M.Ü.Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi
Yıl: 1995, Sayı : 7 Sayfa 125-148

1. ÖLÇEKLERDE GEÇERLİK VE GÜVENİRLİK

Dr.Demirali Y.Ergin

Her ölçme aracında vazgeçilemez iki özelliğin olması aranır: Bunlar geçerlik ve güvenirliktir. Bir ölçme aracı, her şeyden önce ölçülecek özelliği tam ve doğru olarak ölçmeli, başka bir özellikle karıştırmadan ölçmelidir: ölçeğin bu niteliğine geçerlik denir Geçerlik, testin kullanılış amacına uygun hizmet etme derecesini belirler. Bir ölçme aracı ölçtüğü özelliği tutarlı olarak ölçebilmelidir. Aynı şartlar altında tekrarlandığında aynı sonuçları verebilmelidir. Bu ise güvenirlik olarak adlandırılır.

Öncelikle, geçerlik ve güvenirlik çalışmalarının hangi ölçme araçları için söz konusu olduğunu değerlendirmek gerekir. Her- ölçme aracı geçerlik ve güvenirlik özelliğinin her ikisini birden taşımalıdır. Ama ölçme araçlarının ne ölçüde geçerli ve güvenilir' olduğunu belirlemek istatistik biliminden yararlanılarak hile olsa her zaman kolay ve hatta mümkün olmayabilir. Doğa bilimlerinde olduğu gibi olgusal verilerin elde edildiği ölçme araçları için geçerlik ve güvenirlik hesaplanabilir ve hesaplanmalıdır. Özellikle psikolojik temelli bilimlerde olduğu gibi yargısal verilerin elde edildiği çileme araçlarında ise farklı durumlar ortaya çıkmaktadır', İster 'yaparım gibi doğrudan ifadeyle (değişkenin gözlenebilir olgular sayesinde doğrudan ölçülmesi anlamında değil ı ister 'yapılmaktadır' gibi dolaylı ifadeyle alınıyor olsun tüm psikolojik içerikli ölçeklerde ölçülenler tutum genel kalıbı içindedir. Tutumlarda, verilen cevapların hiçbiri için doğru-yanlış söz konusu değildir. Verilen cevap ne olursa olsun o kendine göre algılamasını ifade etmektedir, tutumunu yansıtmaktadır. Tutum ölçerlerden elde edilen ölçümler arası farklar: gerçek farkları ve sadece random hataları (burada farkı doğuran araç kökenli tüm etkenler random nitelikli olduğundan, yanlı hata olmadığından) içerdiğinden geçerlik ve güvenirlik çalışması yapılmalıdır ve anlamlıdır. Tutum ölçeklerde ölçek düzeyinde, faktör düzeyinde veya madde düzeyinde bir puan elde edildiği için de istatistik işlemlerin yapılması için gerekli nitelikte veri elde edilebildiğinden istatistiksel geçerlik ve güvenirlik çalışması yapılabilir'.

Yine yargısal verilerin görüş, düşünce ve durum bildirim seklinde alındığı anketlerde elde edilen veriler tutum ölçeklerinde veya basarı testlerinde olduğu gibi puan cinsinden değil de genellikle sınıflama türündendir. Elde edilen sayısal verilerin de bir bütün olarak anlamı yoktur. Yaşınız sorusuna verilen ' :i:'J ile aylık geliriniz sorusuna verilen 20 milyon arasında (iletilen değişken bakımdan birliktelikten bahsedilmez. Aynı değişkeni ölçmediği açıkça belli olan soruların,

örneğin birbiriyle iç tutarlığı olup olmadığını belirlemeyi düşünmek ne anlama gelir.

Güvenirliği kontrol etmek için uygulamada sıkça yapıldığı gibi aynı soruyu değişik ifadelerle anketin veya tutum ölçerin farklı verinde kontrol sorusu adı altında tekrarlama beklemen amaca hizmet eder mi? Hayır. Çünkü: eğer aynı içerik farklı ifadelerle sunulmuşsa artık farklı sorudur. Benzer' olmaları eş' olmaları anlamına gelmeyebileceği için ortaya çıkan ölçüm farklılığının ifadedeki ölçüm farklılaşmadan mı yoksa başka etkenlerden mi olduğunu ayırt edecek bir teknik bulunmamaktadır.

Bir ölçme aracı olması nedeniyle tabiî ki anketlerin de geçerliği-güvenirliği söz konusudur. Ama sorun 'nasıl? belirlenebileceğidir. Öncelikle, anketten elde edilen sınıflama türündeki verilerin bilinen güveniriik-geçerlik istatistik yöntemlerine uygun olmadığı dikkate alınarak 'anketlerin geçerliği ve güvenirliği hesaplanamaz" değerlendirmesi doğru olabilir.

Verilen cevapların doğru-yanlışlığı (gerçeğe, olgulara uygunluğu veya aykırılığı) söz konusudur. Yaşınız?' sorusuna 30 yaşında olduğu halde **15** cevabını vermesi olguyu yansıtmaması bakımından yanlış bir cevaptır. Böyle yanlış cevapların alınabildiği anketten elde edilen ölçüm sonuçlarının içine yanlı hata karışmış olmaktadır. Ölçülen nesneler arasındaki ölçüm farkları sadece ölçülenden kaynaklanan fark olmaktan çıkıp yanlı hatayı da içerecektir. Oysa istatistik ister sabit, ister değişken olsun yanlı hataları hesaplayamaz. İstatistik yanlı hatalardan arınmış, random hataları (belki) içeren verilerle çalışır. Diğer taraftan güvenirlik, random hatalardan arınmış olmayı içeren bir kavramdır. Ölçümlerinde sürekli 1 kilo fazla tartan bir basküle çıktığınızda her seferinde sizin ağırlığınızı 60 kilo olarak belirlemesi, bu ölçümleri arasındaki tutardık onun güvenilir olduğunu göstermez. 1 kiloluk yanlı hatayı dikkate almamış olursunuz.

1.1. Geçerlik

Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan, doğru ve tam olarak ölçebilmesidir. Geçerlik, ölçülmek istenen değişkenin ölçülebilmiş olma derecesidir. Geçerliğin yüksek olması, ölçülmek istenen kavramın gözlenebilir nitelikteki değişkenlerle ifade edilebilmesine bağlıdır. Bu nedenle doğrudan ölçmelerde geçerlik, dolaylı ölçmelere göre daha yüksektir.

Bir ölçeğin geçerliği; objektiflik, ayırt edicilik (maddelerin ayırt etme gücü güvenirlik değil bir geçerlik sorunudur,!, kapsamlılık, kolay uygulanabilirlik ve puanlanabilirlikten etkilenir.

Güvenirlik ve geçerlik birbirinden tamamen ayrı iki kavram olup elde edilen istatistik değerleri arasında da hiçbir ilişki yoktur. Elde edilişleri birbirlerinden azadedir. Bir ölçme aracının mutlaka hem geçerli hem güvenilir olması anlamında birbirlerini bütünlerler'. Bunlardan biri diğerinin önkoşulu veya ikamesi değildir. Her ikisinin birlikte olmasıyla ölçek anlam bulur. Örneğin, ölçeğin güvenilir olması onun geçerliğini garantilemez.

Geçerlik konusu "içerik geçerliği ve yapı geçerliği" olmak üzere iki başlık altında incelenebilir.

1.1.1. İçerik Geçerliği (Content Validity)

Ölçme aracının tanımlanan davranış tepki evrenini yeterince temsil edebilmesidir. Önce ölçülmesi istenen kavram ile ilgili davranışlar evreninin çözümlemesi yapılarak içerdiği etkenlerin açık olarak saptanması gerekir. Sonra da, her etkenin davranış evrenindeki oranına göre ölçme aracında temsil edilip edilmediğine bakılır. İçerik geçerliği ölçülen konudaki tüm boyutlardaki olası tüm maddelerden oluşan tepki (soru, madde) evrenini, ölçeğin (sınırlı sayıda madde içermesi nedeniyle bu evrenden bir örnekleme niteliğindedir! Temsil etme gücüdür. Temsil ediciliği sağlamak için her alt boyutu içeren maddeler ölçekte yer almalıdır ve tepki evrenindeki oranı veya önem ağırlığı ölçeğe de yansımalıdır.

Örneğin, matematik bilgisi ölçülmek istendiğinde, öğrenciye sadece toplama işlemini sormak onun dört işlemin diğer öğelerini bilip bilmediğini anlamaya yeterli olmaz. Bu nedenle ölçülmek istenen amacın tümünü kapsayan sorular sorulmalıdır.

İçerik geçerliğini, aynı değişkenle ilgili bir başka ölçeğin sonuçları ile korelasyon arayarak belirlemek, sonuçta istatistik bir değere dayanarak içerik geçerliğinden bahsetmek yaygın bir kanı olmakla beraber yanıltıcıdır. Her şeyden önce 'ölçüt olarak alman ölçeğin içerik geçerliği nasıl hesaplanmıştır?" sorusuna alınacak "bir başka ölçme aracını ölçüt alarak" cevabı kısır döngüyü gözler önüne serip içerik geçerliğini hesaplama hevesini boşa çıkaracaktır.

