**Yönetim Bilimlerinde Araştırma Yöntemleri (Kitap hazırlık notu) (üretim tarihi: 6.7.2017)**

**Kaynak gösterme biçimi**: Hüner Şencan (2017), Yönetim Bilimlerinde Araştırma Yöntemleri (Kitap hazırlık notları), [www.ders.es/](http://www.ders.es/).... ilgili sayfanın adı, çevrimi içi erişim tarihi, …/…..2017

© Prof. Dr. Hüner Şencan

3. Yüzey Geçerliliği

Tezlerde kullanılan ölçeklerin yüzey geçerliliğine sahip olduğunun ifade edilmesi genel bir uygulama haline gelmiştir. Bununla birlikte “yüzey geçerliliği iddiası”, güvenilirlik açısından yeterince sağlıklı bir yaklaşım değildir. Tercih edilme nedeni, yüzey geçerliliğinin kriter ve yapısal geçerliliği destekleyen bir özelliğe sahip olmasıdır. Yüzey geçerliliği iddiası iki tür bulguya dayalı olarak öne sürülür. Birincisi hakemlerin veya Konu İçeriği Uzmanlarının (KİU) yaptığı ön inceleme ve değerlendirme ve ikincisi anket uygulanacak kişilerden seçilen panel grubunun yaptığı değerlendirmedir. Yüzey geçerliliği hakemlerin veya konu içeriği uzmanlarının uzmanlık derecesine, konuya hakim olma derecesine bağlıdır (Kenneth S. Shultz, 2014, 87). Bu nedenle yüzey geçerliliği konusunda bilgi alınacak kişiler dikkatle seçilmelidir. Konu hakkında bilgisi, uzmanlığı, unvanı, deneyimi olmayan kişiler KİU olarak belirlenmez. Hakem değerlendirmesi şu şekilde gerçekleştirilir. Belirlenen hakemlere ölçülmek istenen “kavramsal yapı” hakkında ön bilgi verilir ve daha sonra ölçek maddeleri kendilerine okutulur. Okuma işlemi bittikten sonra hakemlere şu soru sorulur. “Bu ölçeğin maddeleri ölçmeye çalışılan “kavramsal yapıyla” genel olarak ilgili midir?” Değerlendirme genel niteliktedir, maddeler üzerinde derecelendirme yapılmaz ve maddeler tek tek ele alınmaz. Ölçeğin ortaya koyduğu genel algıya bakılır. İkincisi anket uygulanacak kişilerden seçilen panel grubunun yaptığı değerlendirmedir. Onlara da aynı soru sorulur. Panel grubunda dört veya beş üyenin bulunması yeterlidir. Değerlendiricilerin “bu maddelerdeki ifadeler stresle, örgütsel bağlılıkla, örgütsel adaletle veya alaycılıkla ilgilidir” demeleri “yüzey geçerliliği” bulgusu için yeterlidir. Zayıf bulunmuşsa, “yetersiz” değerlendirmesi yapılmışsa ölçeğin yüzey geçerliliği yoktur.

4. İçerik Geçerliliği

İçerik geçerliliğinin iki yönü vardır: “mantıksal geçerlilik” ve hakemlerin belirlenen maddeleri alakalı ve gerekli görmesiyle ilgili olan “uyuşum istatistikleri”. Kapsam geçerliliği, büyük ölçüde uzmanların yaptıkları mantıksal bir değerlendirmedir. Bu değerlendirmeyle ortaya çıkan ölçek maddeleri daha sonra “gereklilik” ve “ilgililik” faktörleri açısından değerlendirmeye tabi tutularak maddelerin ne ölçüde gerekli, ilgili ve anlaşılır olduğu konusunda bir uyuşum oranı elde edilir. Değerlendirici hakemlere ait “uyuşum oranı” değerlerinin yüksek olması ölçüsünde ölçeğin tek tek maddelerinin veya genel olarak ölçeğin bütününün kapsam açısından geçerli olduğuna karar verilir.

**Mantıksal kapsam geçerliliği**. Yargısal geçerlilik olarak da isimlendirilir. Geliştirilen ölçeğin ilk taslağı konunun uzmanı öğretim üyeleri ve iş dünyasından profesyoneller tarafından incelenir. Mantıksal veya yargısal geçerlikte ölçek üç başlık çerçevesinde değerlendirilir. Birincisi maddelerin alansal kapsam yeterliliğidir. İkincisi ifadelerin sözel yeterliği ve üçüncüsü ölçeğin tasarım ve stil olarak düzenleme yeterliliğidir.

*Alansal**kapsam yeterliliği*. Ölçeğin “kavramsal alanı” sağlıklı bir şekilde kapsama yeterliliğini belirlemek için iki konu üzerinde durulur. Birinci konu, kavramsal yapı alanına ilişkin “geniş madde evreniyle” ilgilidir. Ölçekteki maddelerin toplam sayısı, bir boyut altında yer alan maddelerin sayısı, belirlenen maddelerin ölçeği veya ilgili boyutu yeterli ölçüde temsil etme ve kapsama derecesidir. İlk defa geliştirilen ölçeklerde alansal kapsam yeterliliği önemli bir konudur. Bir boyut altında yer alan madde sayısı ne çok ne de az, o boyutun özünü ortaya çıkaracak nitelikte olmalıdır. *Alansal kapsam yeterliliği* asıl araştırmanın sonuna kadar devam eden bir süreçtir. İlk aşamada belirlenen hakemlerin yargısal incelemesiyle değerlendirilir. İkinci aşamada pilot araştırma sonuçlarına bakılır. Son aşamada ise asıl alan araştırması verileri dikkate alınarak “gerçek madde örneklemi” ortaya çıkarılır. Yabancı ülkelerde geliştirilmiş olan ölçeklerin alansal kapsam yeterliliği Türkiye uygulamasında çoğunlukla yeniden şekillenir. Alansal kapsam yeterliliğinin ikinci konusu, katılımcı evreninde yer alan kişilerin maddeleri makul ve uygun bulmalarıdır. Hedef kitledeki kişiler kavramsal yapıyla ilgili olarak kendilerini belirlenen ifadelerde bulabilmeli, dereceleme ve etiketlemeyi kabullenmeli, ifadeler kendileriyle, tutum ve düşünceleriyle ilgili değilse bunu yansıtacak ayrı bir şıkka veya dereceye sahip olabilmelidirler.

