

T.C.
Mersin Üniversitesi
Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

TEKRARLI ÖLÇÜMLERDEN ELDE EDİLECEK VERİLERLE ÇİFT TUTARLIK
İNDEKSİNİN İNCELENMESİ

Ülker Sonsuzay GÖKHAN KUZU

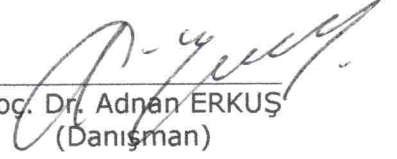
Tez Danışmanı: Doç. Dr. Adnan ERKUŞ

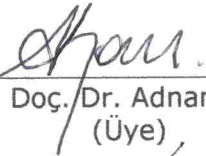
YÜKSEK LİSANS TEZİ

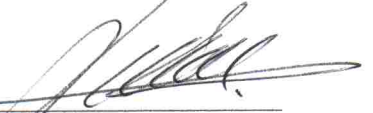
Mersin, 2007

Mersin Üniversitesi Sosyal Bilimler Enstitüsü Müdürlüğüne,

Bu çalışma, jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.


Doc. Dr. Adnan ERKUŞ
(Danışman)

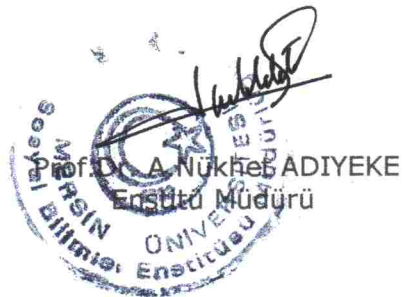

Yrd. Doç. Dr. Adnan KAN
(Üye)


Yrd. Doç. Dr. Leyla ONAT
(Üye)

ONAY

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduklarını onaylım.

27.06/2007



ÖNSÖZ

Yüksek lisansa başlamak ve mezun olmak benim için hayaldi. Bu hayali kurarak başvurmuştum bu çok sevdiğim bölüme. Kabul edilmek ise inanılmazdı. Çok çetin geçen bir eğitim süreci olacağı belliydi, ancak hesaplanmamış durumlar vardı ve hayat devam ediyordu. Ders sürecinde umudumu kaybettiğim zamanlar çok oldu, ancak bu yola beraber baş koyduğum devre arkadaşlarım Uzm. Sibel Bozdağ, Uzm. Fahriye Yergin ve A. Servet Şahin ile birlikte umutsuzluğun inadına yürüdük. Ders aşaması bittiğinde asıl zor olanın tez çalışması olduğunu anlamak uzun sürmedi. Eğer bu önsözü şuan okuyorsanız bu zafer kazanılmış demektir.

Tezimin tüm aşamalarında desteğini esirgemeyen, en umutsuz anımda “Umudunu kaybetme” diyen ve her “olmamış” dediğinde üzülmemeye üzülüp, omzunda ağlamama izin veren değerli hocam ve danışmanım Doç. Dr. Adnan Erkuş’a çok teşekkür ediyorum. Sizin desteğiniz olmadan başaramazdım hocam.

Bu çalışmaya dayanak olan dersleri alarak yetişmemde büyük emeği geçen, başta Doç. Dr. Adnan Erkuş’a, Ana Bilim Dalı Başkanım Prof Dr. Ata Tezbaşaran’a, Yrd. Doç. Dr. Devrim Alıcı’ya, Tez İzleme Komitesinde bulunan Sayın Yrd. Doç. Dr. Adnan Kan ve önerileriyle katkı sağlayan Sayın Yrd. Doç. Dr. Leyla Onat’a, her türlü desteğinden dolayı Araştırma Görevlisi Önder Sünbül’e özellikle teşekkür ediyorum.

Ölçekleri uygulamamda yardımcı olan şuan isimlerini tek tek sayamayacağım akraba ve arkadaşlarıma, sabırla 2 kez yanıtlayan tüm deneklere, çevirilerde yardımcı olan biricik ablam Şanlıay Şahin’e, 4 yıl boyunca süren zahmetli çalışmalarımı sabırla hoş gören tüm Kuzu ailesine, eğitimim için ellerinden geleni yapan ve hiçbir zaman desteğini esirgemeyen canım anneme ve babama şükranlarımı bir borç bilirim.

Bu çalışmayı, tüm yüksek lisans eğitimim boyunca yanımda olan, bana olan güvenini asla kaybetmeyen, her anlamda ve her alanda destek olan değerli eşim Erkan Kuzu’ya ithaf ediyorum.

Ülker Sonsuzay GÖKHAN KUZU

ÖZET

Bu arařtırmada, bir ölçme aracının ne denli tutarlı sınıflama yaptığını belirlemede alternatif yöntemler bulunması gerekliliđi üzerine, Erkuř (1999) tarafından geliştirilen Çift Tutarlık İndeksinin (P_{CT}), aynı ölçęin 1-0 veya çoklu uygulanması arasında nasıl sonuçlar alacađı ve tekrarlı ölçümlerle yapılan irdelemelerin testleri yarıya bölmeyle aynı sonuçlar verip vermeyeceğinin irdelenmesi amaçlanmıřtır.

Bunun için, Mersin Üniversitesi Eğitimde Ölçme ve Deęerlendirme Bilim Dalında bir grup arařtırmacı tarafından geliştirilmiř, henüz yayınlanmamıř, 4 dereceli Likert tipi bir ölçek olan Sorumluluk Ölçeđi kullanılmıřtır. Arařtırmada, P_{CT} indeksinin aynı ölçęin 1-0 ve çoklu puanlanan iki formundaki işleyiři inceleneyeceđi için, ölçęin tepki kategorileri 1-0 puanlanacak şekilde dönüřtürülmüřtür.

1-0'lı ve çoklu puanlanan ölçęin, geleneksel düzeltilmeli P_{CT} ve tekrarlı P_{CT} deęerlerinin içtutarlık katsayılarına oldukça yakın olduđu için güvenilirlik, verilecek kararların tutarlıđına baktıđı için ise geçerlik belirleme yöntemi olduđu sonucuna ulařılmıřtır.

Tekrarlı P_{CT} , sınıflama ve sıralamanın zaman içindeki kararlılıđı; geleneksel P_{CT} ise, yarıya bölmeden dolayı ölçęin içtutarlıđı bilgisini kapsamaktadır. Bu bakımdan, her iki yöntemle elde edilen katsayıların amaca bađlı olarak kullanılması sonucuna ulařılmıřtır.

Çoklu puanlamanın 1-0 puanlamaya göre daha üst ölçme düzeyine karşılık gelmesinden dolayı, bireylerin sınıflanmasının ve sıralanmasının çoklu ölçekte daha duyarlı hale geldiđi gözlenmiřtir.

Anahtar kelimeler: *Eğitimde Ölçme ve Deęerlendirme, Çift Tutarlık İndeksi, Tekrarlı Ölçümler*

INVESTIGATION OF DOUBLE CONSISTENCY INDEX WITH DATA OBTAINED FROM MULTIPLE MEASURES

ABSTRACT

In this study, the aim is to investigate a scale of measurement to determine consistency of classification and the need to find alternative methods Double Consistency Index (P_{DC}), which is improved by Erkuş (1999) the outcome of the same scale when applied between 1-0 or polytomous and the investigation of repeated measurements if they give the same result with split half.

For this study we used a 4 degree Likert type Responsibility Scale which is improved by a group of researcher in Mersin University Educational Measurement and Evaluation Science Section and not yet published. In the study, because we will investigate the working of P_{DC} index in two forms of same scale, 1-0 and polytomous, we transformed the reaction categories of the scales so that they can be scored 1-0.

We concluded that; because 1-0 and polytomous scored measures were very close to internal consistency coefficients of conventional corrected P_{DC} and repeated P_{DC} values; it is reliability and because it shows the consistency of decision; it is validity.

Repeated P_{DC} shows stability of classification and ordering in time; conventional P_{DC} involves information about internal consistency because of split half. Thus, we concluded that the coefficients that are obtained by both method must be used do to our purpose.

Because polytomous scale is giving a better measurement when considered with 1-0 scales we observed that classification and ordering of an individual will be more sensitive in polytomous scale.

Key words: *Educational Measurement and Evaluation, Double Consistency Index, Repeated Measurement*

İÇİNDEKİLER

Önsöz	i
Özet	ii
Abstract	iii
İçindekiler	iv
Tablolar	vii
Grafik	viii
GİRİŞ	1
I.1. Ölçme Araçlarının Güvenirliği	3
I.1.1. Güvenirlik Hesaplama Yöntemleri	5
a. İki Uygulama Gerektiren Yöntemler	5
<i>i. Eşdeğer Formlar Yöntemi</i>	5
<i>ii. Test Tekrar Test Yöntemi</i>	5
<i>iii. Karma Yöntem</i>	5
b. Tek Uygulama Gerektiren Yöntemler	6
<i>i. Yarıya Bölme Yöntemi</i>	6
<i>ii. Madde Kovaryanslarına Dayanan Yöntemler</i>	6
I.1.2. Ölçmenin Standart Hatası	7
I.2. Ölçme Araçlarının Geçerliği	8
I.2.1. Ölçme Geçerliği	10
a. Kapsam Geçerliği	10
b. Yapı Geçerliği	12
c. Ölçüt Bağıntılı Geçerlik.....	12
<i>i. Yordama Geçerliği (Predictive Validity)</i>	13

ii. <i>Zamandaş Geçerlik (Concurrent Validity)</i>	14
I.2.2. Karar Geçerliđi	14
a. Sınıflama Geçerliđi.....	15
b. Sıralama Geçerliđi	16
c. Sınıflama ve Sıralama Geçerliđi.....	17
I.3. İlgili Arařtırmalar	20
I.4. Arařtırmanın Amacı	22
I.5. Sayıtlar	24
I.6. Sınırlılıklar	24
BÖLÜM II. YÖNTEM	25
II.1. Arařtırmanın Türü	25
II.2. Verilerin Elde Edildiđi Grup	25
II.3. Veri Toplama Aracı	26
II.4. Verilerin Analizi	27
BÖLÜM III. BULGULAR	29
III.1. Sorumluluk Ölçeđinin Betimsel İstatistikleri	29
III.2. 1-0 Puanlanan Sorumluluk Ölçeđinin Geleneksel ve Tekrarlı P_{ÇT}'sine İliřkin Bulgular	31
a) Geleneksel Uygulamaya Dayanarak Hesaplanan P _{ÇT} Deđerleri	31
b) Tekrarlı Uygulamaya Dayanarak Hesaplanan P _{ÇT} Deđerleri	32
c) Geleneksel ve Tekrarlı Uygulamalardan Elde Edilen P _{ÇT} Deđerlerinin Karřılařtırılması.....	34

III.3. Çoklu Puanlanan Sorumluluk Ölçeğinin Geleneksel ve Tekrarlı P_{CT}'sine İlişkin Bulgular	35
a) Geleneksel Uygulamaya Dayanarak Hesaplanan P _{CT} Değerleri	35
b) Tekrarlı Uygulamaya Dayanarak Hesaplanan P _{CT} Değerleri	35
c) Geleneksel ve Tekrarlı Uygulamalardan Elde Edilen P _{CT} Değerlerinin Karşılaştırılması.....	37
III.4. 1-0 ve Çoklu Puanlanan Geleneksel ve Tekrarlı Hesaplanan P_{CT} Değerlerinin Geleneksel Uyum Katsayısı (P_o) ile Karşılaştırılması	38
BÖLÜM IV. TARTIŞMA VE YORUM	40
IV.1. 1-0 Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerlerinin İrdelenmesi	41
IV.2. Çoklu Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerlerinin İrdelenmesi	42
IV.3. 1-0 – Çoklu Puanlanan ve Geleneksel - Tekrarlı Hesaplanan P_{CT} Değerlerinin Geleneksel Uyum Katsayısı (P_o) ile Karşılaştırılması	43
IV.4. Geleneksel ve Tekrarlı P_{CT} Değerlerinin Çeşitli Katsayılar İle Karşılaştırılması	44
IV.5. Aynı Örneklem Üzerindeki Tekrarlı P_{CT} ile I. Ve II. Uygulamalardaki Geleneksel P_{CT} Değerlerinin Karşılaştırılması	45
SONUÇLAR VE ÖNERİLER	48
Sonuçlar	48
Öneriler	49
KAYNAKÇA	50
EKLER	54

TABLOLAR

Tablo-1	Erkuş (1999) – Bu Çalışma Farklılıkları	23
Tablo-2	Verilerin Elde Edildiği Gruba İlişkin Veriler	26
Tablo-3	SÖ'nin 1-0 ve Çoklu Uygulamalarından Elde Edilen Betimsel İstatistikleri	29
Tablo-4	SÖ'nin 1-0 ve Çoklu Uygulamalarından Elde Edilen Kolmogorov – Smirnov z Testi Sonuçları	31
Tablo-5	1-0 Puanlanan Ölçeğin Geleneksel P_{CT} Değerleri ve Diğer İçtutarlık Katsayıları	32
Tablo-6	1-0 Puanlanan Ölçeğin Tekrarlı P_{CT} Değerleri ve Diğer İçtutarlık Katsayıları	33
Tablo-7	1-0 Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerleri	34
Tablo-8	Çoklu Puanlanan Ölçeğin Geleneksel P_{CT} Değerleri ve Diğer İçtutarlık Katsayıları	35
Tablo-9	Çoklu Puanlanan Ölçeğin Tekrarlı P_{CT} Değerleri ve Diğer İçtutarlık Katsayıları	36
Tablo-10	Çoklu Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerleri	37
Tablo-11	1-0 ve Çoklu Geleneksel Uyum Katsayısı (P_o) Değerleri	38
Tablo-12	1-0 ve Çoklu Geleneksel Uyum Katsayısı ve P_{CT} Değerleri	39
Tablo-13	Geleneksel ve Tekrarlı Güvenirlik Değerleri	40

GRAFİK

Grafik 1	Geleneksel ve Tekrarlı P _{ÇT} Değerleri	41
-----------------	--	----

GİRİŞ

Bireylerin herhangi bir özelliğın bireyde bulunuş miktarını ölçebilmek için çeşitli araçlardan yararlanmak gerekmektedir. Sosyal bilimlerde kullanılan bu araçların büyük çoğunluğu, psikolojik ölçme araçlarıdır. Psikolojik deęişkenler, doğrudan doğruya gözlenemeyen deęişkenlerdir. Bu nedenle ölçülmek istenilen nitelikle ilişkili olduęu bilinen ve büyüklüğü gözlenebilen dięer bir nicelik ölçülür ve aralarındaki istatistiksel ilişkiye bakılarak psikolojik ölçme aracı veya birey hakkında karar verilir. Psikolojik testler, belli bir özellik açısından bireysel farklılıkları belirleyebilmek ölçmek için uygulanır (Guilford, 1954: 373).

Sosyal bilimlerde kullanılan psikolojik ölçme araçları, bir kuruma eleman veya bir okula öğrenci seçme, seçilen eleman veya öğrencileri uygun programa yerleştirme, çeşitli hastalıklara tanı koyma gibi bireyler hakkında çok çeşitli ve ciddi kararlar vermek için kullanılmaktadır. Bu kararlar bireylerin tüm yaşamlarını etkileyeceğinden, ölçme araçları, psikometrik özellikleri açısından çok iyi irdelenmelidir. Aksi halde birey açısından da, seçme ve yerleştirmeyi yapan kurum açısından da yanlış kararlar verilmiş olacak, bu da çok ciddi sonuçlar doğuracaktır. Böylece psikolojik ölçme araçlarının psikometrik niteliklerini irdelleyen çalışmaların önemi de ortaya çıkmaktadır.

Verilecek kararların doğruluęu ve tutarlıęı ise, elde edilecek ölçümlerin tutarlıęına ve kararların temel alındıęı psikolojik ölçme araçlarının kararlılıęına baęlıdır. Bu kararlılık ise, ölçme araçlarının güvenilirlięi ve geçerlięi ile saęlanır. Güvenilir ve geçerli olmayan ölçme araçlarıyla verilecek kararlar da doğru ve tutarlı olmayacaktır.

Güvenirlik ve geçerlik birbiriyle yakından ilişkilidir. Ölçme aracı güvenilir olabilir ama geçerli olmayabilir. Eđer ölçme aracı geçerliyse aynı zamanda güvenilir de olmaktadır (Campbell, 1993).