1.1.1.1. Yüzeysel Geçerlik (Faee Validity)

Ölçme aracının hangi değişkeni ölçtüğü hakkındaki uzman görüşüdür. Geçerlik seviyesini sayısal değerle belirtme olanağı yoktur. Sadece kanaatlere göre bir kabul söz konusudur. Geliştiren ölçme araçlarında ilk başvurulan geçerlik türüdür. Bir veya birkaç uzmanın görüşüne başvurularak, ölçme aracının kullanılacağı amaç için gerekli veriyi toplayacak durumda olup olmadığı yönünde alınan bilgi yüzeysel geçerlik için yeterli görülmektedir. Yüzeysel geçerlik ölçme aracının hangi değişkeni ölçtüğünü değil, ölçer gibi göründüğünü belirler.

1.1.1.2. Uygulama Geçerliği (Empirical Validity)

Ölçülmeye çalışılan değişkenin gerçek hayattaki gözlemlenebilir belirtileri ile ölçme sonuçları arasındaki uyum. uygulama geçerliğidir. Uygulama geçerliğini belirlerken dış ölçüt ile ilişkisine bakılır. Ölçüt, ölçmek istenen davranışı ne dereceye kadar' yansıtırsa, bulunan ilişki ölçme aracının geçerliği hakkında o derecede sağlam bilgi verir. Ölçüt saptanması ölçme aracının niteliğine ve ölçmenin amacına göre farklılık gösterir. Örneğin okulda alman notlar ile iş hayatındaki başarı arasında yüksek bir ilişki bulunmuşsa okuldaki notların uygulama geçerliği olduğu söylenebilir. Burada ölçüt, iş hayatındaki basarıdır. Atılganlık ölçeğinde yüksek puan alan bireyler, günlük yaşamlarında yakın

çevrelerince pısırık olarak tanınıyorlarsa atılganlığı ölçmek için kullanılan ölçeğin pratik geçerliğinin olduğu düşünülemez.

Ölçeklerde uygulama geçerliği hakkında bazı bilgiler elde etmek için su yaklaşımlar kullanılabilir.

- Ölçülen değişken bakımından bireyleri bu açıdan tanıdıklarına inanılan başka kişiler, onları sıralar veya sınıflar (atılgan-çekingen vb). Ölçülen bireyler hakkında gözlem yoluyla da aynı veriler toplanabilir,. Elde edilen bu veriler ile ölçüm sonuçları arasındaki yüksek uyum kullanılan ölçeğin uygulama geçerliğinin olduğunu ifade edebilir,

- Ölçülen değişken bakımından uç değerlere sahip oldukları hakkında yaygın kanaatin bulunduğu birey ya da gruplara geliştirilen ölçek uygulanır. Hakim olan kanaat ile ölçek sonuçları arasındaki uyum yüksekse uygulama geçerliğinin olduğu, ölçeğin uygun içerikte olduğu yönünde ipucu elde edilmiş olur. Örneğin akademik yetenekler ölçümünde kullanılacak bir test hazırlandığında araç, zihinsel gücü bakımından yukarı (üstün okulları) ve aşağı (alt özel sınıflar, vb) seviyeleri temsil eden bireylere uygulanır. Yukarı seviye diye bilinenler ölçekte de üst puanları almış, aşağı seviye diye bilinenler de ölçekte alt puanları almış ise, ayrıca bu iki karşıt grubun ölçek puanlan karşılaştırıldığında üstünler-yönünde manidar bir farklılık belirmişse ölçeğin uygulama geçerliği olduğu yönünde veri elde edilmiş olur.

Uygulama geçerliği ilgilendiği zaman dilimine göre "halihazır geçerlik" veya 'kestirirsel geçerlik adını alır. Şimdiki zamandaki uygulama geçerliği araştırılıyorsa hâlihazır geçerlik, gelecek zamandaki uygulama geçerliği araştırılıyorsa kestirirsel geçerlik söz konusudur.

Bu uygulama geçerliği tür ve tekniklerinden hangisiyle olursa olsun elde edilen geçerlik, içerik geçerliği hakkında fikir verecekti r. Uygulama geçerliği içerik geçerliğinin dışında bir anlam taşımaz Çünkü uygulama geçerliğinde kullanılan ölçütler ölçülen değişkenle yani içerikle ilgili olmak zorundadır. Uygulama geçerliğini belirlemede biricik sorun objektif veriler- elde ettiğimiz (ya da etmeye çalıştığımızı ölçek sonuçlarının uygulamayla karşılaştırmasının yaygın kanı. Ön kabul gibi sübjektif verilere dayandırılmasıdır.

1.1.1.2.1. Kestirisel Geçerlik (Predictive Validity)

Değişkenler arası ilişkileri anlayarak olayları önceden tahmin edebilip kontrol altına almak bilimin nihai amacıdır. Bilimin ortak amaçlarından olan geleceğe yönelik tahmin, aleme araçları için de söz konusudur-. Sosyal bilimlerde ölçme araçları, genellikle bireylerin gelecekte gösterecekleri davranışları tahmin etmek amacıyla geliştirilirler. Bir zeka testi ilerideki akademik başarıyı, bir yetenek testi meslekteki başarıyı, bir politik tutum ölçeği belli niteliklerdeki kişilerin hangi partiye oy vereceklerini önceden tahmin etmeye olanak sağlayabilmelidir.

Kestirisel geçerliğin saptanmasında izlenen yol şöyledir. Ölçme aracı uygulanarak sonuçlar alınır. Ölçülen niteliğin belirgin olarak görülebileceği yeterli bir süre beklenir. Belirlenen ölçüt acısından da uygulamadaki durumun

değerlendirmeleri yapılır. Geliştirilen testin sonuçları ile ölçüt değişken sonuçları arasında manidar bir ilişki bulunmuşsa testin tahmin geçerliği olduğu kabul edilir.

1.1.1.2.2. Halihazır Geçerlik (Concurrent Validity)

Her şeyden önce belli bir zaman aralığı gerektirdiğinden çoğu kere önce ölçme aracının uygulanması, bir süre sonra ölçüt değişken değerlendirmesinin yapılması mümkün veya uygun olmayabilir. Bu durumlarda halihazır geçerlik ile yetinilir. Halihazır geçerlikte, ölçme aracı kullanıldığında, ölçüt değişken değerlendirmesi de avın zamanda yapılır.

l.l.l.3 Ölçeğin İctutarlığı

Ölçeğin içerik geçerliği hakkındaki değerlendirmelerde enendi ölçütlerden biri de aracın içtutarlıgıdır. İçtutarlık temelde bir güvenilirlik sorunudur. Ölçekleme varsanımlarından tok boy ut tuluğu il karşılanma durumunu belirlemek için hesaplanır. Güvenirlik baslığı altında ayrıntıları verildiği gibi bir faktörü oluşturan maddeler birbiriyle ne kadar yüksek ilişki gösterirse tek boyutlu oldukları, yani ölçülen değişken her ne ise de aynı değişkeni ölçtükleri kararına varılır.

İlişki ne kadar yüksekse o ölçüde tek boyutluluk hakimdir. Güvenilirlik açısından durum böyle olmakla beraber geçerlik açısındansa tepki evreni temsil etmek bakımından maddeler birbirinden ne kadar farklı ise aynı değişkenle ilgili o kadar çok örnekleme yapılmıştır. Ölçeği oluşturan faktörler veya faktörü oluşturan maddeler hem birbiriyle ilişki olacak (içtutarlık için) hem de birbirinden farklı olacak (içerik geçerliği için). Çünkü birbirinden farklı olmayan maddelerin ölçekte bulunmasının cevaplayanı bunaltmaktan başka bir etkisi olmayacaktır, faydası ise hiç. İlk bakışta çelişki gibi gözüken bu durum istatistik teknikler açısından çok normaldir. İçtutarlık için ilişki belirleyen korelasyon tekniklerinin, içerik geçerliği için ise fark belirleyen (t-testi, varyans analizi) tekniklerinin her ikisinin de kullanılması gereğini hatırlatmaktadır. Çünkü istatistiksel olarak iki veri dizisi birbiriyle hem ilişkili olabilir ama aynı zamanda da farklı olabilir

İçtutarlık için beliren bu değerlendirmeye 'yapı geçerliği" başlığında açıklanan faktör analizi tekniği kısmen (çünkü tüm korelasyon-kovaryanslara göredir) kolaylık getirmektedir. Faktör analizinin hesaplarıma tarzı olarak birbiriyle en fazla ilgili maddeler aynı faktör etrafında toplanarak ilişkiyi, birden fazla faktör olması ise farkı yansıtmaktadır.

1.1.2. Yapı Geçerliği (Construct Validity)

Soyut kavramlara yönelik ölçmelerde önce ölçülen kavramı tanımlayan kuramlardan biri tercih edilir. Böylece ölçülmek istenen kavramın yapısı belirlenir. Bu kuramsal yapıya göre gözlenebilir değişkenler ortaya konur. Son olarak bu gözlenebilir değişkenlerini ifadelendiren maddeler yazılarak ölçek hazırlanır. Ölçek geliştirildiğinde maddelerin hangi faktörleri temsilen

yazıldığını araştırmacı bilmektedir. Yani teorik yapıya bağlı ölçek yapısı belirlidir.