*İfadelerin sözel yeterliği*. Sözcükler ve cümlelerle ilgilidir. Ölçekteki (a) cümlelerin anlaşılır olması, (b)kullanılan kelime ve kavramların açık ve herkes tarafından bilinir olması, (c) mahrem nitelikte soru, kelime veya ifadeler içermemesi, (d) teknik dil kullanılmaması, (e) hedef kişilerin belirlenen sorulara kolay cevap verebilmesidir. Değişik bilim adamlarının Türkçeye çevirdiği, adapte ettiği veya kendilerinin geliştirdiği ölçeklerin “sözel yeterliliğe” sahip olduklarını varsaymak doğru değildir. Bu ölçeklerin her defasında gözden geçirilerek sözel yeterlilik açısından incelenmesi ve gerekli düzeltmelerin yapılması gerekir. Bu tür düzeltmeler ölçeğin yapısına müdahale anlamında değildir. Gerekiyorsa maddeler yeniden yazılabilir veya madde çıkarılması yoluna başvu­rulabilir.

*Ölçeğin tasarım ve stil yeterliliği*. Genelde anketin ve özelde yararlanılan ölçeklerin bulunduğu sayfaların düzenleme ve tasarımı ile yazım sitilinin cevap vermeyi kolaylaştırıcı nitelikte olmasıdır. Anket formu bilgilendirmeye ve süjelerin rızalarını almaya yönelik olarak hazırlanan kapak sayfası ile birlikte önceden belirlenen sayfa kenar boşlukları içinde düzenlenir. Kenar boşlukları gözü yoracak bir şekilde zorlanmaz. Stil yazım biçimiyle ilgilidir. Yazı fontu, yazı büyüklüğü, kelime aralarının gereğinden fazla açık olmaması, ifadelerin sola dayalı olarak düzenlenmesi, ifadelere otomatik numara verilmesi, tablo hücreleri içine yazılan ifadelerin sıkışık olmaması, alt ve üst çizgiden önce 2 nokta boşluk bırakılması vb. konular stil ile ilgilidir. Tasarım ve stil ölçeklerin ciddiye alınması, dikkatli ve özenli doldurulması, kolay okunması, kolay işaretlenmesi gibi yararlar sağladığından dolaylı olarak içerik geçerliliğine katkı sağlar.

**Matematiksel ve** **İstatistiksel kapsam geçerliliği**. İçerik geçerliliği esas olarak hakem değerlendirmesine dayanır. Hakem değerlendirmeleri panel tartışmaları, bireysel değerlendirmeler, geri besleme uygulamaları şeklinde yapılabileceği gibi “ilgililik”, “gereklilik”, “anlaşılırlık”, “açıklık” dereceleme etiketleri ve ölçekleri kullanılarak da gerçekleştirilebilir. Bu yaklaşımda “hakemler arası uyuşma ve mutabakat” ön plana çıkar. Hakemler ölçekteki maddelerin ilgili ve gerekli olduğu konusunda belli bir oranın üzerinde mutabakata sahiplerse tek tek ölçek maddelerinin ve genelde tüm ölçeğin geçerli olduğuna karar verilir. Aslında hakemler arası uyuşma geçerliliği “alansal kapsam yeterliliği” koşulunun sağlandığı durumda anlamlıdır. Aksi halde kapsam yeterliliği sağlanmamış bir ölçekteki maddelerin gerekliliğini tek başına matematiksel ve istatistiksel uyuşma indeksleriyle değerlendirmek anlamlı değildir.

Son yıllarda bilimsel dergi editörleri ve hakemler yeni geliştirilen ölçeklerde istatistiksel içerik geçerliliği bulgularına yer verilmesini talep etmeye başlamışlardır. İstatistiksel içerik geçerliliği analizleri yapısal geçerliliği destekleyen temel öğelerden biridir.

Araştırmalarda istatistiksel içerik geçerliliğini kanıtlamaya yönelik olarak değişik hesaplamalara dayanan katsayılardan veya indislerden yararlanılmaktadır. Araştırmacı hesaplama veya istatistik yöntemlerden hangisini uygulayacağına ölçeği ilk defa geliştirme, uyarlama veya mevcut bir ölçeği kullanma durumuna göre kendisi karar verir. İlk defa geliştirilen ölçeklerde istatistiksel içerik geçerliliği mümkün olduğu kadar geniş bir çerçevede analiz edilir. İçerik geçerliliği önceden saptanmış ve/veya alanda sık uygulanmış olan ölçeklerde ise sadece bir veya iki yöntemin uygulanması yeterli görülebilir. Aslında bu tür ölçeklerde içerik geçerliliğinin gerçekten yapılıp yapılmadığı tam olarak bilinmediğinden yeniden içerik geçerliliği analizi yapmakta yarar vardır. Bir başka neden orijinal geliştiricinin ölçeğinde sık aralıklarla izinsiz değişikliklerin yapılması nedeniyle ölçeğin gerçekten “orijinal” olup olmadığının saptanamamasıdır. Bilim adamları, orijinal ölçekte yazarından veya geliştiricisinden izin alınmak suretiyle değişiklik yapılması halinde içerik geçerliliğinin yeniden yapılmasını önermişlerdir (Jennifer R. Gray, 2017, s. 380).

İçerik geçerliliğine kanıt oluşturmak üzere sık başvurulan istatistiksel yöntemler şunlardır: İçerik Geçerliliği İndeksi (İGİ)[[1]](#footnote-1), Değerlendiriciler Arası Uyuşum (DAU), Lawshe İçerik Geçerlilik Oranı (LİGO)[[2]](#footnote-2), Madde-toplam puan korelasyonu (MTPK).