Güvenirlilik ve geçerlik belirleme çalışmalarında kullanılabilir birçok yöntem ve teknik bulunmaktadır. Güvenirlilik belirleme yöntemleri, genel olarak, paralel formlar, test tekrar test, yarıya bölme ve madde kovaryansına dayanan yöntemler şeklinde ele alınabilir (Guilford, 1954; Lord ve Novick, 1968; Anastasi, 1971; Thorndike, 1982; Crocker ve Algina, 1986; Aiken, 2000). Geçerlik konusunda ise, son yıllarda geleneksel geçerlik türleri olan kapsam geçerliği, yapı geçerliği ve ölçüt bağımlı geçerliğin (Anastasi, 1971; Cronbach, 1971; Thorndike, 1982; Crocker ve Algina, 1986) irdelendiği görülmekte; yeni sınıflamalar gündeme getirilmektedir (Murphy ve Davidshopher, 2001). İlerde daha geniş bir şekilde ele alınacağı gibi, Murphy ve Davidshopher (2001) ile Erkuş (2003 ve 2004a) geçerlik belirleme yöntemlerini ölçme geçerliği ve karar geçerliği olarak ikiye ayırmış; ölçme aracının, ölçme geçerliği kadar karar geçerliğinin de sağlanması gerektiğini ileri sürmüşlerdir. Buna göre, psikolojik değişkenleri ölçmede kullanılan araçların geleneksel yollarla geçerlikleri gösterilmiş olsa da, geliştirilmiş olan ölçme aracına dayanarak çok çeşitli kararlar verilmektedir ve bu nedenle verilecek kararların da ne denli geçerli olduğunun irdelenmesi (Murphy ve Davidshopher, 2001; Erkuş, 2003 ve 2004a) son derece önemli olmaktadır.

Hangi psikolojik yapı ölçülürse ölçülsün, ölçme sonuçlarına dayanarak çoğunlukla sınıflama düzeyinde kararlar verilmektedir. Eğitimde kullanılan bir ölçekse geçer/kalır; iş yerine eleman seçmede kullanılan bir ölçekse kabul/red; herhangi bir tutuma ait bilgi toplama amacıyla kullanılan bir ölçekse olumlu/olumsuz tutum; tanı koyma amacıyla kullanılan bir ölçekse hasta/hasta değil kararı verilmekte ve bireylerin bundan sonraki yönlendirmesi bu karara bağlı olarak yapılmaktadır. Bu kararların verilmesi ise, ister istemez, o kararın dayanacağı bir ölçütün belirlenmesini gerektirmektedir. Herhangi bir ölçme sonucu, bir ölçüte göre ele alındığında ve buna göre bir yargıya varıldığında

anlam kazanır (Turgut ve Baykul, 1992) ve bu yargılara göre bireyler hakkında sınıflama veya sıralama kararları verilir. Eğer elde edilen sonuçlar güvenilir ve geçerli ise, testi yanıtlayan hakkında özel kararlar vermek mümkün olacaktır.

Literatürde sınıflama tutarlığına ilişkin çeşitli indeksler geliştirilmiş olmasına rağmen (Berk, 1980), bu indekslerin sınıflamada kullanılan kesme puanının dağılımdaki yerine göre radikal değişiklikler göstermesi (Erkuş, 1999 ve 2000a) hep eleştirilmiştir. Erkuş (1999, 2004b ve 2005) tarafından, bu sınırlılığı gidermek için sabit %27'lik kesme alanlarına dayanan Çift Tutarlık İndeksi (P_{CT}) geliştirilmiştir. Bu indeks, ölçek puanları sıralandıktan sonra puanların iki eşdeğer yarıya bölünüp, her iki yarıdaki %27'lik alt ve üst alanlara bireylerin tutarlı sınıflanıp sınıflanmadığı rasyoneline dayanmaktadır. 1-0 ve çoklu, fakat ayrı iki ölçek ile indeksin işleyişi irdelenmiş (Erkuş, 1999,2004b ve 2005), aynı özelliği ölçen bir tek ölçekle indeksin işleyişi bugüne kadar ele alınmamıştır. Bu çalışmada, hem tek uygulamaya dayalı eşdeğer yarıya bölme yöntemi ile aynı ölçme aracının tekrarlı ölçümlerindeki, hem de aynı ölçeğin 1-0 ve çoklu puanlandığı durumdaki P_{CT} değerleri karşılaştırmalı olarak incelenecektir.

Aşağıdaki kısımlarda, ölçme araçlarının önce güvenilirlikleri, sonra da, kararların geçerliğini de kapsayacak şekilde geçerlikleri ve daha sonra da P_{CT} indeksi ve ilgili çalışmalar ele alınmıştır.

1.1. Ölçme Araçlarının Güvenirliği

Lord ve Novick'e (1968) göre, güvenirlilik genel olarak, ölçmenin doğruluğunun (eşdeğerlik ve kararlılık) göstergesidir. Crocker ve Algina'ya (1986) göre güvenirlilik, test puanlarının mükemmel tutarlılığı veya üretilebilirliğidir. Anastasi'ye (1971) göre güvenirlilik, farklı durumlarda, aynı bireylere aynı testin tekrarlı

uygulanmasıyla elde edilen puanların tutarlıdır. Kline (1986) ise güvenilirliği, testin içsel olarak tutarlılığı olarak tanımlamaktadır.

Klasik güvenilirlik anlayışı Sperman'ın, "test puanı, gerçek puan ve seçkisiz (random) hata olarak iki bileşenin toplamından elde edilir" görüşüne dayanmaktadır (Gulliksen, 1950). Gözlenen puanın, gerçek puan ile hatalar toplamına eşit olması modelin temel sayılılarından biridir (Thorndike, 1982: 144). Eğer elde edilen test puanından ölçmenin hatası temizlenirse, gerçek ölçme, yani gerçek puan elde edilecektir (Lord ve Novick, 1968: 27).

$$X_{\text{gözlenen}} = X_{\text{gerçek}} + X_{\text{hata}} \quad \dots \text{ (Eşitlik 1)}$$

Ölçme sonucuna karışabilecek hatalar, sabit, sistematik ve seçkisiz (random) hatalardır. Sabit ve sistematik hatalar, güvenilirliğin konusu değildir. Seçkisiz hatalar, kaynağı, miktarı ve yönü bilinmeyen hatalardır. Önemli olan hata puanlarını en aza indirerek gözlenen puanı gerçek puana yakınlştırabilmektir. Gerçek puan teoriktir gözlenemez, ancak gerçek puanın kestirimi, aynı ölçme aracının tekrarlanmasından elde edilen ölçmelerden elde edilecek gözlenen puanların ortalama değeri olarak tanımlanabilir (Thorndike, 1982: 144).

Campbell'e (1993) göre, tüm güvenilirlik kestirimleri klasik gerçek puan modeline dayanmaktadır. Klasik Test Kuramında güvenilirlik, ölçeğin geliştirildiği gruba bağımlıdır; bu nedenle ölçüm yapılan grup değıştiğinde ölçeğin güvenilirlik katsayısı da değışmektedir. Bu sınırlılığı gidermeye yönelik olarak, örneklemden bağımsız kestirimler veren modeller geliştirilmeye çalışılmıştır (Rash, 1960; van der Linden ve Hambleton, 1997; Akt: Campbell, 1993).

I.1.1. Güvenirlik Hesaplama Yöntemleri

Ölçme araçlarının güvenilirliklerini belirlemenin değişik yöntemleri bulunmaktadır. Ölçme aracının özelliklerine, uygulama olanaklarına, karışabilecek hatanın veya bireylerin özelliklerine göre bu yöntemler seçilebilir. Aşağıdaki sınıflama Crocker ve Algina'dan (1986) alınmıştır.

a. İki Uygulama Gerektiren Yöntemler

i. Eşdeğer Formlar Yöntemi: Bu yöntemde, ölçme aracına eşdeğer bir form geliştirilerek, bu iki eşdeğer form aynı deneklere aynı koşullarda uygulanır ve iki formdan alınan puanlar arasındaki korelasyon hesaplanır (Thorndike, 1982). Klasik Test Teorisinin temelini oluşturan eşdeğer formlar yönteminde, iki formun eşdeğer olabilmesi için ortalamalarının, varyanslarının ve kovaryanslarının eşit olması gerekmektedir (Gulliksen, 1950; Akt: Guilford, 1954: 374). Bu yöntemle bulunan katsayı, “eşdeğerlik katsayısı” olarak adlandırılır.

ii. Test Tekrar Test Yöntemi: Aynı testin aynı bireylere, aynı koşullarda iki kere uygulanmasına ve iki uygulama arasındaki korelasyon katsayısının hesaplanmasına dayanmaktadır (Lord ve Novick, 1968: 134). Bulunan katsayı, “kararlılık katsayısı” olarak adlandırılır (Guilford, 1954: 374). İki uygulama arası zamanın ne kadar olacağı, ölçülen özelliğin dinamikliğine bağlıdır.

iii. Karma Yöntem: Bu yöntem, eşdeğer form ve test tekrar test yönteminin birlikte kullanılmasını içermektedir. İki eşdeğer formdan birincisi uygulanır, bir süre beklenir; ikinci form uygulanır (Crocker ve Algina, 1986). Bulunan güvenilirlik katsayısı, iki yöntemin ayrı ayrı kullanılmasından daha düşüktür.

b. Tek Uygulama Gerektiren Yöntemler

Tek uygulamayla hesaplanan güvenilirlik katsayıları, “içtutarlık katsayıları” olarak anılmaktadır.

i. Yarıya Bölme Yöntemi

Geleneksel yöntemle, testteki maddeler iki farklı yolla yarıya bölünebilmektedir. Birincisi tek ve çift numaralı maddeler şeklinde, ikincisi ise seçkisiz olarak testi yarıya bölmektir. Test iki yarıya ayrıldıktan sonra, iki yarı iki ayrı eşdeğer testmiş gibi kabul edilerek uygun korelasyon tekniğiyle aralarındaki ilişki bulunur (Guilford, 1954). Elde edilen katsayıya Spearman-Brown düzeltmesi yapılarak, tüm testin güvenilirliği elde edilmiş olur. Yarıya bölme yöntemlerinden olan Rulon, Flanagan ve Guttman yöntemleriyle ise, Spearman-Brown düzeltmesi gerektirmeden güvenilirlik hesaplanabilmektedir.

ii. Madde Kovaryanslarına Dayanan Yöntemler:

Bu yöntemlerin Nunnally (1970) tarafından, testi yarıya bölmeye dayanan yöntemlere göre daha üstün olduğu ileri sürülmektedir (Akt: Campbell, 1993). Bu tür yöntemlerle güvenilirlik kestirimi, madde homojenliğine veya aynı kapsamı ölçen test maddelerinin ilişkisine dayanmaktadır (Henson, 2001).

➤ Kuder Richardson 20 Yöntemi: Kuder-Richardson’un geliştirmiş olduğu birçok güvenilirlik formülü bulunmaktadır; ancak bunlardan en sık kullanılanı KR-20 yöntemidir. KR-20, testteki her maddenin aynı özelliği ölçtüğü sayılımasına dayanmakta; eğer testteki madde sayısı, her maddenin güçlüğü ve testin varyansı biliniyorsa, test güvenilirliği KR-20 ile hesaplanabilmektedir (Gulliksen, 1950). KR-20, süreksiz olduğu bilinen, 1-0 puanlanan maddelerden oluşan ölçeklere uygulanmaktadır.

Kuder-Richardson 20 formülü (Lord ve Novick, 1968);

$$\alpha_{20} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{\sigma_x^2} \right) \quad (\text{Eşitlik 2})$$

➤ Cronbach Alpha Yöntemi: Bu yöntem, ağırlıklı olarak puanlandırılmış maddeli testler için Cronbach (1951) tarafından geliştirilmiştir. Cronbach α , sürekli veya sürekli olduğu varsayılan çoklu puanlanan (ağırlıklı) maddelerden oluşan ölçeklere uygulanır ve bütün maddelerin birbiriyle uyumuna bakılması esasına dayanır.

Cronbach Alpha formülü (Cronbach, 1951; Thorndike, 1982: 176);

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right) \quad (\text{Eşitlik 3})$$

I.1.2. Ölçmenin Standart Hatası

Güvenirlik hatasızlığın ölçüsüdür, test puanlarına karışan hatanın ölçüsü ise test puanları cinsinden ölçmenin standart hatası ile elde edilir. Ölçmenin standart hatası, gerçek puanların yerine gözlenen puanları kullanmadaki hata olarak da tanımlanabilir (Guilford, 1954: 389). Ölçmenin standart hatası ve güvenilirlik indeksi, test güvenilirliğinin alternatif bir yoludur (Anastasi, 1971: 95).

Ölçmenin standart hatası (S_{em}), Klasik Test Teorisi'nde ölçeğin geliştirildiği gruptan elde edilen güvenilirlik katsayısı ile standart sapmaya bağlı olarak aşağıdaki formülü ile verilebilir.

$$S_{em} = S_x \sqrt{1 - r_{XX}} \quad (\text{Eşitlik 4})$$

Klasik Test Teorisinde ölçmenin standart hatası her birey için aynıdır. Modern test teorisinde ise, standart hata, her yetenek düzeyine göre bireyden bireye değişebilmektedir.

Bir ölçme aracının güvenilirliğinin yanında, S_{em} 'in de rapor edilmesi, bireylerin puanlarının gerçek sınırlarının kestirilmesinde yarar sağlayacaktır.

I.2. Ölçme Araçlarının Geçerliği

Anastasi'ye (1971) göre testin geçerliği, testin neyi ölçtüğü ve bunu doğru yapıp yapmadığıyla ilişkilidir. İlk olarak, L. J. Cronbach öncülüğünde, APA tarafından (1954) geçerliği irdelemeye ilişkin olarak getirilmiş olan teknik öneriler, geçerliği 4 farklı grup altında ele almıştır: Kapsam geçerliği, Yapı geçerliği, Yordama geçerliği (Predictive Validity) ve Zamandaş geçerlik (Concurrent Validity). Bu geçerlik tiplerinin yıllarca farklı amaçlar için kullanıldığı düşünülmüş, ancak zamanla bunların test puanları temelinde yatan 4 farklı geçerlik stratejisi olduğu anlaşılmıştır (Murphy ve Davidshofer, 2001).

Geçerlik, bir ölçme aracının geliştirilme ve kullanılma amacına hizmet etme derecesidir (Erkuş, 2006). Bu tanım, “amaca hizmet etme derecesi” olarak tanımlanan geçerliği ele alırken, karşımıza iki tür amacı çıkarmaktadır: 1) Testin, “ölçmeyi” amaçladığı şeyi ölçüp ölçmediği ve 2) testin, uygulandığı bireyler hakkında verilecek kararların doğruluğu (Murphy ve Davidshofer, 2001: 74; Erkuş, 2004a). Amacın bu şekilde ayrıştırılmasına göre, örneğin, depresyonu ölçmek amacıyla geliştirilen bir ölçeğin “ölçmeyi amaçladığı (depresyonu) özelliği” ne kadar geçerli bir şekilde ölçtüğünün öncelikle gösterilmesi, sonra da bu ölçekten alınan puanlara göre bireylerin hakkında verilecek kararların (yüksek depresyonlu/düşük depresyonlu gibi) ne derece doğru olduğunun gösterilmesi gerekmektedir. Bu durum, aslında “ölçme” ile “değerleme (assessment)” arasındaki farkı ön plana çıkarmaktadır. Psikolojik ölçme araçlarının asıl

amacı, test puanlarına dayanarak bireyler hakkında çeşitli kararlar vermek, o puana bir değer biçmek yani bir değerlemede kullanmaktır: Geçer/kalır, depresyonlu/değil, kabul/red gibi. Bu açıdan bakıldığında, ölçme aracının “ölçmeyi amaçladığı özelliği” ölçüp ölçmediğinin gösterilmesi, “ölçme” boyutundaki bir işlemdir; bu gösterildikten sonra “verilecek kararların doğruluğunun” saptanması ise değerlendirme, yani “karar verme” boyutundaki bir işlemdir. Bu iki işlem ise geçerliğin, “Ölçme Geçerliği” ve “Karar (veya değerlendirme) Geçerliği” şeklinde incelenmesine işaret etmektedir.

Murphy ve Davidshofer’a (2001) göre, geleneksel geçerlik türlerinden kapsam ve yapı geçerliği, ölçme geçerliği için, yordama ve zamandaş geçerlik, karar geçerliği için kanıt toplamada kullanılır. Erkuş’a (2003) göre ise bu geleneksel dört farklı geçerlik stratejisinin dördü de ölçme geçerliği başlığı altında ele alınmaktadır. Ancak bunlardan yordama geçerliği hem ölçeğin ölçmeyi amaçladığı şeyi ölçüp ölçmediğine (ölçme geçerliği) hem de ileriye dönük yapılan kararların tutarlı yapılıp yapılmadığına (karar geçerliği) bir kanıt oluşturduğundan; yordama geçerliğini her iki geçerlik grubu arasına sınıflamak daha uygun görünmektedir (Erkuş, 2004a).

