

Eşdeğer Yarılar Güvenirliğinin Farklı Homojenlik Düzeylerindeki Örneklem Büyüklüklerinde, Test Uzunluğuna, Yarıya Bölme Yöntemlerine ve Güvenirlik Kestirme Tekniklerine Göre İncelenmesi*

Examining Split-Half Reliability for Sample Sizes In Different Homogeneity Levels According to Test Length, Split-Half Methods And Reliability Estimation Techniques

Yeşim Beril SOĞUKSU**, Devrim ALICI***

Öz: Bu araştırmanın amacı, farklı madde uzunluklarına sahip testlerin eşdeğer yarılar güvenilirliklerini, düşük, orta ve yüksek düzey homojenlikteki gruplarda, farklı yarıya bölme yöntemlerine ve güvenirlik kestirme tekniklerine dayalı olarak incelemektir. Araştırmada yer alan evrenlere ilişkin veri üretimi Wingen 3 programında, tek boyutlu Madde Tepki Kuramı modellerinden 2 parametrelili lojistik modele dayalı olarak gerçekleştirilmiştir. Örneklem büyüklüğünün eşdeğer yarılar güvenirliğine olan etkisinin incelenmesi için her evrenden basit seçkisiz örnekleme ile 50'şer adet 50, 300 ve 500 birimlik örneklem çekilmiştir. Evrenlere ve örneklemelere ilişkin güvenirlik ve standart hata hesaplamaları için, MATLAB 6.5 programlama dili kullanılmıştır. Sonuç olarak, çalışma grubunun heterojenleşmesi, örneklem büyüklüğü ve test uzunluğunun artışı ile eşdeğer yarılar güvenirliklerinin arttığı, güvenirliklere ilişkin standart hataların ise azaldığı gözlenmiştir. Tekler-çiftler ve madde gücüne göre tekler-çiftler yöntemleriyle yarıya bölünen testlerden elde edilen eşdeğer yarılar güvenirliklerinin daha yüksek, standart hataların ise daha düşük çıktığı belirlenmiştir. Tüm homojenlik düzeyindeki gruplarda, kullanılan güvenirlik kestirme tekniklerinin, eşdeğer yarılar güvenirlik ortancaları ve standart hataları arasında neredeyse yok denecek kadar az bir fark oluşturduğu gözlenmiştir. Araştırmada yer alan gruplar heterojenleştikçe, örneklem büyüklüğü ve test uzunluğu arttıkça; tekler-çiftler ve madde gücüne göre tekler-çiftler yöntemlerinin kullanıldığı durumlarda evren ve örneklem arasındaki uyumun arttığı gözlenmiştir.

Anahtar Kelimeler: Güvenirlik, eşdeğer yarılar güvenirliği, homojenlik, Spearman-Brown düzeltme formülü, Rulon formülü, Flanagan formülü

Abstract: The purpose of the study is to examine split-half reliabilities of tests in different lengths for low, medium and high homogeneity level groups based on different split-half methods and reliability estimation techniques. The unidimensional data simulation of the study was generated on Wingen 3 software depending on 2 Parametric Logistic Model of Item Response Theory. In order to determine how split-half reliabilities is affected by sample size, 50 samples for 50, 300 and 500 sample sizes were drawn from populations by using simple random sampling. The split-half reliabilities and the standard errors of sample reliabilites were computed by codes written in MATLAB 6.5 programming language. On examining the results, it was found that the more heterogeneous is the sample, the more higher the sample and test length, the higher is the split-half reliabilities and the standard errors of reliabilities were decreased. When the tests split into halves with odds-evens method and odds-evens according to item difficulties method, the split-half reliabilities were higher; the standard errors were lower. The difference between the standard errors and the medians of split-half reliabilites attained by reliability estimation techniques in all homogeneity groups was scarcely any. It was found that the consistency between populations and samples was higher in the cases where the groups were heterogenous; the test length and the sample sizes were high and odds-evens method and odds-evens according to item difficulties method were used.