İçerik geçerliliği analizlerini yaparken ölçek maddeleri hakemlere orijinal anket formunun etiketlerinden farklı etiketlerle sunulur. Bu etiketlerin her biri alan yazında “ölçek” olarak adlandırılır. Buradaki “ölçek” sözcüğü klasik anlamda tutum ölçeği olmayıp “derece etiketleri” anlamındadır. Bu şekilde belirlenen derece etiketleri veya ölçekler şunlardır: ilgililik, açıklık, basitlik, açıklık ve belirsizlik. Genelde dört dereceli olarak düzenlenen ölçeklerin tamamını bir araştırmada kullanmak pratik bir yarar sağlamayabilir. Araştırmacı baştan hangi ölçekleri ve hangi değerlendirme yöntemlerini kullanacağını belirleyerek bir plan hazırlar ve bu plana göre hareket eder. Genelde tek veya iki ölçek üzerinden değerlendirme yapmak yeterlidir. Bu çalışmada “ilgililik” ve “anlaşılırlık” ölçekleriyle ilgili tablolara yer verilmiştir.

Bilim adamları istatistiksel içerik geçerliliği analizleri için en azından üç hakem saptanmasını önermişlerdir. Bu sayının 10’dan fazla olmasının da fazla bir fayda sağlamayacağı bilinmektedir. Kimi yazarlar ise hakem sayısının 2 ila 20 arasında değişebileceği görüşünü öne sürmüşlerdir. Önerimiz yüksek lisans tezleri için 5; doktora tezlerinde 10 hakemden yararlanılması yönündedir. Bu hakemlerin bir bölümü iç uzman bir bölümü ise dış uzman niteliğindedir. İç uzmanlar ilgili bilim dalında yüksek lisans veya doktora yapmış öğretim üyeleri veya ilgili branşta iş dünyasında çalışan uzman kişilerdir. Dış uzmanlar ise, bilim dalı dışındaki öğretim üyeleri veya ölçüm konularında bilgi ve uzmanlık sahibi olan diğer kişilerdir. Araştırmacı hakemlerin veya KİU’ların cinsiyet, eğitim, unvan ve kurum olarak dağılımları hakkında araştırmasında ve yüzde dağılımları hakkında genel olarak bilgi verir. Hakemlerin yüzde kaçı üniversite öğretim üyelerinden yüzde kaçı alandan belirlenmiştir ve uzmanlık kıdemleri nedir bu konularda okuyucuları bilgilendirir.

*İGİ* *Geçerlilik analizleri*. İçerik geçerliliği İndeksi (İGİ) Waltz ve Bausell (1983) tarafından geliştirilmiştir (Bertea & Zait, 2013). Yeni geliştirilen ölçeklerde İGİ analizleri maddeler bazında ve genel ölçek bazında yapılmaktadır. Maddeler bazında yapılan geçerlilik (M-İGİ) [[3]](#footnote-3) ve ölçek bazında yapılan ise Ö-İGİ[[4]](#footnote-4) olarak nitelendirilmektedir. İGİ analizlerini yapmak için 1 = madde kavramsal yapıyla ilgili değil, 2= bir ölçüde ilgili, 3= büyük ölçüde ilgili, 4= tam ilgili dereceleme şıklarından yararlanılır. Değerlendirmede 3 ve 4 puanı veren hakem sayısı toplam hakem sayısına bölünür. Sonuç ilgililik konusunda uyuşma oranını verir. Dört veya beş hakemden yararlanılması durumunda 0,80 uyuşma oranı kabul edilebilir bir değer olarak görülür (Denise F. Polit, 2008, s. 459). Bir başka çalışmada ise bu oran 0,78 olarak öngörülmüştür (Rajeev Bali, 2013, s. 355). Düşük uyuşma oranına sahip maddeler ölçekten çıkarılır. Rahayu (2017) bu konuda otorite olan Lynn (1986)’den aktararak İGİ değeri >0,60 olan maddelerin ölçekten çıkarılmasını, 0,60 - 0,80 arasında olan maddelerin gözden geçirilmesini önermiştir (Rahayu, 2017). Ölçek düzeyinde uyuşma oranın belirlerken iki yaklaşım söz konusudur. Birincisi “tam uyuşma”[[5]](#footnote-5) halidir. Bu uygulama kısaltma olarak Ö-İGİ/TU sembolü ile gösterilir. Bu uygulamada 4 hakemin kaç madde üzerinde tam bir mutabakat içinde aynı puanları verdiklerine bakılır. Diyelim ki 4 hakem 10 maddeli bir ölçekte 7 maddeye 4 puanını vermiş olsunlar. Böyle bir durumda Ö-İGİ/TU= 0,70 olarak belirlenir. Fakat hakemlerin değerlendirme sonucu ortaya çıkan uyuşumları veya uyuşmazlıkları tesadüfen[[6]](#footnote-6) de ortaya çıkmış olabileceğinden bu değerlendirme “uzman yanlılığı” veya “yanlış anlaşılma” hata paylarını içeriyor olabilir. Çok kesin ve katı olan bu yaklaşım genelde tercih edilmez (Denise F. Polit, 2008). İkinci yaklaşım, Ö-İGİ’yi M-İGİ değerlerinin ortalamasını alarak hesaplamaktır. Bu hesaplama biçimi Ö-İGİ/ort simgesiyle gösterilir ve iyi bir içerik geçerliliği için kriter olarak 0,90 oranı önerilmiştir. Ölçeğin maddelerinde ve genel ölçeğin tamamında kriter puanları yakalama konusunda sorunlar yaşandığı durumda ifadelerde gerekli gözden geçirmeler yapılır ve hakemlere yeniden derecelendirme yaptırılarak gerekli oranlar bulununcaya kadar iyileştirme çalışmalarına devam edilir.