Yine Erkuş’a (2004a ve 2005) göre, Murphy ve Davidshofer’ın (2001) karar geçerliği sınıflamasının altında yer alan geçerlik türleri (kısmen yordama geçerliği) işlevsel değildir. Herhangi bir eğitim kararı, zaten güvenilir ve geçerli olan bir ölçme aracından elde edilen bir ölçme sonucunun uygun bir ölçüt ya da ölçüt takımına dayanarak verilir; verilmesi gerekir. Ölçme sonuçlarına dayanarak, bireyler hakkında “kurumsal” amaçlara bağlı olarak seçme, yerleştirme, tanı koyma, rehberlik gibi çeşitli kararlar verilmektedir; oysa, hangi kurumsal karar olursa olsun, aslında ölçme sonuçlarına dayanarak “istatistiksel” olarak ya sınıflama, ya sıralama ya da her iki karar birden verilmektedir. Dolayısıyla Karar Geçerliği altında, bu istatistiksel kararlara göre, Sınıflama Geçerliği,

Sıralama Geçerliđi ile Sınıflama ve Sıralama Geçerliđinin yer almasının daha uygun olacađı ileri sürülebilir. Bu bakımdan, Murphy ve Davidshofer'ın (2001) ana geçerlik sınıflamasının; dođru olmakla birlikte, "karar"ın "deđerleme" (assessment) ile eř bir şekilde ele alınmamasından kaynaklanan bir eksiklik tařıdıđı; sadece varolan ölçüt bađıntılı geçerliđin karar geçerliđi altında ele alınmasının bu sınıflamayı işlevsiz kıldıđı ileri sürülebilir. Sınıflamanın, karar geçerliđi kategorisi, deđerleme türlerine dayanarak yeniden tanımlandıđında sorun çözülr görünmektedir.

Ařađıdaki kısımlarda geçerlik, Erkuř'un (2004a ve 2005) sınıflaması dikkate alınarak irdelenecektir. Bu sınıflamanın temel alınmasının bir nedeni de bu çalıřmada Erkuř'un (1999) geliřtirdiđi "Çift Tutarlık İndeksi (P_{CT})"nin incelenecek olmasıdır.

I.2.1. Ölçme Geçerliđi

Murphy ve Davidshofer'a (2001) göre ölçme geçerliđi stratejileri, ölçülen nitelik hakkında ölçülmek istenilen kapsamı ve testin ölçmeyi amaçladıđı yapıyı ölçüp ölçmediđini içermektedir.

a. Kapsam Geçerliđi:

Anastasi'ye (1976) göre, "psikolojik test, belli bir özelliđe ait davranıř örnekleminin standart kořullar altında objektif ölçümüdür". Buna göre, psikolojik ölçme araçları, ölçülen özelliđin davranıř örneklemini yansıtmak durumundadır. Bu nedenle, ölçme aracını oluřturan maddelerin bu örneklemini (kapsamı) yansıtıp yansıtmadıđının irdelenmesi, kapsam geçerliđi hakkındadır.

Murphy ve Davidshofer'a (2001) göre, ölçülmek istenen özelliđe iliřkin davranıřları temsil eden tüm maddeler testte mevcutsa, kapsam geçerliđi sađlanmış demektir. Ancak detaylı kanıt için kapsam alanlarına bakmak gerekmektedir. Bir kapsam

alanı, ölçülmek istenen belirli bir özelliğin ölçümünde kullanılan, özelliğe ilişkin davranış şekillerini temsil eder.

Cronbach'a (1971) göre ise, kapsam geçerleme, testin kapsamının yeterli yargısıdır ve kapsamın yeterliği, davranış evreninin uygun tanımlanmasıyla elde edilmektedir. Ölçme aracını oluşturan maddelerin, ölçülmek istenilen davranış alanını yansıtır özellikte olması gerekmektedir.

Kapsam geçerliğini değerlendirmek için, kapsam alanının sınırlarının ve yapısının tanımlanması gerekmektedir. Kapsam geçerliği, bir testin, belli bir kapsam alanını temsil edebilme derecesi hakkında verilen yargıyı temsil etmektedir (Murphy ve Davidshofer, 2001). Kapsam geçerliği, test maddelerinin, ilgilenilen yapıyı uygun bir şekilde temsil edip etmediğinin değerlendirilmesidir (Erkuş, 2003). Kapsam geçerliği, başarı testlerinin değerlendirilmesinde yaygın olarak kullanılmaktadır (Anastasi, 1971; Crocker ve Algina, 1986).

Kapsam geçerliği, genellikle mantıksal yollarla irdelenir (Lord ve Nocivk, 1968; Crocker ve Algina, 1986); ancak genellikle yargıcılara dayanarak da olsa, görgül ve istatistiksel olarak çeşitli yollara başvurmak (Erkuş, 2006) olanaklıdır. Aslında madde örnekleme işlemleri, ölçülmesi amaçlanan yapının kavramsal – kuramsal tanımına dayanır ve bu yapının göstergelerinin bulunmasıyla devam eder. Bu göstergelerin davranışa dönüştürülmesi ise, madde-tepki formatıyla gerçekleştirilir. Bu bakımdan, kapsamın geçerli olmasının önleminin alınması bu aşamada, eğer davranışlar bir eğitimin hedeflerinde olduğu gibi belirginse “belirtke tablosu” oluşturularak; yine davranışlar örtük değil de, bir motorun parçalarının takılması gibi bir işteki sergilenen performans ise, “iş analizi” yapılarak gerçekleştirilir.

b. Yapı Geçerliđi

Yapı geçerliđini tanımlamak için öncelikle psikolojik yapıların doğası bilinmelidir. Psikolojik özellikler, fiziksel özelliklerden farklı olarak doğrudan gözlenemez, psikolojik uzaydaki konumları bilinemez. Bir yapı doğrudan gözle görülemez de gözle görülebilir davranışlardan vardanabilir. Doğrudan gözlenemeyen yapılar, bireylerin test maddelerine verdikleri tepkilerden dolayı olarak ölçülmektedir.

Yapı geçerliđinin amacı, test puanlarının belirli bir yapı için iyi bir ölçüm sağlayıp sağlamadığını göstermektir. Yapı geçerliđi çalışması, ölçme aracının, gerçekten araştırılan söz konusu yapıyı ölçüp ölçmediğini anlamak için yapılmaktadır. Bu nedenle, aslında ölçülmesi amaçlanan yapı hakkında bir açıklama getiren bir denence ileri sürülür ve yapı geçerliđi test edilir.

Bu bakımdan, psikolojik ölçme araçları için yapılan ve yapılacak her şey yapı geçerliđine kanıt toplamayı içerir. Yapı geçerliđi için faktör analitik teknikler başta olmak üzere, yapının özelliđine göre, farklı deđişkenlere göre (yaş, sınıf, eğitim düzeyi gibi) ortaya çıkan farklılıkların incelenmesi, benzer yapılarla ilişkilerin incelenmesi gibi pek çok işlemsel ve istatistiksel yolla kanıt toplanmaya çalışılır. Geliştirilen ölçeđe dayanarak daha sonra yapılan diđer tüm görgül çalışmalar, ölçeđin yapı geçerliđine bir kanıt ya da kanıtsızlık sunarlar.

c. Ölçüt Bađıntılı Geçerlik

Psikolojik ölçme araçları ile doğrudan ölçülemeyen bireyin belli bir özelliđe ait davranışları, test puanlarıyla vardanmak istenildiğinde, ölçme aracının uygun bir ölçütle korelasyonuna bakılarak, ölçüt bađıntılı geçerlik kanıtları toplanmaktadır (Crocker ve Algina, 1986: 224). Ölçüt bađıntılı geçerlik, ölçüt ile ölçme aracı arasındaki zamansal ilişki temel alındığında, yordama geçerliđi ve zamandaş geçerlik olmak üzere ikiye

ayrılmaktadır (Anastasi, 1971: 105). Ölçüt puanlar, yordayıcıdan sonra elde edilirse yordama geçerliği, yordayıcı puanlarla aynı zamanda veya önce elde edilmişse zamandaş geçerlik belirlenmiş olur.

i. Yordama Geçerliği (Predictive Validity)

Ölçütle ilgili puanlar gelecekteki bir zamanda elde edileceğinde, alınan kararlar, ölçme aracından elde edilen puanlardan, bireylerin beklenen gelecek performanslarının yordanmasına dayanır (Cronbach, 1971). Yordama geçerliğinin amacı, karar vermeden önce elde edilen ölçek puanları ile karar verdikten sonra elde edilen ölçüt puanları arasındaki korelasyona dayalı olarak, ölçeğin sınıflama ve sıralama geçerliğini belirlemektir (Murphy ve Davidshofer, 2001). Örneğin, üniversite sınavında alınan puanın (ÖSS), üniversite başarısını yordayacağı vardanmaktadır. Üniversite başarısı gelecekteki bir tarihte elde edildiğinde ve ÖSS’den alınan puanla, bireyin üniversite başarısı arası ilişkiye bakıldığında, ÖSS’nin yordama geçerliği kanıtları elde edilmiş olacaktır.

Yordama geçerliğinin en sorunlu yanlarından biri, ölçüt ölçülerinin uygunluğu ve aradaki zamandır; çünkü ölçek puanları ile ölçüt arası zaman arttıkça ölçütün yordama gücü azalmaktadır (Cronbach, 1971: 127). Diğer bir sorun da, ölçüt ölçüsünün daha önceden seçilmiş homojen bir grup üzerinden elde edilmesinden kaynaklanan “ranj/daralması”dır (Thorndike, 1982; Aiken, 2000).

Yordama geçerliği, testin hem ölçmeyi amaçladığı şeyi ölçüp ölçmediğine (ölçme geçerliği), hem de ileriye dönük kararların (örneğin ÖSS’ye dayanarak bireylerin üniversitedeki programlara yerleştirme) doğru yapılıp yapılmadığına (karar geçerliği) bir kanıt oluşturduğundan; yordama geçerliğini, her iki geçerlik grubu arasına sınıflamak daha uygun görünmektedir (Erkuş, 2004a: 117).

ii. *Zamandaş Geçerlik (Concurrent Validity)*

Ölçme aracı ile ölçüt puanlar eş zamanda elde edildiyse, kanıt aranan geçerlik zamandaş geçerliktir. Daha önceden seçilmiş bir kaynaktan ölçüt puanların alınmasına ve ölçme aracından elde edilen puanlarla ölçüt puanlar arası korelasyona bakılmasına dayanır (Murphy ve Davidshofer, 2001). Ortaöğretim giriş sınavında kullanılan testler için ilköğretim başarı ortalaması ölçüt olarak alındığında ya da ÖSS testi için lise başarı ortalaması ölçüt olarak alındığında aralarındaki ilişkiye bakılarak, zamandaş geçerlik kanıtları elde edilmiş olacaktır.

I.2.2. Karar Geçerliği

Ölçme aracından elde edilecek sonuçlara dayalı olarak seçme, yerleştirme, tanı koyma gibi ciddi kurumsal kararlar verilmektedir. Ölçme araçlarının, sınıflamayı ya da sıralamayı ne kadar doğrulukla yaptığının belirlenmesi, verilecek kararların doğruluğunun göstergesidir. Verilecek kararların doğruluğu, bireylerin doğru sınıflandırılmasına olanak sağlar.

Bir ölçme sonucu, tek başına bir şey ifade etmez; onun anlamlandırılması ancak bir karar verme süreciyle olur. Kullanılacak ölçüte göre değerlendirme şekli değişmektedir. Testten alınabilecek maksimum puan ölçüt olarak alınıyorsa, maksimum puana dayalı değerlendirme; geçti kararı verebilmek için minimum bir standart ölçüt olarak alınıyorsa, kesme puanına dayalı değerlendirme; bireyin içinde bulunduğu grubun değerleri ölçüt olarak alınıyorsa, norm dayanaklı değerlendirme sınıflaması yapmak mümkündür (Erkuş, 2006).

Karar geçerliği kavramı ilk kez Murphy ve Davidshofer (2001) ile Erkuş'ta (2003 ve 2004a) karşımıza çıkmaktadır. Murphy ve Davidshofer (2001) karar geçerliğini, yordama geçerliği ve zamandaş geçerlik olarak sınıflandırmıştır. Ancak Erkuş'a (2003 ve

2004a) göre, ölçme sonuçlarına dayanarak istatistiksel kararlar verilmekte, bu nedenle karar geçerliği Sınıflama geçerliği, Sıralama Geçerliği, Sınıflama ve Sıralama geçerliği olarak sınıflandırılmaktadır. Bu araştırmada irdelenecek olan Erkuş'un (2000b) önermiş olduğu Çift Tutarlık İndeksi (P_{CT}), yine Erkuş'un (2000b) önermiş olduğu Sınıflama ve Sıralama Geçerliği başlığı altında incelenecektir.

a. Sınıflama Geçerliği

Glaser'in (1963) ölçüt dayanaklı değerlendirme kavramsallaştırmasından sonra, ölçme araçlarının bir ölçüte göre ne denli tutarlı sınıfladığına ilişkin yapılan çalışmalar çerçevesinde birçok indeks geliştirilmiştir (Akt: Erkuş, 2000b: 64). Sınıflama kararlarının tutarlığına ilişkin geliştirilmiş bu indekslerin, hangi başlık altında ele alınacağı noktasında kavram karmaşası yaşanmıştır. Bu karmaşanın Erkuş'un (1999) da belirttiği gibi yeni bir kavramlaştırmayla çözülebileceği ve bahsi geçen indekslerin sınıflama geçerliği indeksi olarak sınıflandırılabilmesi açıktır. Erkuş'un (2000a) kavramlaştırmasına dayanarak sınıflama geçerliği kararı veren en çok kullanılan indekslerin gelişimi aşağıda kısaca ele alınacaktır.

Hambleton ve Novick (1973), sınıflama kararlarının tutarlığını belirlemede çok basit bir hesaplamayla, bir formun iki kez uygulanmasına ya da klasik paralel testlere dayanan, dağılım sayılıtsı gerektirmeyen karar tutarlığı indeksi olarak, "geleneksel uyum katsayısı"nı (P_o) önermişlerdir. Tutarlı sınıflama indeksi (basit uyum katsayısı) aşağıdaki şekilde hesaplanmaktadır.

$$P_o = \frac{a + d}{N} = P_a + P_d \quad (\text{Eşitlik 5})$$

P_a = 1. uygulamada da 2. uygulamada da başarılı olan birey oranı

P_b = 1. uygulamada da 2. uygulamada da başarısız olan birey oranı

P_o , bir oran olduğu için 0.00 ile 1.00 arasında yer alır; indeks değeri 1.00'e yaklaştıkça sınıflama kararlarının tutarlılığını, 0.00'a yaklaştıkça da tutarsızlığını gösterir.

Swaminathan, Hambleton ve Algina (1974) ise, geleneksel uyum katsayısı yerine, şanstın arındırılmış uyum katsayısı olan Cohen'in Kappa'sını (κ) 2x2'lik durumlar için önermiştir. κ 'yı hesaplamak için geleneksel basit uyum katsayısı (P_o) ve şansla beklenen uyumun hesaplanması gerekir. κ katsayısı aşağıdaki şekilde hesaplanmaktadır (Akt: Erkuş, 2000a).

$$\kappa = (P_o - P_{\text{şans}}) / (1 - P_{\text{şans}}) \quad \dots \text{ (Eşitlik 6)}$$

Kappa, şansla beklenenden arınık, gözlenen tutarlı sınıflamaların oranını verir.

Subkoviak (1976) ve Huynh (1976) ise, P_o ve κ 'yı tek uygulamaya dayanarak hesaplama yöntemleri geliştirmişler; Bayer ve Lewis (1994) ise, içtutarlık katsayısının hesaplanmasını gerektirmeyen ve hem 1-0 hem de çoklu puanlanan maddeli testlere uygulanabilen indeksi geliştirmişler; Livingston ve Lewis (1995) ise, testin minimum ve maksimum puanlarından hareketle hesaplanan yöntemleri geliştirmişler; Berry ve Mielke'nin (1997) önerdiği indeks ise, Spearman'ın "footrule" katsayısına dayanmaktadır (Akt: Erkuş, 2000a).

b. Sıralama Geçerliği

Test puanları açısından bireylerin birbirlerine göre sıralanmasına dayanarak da kararlar verilmektedir. Sıralama kararlarının tutarlılığının, Spearman'ın sıra farkları korelasyon tekniğiyle çözülebileceği, sonucun sıralama geçerliği hakkında kanıt sunacağı ileri sürülebilir (Erkuş, 2004a). Spearman'a dayanarak Berry ve Mielke'nin (1997), önerdiği indeksin de, Spearman'ın Footrule katsayısına dayanması nedeniyle, sıralama tutarlılığı bilgisi verildiği söylenebilir (Akt. Erkuş, 1999).

Bireylerin herhangi bir testten önce ve sonra aynı yöntemle derecelendirildiği durumlarda, yapılan iki derecelendirme arasındaki uyumun derecesi sıralama geçerliği hakkında bilgi vermektedir. Uyumun derecesini veren istatistikler ise, Spearman'ın rho (ρ)'su, Kendall'ın tau (τ)'su, Goodman ve Kruskal'ın gamma (γ)'sı, Cicchetti'nin C istatistiği ve Cohen'in Kappa (κ)'sıdır (Erkuş, 2000a).

c. Sınıflama ve Sıralama Geçerliği

Bir ölçme aracı hangi ölçek düzeyinde ölçme yaparsa yapsın, sonuçta bireyler hakkında karar vermek, bireyleri sınıflamak ve birbirlerine göre sıralamak için kullanılır. Bir programa öğrenci alınacağına öncelikle kesme puanına göre bireyler sınıflandırılır ve sonra başarılı öğrenciler almış oldukları puanlara göre sıralanırlar. Örneğin yüksek lisans programına başvurmuş öğrencilerden 65 puan alanlar geçer, ancak geçen öğrencilerden sadece ilk 5 kişi yüksek lisans programına kabul edilir. Yüksek öğrenim sınavlarında bunun tersi mevcuttur. Bireyler öncelikle almış oldukları puana göre sıralanır sonra da programa kabul edilecek birey sayısına göre sınıflanarak bireylere kabul veya red cevabı verilir.