Kavram ile değişkenin özdeş olmadığı (ki sosyal bilimlerdeki ölçmelerin tümünde kavram ve değişken özdeş değildir), bunun yanında ölçmenin gözlenebilir değişken değil de onun belirtileri üzerinden yapıldığı durumlarda, ölçülenin ölçülmek istenen kavramın belirtileri olup olmadığının belirlenmesi gerekir. Yapı geçerliği, bu belirleme çalışmalarının sonucudur Diğer taraftan yapı geçerliği, ölçeğin temelindeki kuramların geçerliği ile de ilgilidir.

Ölçeğin uygulanmasından elde edilen veriler, "faktör analizi" istatistik tekniği ile işlenir. Çeşitli faktör analizi teknikleri olmakla beraber (varimax rotated faktör analizi en yaygın kullanım bulanıdır) temel mantık aynıdır. Maddeler arasındaki olası tüm ilişkiler hesaplanıp bir korelasyon matrisi oluşturulduktan sonra aralarında beliren ilişki yapılarına göre farklılık gösterdikleri oranda faktör adedi belirir, her maddenin her faktör içindeki ağırlığını ifade eden katsayılardan (maddenin faktörle ilişkisi olarak yorumlanabilir) oluşan bir tablo elde edilir. Bu tablo üzerinde yapılan değerlendirmede her madde en yüksek ağırlığı hangi faktörde bulmuşsa o faktörün kapsamında olmasına karar verilir. Bu maddenin her-hangi bir faktöre girebilmesi için ulaşması gereken asgari bir değer istatistiksel olarak söz-konusu değildir, konulan barajlar keyfîdir. Sanıldığının aksine faktör analizi sonunda şu maddeler şu faktörü oluşturur gibi kesin bir yargıyı gerektirecek istatistik karar çıkmaz. Faktör-madde ilişkisi pratikte, ideal değerini bulmadığından bir madde birden fazla faktörle yüksek ilişki gösterebilir. Bazı yabancı test uygulamalarında görüldüğü gibi bir madde birden fazla faktörde yer alabilir. Bu durumun uygulamada bazı değerlendirme ve yorumlama güçlüklerine yol açabileceğini de göz önünde tutmak gerekir.

Ölçeğin başlangıçta belirlenen teorik yapısı ile uygulama sonuçlarına dayanan faktör analizi işlemi sonrasında beliren pratik yapı birbirine uyumlu ise ölçeğin yapı geçerliği vardır. Hem ölçeğin dayandığı kuramsal yapı doğrudur hem de ölçek bu kuramsal yapıyı ölçebilecek niteliktedir.. Teorik yapı ile pratik yapıyı karşılaştırmaya, ikisi arasındaki uyumu belirlemeye hizmet eden bir istatistik teknik henüz maalesef yok.

Öğrencilerin ders çalışma alışkanlıklarını belirlemeye çalışan araştırmacının "zamanı kullanma" boyutu ile ilgili olduğunu düşünüp yazdığı maddeler faktör analizi sonuçlarında da bir faktör olarak belirmişse yapı geçerliğinin olduğu ve boyutun "zamanı kullanma" kavramını ölçtüğü düşünülebilir. Ama bazı maddeler başlangıçta düşünülenin dışında maddelerle bira raya gelmişse yapı geçerliği yoktur. Böyle bir durumda ya ölçek, önerilen kuramsal yapıyı ölçmekten uzaktır ve/veya ölçeğin dayandırıldığı kuramsal yapı yanlıştır.

Ölçeğin yapı geçerliği için faktör analizinin yapılması şarttır. Bunun yanında sonuçları desteklemek amacıyla, daha önceden yapı geçerliği faktör analiziyle belirlenmiş ve aynı kuramsal yapıya dayanan bir başka ölçme aracı ile birlikte uygulanarak sonuçlar arasındaki uyum değerlendirilebilir

Faktör analizi ile ilgili iki yanlış kanı vardır. Birincisi; herhangi bir kuramsal yapı belirlemeden faktör analizi sonuçlarını olduğu gibi kabul edip ölçeği yapılandırmak. Bu elbette kolaylık sağlamaktadır, ama başlangıçta belirlenen bir kuramsal yapı olmadığından bu tarz hareketin yapı geçerliğini sağladığını düşünmek yanlış olur. herhangi bir yapı belirlenmemişti ki geçerli olup olmadığına karar verilmiş olsun. İkincisi; yabancı dilde ve ülkede geliştirilen bir ölçek Türkçeye çevrilirken (dilsel adaptasyon,) bu ölçeği oluşturan maddelerin içeriği aynı kalırken faktör analizi yapmakla yeniden yapılandırarak, yapısal adaptasyon hakkını da kendinde görmektir. Öncelikle, ölçeği geliştiren biz olmadığımız halde geliştirenin belirlediği yapıyı değiştirmek, bozmak bilimsel etik bakımından ne ölçüde uygundur? Çünkü ölçeğin yapılandırılması belli bir kuramsal yapıyı yansıtır, başkası adına bu kuramsal yapıyı biz nasıl belirleyebiliriz. Adapte eden biz olsak bile ölçeğin yasal ve bilimsel sahibi orijinalini geliştirendir. Var olan orijinal kuramsal yapı eğer beğenilmiyorsa dilsel uygulama çalışmasına girmek gereksiz, yok eğer beğeniliyorsa niye yapı değiştirilmek istensin. Faktör analizi ölçek geliştirilmesinde kullanılabilecek bir tekniktir, adaptasyonunda değil. Uyarlaması yapılan ölçeğin orijinal ile eşdeğer olduğunu iddia edebilmek için gerekli koşullar eşdeğerlik başlığı altında incelenmiştir Ölçeğin yapısının değiştirilmesi eşdeğerliği ortadan kaldırır, genel hatlarıyla aynı değişkeni ölçen farklı bir ölçek ortaya çıkmış olur. Dolayısıyla bu iki ölçekten elde edilen sonuçlar hiçbir şekilde karşılaştırılamaz. Bu ise bilimin evrenselleşmesi doğrultusundaki kapsayıcı genellemelere ulaşılmasına bir engeldir,

1.1.3. Ayırt etme Gücü

Bir maddenin ayırt etme gücü, ölçülen değişken bakımından birimler arası farklılığı ne ölçüde ortaya çıkarabildiği ile ilgilidir. Ölçmenin temel amacı ölçülen nesnelerdeki farkı yakalayabilmek olduğuna göre ayırt etme gücü ayrı bir önem kazanmaktadır. Bir maddeye, herkes aynı cevabı vermiş ise diğer özellikleri ve önemi ne olursa olsun kimseyi diğerinden ayırt etmediği için maddeyi ölçekte tutmanın bir yararı yoktur. Ayırt etme gücü zayıf maddelerin ayıklanması ile ölçek daha kısa ama etkili bir hale getirilmiş olur.

Maddelerin ayırt etme gücünün analizi için bireylerin, ölçekten aldıkları toplam puanı belirlenir ve bu toplam puana göre en büyükten en küçüğe doğru sıralanırlar, O maddeyi cevaplayan bireylerin %27 sinin kaç kişi olduğu belirlenir. Sıralamanın en üstündeki %27'lik grup ($n_{ü}$) ile en altındaki % 27'lik grup (n_{a}) belirlenir. Bu işlem öncesi hazırlıktan sonra ölçek değer ayrımı veya t-testi ile gereken sınama yapılır. İstatistik yetkinliği bakımından elbette t-testi tercih edilir. Ayırt etme gücünü belirleme tekniklerindeki temel yaklaşım, testin toplamında yüksek puan alanların incelenen madde de yüksek puan almaları gereğinin karşılanıp karşılanmadığıdır. Aynı şekilde ölçeğin toplamında düşük puan alanlar grubunda yer alan bireylerin madde puanlarının da düşük olması gerekir. Üst ve alt çeyrekte yer alan bireylerin her madde için bu konumlarını koruyup korumadıkları test edilir. Başarı testlerinden örneklemek gerekirse çok kolay olduğu için hemen herkes tarafından doğru cevaplandırılan veya çok zor

olduğu için hemen hiç kimse tarafından doğru cevaplandırılmayan sorular ayırt etme gücü bakımından zayıftır.

Tıpkı ölçeğin içtutarlılığındaki gibi ayırt etme gücü için de geçerlik-güvenirlik ikilemi vardır İçtutarlık maddeler arası, ayırt etme gücü ise madde içinde bireyler arası değişkenlik ile ilgilidir. Geçerlik açısından ayırt etme gücünün yüksek olması gerekir. Yani bireyleri puanlama bakımından birbirinden ne ölçüde farklılaştırabiliyorsa o ölçüde ayırt edicidir. Bu ise büyük ölçüde istatistiksel değer olarak varyansın büyümesine yol açar. Diğer taraftan ölçmenin standart hatası konusunda görüldü ki, varyans büyüdükçe standart hata da büyümektedir. Yani güvenirlik ölçütlerinden birine göre güvenirlik düşmektedir. Hatanın, güvenirliği yükselttiğini düşünmek mümkün mü? Bazı kaynaklarda rastlandığı gibi ayırt etme gücü artıkça güvenirlik artmaz, zincirleme olarak standart sapma büyür, standart bata büyür ve güven aralığının sınırları genişler. Güven aralığının sınırlarının genişlemesi güvenirliğin arttığına delalet etmez, sadece gerçek ölçüm sonucunu tahmin gücünün zayıfladığının ifadesidir.