Tablo 1. Manevi Değerler Ölçeği İçerik Geçerliliği İndeksi Değerleri

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| İlgililik | | | | Anlaşılırlık | | | |
| Madde no | hakem sayısı | toplam | I-CVI | Madde no | hakem sayısı | toplam | I-CVI |
| 1 | 4 | 4 | 1,00 | 1 | 4 | 4 | 0,05 |
| 2 | 4 | 4 | 1,00 | 2 | 4 | 4 | 1,00 |
| 3 | 4 | 4 | 1,00 | 3 | 4 | 4 | 1,00 |
| 4 | 4 | 4 | 1,00 | 4 | 4 | 4 | 1,00 |
| 5 | 4 | 4 | 1,00 | 5 | 4 | 4 | 1,00 |
| 6 | 4 | 3 | 0,75 | 6 | 4 | 4 | 1,00 |
| 7 | 4 | 4 | 1,00 | 7 | 4 | 2 | 0,50 |
| 8 | 4 | 2 | 0,50 | 8 | 4 | 3 | 0,75 |
| 9 | 4 | 2 | 0,50 | 9 | 4 | 2 | 0,50 |
| 10 | 4 | 3 | 0,75 | 10 | 4 | 4 | 1,00 |
| 11 | 4 | 3 | 0,75 | 11 | 4 | 4 | 1,00 |
| 12 | 4 | 2 | 0,50 | 12 | 4 | 4 | 1,00 |
| 13 | 4 | 4 | 1,00 | 13 | 4 | 4 | 1,00 |
| 14 | 4 | 3 | 0,75 | 14 | 4 | 3 | 0,75 |
| 15 | 4 | 3 | 0,75 | 15 | 4 | 4 | 1,00 |
| 16 | 4 | 4 | 1,00 | 16 | 4 | 4 | 1,00 |
| 17 | 4 | 3 | 0,75 | 17 | 4 | 3 | 0,75 |
| 18 | 4 | 2 | 0,50 | 18 | 4 | 3 | 0,75 |
| 19 | 4 | 3 | 0,75 | 19 | 4 | 4 | 1,00 |
| 20 | 4 | 2 | 0,50 | 20 | 4 | 2 | 0,50 |
| S-CVI |  |  | 0,79 | S-CVI |  |  | 0,83 |

S-CVI of .80 oldukça yüksek bir iç geçerlilik oranıdır.

Tablo 2. Farklılıkların Yönetimi Ölçeği İçerik Geçerliliği İndeksi Değerleri

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| İlgililik | | | | Anlaşılırlık | | | |
| Madde no | hakem sayısı | toplam | I-CVI | Madde no | hakem sayısı | toplam | I-CVI |
| 1 | 4 | 4 | 1,00 | 1 | 4 | 4 | 1,00 |
| 2 | 4 | 3 | 0,75 | 2 | 4 | 4 | 1,00 |
| 3 | 4 | 4 | 1,00 | 3 | 4 | 4 | 1,00 |
| 4 | 4 | 4 | 1,00 | 4 | 4 | 4 | 1,00 |
| 5 | 4 | 4 | 1,00 | 5 | 4 | 3 | 0,75 |
| 6 | 4 | 3 | 0,75 | 6 | 4 | 4 | 1,00 |
| 7 | 4 | 3 | 0,75 | 7 | 4 | 3 | 0,75 |
| 8 | 4 | 3 | 0,75 | 8 | 4 | 4 | 1,00 |
| 9 | 4 | 4 | 1,00 | 9 | 4 | 4 | 1,00 |
| 10 | 4 | 2 | 0,50 | 10 | 4 | 2 | 0,50 |
| 11 | 4 | 4 | 1,00 | 11 | 4 | 3 | 0,75 |
| 12 | 4 | 4 | 1,00 | 12 | 4 | 4 | 1,00 |
| 13 | 4 | 3 | 0,75 | 13 | 4 | 4 | 1,00 |
| 14 | 4 | 4 | 1,00 | 14 | 4 | 3 | 0,75 |
| 15 | 4 | 4 | 1,00 | 15 | 4 | 4 | 1,00 |
| 16 | 4 | 3 | 0,75 | 16 | 4 | 4 | 1,00 |
| 17 | 4 | 3 | 0,75 | 17 | 4 | 4 | 1,00 |
| 18 | 4 | 4 | 1,00 | 18 | 4 | 4 | 1,00 |
| 19 | 4 | 3 | 0,75 | 19 | 4 | 4 | 1,00 |
| 20 | 4 | 3 | 0,75 | 20 | 4 | 4 | 1,00 |
| 21 | 4 | 3 | 0,75 | 21 |  | 4 | 1,00 |
| 22 | 4 | 4 | 1,00 | 22 |  | 4 | 1 |
| 23 | 4 | 4 | 1 | 23 |  | 4 | 1,00 |
| 24 | 4 | 3 | 0,75 | 24 |  | 4 | 0,75 |
| 25 | 4 | 3 | 0,75 | 25 |  | 3 | 1,00 |
| 26 | 4 | 4 | 1,00 | 26 |  | 4 | 1 |
|  |  |  | 0,87 |  |  |  | 0,93 |

*Değerlendiriciler Arası Uyuşum (DAU).* Belli sayıda hakemin bir ölçeğin maddelerine verdikleri puanlar arasındaki uyuşumu belirleyen oran Değerlendiriciler Arası Uyuşum (DAU) olarak adlandırılır (inter-rater agreement - IRA). Alan yazında Değerlendiriciler Arası Uyuşum (DAU) ile Değerlendiriciler Arası Güvenilirlik (DAG) (inter-rater reliability - IRR) kavramları çoğu kez eş anlamda kullanılır. Bununla birlikte iki kavram arasında teknik farklar vardır (Gisev, Bell, Simon, & Chen, 2013). Uyuşum, uzmanların verdikleri puanların birbirine benzer veya aynı olmasını ifade ederken; “güvenilirlik” maddelere verilen puanlardaki değişkenlik ve maddelerin içerdiği hata paylarıyla ilintilidir (Gisev, Bell, Simon, & Chen, 2013). Güvenilirlikte, bir hakemin ölçeğin değişik maddelerine verdiği puanların birbiriyle tutarlı (consistency)[[7]](#footnote-7) ve mantıklı olması önemlidir. Buna “puanlama trendi” adı verilir. Puanlama trendi, ortaya çıkan bir tür puanlama desenidir ve puanlamanın mantıklı olup olmadığını belirler. Uyuşumda ise değişik hakemlerin bir maddeye aynı puanı verip vermediğine bakılır. Gisev ve diğerleri (2013) DAU/DAG hesaplamaları arasında önemli bir farklılık olmadığını belirtmişlerdir. Burada önemli olan araştırmacının “güvenilirliği” mi yoksa “uyuşmayı” mı öne çıkardığıdır.