Erkuş (1999) tarafından geliştirilen indeks, keyfiyete dayalı bir kesme puandan etkilenen indekslere alternatif olarak, kesme alanlarına ve tek uygulamaya dayanması, karmaşık işlemler gerektirmemesi ve hem 1-0 hem de çoklu puanlanan ölçeklere uygulanabilirliği nedeniyle önerilmiştir (Erkuş, 2000b).

Kesme puanı, çeşitli kararlara göre belirlenen ve minimum yeterlik düzeyini gösteren puandır. Ölçüt dayanaklı değerlendirme kapsamında geliştirilen sınıflama indeksleri, keyfiyete dayalı bir kesme puanıyla bireyleri sınıflamaktadırlar. Bu hem büyük bir bilgi kaybına yol açar hem de yanlış sınıflama olasılıklarını da beraberinde getirir. Bu indekslerin çoğu iki uygulamaya, ancak P_{CT} tek uygulamaya dayanır. İndeksin dayandığı

yöntem, bilinen birkaç yöntemin bir araya getirilmesine dayandığından anlaşılması da zorluk içermemektedir.

P_{CT} , ölçüt dayanaklı değerlendirme kapsamında geliştirilen sınıflama indeksleri gibi karmaşık hesaplama işlemleri gerektirmemektedir. Bugüne kadar geliştirilen indekslerden ikisi dışında (P_0 ve κ) tümü çok karmaşık hesaplama işlemleri gerektirir.

P_{CT} , hem 1-0 hem de çoklu puanlanan ölçüklere uygundur. İndeks, her iki yarıdaki toplam puanları dikkate aldığından, hem 1-0 hem de çoklu puanlanan, her türlü psikolojik ölçme aracına uygulanabilir.

P_{CT} , kesme puanı (minimum bir standart ölçüt) kullanmadan sınıflama yapması, tek uygulama gerektirmesi, karmaşık hesaplamalar gerektirmemesi ve her türlü puanlanan ölçüklere uygulanabilmesi nedeniyle, her türlü ölçme aracının irdelenmesinde kullanılabilir.

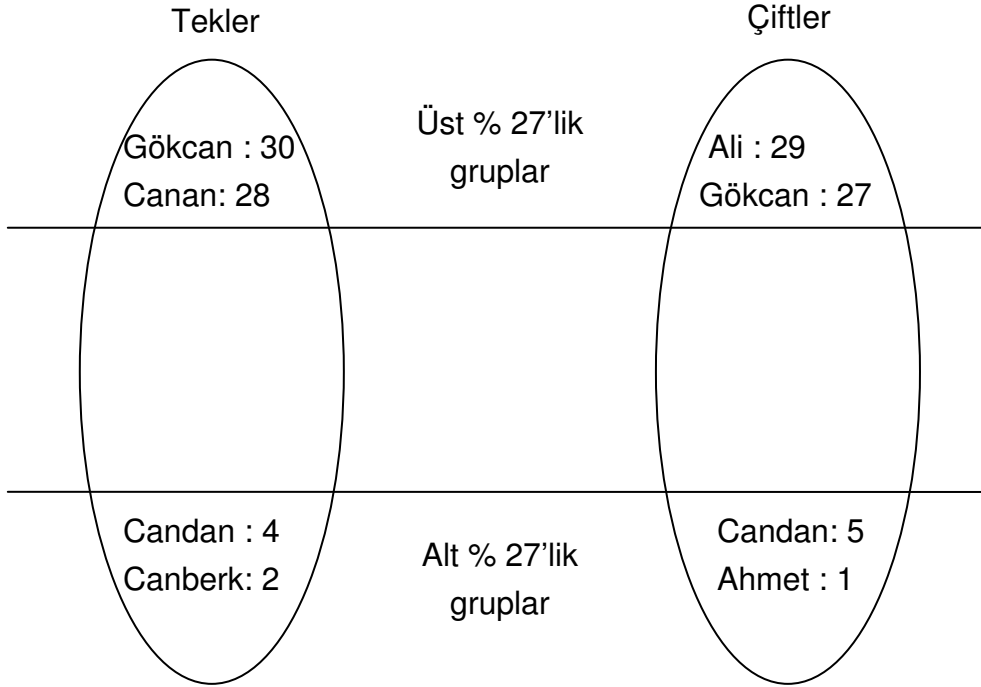
P_{CT} 'nin mantıksal temeli; testin iki eşdeğer yarıya bölünmesine ve eğer test tutarlı sınıflama yapıyorsa (bir anlamda tutarlı ayırt ediyorsa), testin ilk yarısında üst grupta (başarılı) sınıflanan bireylerin ikinci yarıda da üst grupta; ilk yarıda alt grupta (başarısız) sınıflanan bireylerin ikinci yarıda da alt grupta sınıflanmasının beklenmesine dayanmaktadır (Erkuş, 2000b).

Kesme puanını bertaraf eden bir sınıflama tutarlığı indeksine olan gereksinimden dolayı geliştirilen P_{CT} , halen psikometride kullanılmakta olan bazı yöntemlerin bir araya getirilmesine dayanmaktadır (Erkuş, 2000b).

Önerilen yöntem;

- a. Test maddelerinin tekler ve çiftler şeklinde iki eşdeğer yarıya bölünmesine,
- b. İki eşdeğer yarı testte bireylerin toplam puanlarının bulunarak büyükten küçüğe doğru sıralanmasına,

- c. Testin her iki yarısı için de ortak-sabit alt-üst %27'lik alanların belirlenmesine ve
- d. Her iki yarıda da bireylerin alt ve üst alanlarda tutarlı sınıflanıp sınıflanmadığının bakılmasına dayanmaktadır.



Eğer birey yarı testteki toplam puanlara göre, teklerde %27'lik üst grupta yer almışsa, çiftlerde de üst grupta; teklerde %27'lik alt grupta yer almışsa, çiftlerde de alt grupta sınıflanması beklenir. Eğer birey teklerde üst grupta, çiftlerde ise orta veya alt grupta yer almışsa, tutarsız sınıflama yapılmış demektir.

$P_{ÇT}$, bireylerin birinci yarı ile ikinci yarı arasındaki uyumunu dikkate aldığından, bu iki yarıdaki frekans farklarına dayanarak aşağıdaki eşitlik ile verilebilir (Erkuş, 2004b: 71):

$$P_{ÇT} = 1 - \left[\frac{(f_{ÜÇ} - f_{ÜT}) + (f_{AÇ} - f_{AT})}{N_{Ü+A}} \right] \quad \dots \text{(Eşitlik 7)}$$

$$f_{ÜÇ} = \text{Üst - Çift Frekansı} \quad f_{ÜT} = \text{Üst - Tek Frekansı}$$

$$f_{A_C} = \text{Alt} - \text{Çift Frekans} \quad f_{A_T} = \text{Alt} - \text{Tek Frekans}$$

$$N_{\bar{U}+A} = \text{Grubun \%54'üne karşılık gelen birey sayısı}$$

Formülasyonun mantığı, Hambleton ve Novick'in (1973) önerdiği geleneksel uyum indeksine benzemesine rağmen önemli farklılıklar da içerir (Erkuş, 2000b). Yeni yöntem bir yarı testin temel (anchor) alınmasına dayandığından, temel alınan yarının sabit uç birey sayılarından, ikinci yarının (ölçüt yarı) tutarlı birey sayısının çıkarılması yoluyla tutarlı sınıflama sayıları (frekans farkları) belirlenebilir. P_{CT} 'nin 0.00 ile 1.00 arasında değiştiği, 0.00'a yaklaştıkça tutarsız sınıflama, 1.00'e yaklaştıkça ise tutarlı sınıflama bilgisi verdiği, bu bakımdan da yorumunun geleneksel indekslere benzer olduğu görülmektedir.

Sağa ya da sola çarpık ve normal dağılım gösteren, hem 1-0 hem çoklu puanlanan ölçeklerden elde edilen veriler üzerinde P_{CT} değerinin değişiminin belirlenmesi ve indeksin dağılım fonksiyonuna ilişkin bulgular edinilmesi, P_{CT} için yarar sağlayacaktır.

I.3. İlgili Araştırmalar

Erkuş (1999), ölçüt dayanaklı değerlendirme bağlamında geliştirilen geleneksel sınıflama tutarlılığı indekslerinin, kesme puanlarının dağılım içindeki pozisyonlarından nasıl etkilendiğini ortaya koymak amacıyla, önerdiği alternatif indeksin bu sakıncayı ortadan kaldırıp kaldırmadığını çeşitli parametreler açısından incelemiştir.

Erkuş (1999) araştırmasında P_{CT} 'nin, çeşitli madde sayısı ve denek sayısı kombinasyonlarını, en uygun alt üst uç grupların değişimini, 1-0 ya da çoklu puanlanan ölçeklerdeki değişimini ve indeksin bazı indekslerle olan ilişkisini incelemiştir. Araştırmada, en ayırt edici olarak %27'lik grup olduğu; P_{CT} 'nin, birey sayısı azaldıkça

çoklu puanlanan ölçekte arttığı, 1-0 puanlanan ölçekte azaldığı; madde sayısı azaldıkça ise, P_{CT} 'nin hem 1-0 puanlanan hem de çoklu puanlanan ölçekte azaldığı saptanmıştır.

İlgili çalışmada indeksin türetilmesi ve uygulaması irdelenmekte, sonuç olarak P_{CT} 'nin bireylerin sabit %27'lik alanlardaki uyumuna baktığı için sınıflama tutarlığı, bireylerin birbirine göre toplam puanlar açısından sıralanmasına dayandığı için de sıralama tutarlığı hakkında bilgi vermekte, bu durum, bireyler hakkında verilecek istatistiksel kararların geçerliği açısından bir üstünlük sağlamaktadır (Erkuş, 2004b). Aynı zamanda indeksin, testin iki eşdeğer yarısı arasındaki uyuma baktığı için ölçme aracının güvenilirliği; verilecek kararlardaki tutarlığa baktığı için de geçerliği hakkında bilgi verdiği ileri sürülebilir.

Erkuş'un (2004b) diğer bir çalışmasında, indeksin yüzde gruplarına (%5, 10 ve 27), örneklemdaki birey (n) ve testteki madde (k) sayısına, test maddelerinin puanlanma biçimine (1-0 ve çoklu) göre değişimi irdelenmiştir. Yapılan analizler sonucunda, indeks değerinin hem 1-0 hem de çoklu puanlanan ölçek için yüzde grupları küçüldükçe düştüğü; birey sayısı azaldıkça, çoklu veriler için düştüğü, 1-0 puanlanan veriler için arttığı; madde sayısı azaldıkça hem 1-0 hem de çoklu puanlanan ölçek için düştüğü; %27'lik gruplarda en yüksek değerlerini aldığı gözlenmiştir. Bu bulgular, araştırmacının 1999'daki çalışmasını desteklemiştir.

Erkuş'un (2005) başka bir çalışmasında ise, P_{CT} indeksinin, 1-0 ve çoklu puanlanan ölçeklerde, değişik birey sayısı, madde sayısı ve % manipülasyonlarında ne tür değişim gösterdiği, ρ (rho) yarıya bölme katsayısı ve κ ile P_0 gibi sınıflama geçerliği katsayılarıyla ilişkisi irdelenmiştir. Analiz sonuçları, daha önceki çalışmalarda olduğu gibi, indeks için en iyi kesme alanının %27'lik alan olduğu; 1-0 puanlanan ölçek için birey sayısı azaldığında indeksin de azaldığı, çoklu puanlanan ölçek için ise tersine arttığı;

madde sayısı azaldığında ise birey sayısı manipülasyonunun tam tersi bir değişim olduğu gözlenmiştir. Bu çalışmada P_{CT} indeksine Spearman-Brown düzeltmesi yapılarak geleneksel yarıya bölme katsayısı, ρ , κ ve P_0 ile karşılaştırılmış ve indeksin değerinin iki katsayı arasında yer aldığı gözlenmiştir. Analiz sonuçları, P_{CT} indeksine düzeltme yapılması gerektiğini, indeksin hem sınıflama hem de sıralama geçerliği indeksi olarak kullanılabileceğini, ayrıca ölçme aracının güvenilirliğine ilişkin de bilgi verebileceğini ortaya koymuştur.

I.4. Araştırmanın Amacı

Bu araştırmada, bir ölçme aracının ne denli tutarlı sınıflama yaptığını belirlemede alternatif yöntemler bulunması gerekliliği üzerine, Erkuş (1999) tarafından geliştirilen Çift Tutarlık İndeksinin (P_{CT}), aynı ölçeğin 1-0 veya çoklu uygulanması arasında nasıl sonuçlar alacağı ve tekrarlı ölçümlerle yapılan irdelemelerin testleri yarıya bölmeyle aynı sonuçlar verip vermeyeceği, yapılacak karşılaştırmaları daha anlamlı duruma getirmek amacıyla irdelenmiştir.

Daha önceki çalışmalarda indeks, tekrarlı ölçümler kullanılmadan ve 1-0 ile çoklu puanlanan farklı ölçekler kullanılarak çalışılmış, aynı özelliği ölçen hem 1-0 hem de çoklu puanlanan bir tek ölçekle P_{CT} indeksinin işleyişi bugüne kadar ele alınmamıştır. 1-0'lı ve çoklu puanlanan ölçeklerin farklı özellikleri ölçen ölçekler olduğu göz önüne alındığında, Erkuş'un önceki çalışmalarından daha tutarlı sonuçlar elde edileceği beklenebilir. Bu çalışmanın amacı, sınıflama ve sıralama geçerliği indeksi olarak önerilen P_{CT} indeksini, *“aynı ölçeğin 1-0 ve çoklu puanlanan formlarını, tekrarlı ölçüm kullanarak”* irdelemek ve literatüre katkıda bulunmaktır.

Erkuş'un çalışmaları ile bu çalışma arası farklılığı içeren açıklamalar Tablo-1'de verilmiştir.

Tablo-1: Erkuş (1999) – Bu Çalışma Farklılıkları

Erkuş (1999)		Bu çalışma	
Tek Uygulama		Tekrarlı Uygulama	
Tekler	Çiftler	1. Uygulama	2. Uygulama
1-0'lı ve Çoklu Farklı Ölçekler (Farklı Örneklem Üzerinde)		1-0'lı ve Çoklu Aynı Ölçek (Farklı Örneklem Üzerinde)	

Bu amaca ulaşabilmek için aşağıdaki sorulara yanıt aranmaya çalışılmıştır.

1. 1-0 puanlanan ölçeğin;
 - a) Geleneksel yarıya bölme yöntemi kullanılarak, düzeltilmemiş P_{CT} değeri ile düzeltilmiş P_{CT} değerlerinin değişimi nasıldır?
 - b) Tekrarlı uygulanmasından elde edilen P_{CT} değerinin diğer katsayılara göre değişimi nasıldır?
 - c) Geleneksel yarıya bölme yöntemi ile tekrarlı uygulamadan elde edilen P_{CT} değeri ile tekrarlı uygulamadan elde edilen P_{CT} değerinin değişimi nasıldır?
2. Çoklu puanlanan ölçeğin;
 - a) Geleneksel yarıya bölme yöntemi kullanılarak, düzeltilmemiş P_{CT} değeri ile düzeltilmiş P_{CT} değerinin değişimi nasıldır?
 - b) Tekrarlı uygulanmasından elde edilen P_{CT} değerinin diğer katsayılara göre değişimi nasıldır?
 - c) Geleneksel yarıya bölme yöntemi ile tekrarlı uygulamadan elde edilen P_{CT} değeri ile tekrarlı uygulamadan elde edilen P_{CT} değerinin değişimi nasıldır?

3. 1-0 ve Çoklu Puanlanan geleneksel ve tekrarlı hesaplanan P_{CT} değerlerinin Geleneksel Uyum Katsayısı (P_0)'na göre deęişimi nasıldır?

I.5. Sayılılar

- 1) Denekler ölçekleri içtenlikle yanıtlamışlardır.
- 2) İndeks, kullanacak ölçeğin tek boyutlu olması sayılıısına dayanmaktadır.

I.6. Sınırlılıklar

- 1) Üzerinde çalışılan ölçek verileri, geliştirildikleri örneklemlerle sınırlıdır.

BÖLÜM II

YÖNTEM

II.1. Araştırmanın Türü

Bu araştırma, aynı özelliği ölçen hem 1-0 hem çoklu puanlanan bir tek ölçekle tekrarlı ölçümler kullanarak P_{CT}'yi irdelemek amacıyla yapılmıştır. Bu nedenle temel nitelikteki bir araştırmayı farklı açıdan irdeleyen bir araştırmadır.