* Bu çalışma, Yeşim Beril SOĞUKSU tarafından Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü'nde Yrd. Doç. Dr. Devrim ALICI danışmanlığında yapılan yüksek lisans tez çalışmasının bir bölümüdür.

** Uzm., Mersin Üniversitesi, Eğitim Fakültesi, Mersin-Türkiye, e-posta: b_emrealp@hotmail.com

*** Doç. Dr., Mersin Üniversitesi Eğitim Fakültesi, Mersin-Türkiye, e-posta: devrimo.alici@gmail.com

Keywords: Reliability, split-half reliability, homogeneity, Spearman-Brown prophecy formula, Rulon formula, Flanagan formula

Giriş

Eğitimde ve psikolojide ölçme, bireyler hakkında kararlar verme ve değerlendirmede vazgeçilmez bir unsurdur. Yapılan ölçme işlemi ile bireylerin doğrudan gözlenemeyen psikolojik yapılarına ilişkin kararlar verilir. Dolayısıyla bireyler hakkında verilen kararların, yapılan değerlendirmelerin doğru olabilmesi için, kullanılan ölçme araçlarının hatasız ölçümler yapması gerekliliği doğmaktadır. Ölçme araçlarının hatasız ölçüm yapma gücü güvenilirlik olarak tanımlanır. Crocker ve Algina (1986) güvenilirliği, aynı bireylerin benzer koşullarda test edilmesiyle elde edilen puanların tutarlılığı olarak tanımlamıştır. Baykul (2000) ise güvenilirliği, aynı araçla ve aynı bireyler üzerinde yapılan birden çok ölçmede aynı sonuçların alınabilmesi olarak ifade etmiştir.

Bir testin güvenilirliğini etkileyen çeşitli faktörler vardır. Bunlar: Test uzunluğu, testin süresi ve grubun homojenliği şeklinde sıralanabilir. Testin uzunluğu, hem gözlenen puan varyanslarını hem de gerçek puan varyanslarını etkiler. Bir testteki madde sayısı n katı kadar arttırıldığında, gerçek puan varyansı n^2 kadar artacaktır. Bunun yanı sıra testteki madde sayısı n katı kadar arttırıldığında, hata puanı varyansı da n kadar artacaktır (Magnusson, 1968). Burnett (1974), teste eklenecek olan maddelerin güçlük, iç korelasyon ve içerik bakımından orjinal testteki maddelerle benzer olması gerekliliğinin altını çizmiştir. Bir testin, cevaplayıcıların tamamı ya da büyük çoğunluğu tarafından cevaplanabilmesi için yeterli süre verilmesi gerekir. Aksi takdirde, güvenilirlik katsayısı yapay bir şekilde yükselecektir. Cureton (1958), güvenilirlik katsayıları KR-20 veya eşdeğer yarılar yöntemleri ile hesaplanacaksa, cevaplayıcıların testte yer alan tüm maddeleri cevaplayabilecekleri kadar zaman verilmesi gerektiğini, aksi takdirde güvenilirlik kestirimlerinin olduğundan yüksek çıkacağını ifade etmiştir. Fiske (1973), grup homojenliğini, bireylerin psikolojik özelliklerine ilişkin benzerliğin derecesi olarak tanımlamıştır. Miller (1995), test güvenilirliğinin belli koşullar altında uygulandığı gruba bağlı olduğunu ifade etmiştir. Güvenirlik katsayısı, testi cevaplayan grubun puanlarına ait bir özellik olduğuna göre, grubun homojen ya da heterojen bir yapıya sahip olması, güvenilirlik katsayılarını etkileyecektir. Crocker ve Algina'ya (1986) göre, bir test heterojen bir gruba uygulandığında, gözlenen puan ve gerçek puan varyansları daha yüksek olacağından ve hata varyansları değişmeyeceğinden, daha yüksek güvenilirlik katsayısına sahip olması beklenir.

Bir testin güvenilirliği farklı yöntemlerle hesaplanabilir. Bu yöntemler; birden çok uygulamaya dayalı yöntemler ve tek uygulamaya dayalı yöntemler olmak üzere iki grupta incelenebilir (Crocker ve Algina, 1986).