Değerlendiriciler arası uyuşumu/güvenilirliği belirlemek için hangi “indis”[[8]](#footnote-8) veya türevinin kullanılmasının daha doğru olacağı konusunda literatürde istatistiki tartışmalar vardır (Gisev, Bell, Simon, & Chen, 2013). Uyuşum indislerinin sayısı oldukça fazladır. Yapılan bir incelemede 16 farklı “hakemler arası uyum indisinin” var olduğu saptanmıştır (Tinsley & Brown, 2000). Araştırmacılar ölçüm verilerinin niteliğine ve hakem sayısına ağlı olarak değişik indislerden yararlanmaktadırlar. DAU ölçümü için Bland-Altman grafiği üzerinde duran araştırmacılar olduğu gibi DAU ve DAG için birlikte kullanılabilen Cohen Kappa yaklaşımını tercih eden bilim insanları da bulunmaktadır. İçerik geçerliliğinde “uyuşum” konusu üzerinde odaklanıldığından araştırmacı hakemler arası uyuşumu hakem sayısı ve ölçüm verilerinin niteliğini göz önünde bulundurarak belirlemedir. Aşağıdaki paragraflarda dokuz yaklaşıma ilişkin özet bilgiler verilmiş, yüksek lisans ve doktora tezleri için çoklu hakem-sıralı ölçek verilerine uygun dört yöntemin kullanılması önerilmiştir.

Birinci yöntem, manuel hesaplama yöntemidir. Bunu Değerlendiriciler Arası Basit Uyuşma İndisi (DA-BUİ) olarak isimlendirebiliriz. Bu yöntemin zayıf yöntemi hakemlerin bir maddeye “tesadüfen” aynı puanı vermiş olabileceklerini dikkate almamasıdır. Buna “rassal uyuşum” (chance agreement) adı verilir. Literatürde DA-BUİ “yüzdesel uyuşum” (percent agreement) olarak da adlandırılmıştır. Yaklaşımın değişik hesaplama yöntemleri vardır. Yüksek lisans öğrencileri için nispeten daha basit olan yaklaşım seçilmiştir. Bu nedenle daha sağlıklı olanı Cohen kappa veya diğer yöntemlerin uygulanmasıdır. Geliştirilen ölçeklere yönelik olarak DA-BUİ analizleri hakemlerin ikili gruplandırılmaları suretiyle yapılmıştır ve sonuçları Tablo 3 ve Tablo 4’te verilmiştir.

İkinci yöntem Cohen kappa [Kohen kapa] yaklaşımıdır. “Nominal veri” veya “sıralı veri” türünde iki değerlendirici veya hakem için uygulanır. SPSS’te uygulama yaparken sütunlara hakemlerin puanları ve satırlara ise ölçek maddelerine verilen puanlar girilir. Transform bölümüne girilerek iki değerlendirici arasındaki fark hesaplanır. Farklılık= Hakem1-Hakem 2. Analiz bölümünde tanımlayıcı/frekans menüsüne girilerek sadece “farlılık” değişkenin hesaplaması yaptırılır. Çıktılarda 0 değerinin frekansına bakılır ve bu değer “yüzdesel uyuşum” olarak değerlendirilir. Tesadüfi uyuşumu belirlemek için Analiz bölümüne girilir Crosstabs menüsü altında Row ve Column sütunlarına hakemlerin değerlendirme değişkenleri alınır. İstatistik bölümünde Kappa şıkkı seçili hale getirilir. Cells düğmesi altında Observed ve Expected şıkları seçilir ve OK tuşuna bakılarak hesaplama yapılır. Çıktılarda iki tabloya bakılır birincisi Crosstab ve ikincisi Symmetric Measures tablosudur. Crosstab tablosunda “Expected Count” değeri şans veya tesadüfilik faktörü dikkate alınarak elde edilmiştir. İkinci aşamada Symmetric Measures tablosu incelenir. Buradaki Kappa değerinin yüzdesel uyuşum değerinden daha düşük olduğu görülür. Bunun nedeni hesaplamada “şans” faktörünün dikkate alınmasıdır. Tablodaki *p* değeri hakemler arasındaki uyuşmanın anlamlı olup olmadığını gösterir.

**Ağırlıklandırılmış Kappa Syntax’ı**

matrix.

GET x /var=x1 to x4. (matristeki sütun veya hakem sayısına sayısına göre değişir.)

compute wt1=make(nrow(x),ncol(x),0).

compute wt2=make(nrow(x),ncol(x),0).

compute prop=x/msum(x).

loop i=1 to nrow(x).

loop j=1 to ncol(x).

compute wt1(i,j)=1-(abs(i-j)/(nrow(x)-1)).

compute wt2(i,j)=1-((i-j)/(nrow(x)-1))\*\*2.

end loop.

end loop.

compute wk1num=msum(wt1&\*prop)-msum(mdiag(rsum(prop))\*wt1\*mdiag(csum(prop))).

compute wk1den=1-msum(mdiag(rsum(prop))\*wt1\*mdiag(csum(prop))).

compute wk1=wk1num/wk1den.

print wk1.

compute wk2num=msum(wt2&\*prop)-msum(mdiag(rsum(prop))\*wt2\*mdiag(csum(prop))).

compute wk2den=1-msum(mdiag(rsum(prop))\*wt2\*mdiag(csum(prop))).

compute wk2=wk2num/wk2den.

print wk2.

end matrix.