TABLO 1. AYIRT ETME GÜCÜ HESAPLANIŞI (1. ÖRNEK)

Öğrenci	2,soru	Ölçek Toplamı	2. sorunun istatistik değerleri n	$\sum x$	$\sum x^2$	μ	α	α^2
A	3	40 ü	3	10	34	3.33	0.577	0.333
B	4	40 ü	üst çeyteklik					
C	3	35 ü						
D	2	33						
E	4	30						
F	3	29						
G	4	28						
H	2	26 a	3	4	6	1.33	0.577	0.333
I	1	16 a	alt çeyreklik					
İ	1	16 a						
Toplam			10	27	85	2,70	1.160	1.344

1.1.3.1. Ölçek Değer Ayrımı (D_a)

Ölçek değer ayrımı ölçek toplam puanına göle üst çeyrekte yer alanların aritmetik ortalamasından alt çeyrekte yer alanların aritmetik ortalamasının çıkarılmasıyla elde edilen katsayıdır.

$\mu_ü$: üst çeyreklikte yer alanların aritmetik ortalaması

μ_a: alt çeyreklikte yer alanların aritmetik ortalaması

D_a: maddenin ölçek değer ayrım katsayısı

$$\mathbf{D_a} = \mu_ü - \mu_a \qquad [11]$$

Tablo l'deki örneğe göre 10 kişilik bir grubun üst çeyreğinde yer alan 3 kişinin incelenen 2. maddeden aldıkları puanların aritmetik ortalaması 3.3 (10/3), alt grubun 1.33 (4/3)'dür. Bu maddenin ölçek değer ayrımı $D_g = 3.33 - 1.33 = 2$'dir.

Aynı çözümleme tekniğini bir başka form içinde örnekleyen Tablo 2'deki verilere göre 152 kişilik bir gruba uygulanan bir ölçekteki bir maddenin ölçek değer ayrımı 0.64 bulunmuştur.

Ölçek değer ayrım katsayısı ne kadar büyükse ilgili madde o ölçüde ayırıcı değere sahiptir. Ancak bununla ilgili istatistiksel bir manidarlık sınaması söz konusu değildir. Ne kadarlık miktardaki ölçek değer ayrımı, ayırt ediciliğin var olduğunu ifade ederin kararı objektif ölçütlere bağlanmamıştır. Ayrıca dizilerde sadece merkezi eğilim değerini dikkate alıp dağılım değerleriyle ilgilenmemesi nedeniyle grupları karşılaştırmada istatistiksel yetkinliği yüksek değildir. Bu nedenle aynı amaca hizmet eden ve istatistik açıdan manidarlık sınamasının yapıldığı t-testini, ayırt ediciliği belirlemede kullanmak çok daha yararlı olacaktır.

TABLO 2. AYIRT ETME GÜCÜ HESAPLANIŞI (2.ÖRNEK)

	Ağırlık	Üst Çeyreklik			Alt Çeyreklik		
	x	f	fx	fx^2	f	fx	fx^2
Kesinlikle katılıyorum	5	5	25	125	l	5	25
Katılıyorum	4	10	40	160	5	20	80
Kararsızım	3	20	60	180	20	60	180
Karşıyım	2	5	10	20	10	20	40
Kesinlikle Karşıyım	1	1	1	1	5	5	5
Toplam		**n**	**Σfx**	**$Σfx^2$**	**n**	**Σfx**	**$Σfx^2$**
		41	**136**	**486**	**41**	**110**	**330**
$\mu=\bar{x}=\frac{\Sigma X}{n}$		$\mu_{ü}=\frac{136}{41}=3.32$			$\mu_{a}=\frac{110}{41}=2.68$		
$\sigma^2=\frac{n\Sigma X^2-(\Sigma X)^2}{n(n-1)}$		$\sigma^2_{ü}=\frac{41 \times 486-136^2}{41 \times 41}=0.851$			$\sigma^2_{a}=\frac{41 \times 330-1106^2}{41 \times 41}=0.851$		
$n \geq 30$ ise $(n-1)$ yerine (n)							
$\Sigma fx=\Sigma x$ $\Sigma fx^2=\Sigma x^2$		Madde Ölçek Değer Ayrımı: 3.32-2.68=0.64					

1.1.3.2. t-testi (Kritik Oran)

İstatistikteki en yaygın tekniklerden olan t-testinin, karşılaştırılan grupların bazı özelliklerine göre farklı modelleri vardır. Maddelerin ayırt ediciliğini belirlemede kullanılacak olan t-testleri ilişkisiz gruplar için kullanılanıdır. T-testi genel form u şöyledir.

$$t=\frac{\overline{X_1}-\overline{X_2}}{\sigma_F} \qquad [1.2]$$

Farkın standart hatası (α_F) örneklem büyüklüğüne bağlı olarak farklı hesaplanmaktadır. Eğer n >=30 ise

$$\sigma_F = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad [1.3]$$

Eğer n<30 ise

$$\sigma_F = \sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}} \qquad [1.4]$$

t-testinin bu şekilde hesaplanmasında aritmetik ortalama ve standart sapma hesaplamalarında kesirli kesimdeki yuvarlamalardan ötürü sonuçta küçük farklılıklar görülebilir. Daha duyarlı bir (t) değerine ulaşılmak isteniyorsa formül [1.5] kullanılabilir. Eğer büyük örneklemlere göre işlem yapılacaksa formüldeki (n_1-1) ve (n_2-l) yerine (n_1) ve (n_2) kullanılacaktır.

$$t = \frac{\Sigma X_1 - \Sigma X_2}{\sqrt{\frac{n_1 \Sigma X_1^2 - (\Sigma X_1)^2}{n_1 - 1} + \frac{n_2 \Sigma X_2^2 - (\Sigma X_2)^2}{n_2 - 1}}} \qquad [1.5]$$

Tablo 1'deki veriler formül [1.4]'e uyarlandığında $\sigma_{\bar{F}} = \sqrt{\frac{0.333}{3-1} + \frac{0.333}{3-1}} = 0.577$ bulunur. Işleme formül [1.2] uygulanarak devam edilince $t = \frac{3.33 - 1.33}{0.577} = 3.46$ bulunur. $Sd=(n_1-1) + (n_2-1) = 4$ için bir yönlü olarak manidarlık sınaması yapıldığında 3.747>3.46>2.132 olduğundan sözkonusu maddenin 0.95 güven düzeyiyle üst ve alt çeyreklikleri ayırt ettiği, t katsayısının pozitif işaretli olması nedeniyle bu farklılığın beklenen yönde olduğu, bütün bu istatistik analizler sonucu maddenin ayırt edici özelliği olduğu ve bu ölçüte göre ölçekte yer alabileceği kararına varılır. Manidar çıkmayan maddeler ölçekten kesinlikle çıkarılır. Eğer ölçmeye çalıştığı ayrıntıyı ölçen başka maddeler ölçekte bulunmuyorsa aynı amaca yönelik başka ifadeler, başka maddeler hazırlanıp, ölçek yeniden yapılandırılarak geçerlik güvenirlik işlemleri yenilenir. Geçerlik ve güvenirlik işlemleri, bu açılardan aranan ölçütler karşılanıncaya kadar devam ettirilir. İşlemler sonucu ölçekten bazı maddelerin atılması ile ölçek yeni bir yapıya kavuşmuştur. Bizim yaptığımız geçerlik güvenirlik işlemleri eski ölçek için geçerliydi, ölçeğin işlemler sonucu aldığı son yapı içindeki geçerlik-güvenirliği nedir? sorusunu araştırmacı olarak merak etmez misiniz? t testine göre manidar çıkıp ayırt edici olduğuna karar verilen bir madde ölçekte mutlaka yer alacak demek değildir. Bu sadece ayırtedicilik bakımından uygun olduğunun göstergesidir, diğer ölçütler örneğin içtutarlık sonuçları olumlu çıkmamışsa ayırt edici olmasına rağmen bu madde ölçekten ayıklanacaktır. Bazen t değeri istatistiksel bakımdan manidar ve fakat negatif çıkabilir, bu alt çeyreklikte yer alanların üst çeyreklikte yer alanlardan daha yüksek puan aldıklarını gösterir, yani beklenen durumun tam tersine bir sonuçla karşılaşılmıştır. Aslında sözkonusu madde alt ve üst çeyreklikleri ayırt edicidir, ama ters yönde. Bu, böyle maddelerin puanlandırılmasında hata olduğu şeklinde yorumlanmalıdır. Bu maddenin puanlamasının ters çevrilebilme imkanı varsa, örneğin Likert türü bir ölçek ise ve "her zaman" seçeneği" 1 puan "hiçbir zaman" seçeneği 5 puan şeklinde değerlendirilmiş iken "her zaman seçeneği 5 puan "hiçbir zaman seçeneği 1 puan alacak şekilde değerlendirmenin yönü deriştirilir. Bu işlem ayırt ediciliği beklenen yöne çevirecek ama toplam puan üzerinde yapacağı etki nedeniyle diğer tüm geçerlik-güvenirlik hesaplarında da büyük bir ihtimalle değişikliğe yol açacaktır. Analizlerin bitiminden sonra geçerlik-güvenirlik çalışmalarının yinelenmesi gereği belirmektedir.

