ölçeğinin güvenirliliğinden daha yüksek güvenirlilik sağlanmıştır. Üç ayrı madde seçme tekniği arasında da elde edilen güvenirlilik katsayıları açısından manidar bir fark olmayışı da elde edilen hesaplamaların hangisine ulaşılabilir ise, o tekniğe başvurulmasının daha kullanışlı olacağı sonucunu ortaya koymaktadır. 1938'de Likert ile Murphy'nin (akt., Edwards, 1957), alt ve üst grup ortalamaları arasındaki farkın büyüklüğüne göre sıralanış ile madde puanları ile ölçek puanları arasındaki korelasyonların büyüklüğüne göre maddelerin sıralanışın arasındaki korelasyonu 0,91 olarak

bulunmuştur. Bu yüzden Edwards (1957) da korelasyonlarla analiz yapmanın oldukça zahmetli olduğunu, alt ve üst grup ortalamaları farkının test edilmesine dayalı olarak madde analizi yapılmasının daha kolay ve pratik olacağını belirtmiştir. Bu araştırmanın sonuçları Likert ile Murphy'nin 15 maddelik sınav bulgularını desteklemektedir. Diğer yandan, günümüzde bilgisayarların kullanıma girmiş olması Edwards'ın (1957) ileri sürdüğü güçlükleri ortadan kaldırmış bulunmaktadır. Analiz işlemlerinin bilgisayarlardaki istatistik paket programlarıyla ya da hesap tablolarıyla yapıldığı düşünülürse, t testine dayalı analiz için verilerin düzenleniş ek işlemler gerektirmekte ve orta gruptan sağlanabilecek bilgilerden de vazgeçilmektedir.

Bu çalışmada incelenen madde seçme tekniklerinden korelasyon katsayılarına dayalı olarak yapılan madde seçme işlemlerinde, ölçek puanları bir iç tutarlık ölçütü olarak kabul edilmektedir. Bu tekniğin diğer sayıltılarından biri de madde puanları dizisi ile ölçek puanları dizisi arasındaki ilişkinin doğrusal olduğu şeklindedir. Korelasyon tekniği ile alt ve üst grup ortalamaları farkına dayalı olan teknikte elde edilen güvenilirlik katsayıları arasındaki farkların manidar olmayışı, bu tekniklerden biriyle elde edilen istatistiklerden diğerlerinin yordayabileceğine ilişkin bir işaret olduğu şeklinde yorumlanabilir. Böyle olup olmadığı, örneğin, madde - ölçek korelasyonlarından t değerlerinin doğrusal olarak ne ölçüde yordayabildiği araştırılabilir. Ayrıca, madde puanları ile ölçek puanlarının doğrusal bir ilişki içinde olması beklendiğinden, seçilecek madde puanlarından ölçek puanları da yordayabilmelidir. Bu araştırmadaki değişkenlerle belirtilecek olursa, korelasyon katsayıları madde puanları ile ölçek puanlarının birlikte değişiminin ölçüsünü vermektedir. İki değişken arasındaki birlikte değişim ölçüsünü elde etmenin ötesinde birinden diğerindeki değişimleri de yordayabiliriz (Arıcı, 1992). Madde puanlarının ölçek puanlarını yordayıp yordayamadığı regresyon tekniği ile

araştırılabilir. Madde puanları ile ölçek puanları arasında beklenen ilişki biçimi doğrusal olduğundan, basit doğrusal regresyon tekniği ile deneme ölçeğinden test puanlarını yordayabilen maddeler seçilerek ölçek oluşturulabilir ve korelasyonlara dayalı madde seçme işlemleri ile karşılaştıran bir araştırma yapılabilir.