Üçüncü yöntem Cohen kappa yaklaşımının 3 veya daha fazla değerlendirici için uyarlanmış halidir. Değerlendirici sayısı arttığı ve sıralı ölçek verileri kullanıldığı zaman bu yöntemden yararlanılır. Likert ölçekleri sıralı veri niteliğine sahip olduğundan normal Kappa hesaplaması için uygun değildir. Onun yerine Ağırlıklandırılmış Kappa hesaplaması yapılır. Bunun için SPSS’te veri tablosuna birinci sütuna hakemlerin kodları girilir ikinci sütuna ise 5 dereceli ölçekte verdikleri puanlar yazılır. Daha sonra Descriptive Statistics menüsü altında Crosstabs analizi yapılır ve sadece gözlem değerleri hesaplatılır. Çıktılardaki x1 \*x2 Crostabulation iç değerleri kopyalanarak x1, x2, x3, x4 şeklinde yeni değişkenler atanır ve Ağırlıklandırılmış Kappa için yazılmış olan Syntax çalıştırılır. Hesaplama sonucunda iki değer elde edilir. Birincisi WK1 ve İkincisi WK2 değeridir. WK1 doğrusal ağırlık kullanılmak suretiyle hesaplanan ağırlıklandırılmış kappa değeri iken WK2 quadratik ağırlık değerini gösterir. Hesaplamada p değerleri verilmez. Quadratik ağırlık değerinin ICC değerine eşit olduğu bildirilmiştir.

Dördüncü yaklaşım ICC veya Küme İçi Korelasyon analizi yönetimidir (KİK). KİK genellikle “eşit aralıklı” veya “oranlı” veriler için uygundur, fakat bazı araştırmacılar KİK’in “sıralı veriler” için de uygulanabileceğini belirtmişlerdir. Tinsley ve Brown (2000) KİK’in sıralı ve eşit aralıklı verilerde “değerlendiriciler arası güvenilirlik” için en iyi ölçüm yöntemi olduğunu belirtmiş nominal verilerin ise “değerlendiriciler arası uyuşum” için kullanılabileceğini ifade etmişlerdir (Tinsley & Brown, 2000, s. 117). Bu görüşe göre, nominal veriler “uyuşum” için kullanılabiliyorsa sıralı veriler de “uyuşum” ölçümü için rahatlıkla kullanılabilmelidir. Bu yaklaşımda 5 ile 10 arasında değişen bir grup değerlendirici “küme” olarak ele alınır ve bu kişilerin verdikleri puanların güvenilirliği veya uyuşumu hesaplanır. Güvenilirlik hesaplamasında Likert ölçeğinin “madde puanları” yerine “eşit aralıklılığı” temsil eden “ölçeğin boyutsal toplam/ortalama puanları” dikkate alınarak hesaplama yapılır. Uyuşum hesaplamasında ise doğrudan Likert ölçeğinin madde derecelendirmeleri temel alınır. SPSS’te KİK hesaplaması için Scale menüsünden yararlanılır. İstatistik bölümünde Two-way random veya Two-way mixed şıklarından biri seçili hale getirilir. Hakemler, alandaki “uzmanlar havuzundan” tesadüfi olarak seçilmişlerse “random” yöntemi; havuz oluşturulmadan belirli özellikler dikkate alınarak hakemler bilinçli olarak seçilmişlerse Two-way mixed şıkkı ile çalışılır. *Consistency* ve *absolute agreement* şıklarından ikincisi seçilir. Birincisi ölçüm verilerinin “güvenilirliğini” gösterirken ikincisi hakemlerin “uyuşum” oranı hakkında bilgi verir. Analiz, eğer “tek bir hakemin verileri” veya tek bir hakemin değerlendirmesi üzerinde yapılmışsa çıktılarda *single measure* değeri temel alınır. Birden fazla hakemin verileri üzerinde yapılmışsa *avarage measure* değerine bakılır. Uyuşum çoğunlukla çoklu hakem üzerinde sınandığından yorumlar *avarage measure* değeri dikkate alınarak gerçekleştirilir. *Two-way random* yöntemine göre *Two-way mixed* yöntemi genelleme yapmaya karşı biraz daha kısıtlayıcıdır. Çünkü “hakemler arası uyuşum” sadece bilinçli olarak seçilen hakemler grubu için geçerlidir. Bu hakemler tesadüfi olarak seçilmediklerinden uyuşum bulgusu diğer hakemlere genellenemez (Eye & Mun, 2012).

Beşinci yaklaşım Krippendorff *K-alfa* yöntemidir. Son yıllarda Likert tipi ölçek maddelerine uygun olması, ayrıca değişkenlerdeki “eksik veri” sorunlarını da başarıyla ele alabilmesi nedeniyle “uyuşum” konusunu belirlemek için “Krippendorff alfa” tekniği kullanılmaya başlanmıştır. Krippendorff alfa tekniği SAS, SPSS ve R istatistik yazılımlarında hesaplanabilir R yazılımında değerlendiriciler arası güvenilirliği değişik yöntemlere göre hesaplamak için *irr* adlı bir paket hazırlanmıştır.  Ayrıca İnternet ortamında ReCal OIR adlı veri tipine uygun çevrim içi hesaplama yapabilen bir modül bulunmaktadır. Krippendorff alfa eşitliğinin temel formu aşağıdaki gibidir:

Do gözlemlenen uyuşmazlık ve De ise şans faktörünü dikkate alan beklenenuyuşmazlıktır. Krippendorff alfa nominal, sıralı, eşit aralıklı ve oranlı veri tiplerinin her biri için rahatlıkla kullanılabilmektedir.

Altıncı yaklaşım literatürde Finn-katsayısı olarak bilinir. Finn katsayısı, varyansı kısıtlı veya çarpık veri durumları için daha uygun bir yöntem olarak değerlendirilmiştir. KİK yaklaşımının verilerin dağılım biçimine duyarlı olması, hakemlerin %80 oranında tam bir uyuşum içinde olmasına karşılık veri dağılımının çarpık olması nedeniyle KİK katsayısının sıfıra yakın çıkabildiği, bu yüzden bu gibi durumlarda *Finn r* katsayısının kullanılmasının daha doğru olacağı ifade edilmiştir (Raan, 1988, s.201). Finn (1970) sıralı ölçek verilerinde “değerlendirici içi varyans” değeri[[9]](#footnote-9) düşük olduğu zaman değerlendiriciler arası güvenilirlik için aşağıdaki formülün kullanılmasını önermiştir (Tinsley & Brown, 2000):

MSe tek yönlü varyans analizinden elde edilen “mean square” değeridir. K ölçekteki derece sayısı ve *i* indisi tek bir değerlendiricinin güvenilirliğini gösterir (Tinsley & Brown, 2000). *Finnr* katsayısı 0.50’nin üzerinde olan değerler tatmin edici olarak değerlendirilirken 0.70’in üzerindeki değerler “iyi” olarak tanımlanır (Rauner, 2010, s.141). Finn katsayısı hakemlerin yaptıkları değerlendirmelerin değişkenliklerine veya puanlama örüntüsüne bağımlı değildir.