Tablo l'deki veriler formül [1.5]'e göre çözümlendiğinde $t = \frac{10-4}{\sqrt{\frac{3x34-10^2}{3-1}+\frac{3x6-4^2}{3-1}}} = 4.24$ olarak bulunur ve manidarlık tablosu değeri olan 3.747 den büyük olduğu için .99 güven düzeyiyle ayırt edicidir.

Tablo 2'deki veriler formül [1.3]'e uyarlandığında $\sigma_t = \sqrt{\frac{0.851}{41}+\frac{0.851}{41}} = 0.204$ bulunur.

Tablo 2'deki veriler formül [1.3]'e uyarlandığında $\sigma_t = \sqrt{\frac{0.851}{41}+\frac{0.851}{41}} = 0.204$ bulunur.

İşleme, formül [1.2] uygulanarak devam edilince $t = \frac{3.32-2.68}{0.204} = 3.137$ bulunur. Sd=(n_1-1) + (n2-1) = 80 için bir yönlü olarak manidarlık sınaması yapıldığında .01 düzeyinde manidar olduğu bulunur. Bu maddenin ayırt etme gücü yüksektir kararına varılır.

Tablo 2'deki veriler formül [1.5]'e göre çözümlendiğinde $t = \frac{136-110}{\sqrt{\frac{41x486-136^2}{41}+\frac{41x330-110^2}{41}}} = 3.113$ olarak bulunur ve .99 güven düzeyiyle ayırt edicidir.

Bazı kaynaklarda kritik oran adıyla verildiği şekil Tablo 2'deki verilere uygulanırsa

$$\text{Kritik Oran} = \frac{\Sigma fx_1 - \Sigma fx_2}{\sqrt{\Sigma fd_1^2 + \Sigma fd_2^2}}$$

$$\Sigma fd_1^2 = \Sigma fx_1^2 - n_1(\overline{X}_1)^2$$

$$\Sigma fd_2^2 = \Sigma fx_2^2 - n_2(\overline{X}_2)^2$$

$$\text{Kritik Oran} = \frac{136-110}{\sqrt{(486-(41(3.32^2))+(330-(41(2.68^2))}} = 3.12$$

sonucun yine aynı olduğu kolaylıkla görülür.

Bir maddenin ayırt ediciliğini belirlemek için elde edilen t katsayısının (kritik oranın) ne kadar büyük olduğuna bakarak değil, serbestlik derecesine göre manidar olup olmamasına göre karar verilir, Manidarlık düzeyi ne kadar yüksekse ayırıcı özelliği de o ölçüde yüksektir. Hatırlanması gereken nokta bir maddenin ayırt edici olması onun güvenilir olmasını (örneğin içtutarhğı) garantilemez.

1.2. Güvenirlik

Bir ölçme aracıyla farklı zamanlarda elde edilen ve aynı nesnelerle ilgili olan bir grup ölçümle ikinci grup ölçüm arasındaki tutarlık eğilimine o aracın güvenirliği denir. Güvenirlik; aynı değişkenin bağımsız ölçümleri arasındaki kararlılıktır, aynı süreçlerin izlenmesi ve aynı ölçütlerin kullanılması ile aynı sonuçların alınmasıdır, ölçmenin random hatalardan arınmış olmasıdır.

Güvenirliğin yüksek olabilmesi, ölçmede izlenen süreçler ile kullanılan ölçütlerin ayrıntılı olarak belirlenebilmesine bağlıdır. Dolaylı ölçmelerin yaygın olarak kullanıldığı sosyal bilimlerde güvenirliği yükseltmek için çok sayıda ölçüt

kullanılmaya çalışılır, madde veya soru sayısı arttırılır. Böylece random hataların birbirini dengelemesiyle güvenirliği yüksek sonuçlar alınabilir.

Birbirini izleyen ölçmelerde, bireyin grup içindeki durumundaki tutarlık aranır. Bu yaklaşımda aynı nesnelerle ilgili iki ölçüm arasında korelasyon hesaplanır ve bulunan korelasyon katsayısı, güvenirlik katsayısı olarak adlandırılır.

Güvenirlik katsayısı, gerçek ölçümlerin varyansının gözlenen puanların (gerçek ölçüm ve hata) varyansına oranıdır.

$\sigma^2 g$: gerçek puanlar dağılımı varyansı

$\sigma^2 t$: gözlenen toplam puan dağılımının varyansı

$\sigma^2 h$: hata varyansı (standart hata karesi)

r_r : elde edilen puanların güvenirlik katsayısı

$$r_r = \frac{\sigma^2 g}{\sigma^2 t} \qquad [2.1]$$

$\sigma^2 t = \sigma^2 g + \sigma^2 h$

$\sigma^2 g = \sigma^2 t - \sigma^2 h$ olacağından Formül 1,

$$r_r = \frac{\sigma^2 t - \sigma^2 h}{\sigma^2 t} \qquad [2.2]$$

halini alır.

Hatasız gerçek ölçümlerin bilinmesi mümkün olmadığından güvenirlik dolaylı yoldan belirlenmeye, tahmin edilmeye çalışır. Güvenirlik tahmininde kullanılan iki yaklaşım vardır: Ölçümler arasındaki ilişkiyle ilgilenen **güvenirlik katsayıları** ve ölçme hatalarının büyüklüğüyle ilgilenen **ölçmenin standart hatası** güvenirlik konusundaki iki temel yaklaşımı oluşturur. Güvenirlik katsayıları ölçülen bilimin grup içindeki konumundaki kararlılığı-136 nı dikkate alır. Ölçmenin standart hatası ise ayrıca; ya farklı birimlerin ölçümleri yada aynı birimin farklı ölçümleri arasındaki değişkenliği (kararlılığı) gözetir.

1.2.1. Ölçmenin Standart Hatası

Güvenirlik katsayısının hesaplanabilmesi için; ölçüm sonuçlarının ne kadarının gerçek varyansı ne kadarının hata varyansı olduğunu bulabilmek, bunun için ise gözlenen değerle birlikte gerçek değeri de bilmek gerekir. Bu söz konusu olamayacağına göre hatanın büyüklüğü ve güvenirlik katsayısı bazı tekniklerle

tahmin edilmeye çalışır. Random hata miktarının tahmini için aynı birey hakkında aynı ölçme aracından elde edilmiş birden fazla veriye veya aynı ölçme aracından birden fazla birey için elde edilmiş veriye gerek vardır. Bu anlamda, elde edilen ölçümler birbirinden farklı olabilir ve bunlar belli bir dağılım gösterir. Standart sapma olarak ifade edilen, ölçümlerdeki bu değişme, gerçek hata payının standart hata olarak hesaplanmasında kullanılır. Puanların gösterdiği dağılımın standart sapmasından, puanların ortalamasının standart hatası hesaplanabilir. Standart hata küçüldükçe ölçmenin güvenirliği artmış olacaktır. Ölçmenin standart hatası, özellikle, bir testteki çeşitli puanların ve puanlar arasındaki farkların güvenirliği konusunda verilecek kararlarda kullanılır.

Ölçmenin standart hatası formülü şöyledir.

σ_m = $[8\sqrt{(1-.92)}]$ = 2.26 olacaktır.

z = 2.58 (0.99 güven düzeyi için normal dağılım puanı)

x = 60 (bireyin aldığı puan)

değerleri Formül [2.4]'e yerleştirildiğinde

R = 60 ± (2.58 * 2.26) = 50 ± 5.83

σm : ölçmenin standart hatası

σ : elde edilen puanların standart sapması

r_r : elde edilen puanların güvenirlik katsayısı

$$\sigma_m = \sigma\sqrt{(1 - r)} \qquad [2.3]$$

Standart hata gerçek ölçümlerin belli olasılıklarla içinde olabileceği puan aralığının belirlenmesinde kullanılır.

σm = : ölçmenin standart hatası

z : normal dağılım değeri

X : elde edilen puan

R : elde edilen puanın aralığı

$$R = X \pm (z \times \sigma_m) \qquad [2.4]$$

Test puanının güvenirliğini belirlemeyi bir örnekle açıklama için;

σ = 8 (elde edilen puanların standart sapması)

r_r = 0.92 (elde edilen puanların güvenirlik katsayısı)

değerleri Formül [2.3]'e yerleştirildiğinde

σ_m = $[8\sqrt{(1-.92)}]$ = 2.26 olacaktır.

z = 2.58 (0.99 güven düzeyi için normal dağılım puanı)

x = 60 (bireyin aldığı puan)

değerleri Formül [2.4]'e yerleştirildiğinde

R = 60 ± (2.58 * 2.26) = 50 ± 5.83

olacaktır. Diğer bir ifadeyle bu ölçüm sonuçlarına göre, bireyin gerçek puanı, 9c99 olasılıkla 54,17 ve 65.83 arasındadır. Güven aralığının sınırları birbirine ne kadar yaklaşırsa pratik değeri o ölçüde artar. Bir ölçüm dizisinin standart hatası, elde edilen varyansm hangi oranda bireyler arası gerçek farklılığı yansıttığını gösterir.