### Kaynaklar

- Abacus Concepts, Inc. (1992). *StatView for Apple Macintosh*. Berkeley, CA.
- Akhun, İ. (1991). *İstatistiklerin Manidarlığı ve Örneklem*. 3. Baskı. Ankara
- Arıcı, H. (1992). *İstatistik: Yöntemler ve Uygulamalar*. Ankara: Meteksan Yayınları no:13, 5. Baskı
- Clarıs C. (1991). *ClarısWorks for Apple Macintosh*. Santa Clara, California
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* 16 (3), 297 - 333.
- Edwards, A. L. (1957). *Techniques of Attitude Scale Construction*. New York: Appleton-Century-Crofts., Inc.
- Fraenkel, J. R., & Wallen, N. E. (2003). *How to Design and Evaluate Research in Education*. New York: McGraw-Hill Companies, Inc. Fifth Ed.
- Gamache, L. M. (1983). *Comparison of Traditional and Latent Trait Procedures in Analysis and Selection of Rating Scale Items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983). US.; Iowa
- Horst, P. (1966). *Psychological Measurement and Prediction*. Belmont, California: Wadsworth Publishing Com., Inc.
- Judd, C. M., Eliot, E. R., & Kidder, L. H. (1991). *Research Methods in Social Relations*. New York: Harcourt Brace Jovanovich College Publishers
- Kutsal, A., & Muluk, F. Z. (1972). *Uygulamalı Temel İstatistik*. Ankara: Hacettepe Üniv. Yayınları A2
- Likert, R. (1932). The Method of Constructing an Attitude Scale. In Fishbein, M. (Ed) (1967). *Readings in Attitude Theory and Measurement*. New York: John Willey & Sons, Inc., pp 90 - 95.
- McIver, J. P., & Carmines, E. G. (1982). *Unidimensional Scaling*. Sage University Paper Series on Quantitative Application in the Social Sciences, series no. 07-024, Beverly Hills and London: Sage Pub.

- Microsoft Corp. (1993). *Microsoft Excel for Apple Macintosh*. USA.
- O'Neal, M. R., & Chissom, B. S. (1993). *A Comparison of Three Methods for Assessing Attitudes*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 10-12, 1993). US.; Alabama
- Oppenheim, A. N. (1979). *Questionnaire Design and Attitude Measurement*. London: Heinemann
- Sommer, R., & Sommer, B. (2002). *A Practical Guide to Behavioral Research: Tools and Techniques*. Oxford: Oxford University Press, Inc. Fifth Ed.
- Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: John Wiley & Sons, Inc.
- Turgut, M. F., & Baykul, Y. (1992). *Ölçekleme Teknikleri*. Ankara: ÖSYM Yayınları



## *Summary*

# A Comparison of Conventional Item Analysis Techniques to Construct Likert Type Scales

A. Ata Tezbaşaran\*  
Hacettepe Üniversitesi

In this study two criteria and three techniques related with item selection procedures for construction of Likert Type scales were investigated. One of these criteria was item selection via fixed significance level of item statistics ( $p \leq 0.01$ ), and the other criterion was selection via fixed number of items ( $K = 20$  items). Three techniques for item selection procedures were examined under these criteria. The first technique depends on calculation of correlation coefficients between item scores and scale scores. The second technique depends on calculation of correlation coefficients between item scores and scale scores obtained with the related item score excluded. The third technique depends on significance of difference between averages of upper and lower group item scores. Five separate attitude scales were prepared for investigative purposes. The first, second and third attitude scales have 60 items and was applied 100 respondents. The fourth attitude scale has 50 items and was applied 100 respondents and the fifth attitude scale has 60 items and was applied 120 respondents.

Depending on the criterion of fixed significance of statistics ( $p \leq 0.01$ ), three techniques were used to select items for final scale. The amounts of selected items by using these techniques are given below in the order of from first scale to fifth. "The technique of item-total correlation coefficients obtained with the related item score excluded (RmpH)" provided 26 of 60, 26 of 60, 41 of 60, 50 of 50, 53 of 60 selectable items. "The technique of

significance of difference between averages of upper and lower group item scores (t-statistics)" provided 29 of 60, 28 of 60, 42 of 60, 49 of 50, 50 of 60 selectable items. "The technique of item-total correlation coefficients obtained with the related item score included (RmpD)" provided 39 of 60, 30 of 60, 47 of 60, 50 of 50, 53 of 60 selectable items.

These results indicate that last two scales have more homogeneous items and the selection rate is more than 80 percent for each scale. First three scales have more heterogeneous items than the last two and there is a cue item selection techniques. It seems that there is a selection pattern depends on amount of selected items:  $RmpH < t\text{-statistics} < RmpD$ . Maximum number of items have been selected by using "the technique of item-total correlation coefficients obtained with the related item score included (RmpD)" because of item scores exaggerates the covariance between item and total scores. These results indicated that using "the technique of item-total correlation coefficients obtained with the related item score excluded" depends on "the criterion of fixed significance level" provides equivalent reliability with less amount of items.

Over all results indicated that using fixed number of items criterion provided equivalent reliability with less amount of items. Reliability coefficients (Cronbach  $\alpha$ ) of obtained final scales via three item selection techniques were not significantly different.

---

\* Address for Correspondence: A. Ata Tezbaşaran, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Beytepe, Ankara, Turkey.  
E-mail: ata@hacettepe.edu.tr