Yedinci yaklaşım “Fleiss kappa” yöntemidir. İkili veriler (0 ve 1); nominal ölçek verileri ve sıralı ölçek verileri için uygun olduğu bildirilmiştir. Cohen kappa’nın tersine ikiden fazla hakemin kategorik ve sıralı veriler üzerinde değerlendirme yapmasına uygundur. Fleiss kappa *Rstudio* yazılımıyla hesaplanabilir. Ancak çok aşamalı hesaplama yöntemi bir ölçüde karmaşık gibi gözükebilir.

Sekizinci yaklaşım Korelasyon katsayıları yöntemidir. Bu yaklaşımda Pearson r, Kendall r ve Spearman r yöntemlerinden yararlanılabilir. Pearson yaklaşımında veriler sürekli niteliğinde iken Kendall ve Spearman yaklaşımlarında sıralı niteliktedir. Eğer ikiden fazla değerlendiriciden yararlanılmışsa korelasyon katsayılarının ortalamalarına bakılarak karar verilir.

Dokuzuncu yaklaşım Bland-Altman Grafiği yöntemidir. İki değerlendirici olması ve verilerin sürekli niteliğe sahip olması halinde uygulanır. Daha çok sıralı verilerle çalışılan yönetim araştırmalarında daha az kullanılır.

Görüldüğü gibi değerlendiriciler arası uyuşumu belirlemeye yönelik olarak farklı yöntemler söz konusudur. Fakat sıralı veriler için bunların içinde daha uygun olan dört yöntem dikkati çekmektedir. Bunlar (1) değerlendiriciler arası basit uyuşma indisi, (2) ağırlıklandırılmış kappa, (3) KİK yöntemi ve (4) Krippendorff alfa uyuşum katsayısıdır. Fleiss kappa yöntemi de uygun olmakla birlikte yüksek lisans ve doktora öğrencilerinin bu konuda istatistik destek almalarına ihtiyaç vardır.

Tablo 3. Maneviyat Ölçeği Değerlendiriciler Arası Uyuşma (DAU)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| İlgililik | | | | Anlaşılırlık | | | |
| Madde sayısı | hakem sayısı | Hakemler | HAU | Madde sayısı | hakem sayısı | Hakemler | HAU |
| 20 | 2 | 1-2 | 0,45 | 20 | 2 | 1-2 | 0,40 |
| 20 | 2 | 1-3 | 0,30 | 20 | 2 | 1-3 | 0,35 |
| 20 | 2 | 1-4 | 0,60 | 20 | 2 | 1-4 | 0,90 |
| 20 | 2 | 2-3 | 0,30 | 20 | 2 | 2-3 | 0,50 |
| 20 | 2 | 2-4 | 0,15 | 20 | 2 | 2-4 | 0,40 |
| 20 | 2 | 3-4 | 0,25 | 20 | 2 | 3-4 | 0,35 |
| Ortalama |  |  | 0,34 | Ortalama |  |  | 0,48 |

HAU: Hakemler arası uyuşum.

DAU oranı ilgililikte 0,34 anlaşılırlık ölçeğinde 0,48 çıkmıştır. Alan yazında 0.40 ila 0,60 arasındaki değerler orta derecede uyuşum olarak yorumlanmaktadır. İlgililik boyutu “makul” ölçülerdedir. Maneviyat ölçeğinde değerlerin düşük çıkması bu tür ölçeklerin Türkiye’de çok yoğun kullanılmamasından kaynaklanıyor olabilir. Bununla birlikte ölçeğin özellikle düşük puan alan dört maddesi yeniden gözden geçirilmiş ve ilgililiğini artırmaya yönelik olarak bazı iyileştirmeler yapılmıştır.

Tablo 4. Farklılıkların Yönetimi Değerlendiriciler Arası Uyuşma (DAU)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| İlgililik | | | | Anlaşılırlık | | | |
| Madde sayısı | hakem sayısı | Hakemler | HAU | Madde sayısı | hakem sayısı | Hakemler | HAU |
| 26 | 2 | 1-2 | 0,58 | 26 | 2 | 1-2 | 0,50 |
| 26 | 2 | 1-3 | 0,19 | 26 | 2 | 1-3 | 0,46 |
| 26 | 2 | 1-4 | 0,85 | 26 | 2 | 1-4 | 0,100 |
| 26 | 2 | 2-3 | 0,35 | 26 | 2 | 2-3 | 0,46 |
| 26 | 2 | 2-4 | 0,54 | 26 | 2 | 2-4 | 0,50 |
| 26 | 2 | 3-4 | 0,19 | 26 | 2 | 3-4 | 0,46 |
| Ortalama |  |  | 0,45 | Ortalama |  |  | 0,56 |

HAU: Hakemler arası uyuşum.

*Lawashe içerik geçerliliği oranı (LİGO)***.**  İçerik Geçerliliği Oranı yaklaşımı İçerik Geçerliliği indeksine benzer. Madde düzeyli bu analizde geçerlilik çalışması yapacak hakemlerin sayısının en az 5 olması önerilmiştir (Norashady Mohd Noor, 2017). Geliştirilen ölçeklerde Lawshe (1975) “bireysel madde düzeyli” geçerlilik analizleri formül yardımı aracılığıyla hesaplanmıştır ve sonuçları Tablo 5 ve Tablo 6’de verilmiştir. Geçerlilik analizi formülü şu şekildedir: LİGO = (ne – N/2) / (N/2). N = Toplam hakem sayısı, ng[[10]](#footnote-10) =gerekli veya “ilgililik” etiketi kullanılmışsa 3 ve 4’ü işaretleyen hakem sayısı. Lawashe yaklaşımında “dereceleme etiketi” olarak “gerek­lilik” temasından yararlanılır. Fakat hakemlere ayrı bir gereklilik ölçeği tanım­lanmadığından ilgililik ölçeği aynı zamanda “gereklilik ölçeği” gibi değerlendi­rilebilir. “Gereklilik” derece etiketleri üç düzeyli olarak şu şekilde belirlenir. 1= gerekli değil 2=yararlı, fakat gerekli değil, 3= gerekli.