Birden çok uygulamaya dayalı, 2 yöntem vardır. Bu yöntemler, test-tekrar test yöntemi ve eşdeğer formlar yöntemidir. Test-tekrar test yönteminde, tek bir test aynı cevaplayıcı grubuna, iki farklı zamanda uygulanır ve bu iki uygulamadan elde edilen puanlar arasındaki korelasyon hesaplanır (Baykul, 2000). Test-tekrar test yöntemi ile elde edilen katsayıya, "kararlılık (stability) katsayısı" adı verilir. Eşdeğer formlar yönteminde, testin eşdeğer formları oluşturulur ve oluşturulan eşdeğer formlar aynı gruba farklı zamanlarda uygulanır. Gulliksen (1967), iki testin eşdeğer olabilmesi için, ortalamalarının, varyanslarının ve iç korelasyonlarının aynı olması gerektiğinin altını çizmiştir.

Tek uygulamaya dayalı yöntemler, tek bir testin, bir defada bir grup cevaplayıcıya uygulanmasına dayanır. Bu yöntemlerde hesaplanan güvenilirlik katsayıları, "iç tutarlık katsayıları" olarak adlandırılır. Miller (1995), tek uygulamaya dayalı yöntemlerde, testteki maddelerin homojen ve testin tek boyutlu olması sayılıtlarının olduğunu belirtmiştir. Tek uygulamaya dayalı, 3 farklı yöntem vardır. Bunlar; Cronbach Alfa, Kuder Richardson ve eşdeğer yarılar yöntemleridir.

Cronbach Alfa, çoklu puanlanan test maddelerinin iç tutarlılığını hesaplamada sıklıkla kullanılan bir katsayıdır. Cortina (1993), testteki maddelerin standart sapmalarının birbirine eşit olması ve teste ait eşdeğer yarılar güvenilirliğini hesaplamada Rulon veya Flanagan

formülü kullanılması durumunda, Cronbach Alfa'nın, olası bütün yarıya bölme yöntemleri ile hesaplanacak olan eşdeğer yarılar güvenirlüklerinin ortalamasına eşit olacağını ifade etmiştir.

Kuder-Richardson Yönteminde, testte yer alan her maddenin birbirine paralel olduğu kabul edilerek işlemler yapılır. Dolayısıyla her maddenin ortalaması ve varyansının aynı olması gerekmektedir. Bu yöntemde, en çok kullanılan işlemler Kuder Richardson 20 (KR-20), Kuder Richardson 21 (KR-21) işlemleridir. Kuder Richardson (KR-20) katsayısı, Cronbach Alfa katsayısı ile çok benzerlik göstermekle beraber, ikili puanlanan test maddelerinde kullanılabilir. Her maddenin eşit güçlükte olması durumunda, Kuder Richardson maddelerin varyansının hesaplanmasını gerektirmeyen KR-21 formülünü geliştirmiştir.

Eşdeğer yarılar yönteminde, test bir grup cevaplayıcıya uygulanır. Test puanlanmadan önce, testin maddeleri iki eşdeğer form olacak şekilde yarıya bölünür. Her form, orjinal testin yarısı kadar madde sayısına sahip olur. Bu yöntem ile oluşturulacak iki yarı formun eşdeğer olması gerekmektedir (Crocker ve Algina, 1986). Anastasi ve Urbina'ya (1997) göre, eşdeğer yarılar güvenirlüğünü hesaplamada çözülmesi gereken ilk problem, iki eşdeğer form oluşturabilmek için testin nasıl yarıya bölüneceğidir. Crocker ve Algina (1986), testi yarıya bölmek için 4 yöntem önermiştir. Bunlar; tekler-çiftler (odds-even) yöntemi, madde güçlüğüne göre tekler-çiftler yöntemi, seçkisiz atama yöntemi ve içerik eşleştirme yöntemidir.