II.2. Verilerin Elde Edildiği Grup

Mersin ilinde çeşitli eğitim düzeylerine sahip 439 kişilik örneklem üzerinden elde edilen veriler, araştırma sorularımıza yanıt bulmak amacıyla irdelenmiştir. Bu bireylerin, yaş ranjı 42 (18-60), öğrenim durumları %5.5 ilkokul, %3 ortaokul, %60.7 lise, %30.8 üniversite ve medeni halleri %41.4 evli, %56.3 bekar, %2.3 dul olarak saptanmıştır. Bu verilere, ilerde yapılacak araştırmalara ışık tutması amacıyla değinilmiştir.

Bireylerden alınan bu veriler 3 farklı etkiyi irdelemek amacıyla 3 farklı şekilde toplanmıştır. Birincisi, 150 bireyden alınan veriler, sıra etkisini (order effect) azaltmak amacıyla, 75 bireye önce 1-0 sonra çoklu puanlanan ölçek şeklinde, 75 bireye de önce çoklu sonra 1-0 puanlanan ölçek şeklinde uygulanmıştır; bir başka deyişle, aynı bireylerden hem 1-0'lı hem de çoklu forma ilişkin veriler elde edilmiştir. İkincisi, 136 bireyden alınan veriler, 1-0 puanlanan ölçeğin tekrarlı uygulamasıyla elde edilmiştir. Üçüncüsü ise 153 bireyden alınan veriler, çoklu puanlanan ölçeğin tekrarlı uygulamasıyla elde edilmiştir.

Verilerin elde edildiği gruba ilişkin veriler Tablo-2'de verilmiştir.

Tablo-2: Verilerin Elde Edildiği Gruba İlişkin Veriler

Toplam Birey Sayısı (n)	439
1-0 - Çoklu Ölçek Yanıtlayanlar (n)	150 (75 birey ilk 1-0 – son Çoklu, 75 birey ilk Çoklu – son 1-0)
1-0 - 1-0 Ölçek Yanıtlayanlar (n)	136
Çoklu - Çoklu Ölçek Yanıtlayanlar (n)	153
Bireylerin Yaş Ranjı	42 (18-60)
Öğrenim Durumu	%5.5 ilkokul, %3 ortaokul, %60.7 lise, %30.8 üniversite
Medeni Hali	%41.4 evli, %56.3 bekar, %2.3 dul

II.3. Veri Toplama Aracı

İndeksin işlerliğini sınamak amacıyla, hem 1-0 hem de çoklu olarak tasarlanmış olan bir ölçme aracı üzerinden elde edilen puanlar kullanılmıştır. Bu ölçeğin psikometrik özellikleri aşağıdaki kısımda verilmiştir.

Sorumluluk Ölçeği (SÖ)

Araştırmada kullanılan Sorumluluk Ölçeği (SÖ), Mersin Üniversitesi Eğitimde Ölçme ve Değerlendirme Bilim Dalında bir grup araştırmacı tarafından geliştirilmiş ve geçerlik çalışmaları halen devam eden, henüz yayınlanmamış bir ölçektir. Bireylerin günlük yaşam alışkanlıklarına bakarak sorumluluk düzeylerini belirlemek amacıyla ve çoklu puanlanacak şekilde geliştirilmiş, 21 maddeden oluşan 4 dereceli (“kararsızım” kategorisi yok) Likert tipi bir ölçektir (EK 1).

Başkalarına ve kendine karşı sorumluluk olarak iki alt bileşeni bulunan SÖ'nin, oblimin döndürmesiyle elde edilen “kendine karşı sorumluluk” alt ölçeğinin Cronbach Alfa içtutarlık katsayısı .770, “başkalarına karşı sorumluluk” alt ölçeğinin Cronbach Alfa içtutarlık katsayısı ise .860'dır. Birbiriyle ilişkili alt boyutlar toplanarak “genel sorumluluk” puanı da hesaplanabilen tüm ölçeğin Cronbach Alfa içtutarlık katsayısı ise .890 olarak bulunmuştur. SÖ'nin geçerlik çalışmaları halen devam etmektedir.

Bu tez çalışmasında, P_{CT} indeksinin aynı ölçeğin ikili (1-0) ve çoklu puanlanan iki formundaki işleyişi inceleneceği için, ölçeğin tepki kategorileri 1-0 puanlanacak şekilde “Evet – Hayır”a dönüştürülmüş (EK 2) ve 219 bireylik örneklem üzerindeki KR-20 içtutarlık katsayısı .805 olarak bulunmuştur.

Tek faktörlü iki bileşenli SÖ'nün, aynı maddelerden oluşan 1-0 ve çoklu puanlanan iki ayrı formu, P_{CT} indeksini incelemeye uygun görünmektedir; ancak ölçeğin iki bileşenli bir yapıda olmasının P_{CT} değerlerini etkileyeceği de beklenmelidir.

II.4. Verilerin Analizi

SÖ'nden elde edilen veriler 1-0 ve çoklu puanlananlar şeklinde ikiye ayrılarak geleneksel P_{CT} değerleri irdelenmiştir. Tekrarlı P_{CT} değerlerini irdellemek için, iki uygulamada da aynı ölçeği alanlardan elde edilen veriler kullanılmıştır.

SÖ'nden elde edilen verilerin bilgisayar analizlerinde SPSS 12.0 paket programı ve Excel 2003 Office programı kullanılmıştır. Veriler SPSS paket programına işlenerek ters puanlanması gereken maddeler ters puanlanmış, toplam puan bulunmuş, toplam puana göre azalan sırada bireyler sıralanmıştır. Uygulamalardan elde edilen veriler üzerinde P_{CT} indeksinin değerlerini hesaplayabilmek için Excel programında, araştırmacı tarafından formüle uygun program yazımı gerçekleştirilmiştir (EK 3).

Geleneksel yarıya bölme yöntemi, Guttman katsayısı ve Cronbach α katsayısı, Kolmogorov Smirnov z testi, SPSS paket programında hesaplanmış; KR-20 katsayıları ve ölçmenin standart hatası (S_{em}), Excel programında, araştırmacı tarafından formüle uygun yazılan programla saptanmıştır. 1. uygulamadan ve 2. uygulamadan elde edilen sorumluluk puanları medyan değerine göre irdelenmiş, Excel programında araştırmacı tarafından formüle uygun yazılan programla geleneksel uyum katsayısı (P_o) elde edilmiştir.

P_{CT} indeksinin, geleneksel yarıya bölme katsayılarıyla karşılaştırılmasında ortaya çıkacak yanlışlığı gidermek için, test iki yarıya ayrıldığından, eşdeğer iki yarı için eşdeğerlik sayılısıyla Spearman-Brown düzeltmesi aşağıdaki formül yardımıyla yapılmıştır (Erkuş, 2005).

$$r_{kk} = \frac{2.P_{CT}}{1 + P_{CT}} \quad (\text{Eşitlik 8})$$

P_{CT} indeksinin tekrarlı bireyler için hesaplanmasında, geleneksel güvenilirlik katsayılarıyla karşılaştırılmasında, ölçek yarıya bölünmediğinden Spearman-Brown düzeltmesi yapılmamıştır.

BÖLÜM III

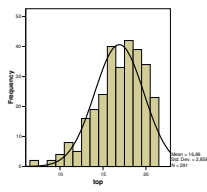
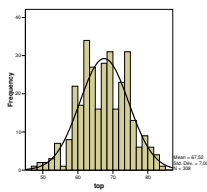
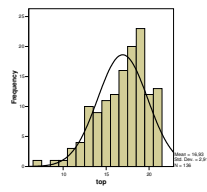
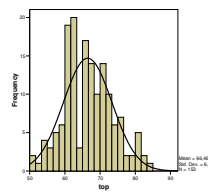
BULGULAR

Aşağıdaki bölümde, araştırma sorularındaki sıraya uygun bir şekilde bulguların ele alınmasına geçmeden önce, aynı ölçeğin 1-0 ve çoklu puanlanan formlarından elde edilen ve analizlerde temel alınan verilere ilişkin betimsel istatistikler verilecektir.

III.1. Sorumluluk Ölçeğinin Betimsel İstatistikleri

SÖ'nin 1-0 ve çoklu uygulamalarından elde edilen verilere ilişkin betimsel istatistikler Tablo-3'te verilmiştir.

Tablo-3: SÖ'nin 1-0 ve Çoklu Uygulamalarından Elde Edilen Betimsel İstatistikleri

	Geleneksel		Tekrarlı	
	1-0	Çoklu	1-0	Çoklu
n	291	308	136	153
			1. Uyg / 2. Uyg	1. Uyg / 2. Uyg
Ortalama	16.890	67.520	16.930 / 16.980	66.460 / 66.120
S ²	8.167	49.097	8.483 / 10.200	47.908 / 54.828
Çarpıklık	-.743	-.097	-.722 / -1.056	.240 / -.271
Basıklık	.362	-.296	.194 / 1.343	-.246 / .147
Güvenirlilik Katsayısı	KR-20 .682	Cronbach α .752	KR-20 .708 / .765	Cronbach α .758 / .816
Test-Tekrar Test	-	-	.759	.715
S _{em}	1.612	3.489	1.573 / 1.548	3.405 / 3.176
Histogram				

Tablo-3'teki güvenilirlik katsayıları incelendiğinde, 1-0 ve çoklu puanlanan ölçeklerin tekrarlı uygulanmalarından elde edilen ikinci katsayının ilk uygulamaya göre yüksek olduğu görülmektedir. 1-0 şeklinde puanlanan ölçeğin tekrarlı uygulanmasından elde edilen test tekrar test güvenilirlik katsayısının, ayrı ayrı hesaplanmış olan güvenilirlik katsayıları arasında bir değer aldığı görülmektedir. Çoklu puanlanan ölçeğin tekrarlı uygulanmasından elde edilen test tekrar test güvenilirlik katsayısının ise, ayrı ayrı hesaplanmış olan güvenilirlik katsayılarından görece düşük olduğu görülmektedir. Test tekrar test katsayılarından da anlaşılacağı gibi iki ölçeğin de zamana karşı kararlılık gösterdiği söylenebilir.

Tablo-3'teki 1-0 puanlanan ölçeğin geleneksel ve tekrarlı S_{em} katsayılarına bakıldığında, bireylerin gerçek puanlarının testten aldıkları puandan en fazla 1,610 saptığı; çoklu puanlanan ölçeğin geleneksel ve tekrarlı S_{em} katsayılarına bakıldığında ise, bireylerin gerçek puanlarının testten aldıkları puandan en fazla 3,490 saptığı görülmektedir.

Tablo-3 incelendiğinde, geleneksel hesaplanmış çoklu puanlanan ölçeğin çarpıklık değerlerinin sıfıra yakın olduğu, geleneksel hesaplanmış 1-0 ve tekrarlı hesaplanmış 1-0'lı ölçeğin sola çarpık olduğu, tekrarlı hesaplanmış çoklu puanlanan ölçeğin ise sağa çarpık olduğu söylenebilir. Basıklık katsayılarına bakıldığında ise, çoklu puanlanan ölçeklerin 1-0'lı puanlanan ölçeklere göre basık olduğu söylenebilir. Çarpıklık ve basıklık katsayılarının 1.00 ile -1.00 arasında yer aldığı görülmektedir.

Her iki ölçeğin her iki uygulamasında da ortalamaların büyük ölçüde benzer olduğu görülmektedir. 1-0 ve çoklu puanlanan ölçeğin iki uygulaması arasında yapılan tekrarlı ölçümler için t-testi sonucunda elde edilen puanlar arasında anlamlı bir farklılık bulunamamıştır. 1-0 ve çoklu puanlanan ölçekten elde edilen verilerin normal dağılıma uygunluğu Kolmogorov – Smirnov z testi ile kontrol edilmiş, çoklu puanlanan ölçeğin

uygulandığı grubun verilerinin normal dağılıma uyduğu, ancak 1-0 puanlanan ölçeğin uygulandığı grubun normal dağılıma uymadığı saptanmıştır. SÖ'nin 1-0 ve çoklu uygulamalarından elde edilen verilere ilişkin Kolmogorov – Smirnov z testi sonuçları Tablo-4'de verilmiştir.

Tablo-4: SÖ'nin 1-0 ve Çoklu Uygulamalarından Elde Edilen Kolmogorov – Smirnov z

Testi Sonuçları

	1-0	Çoklu
Kolmogorov – Smirnov z	2,139	.976
P	.000	.296

Aşağıdaki kısımda, bulgular, araştırma sorularının sırasına uygun bir şekilde ele alınacaktır.

III.2. 1-0 Puanlanan Sorumluluk Ölçeğinin Geleneksel ve Tekrarlı P_{CT} 'sine İlişkin Bulgular

a) Geleneksel Uygulamaya Dayanarak Hesaplanan P_{CT} Değerleri

SÖ'nin, 1-0 puanlanan formunun uygulanmasından elde edilen verilere ilişkin geleneksel yarıya bölme yöntemi kullanılarak elde edilen P_{CT} değerleri ile diğer içtutarlık katsayıları Tablo-5'te verilmiştir.

Tablo-5: 1-0 Puanlanan Ölçeğin Geleneksel P_{ÇT} Değerleri ve Diğer İçtutarlık Katsayıları

1-0 Puanlanan Ölçek	
Birey Sayısı (n)	291
Geleneksel Düzeltmesiz P _{ÇT}	.622
Geleneksel Düzeltmeli P _{ÇT}	.767
KR-20	.682
Geleneksel Yarıya Bölme Güvenirliği (S-B)	.650
Guttman Katsayısı	.608

Tablo-5 incelendiğinde, 1-0'lı puanlanan ölçeğin, KR-20 değerinin düzeltmesiz P_{ÇT} ile düzeltmeli P_{ÇT} arasında yer aldığı görülmektedir. Aynı zamanda yarıya bölme güvenirliliğinin de düzeltmesiz P_{ÇT} ile düzeltmeli P_{ÇT} arasında olduğu görülmektedir.

b) Tekrarlı Uygulamaya Dayanarak Hesaplanan P_{ÇT} Değerleri

SÖ'nin, 1-0 puanlanan formunun tekrarlı uygulanmasından elde edilen verilere ilişkin elde edilen P_{ÇT} değerleri ve diğer içtutarlık katsayıları Tablo-6'da verilmiştir.

Tablo-6: 1-0 Puanlanan Ölçeğin Tekrarlı P_{CT} Değerleri ve Diğer İttutarlık Katsayıları

1-0 Puanlanan Ölçek		
Birey Sayısı (n)	136	
Tekrarlı P_{CT}	.662	
Test Tekrar Test Güvenirlik Katsayısı	.759	
	1. Uyg	2. Uyg
KR-20	.708	.765
Geleneksel Yarıya Bölme Güvenirliği (S-B)	.653	.729
Guttman Katsayısı	.605	.687
Geleneksel Düzeltmesiz P_{CT}	.730	.730
Geleneksel Düzeltmeli P_{CT}	.844	.844

Tablo-6 incelendiğinde, 1-0'lı puanlanan ölçeğin, KR-20 değerlerinin tekrarlı P_{CT} 'den az da olsa yüksek olduğu görülmektedir. Aynı zamanda test tekrar test güvenirligi katsayısının da tekrarlı P_{CT} 'den az da olsa yüksek olduğu, tekrarlı P_{CT} değerinin ise yarıya bölme güvenirligi ile Guttman katsayısının 1. ile 2. uygulama değerleri arasında olduğu görülmektedir.

KR-20, yarıya bölme (Spearman-Brown) ve Guttman katsayıları incelendiğinde, 1-0'lı puanlanan ölçeğin 1. uygulaması ile 2. uygulaması arasında büyük farkın oluşu sorgulamaya değerdir. Ancak bu farklılık geleneksel P_{CT} ölçümlerinde görülmemiştir. 1. ve 2. uygulamada da bireylerin tutarlı sınıflanıp sıralandığı görülürken, klasik güvenirlilik katsayılarının tümünün 2. uygulamada daha yüksek değer almaları ilginçtir. 2. uygulamadan elde edilen verilerin varyansı (10.200), ilk uygulamadan elde

edilene göre (8.483) biraz artmıştır. Bu farklılık klasik güvenilirlik katsayılarındaki artışı kısmen açıklamaktadır. P_{CT} ise varyans değişiminden etkilenmemiştir.

c) Geleneksel ve Tekrarlı Uygulamalardan Elde Edilen P_{CT} Değerlerinin Karşılaştırılması

SÖ'nin, 1-0 puanlanan formunun geleneksel ve tekrarlı uygulanmasından elde edilen verilere ilişkin elde edilen P_{CT} değerleri Tablo-7'de verilmiştir.

Tablo-7: 1-0 Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerleri

	Geleneksel	Tekrarlı	
	1-0	1-0	
Düzeltilmesiz P_{CT}	.622	.662	
Geleneksel Düzeltmeli P_{CT}	.767	.844	.844

Tablo-7'deki 1-0 puanlanan ölçeğin geleneksel düzeltilmesiz P_{CT} değerleri ve tekrarlı uygulamaya dayanarak hesaplanan P_{CT} değerleri incelendiğinde, tekrarlı P_{CT} değerinin az da olsa (.040) artış gösterdiği görülmektedir. Düzeltmeli geleneksel P_{CT} ise tekrarlı P_{CT} 'den .105 daha yüksek bulunmuştur.