Lawshe LİGO oranı için asgari 0,49 değerini belirlemiştir. Bu rakamın altında “LİGO değerine” sahip bir maddenin içerik geçerliliğine sahip olmadığı varsayılır. LİGO 0,49 değeri hakemlerin %75 oranında uyuşum içinde olduklarını gösterir. Başka yazarlar ise 0,80 gibi daha sıkı uyuşum değerleri önermişlerdir. Lawshe LİGO kritik değerleri hakem sayısına göre değişir ve aşağıdaki gibidir:

|  |  |
| --- | --- |
| Hakem sayısı | Minimum LİGO değeri |
| 5 | 0,99 |
| 6 | 0,99 |
| 7 | 0,99 |
| 8 | 0,75 |
| 9 | 0,78 |
| 10 | 0,62 |
|  |  |

Lawshe’nin yaklaşımı esas olarak maddelerin içerik geçerliliğiyle ilgilidir. Bütün ölçeğin geçerliliği hakkında fazla bir fikir vermez. Bu amaçla bu yöntemin Gregory (1996) ve Topf (1986) yöntemiyle desteklenmesinde yarar vardır. Toplam İçerik Geçerliliği -TİG (%) = B / A + B. Formüldeki B değeri LİGO içerik geçerliliğine sahip maddelerin sayısıdır. A + B ise ölçekteki tüm maddelerin sayısıdır. TİG değerinin 0,79 olması hakemlerin 0,90 oranında uyuşum içinde olduğunu gösterir.

Tablo 5. Lawashe geçerlilik Analizleri sonuçları

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Maneviyat ölçeği (İlgililik/gereklilik) | | | | Farklılıkların Yönetimi Ölçeği (İlgililik/gereklilik) | | | |
| ne \* | N/2 | ne-N/2 | (ne-N/2)/(N/2) | ne \* | N/2 | ne-N/2 | (ne-N/2)/(N/2) |
| 4 | 2 | 2 | 1 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 3 | 2 | 1 | 0,5 |
| 4 | 2 | 2 | 1 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 3 | 2 | 1 | 0,5 |
| 3 | 2 | 1 | 0,5 | 3 | 2 | 1 | 0,5 |
| 2 | 2 | 0 | 0 | 3 | 2 | 1 | 0,5 |
| 3 | 2 | 1 | 0,5 | 4 | 2 | 2 | 1 |
| 3 | 2 | 1 | 0,5 | 2 | 2 | 0 | 0 |
| 3 | 2 | 1 | 0,5 | 4 | 2 | 2 | 1 |
| 2 | 2 | 0 | 0 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 3 | 2 | 1 | 0,5 |
| 2 | 2 | 0 | 0 | 4 | 2 | 2 | 1 |
| 3 | 2 | 1 | 0,5 | 4 | 2 | 2 | 1 |
| 4 | 2 | 2 | 1 | 3 | 2 | 1 | 0,5 |
| 3 | 2 | 1 | 0,5 | 3 | 2 | 1 | 0,5 |
| 2 | 2 | 0 | 0 | 4 | 2 | 2 | 1 |
| 3 | 2 | 1 | 0,5 | 3 | 2 | 1 | 0,5 |
| 2 | 2 | 0 | 0 | 3 | 2 | 1 | 0,5 |
|  |  |  |  | 3 | 2 | 1 | 0,7 |
|  |  |  |  | 4 | 2 | 2 | 0,7 |
|  |  |  |  | 4 | 2 | 2 | 0,7 |
|  |  |  |  | 3 | 2 | 1 | 0,7 |
|  |  |  |  | 3 | 2 | 1 | 0,7 |
|  |  |  |  | 4 | 2 | 2 | 0,7 |
| Ortalama |  |  | 0,58 |  |  |  | 0,70 |

\* 3 ve 4’ü işaretleyenler.

*Madde-Toplam Puan Korelasyonu MTPK*. Bu yaklaşımda ölçeğin maddeleri toplam puanla Pearson korelasyon analizine tabi tutulur ve hesaplama sonucunda *P* anlamlılık (çift yönlü) değerlerine bakılır. Hesaplamada *P* değeri 0,000 < 0,05 çıkmışsa ilgili maddenin “geçerli” olduğuna karar verilir.

1. İngilizce kısaltması CVI. [↑](#footnote-ref-1)
2. İngilizcesi: Lawshe Content Validity Ratio. [↑](#footnote-ref-2)
3. İngilizce kısaltması I-CVI. [↑](#footnote-ref-3)
4. İngilizce kısaltması S-CVI. [↑](#footnote-ref-4)
5. Universal agreement. [↑](#footnote-ref-5)
6. Veya “şans eseri”. [↑](#footnote-ref-6)
7. *Consistency*. tutarlılık, mantıklılık. Ölçek maddelerine yapılan işaretlemelerin toplam puanla ne ölçüde ilişkili olduğu veya ilişkililiğin anlamlı olup olmadığı. Bu konuda P değerine bakarak karar verilir. P değeri 0.05’ten küçük çıkmışsa maddenin tutarlı veya mantıklı olarak işaretlendiğine karar verilir. Hakemin bir maddeye kaç puan verdiği değil, maddelere verilen puanların birbiriyle tutarlı veya yanıtların mantıklı olup olmadığı önemlidir. Yanıtlar birbirini doğrulayan bir niteliğe sahip olmalıdır. [↑](#footnote-ref-7)
8. *İndis*. Belli bir formülle ortaya çıkarılan değer. [↑](#footnote-ref-8)
9. Değerlendirici içi varyans: Mean square within = sums of square within / serbestlik derecesi. [↑](#footnote-ref-9)
10. İngilizce kaynaklarda *ne* “n essential” anlamında. [↑](#footnote-ref-10)