Güvenirlik katsayısı ve ölçmenin standart hatası ilişkisini görmek için yukarıdaki veriler Formül [2,2]'ye yerleştirilince

$$r_r = \frac{8^2 - 2.26^2}{8^2} = 0.92$$

çıkmaktadır. Yani ölçmenin standart hatası ölçümdeki random hata miktarını belirlemektedir ve güvenirlik katsayısının yükselmesi standart hatanın, ölçümlerin varyansı içindeki payının küçülmesine bağlıdır. Dolayısıyla standart hata ne kadar küçük ise ölçmenin güvenirliği de o ölçüde artıyor demektir. Örnekte de görüldüğü gibi güvenirlik katsayısı ve standart hata, güvenirlik konusunda birbirini destekleyen iki ayrı sonuçtur.

1.2.2. Güvenirlik Katsayıları

Güvenirlik katsayıları genellikle korelasyon teknikleri ile hesaplanır. Korelasyon katsayısı -1.00 ile + 1.00 arasında değişmekle birlikte, güvenirlik katsayılarının pozitif değerli olması beklenir, negatif değerli katsayı güvenirsizliğin bir göstergesidir. Değer +l'e yaklaştıkça güvenirliğin yüksek olduğu kabul edilir. Güvenirlik katsayıları tekniklerinin mani-darlıkları ile ilgili özel dağılımlar hazırlanmamış, sınama süreçleri belirlenmemiş olmakla beraber çoğunluğunda, Pearson korelasyon katsayısı gibi sınama, doğruya yakın bir fikir verebilir. Geçerlik ve güvenirlik katsayıları belirlenmesinde aksine bir belirtme yapılmadığı sürece, sözü geçen korelasyon katsayısı, ölçekten elde edilen verilerin yapısı uygun olması halinde Pearson korelasyon katsayısıdır.

1.2.2.1. Devamlılık Katsayısı (Coefficient of Stability). r_{cs}

Devamlı özellikler ile ilgili ölçmelerde aranan güvenirlik, ölçümün devamlılığı açısındandır. Devamlılık katsayısı, ölçme aracının uygulanması, belli bir zaman aralığı sonunda aynı gruba aynı ölçme aracı uygulamasının tekrarlanması ile elde edilen iki seri halindeki sonuçlar arasındaki korelasyonun hesaplanması ile elde

edilir. Bu işlem "test-tekrar test' yöntemi olarak da bilinir. Devamlılık katsayısı, ölçme aracının bireyde kalıcı (devamlı) özellikleri ne ölçüde ölçtüğünü belirler.

Devamlılık katsayısının hesaplanmasında, iki uygulama arasında geçen sürenin ne kadar olması gerektiği sorunu ortaya çıkmaktadır. Bu soruya kesin bir cevap verilmemektedir Bununla beraber, geçecek zaman, ölçeğin ölçtüğü özellik bakımından cevaplayıcıların önemli ölçüde değişmelerine yetecek kadar uzun olmamalıdır. Ayrıca, birinci uygulamada verilen cevapları hatırlamaya yetecek kadar kısa da olmamalıdır. Genellikle bu ara 3-6 hafta olarak belirlenebilir.

İki uygulamadan elde edilen puanlar arasındaki ilişki miktarının hesaplanması tek başına yeterli değildir. Bireylerin her iki ölçme aracı puanları arasında önemli bir fark olmamakla beraber her iki uygulamada cevaplandırılan, puan alman maddeler farklı maddeler olabilir, aynı maddeden farklı puanlar almış olabilir. Bu nedenle, bireylerin her iki uygulamada puan aldıkları maddeler arasındaki tutarlığa da bakmak gerekir. Test toplamı, faktör toplamı ve tüm maddeler için korelasyon hesaplanmalıdır.

Devamlılık katsayısı her ölçek için mutlaka hesaplanmak zorunda değildir. Ölçülen değişken devamlılık gösteren bir yapıdaysa devamlılık katsayısının bir anlamı vardır Sürekli kaygı ölçüldüğünde devamlılık katsayısının belirlenmesi bir zorunluluk olmakla beraber, durumluk kaygının belirlendiği bir ölçekte devamlılık katsayısı anlamsızdır.

1.2.2.2.Eşdeğerlik Katsayısı (Coefficient of Equivalance). r_{ce}

Eşdeğerdik katsayısı, a) farklı uygulayıcıların aynı zamanda kullandıkları aynı ölçü araçları için **uygulamaların eşdeğerliğini** b) farklı, fakat benzer ölçme araçlarının (bu aynı değişkeni ölçen iki ayrı ölçek olabileceği gibi aynı ölçeğin iki ayrı formu da olabilir ve paralel test yöntemi olarak da bilinir) aynı gruba aynı zamanda uygulanmasında **araçların eşdeğerliğini,** c) aynı ölçme aracının farklı iki dildeki iki formunun aynı gruba aynı zamanda uygulanmasında **dilsel eşdeğerliği** belirtir. Bu uygulamalarda elde edilen puanlar arasındaki korelasyon katsayısı, eşdeğerlik katsayısıdır,

Paralel iki testin eşdeğerlik sınamasının yapılabilmesi, her iki ölçme aracı, içindeki madde sayısı, niteliği, kullandığı ölçekleme tekniği (Likert, Thurstone, vb), faktör yapısı ve ölçtükleri davranış bakımından birbirine denk olmalıdır. Ölçek adaptasyon çalışmalarında bile bu eşdeğerlik koşullarına titizlikle uyulması gerekir. Örneğin orijinalinden bazı maddeleri hangi gerekçeyle olursa olsun ölçekten tamamen çıkararak ölçekteki madde sayısı azaltılırsa veya herhangi bir maddenin yerine bizce içerdiği değişken aynıdır diye yorumlanan yeni bir madde yazmak eşdeğerliği anlamsızlaştırır. Aynı şekilde orijinalinde "x", "q", "w" maddeleri "z" boyutu içinden yer alırken faktör analizi sonucunda bile olsa adapte ölçekte örneğin "q" maddesinin "ğ" boyutunda yer alması bu ölçeğin İngilizce-Türkçe formlarının eşdeğerlik sınamasının önkoşulunu iptal eder. Bu iki form arasında ilişki bulunsa bile (ki büyük ihtimalle bulunacaktır) bu ilişki eşdeğerlik anlamında değildir.

Eşdeğer iki form, eğer aynı bireylere aynı zamanda uygulanmışsa eşdeğerlik katsayısı hesaplanır, ama belli bir zaman aralığı ile uygulanmışsa eşdeğerlik ve devamlılık birbirine karışır.

1.2.2.3.İçtutarhk Katsayısı (Coefficient of Internal Consistency), r_{ic}

Ölçme aracı bir gruba uygulandıktan sonra, sonuçların değerlendirilmesi aşamasında belli yöntemlere göre ikiye ayrılması ile bunlar arasında elde edilen katsayıya içtutarlık katsayısı denir. İçtutarlık katsayıları eğer ölçek tek bir boyuttan oluşursa ölçeğin bütününe göre yapılabileceği gibi ölçek faktörlerden oluşursa her faktör ve maddeleri kendi içinde bir bütün olarak kabul edilip aynı teknik yaklaşımla faktör düzeyinde de hesaplanabilir.

Tek bir ölçme aracı formu, tek bir bire}' grubu ve tek bir uygulama gerektirdiğinden en sık kullanılan güvenirlik saptama yöntemidir. Bu yöntemde karşılaşılan sorun, testin iki eşdeğer yarıya bölünmesidir. İki eşdeğer yarıya bölmenin en çok kullanılan yolu tek numaralı sorularla çift numaralı soruları ayrı puanlamaktır. Bu iki yarıdan elde edilen puanlar takımı, ayrı ayrı testlerden elde edilmiş gibi işlem görürler. Katsayı yüksek çıkmış, yani testin bir yansı diğer yarısı ile tutarlı ise, "bu testin içtutarlığı (güvenirliği) vardır" şeklinde değerlendirilecektir. Tek ve çift sorular şeklinde ikiye ayırmada göz önünde tutulması gereken nokta soruların numaralanmasında yanlılığa yol açabilecek herhangi bir düzenlemenin yapılmamış olması, soruların tek ve çift numaraları almasında random oluşmasıdır. Örneğin ölçek tek boyuttan oluşuyor ve tek numaralı sorular ölçülen değişken bakımdan olumsuz (reverse) ifadeli maddeler çift numaralı sorular ise olumlu ifadeli maddeler ise bu ikiye ayırma tekniği doğru sonuçlar vermeyebilir.

İçtutarlık katsayısını hesaplamada, bu temel yaklaşımı kullanmakla beraber farklı istatistik değer ve formülleri kullanan çok sayıda teknik vardır. Veri türleri ve koşullara uygun oldukları sürece, güvenirlik ile ilgili sınamaları pekiştirmek amacıyla bu tekniklerin birden fazlası kullanılabilir.

İçtutarlık katsayılarından Spearman-Brown, Stanley, Rulon, Flanagan, Mossier, Horst gibi bazı teknikleri yan-test (split-half) tutarlık katsayıları olarak adlandırılır. İtem-total korelasyon katsayısı, item-remainder korelasyon katsayısı teknikleri ise parça-bütün arasındaki ilişki yoluyla tutarlığı belirler, her madde için ayrı ayrı hesaplanır. Kuder-Richarson 20 ve 21 ile Cronbach ∞ tekniği ise parçalar arası ortak ilişkiyi dikkate alarak bütün için tek bir tutarlık katsayısı hesaplar.