Aynı bireylere tekrarlı uygulanmış olan 1. ve 2. uygulamanın düzeltilmeli P_{CT} değerlerinin aynı olduğu, tekrarlı uygulamaya dayanarak hesaplanmış olan P_{CT} değerinden (.182) fazla olduğu görülmektedir.

III.3. Çoklu Puanlanan Sorumluluk Ölçeğinin Geleneksel ve Tekrarlı P_{CT} 'sine İlişkin Bulgular

a) Geleneksel Uygulamaya Dayanarak Hesaplanan P_{CT} Değerleri

SÖ'nin, çoklu puanlanan formunun uygulanmasından elde edilen verilere ilişkin geleneksel yarıya bölme yöntemi kullanılarak elde edilen P_{CT} değerleri ve diğer içtutarlık katsayıları Tablo-8'de verilmiştir.

Tablo-8: Çoklu Puanlanan Ölçeğin Geleneksel P_{CT} Değerleri ve Diğer İçtutarlık

Katsayıları

Çoklu Puanlanan Ölçek	
Birey Sayısı (n)	308
Geleneksel Düzeltmesiz P_{CT}	.584
Geleneksel Düzeltmeli P_{CT}	.737
Cronbach α	.752
Geleneksel Yarıya Bölme Güvenirliği (S-B)	.713
Guttman Katsayısı	.709

Tablo-8 incelendiğinde, çoklu puanlanan ölçeğin, Cronbach α değerinin diğer tüm katsayılardan daha yüksek olduğu görülmektedir. Aynı zamanda yarıya bölme güvenirliliğinin ve Guttman katsayısının da düzeltmesiz P_{CT} ile düzeltmeli P_{CT} arasında olduğu görülmektedir. Bu bulgu, P_{CT} formülünün güvenirlilik belirleme yöntemi olarak da kullanılabilirliğini göstermektedir.

b) Tekrarlı Uygulamaya Dayanarak Hesaplanan P_{CT} Değerleri

SÖ'nin, çoklu puanlanan formunun uygulanmasından elde edilen verilere ilişkin tekrarlı ölçümler kullanılarak elde edilen P_{CT} değerleri ve diğer içtutarlık katsayıları Tablo-9'da verilmiştir.

Tablo-9: Çoklu Puanlanan Ölçeğin Tekrarlı P_{CT} Değerleri ve Diğer İçtutarlık Katsayıları

Çoklu Puanlanan Ölçek		
Birey Sayısı (n)	153	
Tekrarlı P_{CT}	.720	
Test Tekrar Test Güvenirlik Katsayısı	.715	
	1. Uyg	2. Uyg
Cronbach α	.758	.816
Geleneksel Yarıya Bölme Güvenirliği (S-B)	.675	.814
Guttman Katsayısı	.674	.813
Geleneksel Düzeltmesiz P_{CT}	.659	.744
Geleneksel Düzeltmeli P_{CT}	.794	.853

Tablo-9 incelendiğinde, çoklu puanlanan ölçeğin Cronbach α değerlerinin, geleneksel düzeltmeli P_{CT} değerlerinin dışındaki tüm katsayılardan yüksek değerler aldığı görülmektedir. Tekrarlı P_{CT} değerinin, yarıya bölme güvenirliği, Guttman katsayısı ve geleneksel düzeltmesiz P_{CT} 'nin 1. uygulaması ile 2. uygulaması arasında olduğu görülmektedir.

Çoklu puanlanan ölçeğin geleneksel P_{CT} katsayıları incelendiğinde, hem düzeltmesiz hem de düzeltmeli P_{CT} değerlerin, 1. uygulama ile 2. uygulama arasında büyük farkın (.085 / .059) oluşu sorgulamaya değerdir. Bunun nedeninin çoklu puanlamadan kaynaklanan tepki kaymasından ve bireylerin 2. uygulamayı daha ciddi cevaplamalarından kaynaklandığı düşünülmektedir.

P_{CT} katsayıları incelendiğinde, tekrarlı ölçümlerden elde edilen P_{CT} değerinin, geleneksel düzeltmesiz P_{CT} değerinin 1. uygulaması ile 2. uygulaması arasında olduğu görülmektedir.

Dikkat edilirse, tüm katsayıların 2. uygulamadan elde edilen değerleri 1. uygulamadan elde edilenlere göre dikkat çekici ölçüde daha yüksektir. 1-0'lı ölçek verilerinde olduğu gibi bunun nedeni olarak 1. ve 2. uygulamanın puan varyansları hesaplanmış ve 1. uygulamanın varyansının (47.908), 2. uygulamanın varyansından (54.828) yine daha düşük olduğu görülmüştür. Bu durum, tüm klasik katsayıların puan varyansından etkilendiği görüşünü (Gulliksen, 1950; Guilford, 1954) desteklemektedir.

c) Geleneksel ve Tekrarlı Uygulamalardan Elde Edilen P_{CT} Değerlerinin Karşılaştırılması

SÖ'nin, çoklu puanlanan formunun geleneksel ve tekrarlı uygulanmasından elde edilen verilere ilişkin elde edilen P_{CT} değerleri Tablo-10'da verilmiştir.

Tablo-10: Çoklu Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerleri

	Geleneksel	Tekrarlı	
	Çoklu	Çoklu	
Düzeltilmesiz P_{CT}	.584	.720	
Geleneksel Düzeltilmeli P_{CT}	.737	.794	.853

Tablo-10'daki çoklu puanlanan ölçeğin geleneksel düzeltilmesiz P_{CT} değerleri ve tekrarlı uygulamaya dayanarak hesaplanan P_{CT} değeri incelendiğinde, geleneksel düzeltilmeli P_{CT} 'nin hem ayrı örneklerde hem de tekrarlı uygulamanın 1. ve 2. uygulamasında, tekrarlı P_{CT} 'den daha yüksek olduğu gözlenmiştir.

Tekrarlı uygulamaya dayanarak hesaplanmış olan P_{CT} değerinin, ayrı ayrı hesaplanmış olan 1. ve 2. uygulamanın az da olsa düşük olduğu, geleneksel düzeltmeli P_{CT} değerinden de az da olsa (.017) düşük olduğu görülmektedir.

III.4. 1-0 ve Çoklu Puanlanan Geleneksel ve Tekrarlı Hesaplanan P_{CT} Değerlerinin Geleneksel Uyum Katsayısı (P_o) ile Karşılaştırılması

SÖ'nin, 1-0 ve çoklu puanlanan formunun tekrarlı uygulanmasından elde edilen verilere ilişkin, medyan değeri temel alınarak elde edilen geleneksel uyum katsayısı değerleri Tablo-11'de verilmiştir.

Tablo-11: 1-0 ve Çoklu Geleneksel Uyum Katsayısı (P_o) Değerleri

		2. Uygulama			
Tekrarlı 1-0	1. Uygulama		Doğru Sınıflama %	0	1
		0	%64.7	44	24
		1	%91.2	6	62
			Toplam	Yanlış Sın. (n)	Doğru Sın.(n)
			%77.9	30	106
Tekrarlı Çoklu	1. Uygulama		Doğru Sınıflama %	0	1
		0	%72.9	54	20
		1	%75.9	19	60
			Toplam	Yanlış Sın. (n)	Doğru Sın.(n)
			%74.5	39	114

Tablo-11 incelendiğinde, tekrarlı hesaplanan 1-0 puanlanan ölçeği alan grubun geleneksel uyum katsayısı $P_o = .779$, tekrarlı hesaplanan çoklu puanlanan ölçeği alan grubun geleneksel uyum katsayısı $P_o = .745$ olduğu görülmektedir. Farklı şekilde puanlanan

iki ölçeğe uygulanan geleneksel uyum katsayısı sonucunda bu yüzdeler dikkate alındığında, yanlış sınıflama oranının %25'dan az olduğu göze çarpmaktadır.

SÖ'nin, 1-0 ve çoklu puanlanan formunun geleneksel uyum katsayısı değerleri ile geleneksel ve tekrarlı P_{CT} değerleri Tablo-12'de verilmiştir.

Tablo-12: 1-0 ve Çoklu Geleneksel Uyum Katsayısı ve P_{CT} Değerleri

	Geleneksel Uyum Katsayısı	Geleneksel Düzeltmeli P_{CT}	Tekrarlı P_{CT}
1-0	.779	.767	.662
Çoklu	.745	.737	.720

Hem 1-0 puanlanan hem de çoklu puanlanan ölçeğin, geleneksel uyum katsayısı değerlerinin, tekrarlı P_{CT} 'nden ve geleneksel düzeltmeli P_{CT} 'den görece daha yüksek değerler aldığı görülmektedir.

BÖLÜM IV

TARTIŞMA VE YORUM

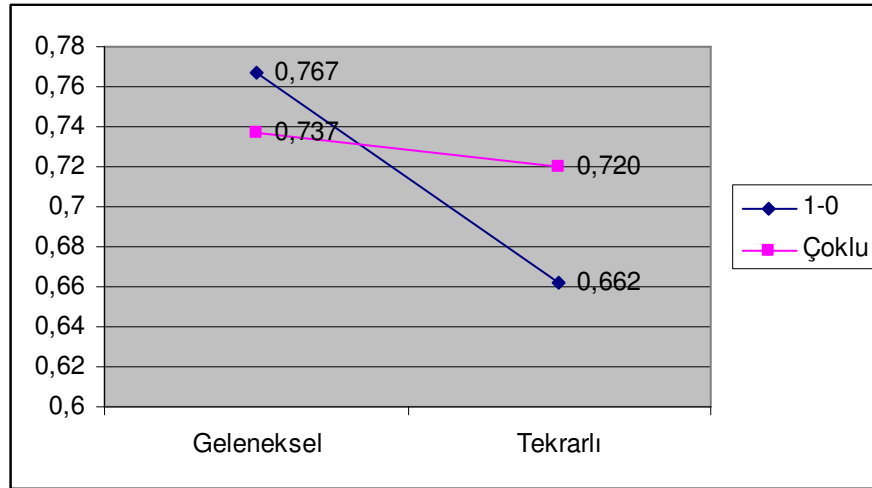
Bu bölümde, Giriş bölümünde belirtilen ilgili çalışmalar ve Bulgular bölümünde özetlenen bilgiler ışığında araştırma soruları teker teker ele alınıp tartışılacak ve yorumlanacaktır.

Yoruma bir temel sağlaması açısından SÖ'nin, 1-0 ve çoklu puanlanan formunun geleneksel ve tekrarlı uygulanmasından elde edilen P_{CT} değerleri ile diğer güvenirlik katsayıları topluca Tablo-13'te verilmiştir.

Tablo-13: Geleneksel ve Tekrarlı Güvenirlik Değerleri

	Geleneksel		Tekrarlı	
	1-0	Çoklu	1-0	Çoklu
Tekrarlı P_{CT}	-	-	.662	.720
Geleneksel Düzeltmeli P_{CT}	.767	.737	.844 - .844	.794 - .853
KR-20 / Cronbach α	.682	.752	.708 - .765	.758 - .816
Yarıya Bölme Güv.	.650	.713	.653 - .729	.675 - .814
Test Tekrar Test Güv.	-	-	.759	.715

SÖ'nin, 1-0 ve çoklu puanlanan formunun geleneksel ve tekrarlı uygulanmasından elde edilen verilere ilişkin elde edilen P_{CT} değerleri grafiği Grafik 1'de verilmiştir.

Grafik 1: Geleneksel ve Tekrarlı P_{CT} Değerleri

Grafik 1 incelendiğinde, hem 1-0 hem de çoklu puanlanan ölçeğin tekrarlı hesaplanan P_{CT} değerlerinde geleneksel hesaplama göre düşüş olduğu, çoklu puanlanan ölçekteki düşüşün (.017), 1-0 puanlanan ölçeğe göre daha az bir düşüş gösterdiği (.105) gözlenmektedir.

Aşağıdaki kısımda, tartışma ve yorum, araştırma sorularının sırasına uygun bir şekilde ele alınacaktır.

IV.1. 1-0 Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerlerinin İrdelenmesi

Tablo-13 incelendiğinde, 1-0 puanlanan ölçeğin geleneksel ve tekrarlı yöntemlerle hesaplanan değerlerinin içinde en yüksek değer geleneksel düzeltilmeli P_{CT} olduğu görülmektedir.

Erkuş'un (1999) önerdiği indeks, Erkuş'un (1999; 2004b; 2005) indeksle ilgili tüm çalışmalarında farklı puanlanan farklı ölçekler kullanılarak irdelenmiştir. Bu çalışmada ise, diğer çalışmalardan farklı olarak, **aynı ölçeğin** 1-0 ve çoklu puanlanan formları

kullanılmış, ayrıca hem geleneksel hem de ölçeği yarıya bölmek yerine **tekrarlı** olarak indeks değerleri elde edilmiştir.

1-0 puanlanan ölçekten geleneksel ve tekrarlı olarak elde edilmiş veriler incelendiğinde, geleneksel hesaplanan P_{CT} değerinin daha yüksek olduğu bulunmuş, bunun nedeninin ise geleneksel hesaplanan ölçeğin Spearman-Brown katsayısı ile düzeltilmiş olmasından veya iki uygulama için farklı örneklem kullanılmasından kaynaklandığı söylenebilir.

Geleneksel P_{CT} , ölçeğin eşdeğer iki yarıya bölünmesine dayanmaktadır; oysa ki, tekrarlı P_{CT} hesaplanmasında eşdeğerlik sayılına gerek yoktur. Bilindiği gibi, eşdeğerlik bozulduğu ölçüde, eşdeğerlik sayılına dayanan güvenilirlik katsayıları da düşer. Bu durumlarda, tekrarlı ölçmeye dayanarak hesaplanan P_{CT} değerlerinin, yarıya bölme yöntemine dayanarak hesaplanan P_{CT} değerlerinden (diğer koşullar aynı kalmak kaydıyla) daha yüksek olması beklenen bir durumdur, ancak bu araştırmada bunun tersi gözlenmektedir. Bunun nedeni olarak Spearman-Brown düzeltmesinin geleneksel P_{CT} indeksinin değerini şişirdiği (inflation) ileri sürülebilir. Öte yandan, tekrarlı P_{CT} ile tek uygulamaya dayalı P_{CT} 'yi karşılaştırmak da pek doğru görünmemektedir. Tekrarlı P_{CT} , sınıflama ve sıralamanın zaman içindeki kararlılığı; geleneksel P_{CT} ise, yarıya bölmeden dolayı ölçeğin içtutarlığı bilgisini de kapsamaktadır. Bu bakımdan, her iki yöntemle elde edilen katsayıların amaca bağlı olarak kullanılmasında yarar vardır. Ayrıca, ölçeğin eşdeğer yarıya dayanan içtutarlık katsayıları da ölçeğin görece homojen olmadığını göstermekte bu da elde edilen P_{CT} değerleriyle paralellik göstermektedir.

IV.2. Çoklu Puanlanan Ölçeğin Geleneksel ve Tekrarlı P_{CT} Değerlerinin İrdelenmesi

Çoklu puanlanan ölçeğin geleneksel yolla hesaplanan P_{CT} değeri ile tekrarlı hesaplanan P_{CT} değeri arasında, 1-0'lı puanlanan ölçekteki düşüşe oranla çok daha az

(.017) bir düşüş söz konusudur. Bunun bir nedeni olarak 1-0'lı durumdaki yoruma ek olarak, ölçeğin eşdeğerlik sayılısını çok da karşılamaması gösterilebilir. Ancak, bu küçük de olsa azalmanın başka nedenleri de olmak durumundadır. Bunun bir nedeni olarak, geleneksel ve tekrarlı uygulamaların örneklemlerinin farklı olması gösterilebilir. Bilindiği gibi, klasik güvenirlik belirleme örnekleme bağlıdır (Thorndike, 1982; Campbel, 1993). Ölçeğin içtutarlık katsayıları incelendiğinde, ölçek aynı olmasına karşın iki örneklemden farklı olduğu göze çarpmaktadır. İlginç olan diğer bulgu ise, P_{CT} indeksinin de örnekleme bağlı (örneklemden etkilenen) olarak klasik güvenirlik katsayıları gibi değişmesidir. Yine aynı şekilde, tekrarlı ve geleneksel P_{CT} 'lerden hangisinin kullanılacağı, ikisi ayrı bilgi verdiği için amaca bağlı olarak değişecektir.

IV.3. 1-0 – Çoklu Puanlanan ve Geleneksel – Tekrarlı Hesaplanan P_{CT} Değerlerinin Geleneksel Uyum Katsayısı (P_o) ile Karşılaştırılması

Geleneksel Uyum Katsayısı ister iki uygulamaya, isterse tek uygulamaya dayansın, bireyleri sınıflamak için kullanılan kesme noktasının dağılım içindeki yerinden radikal bir şekilde etkilenmektedir (Subkoviak, 1980): Dağılımın merkezine yakın olduğunda düşük, uçlara gidildiğinde yüksek P_o değerleri elde edilmektedir. Bu çalışmada, kullanılan kesme puanı medyan olduğu için P_o değerleri de olduğundan düşük çıkmıştır. P_{CT} ise sabit %27'lik alanlara dayanarak hesaplanmaktadır. Bir ölçme aracının, kendisinden bağımsız ve çeşitli keyfi kararlara göre belirlenen bir kesme noktasına göre sınıflama tutarlığının belirlenmesi, bu çalışmada da görüldüğü gibi uygun olmamaktadır. Uygulamada P_o yerine, daha kararlı olan P_{CT} 'nin yeğlenmesi önerilebilir.

IV.4. Geleneksel ve Tekrarlı P_{CT} Değerlerinin Çeşitli Katsayılar İle Karşılaştırılması

Gerek 1-0 veya çoklu puanlansın, gerek tek veya tekrarlı uygulamaya dayanarak hesaplınsın, hesaplanan geleneksel düzeltmeli P_{CT} değerlerinin, içtutarlık güvenilirlik katsayılarından görece yüksek olduğu Tablo-13'ün incelenmesinden anlaşılmaktadır. Ancak, tekrarlı P_{CT} değerleri için ise durum biraz farklıdır. Tekrarlı P_{CT} değerleri, KR-20 veya Cronbach α katsayıları gerek 1-0 gerekse çoklu ölçekler için her durumda yüksek; yarıya bölme güvenilirlik katsayısının ise, 1. ve 2. uygulamadaki değerlerinin arasında yer almaktadır. Önceki yorumlara paralel olarak, bu durumun, gerek iki yarı, gerekse maddelerin eşdeğerliğine dayanan bu yöntemlerin temel sayılıtısının, geleneksel P_{CT} düzeltmesinde ihlal edildiğini ortaya çıkardığını öne sürmek uygun görünmektedir. Geleneksel P_{CT} düzeltmesi, P_{CT} değerinin olduğundan daha yüksek (inflation) çıkmasına yol açmaktadır. Buradan hareketle, düzeltilmiş P_{CT} değerlerinin hesaplamaya katılmayan madde parçasının "mükemmel" eşdeğerliğine dayandığı için yüksek çıktığını ve bu değerlerin içtutarlık katsayılarının üst sınırını verdiğini ileri sürmek mümkündür.

Ölçme araçlarının kararlılığı bilgisini içeren test tekrar test güvenilirlik katsayılarının tekrarlı P_{CT} ile anlamsal ilişki göstermesi beklenen bir durumdur. Çok fazla radikal değişiklik göstermemesine rağmen, tekrarlı P_{CT} değerleri ile test tekrar test güvenilirlik katsayıları arasında bir fark olduğu; üstelik bu farkın 1-0'lı ölçekte tekrarlı P_{CT} 'nin aleyhine, çoklu ölçekte ise lehine olduğu göze çarpmaktadır. Buradan hareketle, tekrarlı P_{CT} 'nin çoklu puanlamaya, test tekrar test güvenilirlik katsayısının ise 1-0'lı puanlamaya karşı daha duyarlı olduğu ileri sürülebilir. Bu durumun nedenlerinin başka çalışmalarda incelenmesinde yarar bulunmaktadır.

IV.5. Aynı Örneklem Üzerindeki Tekrarlı P_{CT} ile I. Ve II. Uygulamalardaki Geleneksel P_{CT} Değerlerinin Karşılaştırılması

Temel araştırma soruları içinde bulunmamasına rağmen, tekrarlı örneklemden elde edilen P_{CT} değeri ile tekrarlı uygulamanın 1. ve 2. uygulamalarındaki geleneksel P_{CT} değerlerinin (Tablo-6, Tablo-9) incelenmesi de yarar sağlayabilir.

a. 1-0 Puanlanan Ölçekte

1-0 puanlanan ve tekrarlı uygulanan durumlarda 1. uygulamadaki geleneksel P_{CT} değeri (.844) ile 2. uygulamadaki geleneksel P_{CT} değeri (.844) değişmemiş; tekrarlı uygulamadan elde edilen P_{CT} değeri ise .662 olarak bulunmuştur. Farklı örneklemlerdeki P_{CT} değişiminin (.767 - .662) burada görülmemesi, geleneksel ve tekrarlı P_{CT} hesaplamalarında örneklemden kaynaklanan değişimi bertaraf ettiği için önemlidir. Üstelik, diğer tüm katsayılarda 2. uygulamadaki varyansın büyüklüğünden kaynaklanan artış P_{CT} 'de görülmemektedir. Bu sonuca göre, tekrarlı yöntemle elde edilen 1. ve 2. uygulamanın geleneksel yöntemle hesaplanan P_{CT} değerlerinin, tekrarlı yöntemle hesaplanana göre yüksek değerler verdiği görülmektedir. Bu bulgu, geleneksel P_{CT} indeksine uygulanan Spearman-Brown düzeltmesinin (Eşitlik 8), düzeltmenin temelini oluşturan eşdeğerlik sayılısının ihlal edilmesinden dolayı indeks değerinin büyümesine (inflation) yol açtığı şeklinde yorumlanabilir. Bu bakımdan, düzeltilmiş ve geleneksel yolla hesaplanmış P_{CT} değerinin, tüm içtutarlık katsayılarına göre, eğer güvenilirlik bilgisi olarak kullanılacak ve yorumlanacaksa, üst sınırı verdiği ileri sürülebilir.

b. Çoklu Puanlanan Ölçekte

Çoklu puanlanan ve tekrarlı uygulamanın 1. uygulamasındaki geleneksel P_{CT} değeri (.794) ile 2. uygulamadaki geleneksel P_{CT} değeri (.853); 1-0 puanlanan ölçeğin aksine artış göstermiş; tekrarlı P_{CT} değerinin ise bu iki değerden de düşük değer aldığı

gözlenmiştir. Örneklem sabit tutulduğuna ve ölçek değişmediğine göre bu durum, ölçek nitelikleri veya indeks işleyişinden öte, cevaplayıcıların bilişsel süreçleri ve tepki kategorileriyle açıklanabilir görünmektedir. Likert tipi ölçek kategorilerinin sıralama düzeyinde olması ve iki farklı zaman diliminde, aynı uyarıcıya (ifadeye) farklı derecelerde katılmaları pek de kabul edilmeyecek bir durum değildir. Bu ön kabulle bakıldığında, bireylerin toplam puan sıralarının ölçmeden ölçmeye değişmesi ve dolayısıyla da bunun indeks değerine yansımaları doğal görünmektedir. 1-0 tepki kategorisinin bu “teпки kayması”na yol açması, sınıflama düzeyinde ölçme özelliğinden dolayı, çok daha güçtür. Bu nedenle, 1-0 ölçeğin 1. ve 2. uygulamasından elde edilen değerlerin kararlı olması şaşırtıcı değildir.

En son olarak;

a. Klasik güvenirlik katsayılarının örneklemden örnekleme değişmesinin yanında aynı örneklemden değişik zamanlarda elde edilen verilerde de değişmesi ilginçtir ve incelenmesi gereken bir durumdur. Burada, ölçeklerin psikometrik niteliklerinden öte, cevaplayıcıların duyuşsal ve bilişsel süreçlerinin karıştırıcı etki yaptığı ve ölçeklerin psikometrik niteliklerinin bu cevaplayıcı özelliklerine bağlı olarak incelenmesinin çok yararlı sonuçlar verebileceği ileri sürülebilir.

b. Bulunan sonuçlar topluca irdelendiğinde, ölçme aracı güvenilir olsa da tutarlı sınıflama ve sıralama yapamayabilir veya tersine görece güvenilir olmayan bir ölçme aracının da tutarlı sınıflama ve sıralama yapabilir olduğu görülmektedir. Bu durum da incelenmeye değerdir. Ancak, bu yorum, geçerlik katsayısının karesinin güvenirlik katsayısına (veya geçerlik katsayısının güvenirlik indeksine) eşit veya küçük olduğu kuramsal eşitlik (Gulliksen, 1950; Guilford, 1954), her durumda doğrulanmaktadır; bu da literatürdeki bu bilgiler açısından ilginç görünmektedir.

c. Bu çalışmanın da gösterdiği gibi, tekrarlı $P_{\text{ÇT}}$ ile geleneksel $P_{\text{ÇT}}$ değerlerinin farklı çıkmış olması, ikisinin farklı bilgiler içeriyor olmasıyla açıklanabilir. Tekrarlı $P_{\text{ÇT}}$, sınıflama ve sıralamanın kararlılığı hakkında, geleneksel $P_{\text{ÇT}}$ ise, ölçeğin içtutarlığı hakkında ağırlıklı bilgi verdiği ileri sürülebilir. Bu nedenle, kararlı sınıflama merak ediliyorsa, tekrarlı $P_{\text{ÇT}}$ 'nin incelenmesi önerilir.

SONUÇLAR VE ÖNERİLER

Bu bölümde, araştırmada elde edilen bulgulara dayalı sonuçlar belirtilmiş ve yeni araştırmalar için bazı önerilerde bulunulmuştur.

Sonuçlar

- 1-0 puanlanan ölçekten geleneksel ve tekrarlı olarak elde edilmiş veriler incelendiğinde, geleneksel hesaplanan P_{CT} değerinin daha yüksek olduğu bulunmuş, bunun nedeninin ise geleneksel hesaplanan ölçeğin Spearman-Brown katsayısı ile düzeltilmiş olmasından veya iki uygulama için farklı örneklemeler kullanılmasından kaynaklandığı söylenebilir. 1-0 puanlanan ölçeğin geleneksel düzeltmeli ve tekrarlı hesaplanan P_{CT} değerlerinin ve KR-20 güvenirlik katsayısının birbirlerine oldukça yakın olduğu, bu nedenle P_{CT} 'nin hem güvenirlik hem de verilecek kararların tutarlığına baktığı için geçerlik belirleme yöntemi olduğu sonucuna ulaşılmıştır.
- Çoklu puanlanan ölçeğin geleneksel yolla hesaplanan P_{CT} değeri ile tekrarlı hesaplanan P_{CT} değeri arasında, 1-0'lı puanlanan ölçekteki düşüş oranla çok daha az (.017) bir düşüş söz konusudur. Bu nedenle çoklu puanlanan ölçeğin geleneksel düzeltmeli ve tekrarlı P_{CT} 'inin hem güvenirlik hem de verilecek kararların tutarlığına baktığı için geçerlik belirleme yöntemi olduğu sonucuna ulaşılmıştır.
- Geleneksel yolla hesaplanan veriler incelendiğinde, 1-0'lı puanlanan ölçeğin P_{CT} değerinin çoklu puanlanan ölçeğe göre (2. uygulama hariç) yüksek olduğu bulunmuştur. Tekrarlı olarak hesaplanan 1-0 ve çoklu puanlanan ölçekler incelendiğinde ise, çoklu puanlanan ölçeğin P_{CT} değerinin 1-0 puanlanan ölçeğe göre yüksek olduğu bulunmuştur.

4. Düzeltilmiş P_{CT} değerlerinin, hesaplama katılmayan maddelerin “mükemmel” eşdeğerliğine dayandığı için yüksek çıktığı ve bu değerlerin içtutarlık katsayılarının üst sınırını verdiği sonucuna ulaşılmıştır.
5. Çoklu puanlanan ölçekte, “tepki kayması”ndan kaynaklanacak kararsızlığın, 1-0’lı ölçekte gözlenmediği; bunun nedeninin ise, çoklu puanlamanın ikiliye göre daha üst ölçme düzeyine karşılık gelmesi ve bireylerin sınıflanmasının ve sıralamasının çoklu ölçekte daha duyarlı hale geldiği sonucuna ulaşılmıştır.

Öneriler

1. Normal dağılan, fakat farklı ortalama güçlükteki başarı testlerinden elde edilecek veriler üzerinde indeksin değişiminin izlenmesi yararlı bilgiler verebilir.
2. Aynı örneklemden (tekrarlı uygulama) iki uygulamada elde edilen içtutarlık katsayılarının neden farklı olduğu incelenmeye değer görünmektedir.
3. P_{CT} indeksi, güvenilirlik bilgisi olarak kullanılacağında, içtutarlık katsayılarının üst sınırını vereceğinden dikkatle yorumlanmalıdır. Ancak, P_{CT} indeksi sınıflama ve sıralama geçerliği olarak güvenle kullanılabilir.
4. Sağa ya da sola çarpık ve normal dağılım gösteren veriler üzerinde indeks değerinin değişimi belirlenerek, indeksin dağılım fonksiyonuna ilişkin yararlı bulgular edinilebilir.
5. Zaman içindeki kararlılık merak ediliyorsa tekrarlı P_{CT} kullanılması önerilebilir.
6. Klasik güvenilirlik katsayılarının örneklemden etkilenmesinin yanında, aynı örneklemin değişik zamanlardaki değerlerinin de farklılaşmasının gerekçelerinin incelenmesi gerekmektedir.

KAYNAKÇA

AIKEN, Lewis R. (2000). *Psychological Testing and Assessment*, Allyn and Bacon.

ANASTASI, Anne (1971). *Psychological Testing Third Edition*, London: The Macmillan Co.

BERK, Ronald A. (1980). "A Consumers' Guide to Criterion-Referenced Test Reliability", *Journal of Educational Measurement*, 17(4), 323-349.

CAMPBELL, Kathleen T. (1993). "Establishing Internal Consistency Reliability of Measurement Data of a New Instrument, the Information Preference Questionnaire", *to the Educational Resources Information Center (ERIC)*, ED356235.

CROCKER, Linda ve ALGINA, James (1986). *Intruduction to Classical and Modern Test Theory*, New York: CBS College Pub Co.

CRONBACH, Linda J. (1951). *Essential of Psychological Testing*, Harper and Row, Tokyo.

CRONBACH, Linda J. (1971). *Essentials of Psychological Testing (Third Edition)*, New York: Harper and Row Pub.

- ERKUŞ, Adnan (1999). “Ölçme Araçlarının Tutarlı Ölçme ve Sınıflama Yapı Yapmadığını Belirlemeye İlişkin Bir Çalışma”. Doktora Tezi, AÜ, SBE.
- ERKUŞ, Adnan (2000a). “Ölçüt Dayanaklı Değerlendirme Bağlanımda Geliştirilen Sınıflama Tutarlılığı İndeksleri ve Bazı Sorunları”, *Türk Psikoloji Yazıları*, Cilt:3, Sayı:5, Haziran.
- ERKUŞ, Adnan (2000b). “Yeni Bir İndeks Önerisi: Çift Tutarlık İndeksi (P_{CT})”, *Türk Psikoloji Dergisi*, Cilt:15, Sayı:46, Aralık.
- ERKUŞ, Adnan (2003). *Psikometri Üzerine Yazılar*, Türk Psikologlar Derneği Yayınları, No:24, Ankara.
- ERKUŞ, Adnan (2004a). “The Proposal of a New Conceptualization for Validity and Criterion-Referenced Assessment”, *Eğitim Araştırmaları Dergisi*, Yıl:4, Sayı:16, Yaz.
- ERKUŞ, Adnan (2004b). “The Proposal of double consistency index (p_{dc}) and its proposition for some variables”, *Eğitim Araştırmaları Dergisi*, Yıl:5, Sayı: 17, Sonbahar.
- ERKUŞ, Adnan (2005). “P_{CT}’nin Değişik n, k ve Yüzdellik Dilimlerindeki Değişimi ile p, Yarıya Bölme ve Değişik İndekslerle Karşılaştırılması”, *Türk Psikoloji Dergisi*, Cilt:20, Sayı:55, Haziran 2005.

ERKUŞ, Adnan (2006). *Sınıf Öğretmenleri İçin Ölçme ve Değerlendirme Kavramlar ve Uygulamalar*, Ekinoks Eğ. Dan. Hiz. ve Bas. Yay. Dağ. San.

GUILFORD, Joy Paul (1954). *Psychometric Methods Second Edition*, New York: McGraw-Hill Book Company.

GULLIKSEN, Harald (1950). *The Theory of Mental Tests*, New York: John Wiley and Sons inc.

HAMBLETON, Ronald K. ve NOVICK, Melvin R. (1973). "Toward an integration of theory and method for criterion-referenced tests", *Journal of Educational Measurement*, 10(3), 159-170.

HENSON, Robin K. (2001). "Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha", *Measurement and Evaluation in Counseling and Development*, 34(3), 177-189.

KLIN, Paul (1986). *a Handbook of Test Construction*, Methuen & Co. Ltd. London.

LORD, Frederic M. ve NOVICK, Melvin R. (1968). *Statistical Theories of Mental Test Scores*, New York: Addison-Wesley Pub. Co.

MURPHY, Kevin R. ve DAVIDSHOFER, Charles O. (2001). *Psychological Testing Principles and Applications 5. Edition*, New Jersey: Prentice Hall.

SUBKOVIK, Michael J. (1980). *Decision-consistency approaches*, Akt. Ronald A.

BERK, Criterion- Referenced Measurement, 129-185, Baltimore: John
Hopkins Univ. Press.

THORNDIKE, Robert L. (1982). *Applied Psychometrics*, Boston: Houghton

Mifflin Company.

TURGUT, M.Fuat ve BAYKUL, Yaşar (1992). *Ölçekleme Teknikleri*, ÖSYM Yayınları,

Ankara.

TURGUT, M.Fuat (1992). *Eğitimde Ölçme Değerlendirme Metotları*, Saydam

Matbaacılık, Ankara.