1.2.2.3.1. Spearman-Brown

Bir gruba uygulanmış testin iki eşdeğer yarıya bölünmesi ve bireylerin iki eşdeğer yarıdan aldıkları puanlar arasındaki Pearson korelasyon katsayısından hareketle Spearman-Brown formülünden de yararlanarak testin bütününün güvenirliği kestirilir.

Spearman-Brown formülü,

r_{ic} : testin güvenirlik katsayısı

r_{xy} : iki yarı arasındaki korelasyon katsayısı

olarak ifade edilirse

$$r_{ic} = \frac{2r_{xy}}{1 + r_{xy}} \qquad [2.5]$$

olur.

Tablo 3'deki verilere göre yan testin güvenirliği 0.468 olarak bulunmuştur. Bu bulgu, testin bütününün güvenirliğini bulmak için Spearman/Brown formülüne [2.5] yerleştirildiğinde

$$r_{ic} = \frac{2^{*}\ (0.468)}{1 + 0.468} = 0.638$$

bulunur.

TABLO 3. SPEARMAN-BROWN TEKNİĞİ İLE İÇTUTARLIK KATSAYISI HESAPLANIŞI

Öğrenci	Tek Numaralı Sorular Toplamı X	Çift Numaralı Sorular Toplamı Y	X^2	Y^2	XY
A	20	20	400	400	400
B	20	20	400	400	400
C	18	17	324	289	306
D	24	9	576	81	216
E	15	15	225	225	225
F	14	15	196	225	210
G	14	14	196	196	196
H	17	9	289	81	153
I	8	8	64	64	64
İ	9	7	81	49	63
10	159	134	2751	2010	2233
n	Sx	Sy	Sx^2	Sy^2	Sxy

$$r_{xy} = \frac{(10^{*}2233) - (159^{*}134)}{\sqrt{[10^{*}2751) - (159)^2] \ ^{*}\ [\ (10^{*}2010) - (134)^2]}} = 0.468$$

1.2.2.3.2. Stanley

Bir testin iki yarısındaki puanlardan test güvenirliğini kestirmede kullanılan Stanley formülü;

Σüx-y : üst çeyreklikte yer alanların x-y toplamı

Σax-y : alt çeyreklikte yer alanların x-y toplamı

Σüx+y : üst çeyreklikte yer alanların x+y toplamı

Σax+y : alt çeyreklikte yer alanların x+y toplamı

$$r_{ic} = 1 - \frac{[(\Sigma üx - y) - (\Sigma ax - y)]^2}{[(\Sigma üx+y) - (\Sigma ax+y)]^2} \qquad [2.6]$$

Ölçeğin güvenirliğini kestirmede uygulanan toplam test sayısının &27 si nq=(n*0,27) bulunur. Ölçeğin her iki yarısındaki puanların toplamına (X+Y) göre üst çeyreklikte yer alanların (ü) puanlarının toplamı (∑üx+y) ve alt çeyrekliltye yer alanların(a) puanlarının toplamı (∑ax+y) bulunur. İşlem benzer olarak ölçeğin her iki yarısı arasındaki farklara (X-Y) göre üst çeyreklikte yer alanların puanlarının toplamı (∑üx-y) ve alt çeyreklikte yer alanların puanlarının toplamı (∑ax-y) bulunur. Elde edilen bu dört değer Stanley formülündeki [2.6] yerine konularak bu yönteme göre içtutarlık belirlenmiş olur.

TABLO 4. STANLEY TEKNİĞİ İLE İÇTUTARLIK KATSAYISI HESAPLANIŞI

Öğrenci	TEK X	ÇİFT Y	(3) X+Y	(4) *X-Y*
A	20	20	40ü	0a
B	20	20	40ü	0a
C	18	17	35ü	1
D	24	9	33	15ü
E	15	15	30	0
F	14	15	29	-1a
G	14	14	28	0
H	17	9	26a	8ü
I	S	8	16a	0
İ	9	7	16a	2ü
10			∑üx+y=115	∑üx-y=24
n			∑ax+y= 58	∑ax-y=- 1

Tablo 3'deki birimlere ait aynı verileri kullanarak hazırlanan Tablo 4 için bu yöntem uygulandığında nq=(10 ¨0.271=3 kişi üst çeyreği 3 kişi alt çeyreği oluşturmaktadır. Tablonun altında hesaplanmış şekilde verilen değerler formüle [2.6] uygulandığında

$$r_{ic} = 1 - \frac{(24 - (-1))^2}{(115-24)^2} = 1 - \frac{25^2}{91^2} = 1 - \frac{625}{8281} = 0.924$$

olarak elde edilir.

1.2.2.3.3. Rulon

Bir testin iki yarısındaki puanlardan test güvenirliğini kestirmede kullanılan Rulon formülü,

TABLO 5. RULON, FLANAGAN, MOSİER VE HORST TEKNİKLERİ İLE İÇTUTARLIK HESAPLANIŞI

Öğrenci	X	*Y*	X+Y	X-Y
A	20	20	40	0
B	20	20	40	0
C	18	17	35	1
D	24	9	33	15
E	15	15	30	0
F	14	15	29	-1
G	14	14	28	0
H	17	9	26	8
I	8	8	16	0
i	9	7	16	2
$\sum X$	159	134	293	25
$\sum X^2$	2751	2010	9227	295
μ	15.90	13.40	29.30	2.50
α	4.98	4.88	8.45	5.08
α^2	24 77	23.82	71.34	25.83
n=10	r_{xy} =0.468	r_{xt}=0.860	r_{yt}=0.854	

Tabloda X (çift numaralı soruları). Y (tek numaralı soruları), X+Y (test toplam puanını), X-Y (her iki yarı arasındaki farkı) ifade etmektedir

Daha önceki tekniklerde de kullanılan verileri içeren Tablo 5'deki sonuçlara göre

$$r_{ic} = 1 - \frac{25.83}{71.34} = 1-0.36 = 0.638$$

olur.

1.2.2.3.4. Flanagan

Bir testin iki yarısındaki puanlardan test güvenirliğini kestirmede kullanılan Flanagan formülü, *o*

σ^2x : Tek numaralı soruların varyansı

σ^2y : Çift numaralı soruların varyansı

σ^2t : toplam puanların varyansı

olarak ifade edilirse

$$r_{ic} = 2 * \left(1 - \frac{\sigma^2_x + \sigma^2_y}{\sigma^2_t} \right) \qquad [2.8]$$

Tablo 5'deki sonuçlara göre

$$r_{ic} = 2 * (1 - \frac{24.77+23.82}{71.34}) = 2 * (1-0.68) = 0.640$$

olur.

1.2.2.3.5.Mosier

Mosier formülü,

r_{xt} : tek numaralı sorularla toplam arasındaki korelasyon katsayısı

σ^2_x : Tek numaralı soruların varyansı

σ^2_t : Toplam puanların varyansı

olarak ifade edilirse

$$r_{ic} = \frac{r_{xt}\sigma_t - \sigma_x}{\sqrt{[\sigma_t^2 + \sigma_x^2 - (2r_{xt}\sigma_x\sigma_t)]}} \qquad [2.9]$$

Tablo 5'deki sonuçlara göre

$$r_{ic} = \frac{0.860*8.447 - 4.977}{\sqrt{[71.344+24.767 - (2*0.860*4.977*8.447)]}} = \frac{2.287}{4.879} = 0.468$$

olur.

Elde edilen değer Spearman-Brown formülüne [2.5] yerleştirilerek testin bütününün güvenirliği bulunur.

1.2.2.3.6.Horst

Şimdiye kadar bahsedilen içtutarlık katsayıları hesaplanan iki varırım eşit sayıda maddeden oluştuğu varsayımına dayanır. Farklı sayıda maddelerden oluşan iki parça arasındaki tutarlığı hesaplamak için buna yönelik düzeltmeleri içeren Horst tekniği kullanılır. Güvenirlik hesabı yapılan ölçeğin iki yansının eşit olmadığı durumlarda kullanılabilmesi bu tekniğin ayırıcı özelliğidir. Bu teknik doğrudan iki yarı arasındaki içtutarlığı hesaplamada kullanılabileceği gibi esas olarak, testin bir alt bölümünün kendi hariç diğer alt bölümlerin toplamına göre tutarlı olup olmadığını sınamada kullanılır. Eğer iki parçadaki madde sayılan eşit ise diğer tekniklerle varılan sonucu verir, ancak madde sayıları arasındaki denge bozuldukça diğer içtutarlık katsayılarından ve korelasyondan farklılık gösterebilir.

Horst formülü,

r_{xy} : iki yarı arasındaki korelasyon katsayısı

p : Birinci yarıdaki soru sayısının toplam soru sayısına oranı

q : l-p

Olarak ifade edilirse

$$r_{ic} = \frac{r_{xy} \{ \sqrt{[(r_{xy}^2 + (4pq(1-r_{xy}^2)))]} - r_{xy} \}}{2pq(1-r_{xy}^2)} \qquad [2.10]$$

Tablo 5'deki sonuçlara göre

$$r_{ic} = \frac{0.468 [(\sqrt{0.468^2 + 4\ (0.50)\ (0.50)\ (1-0.468^2)}) \ -0.468]}{2(0.50)\ (0.50)\ (1-0.468^2)}$$

$$r_{ic} = \frac{0.468 [(\sqrt{1}) - 0.468]}{0.390} = \frac{0.468 * 0.532}{0.390} = \frac{0.249}{0.390} = 0.638$$

olur.

1.2.2.3.7. Cronbach α

Bu tekniğin özelliği Kuder-Richardson tekniğininkine benzer. Testin tüm alt bölümlerinin birbirlerine (hepsi dahil toplama) göre, ya da bir alt bölümün tüm sorularının birbirlerine (hepsi dahil toplam) göre, tutarlı olup olmadığını sınamada kullanılır. Sonuçlar tüm bölümlerin (ya da o bölümdeki tüm soruların) birbiriyle ilgili olduğu şeklinde yorumlanır. Katsayının yüksekliği içtutarlığın yüksekliğini gösterir.

Bir bölümdeki soruların o bölümle ilgisi söz konusu olduğunda Croanbach a alpha formülü,

m : bölümdeki soru sayısı

σ_j^2 : j. sorunun varyansı

σ_t^2 : soruların toplamının varyansı

$\Sigma\sigma_j^2$: soruların varyanslarının toplamı (1. sorunun varyansı +.....+ j. sorunun varyansı)

$$\alpha = \frac{m}{m-1} * (1 - \frac{\Sigma\sigma_i^2}{\sigma_i^2}) \qquad [2.11]$$

olur.

Aynı formüldeki semboller, alt bölümlerin testin bütünüyle ilgisi söz konusu olduğunda,

m : testteki altbölüm sayısı

σ^2_j : j. alt bölümün toplam puanlarının varyansı

σ^2_t : bölüm toplamlarının toplamının varyansı

$\Sigma\sigma^2_j$: bölüm toplam varyanslarının toplamı

anlamında kullanılır.

TABLO 6. CRONBACH, ITEM-TOTAL KORELASYON, ITEM-REMAINDER KORELASYON HESAPLANIŞI

Öğrenci	$İ_1$ 2.soru	$İ_2$ 4.soru	$İ_3$ 6. soru	T toplam	$Tİ_1$ 2.s * tp
A	3	3	5	11	33
B	4	3	5	12	48
C	3	3	4	10	30
D	2	2	1	5	10
E	4	2	5	11	44
F	3	3	3	9	27
G	4	2	2	8	32
H	2	2	1	5	10
I	1	1	2	4	4
İ	1	1	2	4	4
Σx	27	22	30	79	242
Σx^2	85	54	104	713	
μ	2.70	2.20	3.00	7.90	
σ	1.160	0.789	1.633	3.143	
σ^2	1.344	0.622	2.667	9.878	

Tablo 6'deki sonuçlara göre, 2., 4. ve 6. sorulardan oluşan bir alt ölçek için

$$\alpha = \frac{3}{3-1} * \left(1 - \frac{(1.344+0.622+2.667)}{9.878}\right) = 1.5 * (1-0.469) = 0.796 \text{ olur.}$$

1.2.2.3.8. Kuder-Richardson 20-21

Testeki herbir maddeye doğru cevap veren birey yüzdesi hesaplanarak içtutarlık Kuder-Richardson 20 ve 21 no'lu formülleri ile tahmin edilebilir. Bu yolla elde edilen katsayı da iç-tutarhğm bir ölçüsüdür. "

Kuder-Richardson (KR) 20 no'lu formülü şöyledir:

m : testteki madde sayısı

P : bir maddeyi doğru cevaplayanların oranı

q : bir maddeyi doğru cevaplamayanlarm oranı = (1-p)

$\sum pq$ her madde için hesaplanan (p x q)'ların toplamı

$\alpha^2 t$: test toplam puanlarının varyansı

olarak ifade edilirse

$$r_{ic} = \left(\frac{m}{m-1}\right) * \left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}\right) \qquad [2.12]$$

olur.

Bu formül [2.12], doğru cevaplandırılanlara puan verilmesi, yanlış ve boş cevaplara hiç puan verilmediği testlerde uygulanabilir. Cevapların doğruluğu-yanlışlığı mantığına oturduğu için başarı testlerinde kullanılabilmesine karşılık tutum testlerinde kullanılmamalıdır. Eğer testteki maddeler farklı ağırlıklarla puanlanmışsa bu formül uygulanamaz. Bir testteki maddelerin güçlük dereceleri birbirinden önemli ölçüde farklı değilse o testin güvenirliği için KR-21 no'lu formülü kullanılabilir. Bu formül:

p : madde doğru cevap oranlarının (p) aritmetik ortalaması

q : q oranların aritmetik ortalaması

$$r_{ic} = \left(\frac{m}{m-1}\right)\left(\frac{\sigma_t^2 - m\bar{p}\,\bar{q}}{\sigma_t^2}\right) \qquad [2.13]$$

olur. Toplam puanların ortalaması X_t gösterilirse $p = X_t / m$ olduğundan ve $q = (m-X_t)/ m$ olduğundan formül [2.13]

$$r_{ic} = \frac{m\sigma_t^2 - \bar{X}_t (m - \bar{X}_t)}{(m-1)\sigma_t^2} \qquad [2.14]$$

şeklini alır. KR-21 içtutarlığı kestirmede (özellikle de KR-20'ye göre daha) düşük güçtedir.

1.2.2.3.9. İtem-Total Korelasyon Katsayısı r_{ti}

Bu yöntemde her bir sorudan elde edilen puan ile toplam puan arasındaki ilişkiye bakılır Pearson Çarpım Moment korelasyon katsayısı yoluyla her bir madde için elde edilen sonuçların manidarlığına bakılır.

$$r_{ti} = \frac{n\sum Ti - \left(\sum T \sum i\right)}{\sqrt{n\sum T^2 - \left(\sum T\right)^2}\ \sqrt{n\sum i^2 - \left(\sum i\right)^2}} \qquad [2.15]$$

ΣTi : her bireyin toplamdaki ve itemdeki puanların çarpımlarının tüm bireyler için toplamı

ΣT : bireylerin toplam puanlarının toplamı

Σi : bireylerin itemdeki(maddedeki) puanlarının toplamı

ΣT^2 : bireylerin toplam puanlarının karelerinin toplamı

Σi^2 : bireylerin itemdeki (maddedeki) puanlarının karelerinin toplamı

n : birey sayısı

Tablo 6'daki velilerden 2 nolu soru için r_{ti} hesaplamak için veriler formül [2.15]'e uygulandığında;

$$r_{ti} = \frac{10x242 - (79x27)}{\sqrt{10x713 - 79^2}\ \sqrt{10x85 - 27^2}} = 0.875$$

olarak bulunur. Hesaplama formülü [2.15] doğrudan doğruya Pearson formülü olduğundan bu istatistiğe göre manidarlığı sınandığında hesaplanan r=0.875, sd(serbestlik derecesi) = (n-2) = 8 için tablo değeri olan 0.765'den büyük olduğu için .01 düzeyinde manidardır. Bu madde, bu ölçekle ilişkilidir. Bu maddenin ölçtüğü değişken ile ölçeğin tümünün ölçtüğü değişken birbiriyle ilişkilidir. O halde bu madde bakımından içtutarlık vardır. Manidar çıkmayan madde diğer analizler sonuçlarına bakılmaksızın ölçekten atılır. Bazen r_{ti} manidar fakat negatif bir değer olarak çıkabilir. Bu durumda yapılması gereken ayırt edicilik için t-testi açıklanırken belirtildiği gibi maddenin değerlendirme yönünü değiştirmek ve tabiiki geçerlik güvenirlik işlemlerinin yenilemek. Maddenin ölçtüğü değişken ile ait olduğu ölçeğin ölçtüğü değişken "ilişkilidir" ifadesinin kullanılmasına dikkat etmek gerekir, "Aynıdır" ifadesi kullanılamaz, aynı olduğunu ileri sürebilmek için öncelikle iki ölçüm (madde-ölçek) puanlan arasında tam bir ilişki (r=+l) olmalı, ayrıca fark olmadığı bağlantılı grup t-testi ile sınanarak fark bulunmamalıdır.

1.2.2.3.10. Item Remainder Korelasyon Katsayısı r_{ri}

Bu tekniğin özü (j. soru) ile (j. soru hariç ölçeğin tümünden alman puan) arasındaki ilişkiyi, bir maddenin kendi hariç ait olduğu ölçeğin toplamla ilişkisini bulmaktır.

Item-remainder formülü,

r_{ti} : j. soru ile toplam puan arasındaki korelasyon katsayısı

σ^2_j : j. sorunun varyansı

σ^2_t : Toplam puanların varyansı

olarak ifade edilirse

$$r_{ri} = \frac{r_{ti}\,\sigma_t - \sigma_i}{\sqrt{\sigma_t^2 + \sigma_i^2 - 2r_{ti}\,\sigma_t\,\sigma_i}} \qquad [2.16]$$

Tablo 6'daki verilerden 2 nolu soru için r ,hesaplamak için veriler formüle [2.16] uygulandığında

$$r_{ri} = \frac{0.875x3.143 - 1.160}{\sqrt{9.878+1.344 - 2x0.875x3.143x1.160}} = 0.723$$

Hesaplanan r, tablo değeri 0.632'den büyük olduğu için .05 düzeyinde manidardır. Bu madde ölçeği oluşturan diğer 2 madde ile ilişkilidir, ölçtükleri değişkenler ilişkilidir.