EKLER

EK 1:

Çoklu Puanlanan Sorumluluk Ölçeđi

Sayın Katılımcı,

Bu çalışma bireylerin günlük yaşam alışkanlıkları ile ilgili bir ölçek geliştirmek amacıyla yapılmaktadır. Sizden beklenen, aşağıdaki günlük yaşantınız ile ilgili ifadeleri ve yandaki sütunlardaki olası cevapları okuyarak size uygun olan kutucuğun içine çarpı (X) işareti koymanız; hiçbir ifadeyi atlamamanız ve içtenlikle yanıt vermenizdir. Geliştirilecek ölçeğin sağlamlığı, vereceğiniz yanıtların içtenliğine bağlıdır. Bu çalışma sonunda elde edilecek sonuçlar, kesinlikle başka amaçlar için kullanılmayacaktır. İsim yazmanıza gerek yoktur. Ancak ileride ikinci bir ölçüm almak gerektiğinde, sizin kendi yanıtladığınız ölçeği bulabilmemiz açısından TC numaranızın ilk 5 rakamı gerekmektedir.

Katkılarınız için çok teşekkür ederiz.

Ü. Sonsuzay KUZU
Mersin Üniversitesi Eğitimde Ölçme ve
Değerlendirme Abd. Yüksek Lisans Öğrencisi

TC numaranızın ilk 5 rakamı =

— — — — —

İFADELER	Bana çok uygun	Bana biraz uygun	Bana uygun değil	Bana hiç uygun değil
1. Hakkımı aramasını pek beceremem.				
2. Çevre koruma derneklerine destek veririm.				
3. Evimi temiz tutarım.				
4. Yiyecek alırken son kullanma tarihine mutlaka bakarım.				
5. Trafik kurallarına uyarım.				
6. Haksızlığa uğradığımda hakkımı aramak için çaba gösteririm.				
7. Park ve bahçelere zarar vermem.				
8. Bir felaket durumunda yardıma koşarım.				
9. Çok dağınık yaşarım.				
10. “Ülkenin sorunlarını çözmek bana mı kaldı” diye düşünürüm.				
11. Sokağa çöp atmam.				
12. Suyu gereksiz kullananları uyarıp da başıma iş almam.				
13. Doktora gitmeye üşenirim.				
14. Evdeki kir ve toz beni rahatsız eder.				
15. Ana haberleri kaçırmamaya özen gösteririm.				
16. “Bu ülkeyi ben mi kurtaracağım diye düşünürüm.				
17. Beslenmeme pek dikkat etmem.				
18. İşyeri ya da okulda gereksiz yanan lambaları kapatırım.				
19. Parklara zarar verenleri uyarırım.				
20. Sağlığıma pek dikkat etmem.				
21. Trafik kurallarına uymayanları uyarmak bana düşmez.				

Yaşınız

:

Eğitim Düzeyiniz

: İlkokul Ortaokul Lise Üniversite

Medeni Durumunuz

: Evli Bekar Dul

EK 2:

1-0 Puanlanan Sorumluluk Ölçeđi

Sayın Katılımcı,

Bu çalışma bireylerin günlük yaşam alışkanlıkları ile ilgili bir ölçek geliştirmek amacıyla yapılmaktadır. Sizden beklenen, aşağıdaki günlük yaşantınız ile ilgili ifadeleri ve yandaki sütunlardaki olası cevapları okuyarak size uygun olan kutucuğun içine çarpı (X) işareti koymanız; hiçbir ifadeyi atlamamanız ve içtenlikle yanıt vermenizdir. Geliştirilecek ölçeğin sağlamlığı, vereceğiniz yanıtların içtenliğine bağlıdır. Bu çalışma sonunda elde edilecek sonuçlar, kesinlikle başka amaçlar için kullanılmayacaktır. İsim yazmanıza gerek yoktur. Ancak, ileride ikinci bir ölçüm almak gerektiğinde, sizin kendi yanıtladığınız ölçeği bulabilmemiz açısından TC numaranızın ilk 5 rakamı gerekmektedir.

Katkılarınız için çok teşekkür ederiz.

Ü. Sonsuzay KUZU
Mersin Üniversitesi Eğitimde Ölçme ve
Değerlendirme Abd. Yüksek Lisans Öğrencisi

TC numaranızın ilk 5 rakamı =

__ __ __ __ __

İFADELER	EVET	HAYIR
1. Hakkımı aramasını pek beceremem.		
2. Çevre koruma derneklerine destek veririm.		
3. Evimi temiz tutarım.		
4. Yiyecek alırken son kullanma tarihine mutlaka bakarım.		
5. Trafik kurallarına uyarım.		
6. Haksızlığa uğradığımda hakkımı aramak için çaba gösteririm.		
7. Park ve bahçelere zarar vermem.		
8. Bir felaket durumunda yardıma koşarım.		
9. Çok dağınık yaşarım.		
10. "Ülkenin sorunlarını çözmek bana mı kaldı" diye düşünürüm.		
11. Sokağa çöp atmam.		
12. Suyu gereksiz kullananları uyarıp da başıma iş almam.		
13. Doktora gitmeye üşenirim.		
14. Evdeki kir ve toz beni rahatsız eder.		
15. Ana haberleri kaçırmamaya özen gösteririm.		
16. "Bu ülkeyi ben mi kurtaracağım diye düşünürüm.		
17. Beslenmeme pek dikkat etmem.		
18. İşyeri ya da okulda gereksiz yanan lambaları kapatırım.		
19. Parklara zarar verenleri uyarırım.		
20. Sağlığıma pek dikkat etmem.		
21. Trafik kurallarına uymayanları uyarmak bana düşmez.		

Yaşınız

:

Eğitim Düzeyiniz

İlkokul

Ortaokul

Lise

Üniversite

Medeni Durumunuz

Evli

Bekar

Dul

EK 3:

PÇT İndeksi İçin Hazırlanan Excel Programı

toplam n= 308

83üst 83alt (%27)

n=166

sıra no	tekler üst
1	B5
2	B6
3	B7
4	B8
5	B9
6	B10
7	B11
8	B12
9	B13
10	B14
11	B15
12	B16
13	B17
14	B18
15	B19
16	B20
17	B21
18	B22
19	B23
20	B24
21	B25
22	B26
23	B27
24	B28
25	B29
26	B30
27	B31
28	B32
29	B33
30	B34
31	B35
32	B36
33	B37
34	B38
35	B39
36	B40
37	B41
38	B42
39	B43
40	B44
41	B45
42	B46
43	B47
44	B48
45	B49
46	B50
47	B51
48	B52
49	B53
50	B54

çiftler üst
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14
D15
D16
D17
D18
D19
D20
D21
D22
D23
D24
D25
D26
D27
D28
D29
D30
D31
D32
D33
D34
D35
D36
D37
D38
D39
D40
D41
D42
D43
D44
D45
D46
D47
D48
D49
D50
D51
D52
D53
D54

koşul
=EĞERSAY(D5:D87;B5)
=EĞERSAY(D5:D87;B6)
=EĞERSAY(D5:D87;B7)
=EĞERSAY(D5:D87;B8)
=EĞERSAY(D5:D87;B9)
=EĞERSAY(D5:D87;B10)
=EĞERSAY(D5:D87;B11)
=EĞERSAY(D5:D87;B12)
=EĞERSAY(D5:D87;B13)
=EĞERSAY(D5:D87;B14)
=EĞERSAY(D5:D87;B15)
=EĞERSAY(D5:D87;B16)
=EĞERSAY(D5:D87;B17)
=EĞERSAY(D5:D87;B18)
=EĞERSAY(D5:D87;B19)
=EĞERSAY(D5:D87;B20)
=EĞERSAY(D5:D87;B21)
=EĞERSAY(D5:D87;B22)
=EĞERSAY(D5:D87;B23)
=EĞERSAY(D5:D87;B24)
=EĞERSAY(D5:D87;B25)
=EĞERSAY(D5:D87;B26)
=EĞERSAY(D5:D87;B27)
=EĞERSAY(D5:D87;B28)
=EĞERSAY(D5:D87;B29)
=EĞERSAY(D5:D87;B30)
=EĞERSAY(D5:D87;B31)
=EĞERSAY(D5:D87;B32)
=EĞERSAY(D5:D87;B33)
=EĞERSAY(D5:D87;B34)
=EĞERSAY(D5:D87;B35)
=EĞERSAY(D5:D87;B36)
=EĞERSAY(D5:D87;B37)
=EĞERSAY(D5:D87;B38)
=EĞERSAY(D5:D87;B39)
=EĞERSAY(D5:D87;B40)
=EĞERSAY(D5:D87;B41)
=EĞERSAY(D5:D87;B42)
=EĞERSAY(D5:D87;B43)
=EĞERSAY(D5:D87;B44)
=EĞERSAY(D5:D87;B45)
=EĞERSAY(D5:D87;B46)
=EĞERSAY(D5:D87;B47)
=EĞERSAY(D5:D87;B48)
=EĞERSAY(D5:D87;B49)
=EĞERSAY(D5:D87;B50)
=EĞERSAY(D5:D87;B51)
=EĞERSAY(D5:D87;B52)
=EĞERSAY(D5:D87;B53)
=EĞERSAY(D5:D87;B54)

51	B55	D55	=EĞERSAY(D5:D87;B55)
52	B56	D56	=EĞERSAY(D5:D87;B56)
53	B57	D57	=EĞERSAY(D5:D87;B57)
54	B58	D58	=EĞERSAY(D5:D87;B58)
55	B59	D59	=EĞERSAY(D5:D87;B59)
56	B60	D60	=EĞERSAY(D5:D87;B60)
57	B61	D61	=EĞERSAY(D5:D87;B61)
58	B62	D62	=EĞERSAY(D5:D87;B62)
59	B63	D63	=EĞERSAY(D5:D87;B63)
60	B64	D64	=EĞERSAY(D5:D87;B64)
61	B65	D65	=EĞERSAY(D5:D87;B65)
62	B66	D66	=EĞERSAY(D5:D87;B66)
63	B67	D67	=EĞERSAY(D5:D87;B67)
64	B68	D68	=EĞERSAY(D5:D87;B68)
65	B69	D69	=EĞERSAY(D5:D87;B69)
66	B70	D70	=EĞERSAY(D5:D87;B70)
67	B71	D71	=EĞERSAY(D5:D87;B71)
68	B72	D72	=EĞERSAY(D5:D87;B72)
69	B73	D73	=EĞERSAY(D5:D87;B73)
70	B74	D74	=EĞERSAY(D5:D87;B74)
71	B75	D75	=EĞERSAY(D5:D87;B75)
72	B76	D76	=EĞERSAY(D5:D87;B76)
73	B77	D77	=EĞERSAY(D5:D87;B77)
74	B78	D78	=EĞERSAY(D5:D87;B78)
75	B79	D79	=EĞERSAY(D5:D87;B79)
76	B80	D80	=EĞERSAY(D5:D87;B80)
77	B81	D81	=EĞERSAY(D5:D87;B81)
78	B82	D82	=EĞERSAY(D5:D87;B82)
79	B83	D83	=EĞERSAY(D5:D87;B83)
80	B84	D84	=EĞERSAY(D5:D87;B84)
81	B85	D85	=EĞERSAY(D5:D87;B85)
82	B86	D86	=EĞERSAY(D5:D87;B86)
83	B87	D87	=EĞERSAY(D5:D87;B87)
		toplam üst frekans	=TOPLA(F5:F87)

%27 alt-üst birey sayısı	birey sayısı	166
=1-(((190/2-F88)+(190/2-F176))/190)	çift tutarlık	

	tekler alt	çiftler alt	koşul
1	B95	D95	=EĞERSAY(D95:D177;B95)
2	B96	D96	=EĞERSAY(D95:D177;B96)
3	B97	D97	=EĞERSAY(D95:D177;B97)
4	B98	D98	=EĞERSAY(D95:D177;B98)
5	B99	D99	=EĞERSAY(D95:D177;B99)
6	B100	D100	=EĞERSAY(D95:D177;B100)
7	B101	D101	=EĞERSAY(D95:D177;B101)
8	B102	D102	=EĞERSAY(D95:D177;B102)
9	B103	D103	=EĞERSAY(D95:D177;B103)
10	B104	D104	=EĞERSAY(D95:D177;B104)
11	B105	D105	=EĞERSAY(D95:D177;B105)
12	B106	D106	=EĞERSAY(D95:D177;B106)
13	B107	D107	=EĞERSAY(D95:D177;B107)
14	B108	D108	=EĞERSAY(D95:D177;B108)

15	B109
16	B110
17	B111
18	B112
19	B113
20	B114
21	B115
22	B116
23	B117
24	B118
25	B119
26	B120
27	B121
28	B122
29	B123
30	B124
31	B125
32	B126
33	B127
34	B128
35	B129
36	B130
37	B131
38	B132
39	B133
40	B134
41	B135
42	B136
43	B137
44	B138
45	B139
46	B140
47	B141
48	B142
49	B143
50	B144
51	B145
52	B146
53	B147
54	B148
55	B149
56	B150
57	B151
58	B152
59	B153
60	B154
61	B155
62	B156
63	B157
64	B158
65	B159
66	B160
67	B161
68	B162

D109
D110
D111
D112
D113
D114
D115
D116
D117
D118
D119
D120
D121
D122
D123
D124
D125
D126
D127
D128
D129
D130
D131
D132
D133
D134
D135
D136
D137
D138
D139
D140
D141
D142
D143
D144
D145
D146
D147
D148
D149
D150
D151
D152
D153
D154
D155
D156
D157
D158
D159
D160
D161
D162

=EĜERSAY(D95:D177;B109)
=EĜERSAY(D95:D177;B110)
=EĜERSAY(D95:D177;B111)
=EĜERSAY(D95:D177;B112)
=EĜERSAY(D95:D177;B113)
=EĜERSAY(D95:D177;B114)
=EĜERSAY(D95:D177;B115)
=EĜERSAY(D95:D177;B116)
=EĜERSAY(D95:D177;B117)
=EĜERSAY(D95:D177;B118)
=EĜERSAY(D95:D177;B119)
=EĜERSAY(D95:D177;B120)
=EĜERSAY(D95:D177;B121)
=EĜERSAY(D95:D177;B122)
=EĜERSAY(D95:D177;B123)
=EĜERSAY(D95:D177;B124)
=EĜERSAY(D95:D177;B125)
=EĜERSAY(D95:D177;B126)
=EĜERSAY(D95:D177;B127)
=EĜERSAY(D95:D177;B128)
=EĜERSAY(D95:D177;B129)
=EĜERSAY(D95:D177;B130)
=EĜERSAY(D95:D177;B131)
=EĜERSAY(D95:D177;B132)
=EĜERSAY(D95:D177;B133)
=EĜERSAY(D95:D177;B134)
=EĜERSAY(D95:D177;B135)
=EĜERSAY(D95:D177;B136)
=EĜERSAY(D95:D177;B137)
=EĜERSAY(D95:D177;B138)
=EĜERSAY(D95:D177;B139)
=EĜERSAY(D95:D177;B140)
=EĜERSAY(D95:D177;B141)
=EĜERSAY(D95:D177;B142)
=EĜERSAY(D95:D177;B143)
=EĜERSAY(D95:D177;B144)
=EĜERSAY(D95:D177;B145)
=EĜERSAY(D95:D177;B146)
=EĜERSAY(D95:D177;B147)
=EĜERSAY(D95:D177;B148)
=EĜERSAY(D95:D177;B149)
=EĜERSAY(D95:D177;B150)
=EĜERSAY(D95:D177;B151)
=EĜERSAY(D95:D177;B152)
=EĜERSAY(D95:D177;B153)
=EĜERSAY(D95:D177;B154)
=EĜERSAY(D95:D177;B155)
=EĜERSAY(D95:D177;B156)
=EĜERSAY(D95:D177;B157)
=EĜERSAY(D95:D177;B158)
=EĜERSAY(D95:D177;B159)
=EĜERSAY(D95:D177;B160)
=EĜERSAY(D95:D177;B161)
=EĜERSAY(D95:D177;B162)

69	B163
70	B164
71	B165
72	B166
73	B167
74	B168
75	B169
76	B170
77	B171
78	B172
79	B173
80	B174
81	B175
82	B176
83	B177

D163
D164
D165
D166
D167
D168
D169
D170
D171
D172
D173
D174
D175
D176
D177

toplam alt frekans

=EĞERSAY(D95:D177;B163)
=EĞERSAY(D95:D177;B164)
=EĞERSAY(D95:D177;B165)
=EĞERSAY(D95:D177;B166)
=EĞERSAY(D95:D177;B167)
=EĞERSAY(D95:D177;B168)
=EĞERSAY(D95:D177;B169)
=EĞERSAY(D95:D177;B170)
=EĞERSAY(D95:D177;B171)
=EĞERSAY(D95:D177;B172)
=EĞERSAY(D95:D177;B173)
=EĞERSAY(D95:D177;B174)
=EĞERSAY(D95:D177;B175)
=EĞERSAY(D95:D177;B176)
=EĞERSAY(D95:D177;B177)
=TOPLA(F95:F177)