Tekler-Çiftler (Odds – Evens) yönteminde, test cevaplayıcı grubuna uygulandıktan sonra, testteki tek numaralı maddeler formlardan birincisine, çift numaralı maddeler ise formlardan ikincisine yerleştirilir. Gulliksen (1967), tekler-çiftler yöntemini kullanabilmek için testteki maddelerin birbirinden bağımsız olmaları, dolayısıyla bir maddeyi cevaplayamamanın, ondan sonraki maddeyi cevaplama durumunu etkilememesi gerektiğini belirtmiştir. Lord ve Novick (1968), testin tamamının cevaplanabilmesi için yeterli zaman verilmediği durumlarda, testin iki yarısındaki hata puanlarının yüksek korelasyon göstermesi sebebiyle, bu yöntemle hesaplanacak güvenirlük katsayısının, testin gerçek güvenirlük katsayısından daha yüksek olacağını belirtmiştir.

Madde güçlüğüne göre tekler-çiftler yönteminde, test cevaplayıcı grubuna uygulandıktan sonra, madde analizi yapılarak testteki maddelerin güçlük indeksleri hesaplanır. Maddeler, güçlük indekslerine göre kolaydan zora doğru sıralanır ve bu sıraya göre tekrar numaralandırılır. Tek numaralı maddeler formlardan birincisine, çift numaralı maddeler ise formlardan ikincisine yerleştirilir. Magnusson (1968), eğer test homojense, bu yöntemle testin iki yarı formunun eşit güçlükte ve ayırt edicilikte olacağını, aynı zamanda eşit ortalamalara ve varyanslara sahip olacağını, dolayısıyla iki yarı formun eşdeğer olabileceğini belirtmiştir. Fakat testin heterojen olması durumunda, iki yarı formun varyansları ve ortalamaları eşit olsa da formlar içerik açısından aynı olamayacağı için, iki yarı formun eşdeğer olamayacağını vurgulamıştır.

Seçkisiz atama yönteminde, seçkisiz olarak seçilen maddelerden biri eşdeğer formlardan birincisine, diğeri ise ikincisine konulabilir. Seçkisiz atama ile, her maddenin, eşdeğer formlara yerleştirilme olasılığı eşit ve birbirinden bağımsız olacaktır. Magnusson (1968), testin homojen olması durumunda, bu yöntemin kullanılarak eşdeğer iki formun oluşturulabileceğini belirtmiştir.

İçerik eşleştirme yönteminde ise, testteki maddeler yokladıkları davranışlara göre eşleştirilir. Eşlerden biri formlardan birincisine, diğeri formlardan ikincisine yerleştirilir. Magnusson (1968), testin heterojen olması durumunda, test maddelerinin güçlüğü de dikkate alınarak, bu yöntemin kullanılmasının daha doğru olacağını ifade etmiştir.

Test herhangi bir yarıya bölme yöntemi ile bölündükten sonra, eşdeğer yarılar güvenirlüğünü kestirmede, Spearman-Brown düzeltme formülü, Rulon formülü, Guttman formülü ve Flanagan formülünden yararlanılabilir.

Test yarıya bölünüp, iki yarı form oluşturulduktan sonra, her cevaplayıcının iki yarı formdan aldığı puanlar arasındaki korelasyon, Pearson Momentler Çarpımı katsayısı ile hesaplanır. Bu katsayı, yarı testin güvenirlük katsayısıdır (Crocker ve Algina, 1986). Testin tamamına ilişkin güvenirlük katsayısı, Spearman Brown düzeltme formülü yardımıyla hesaplanır. Spearman-Brown düzeltme formülü 1 no'lu eşitlikteki gibidir.

$$r_{tt} = \frac{2r_{12}}{1+r_{12}} \quad (1)$$

r_{tt} : Testin tamamına ilişkin güvenilirlik katsayısı

r_{12} : Yarı teste ilişkin güvenilirlik katsayısı

Lord ve Novick (1968), Spearman Brown düzeltme formülünün, testin iki yarı formunun eşdeğer olması sayılıtısına dayandığını; eğer eşdeğerlik sayılıtısı sağlanmıyorsa, Spearman Brown düzeltme formülünün olduğundan yüksek güvenilirlik katsayıları vereceğini ifade etmiştir.

Eşdeğer yarılar güvenilirliğini hesaplamanın diğer bir yolu, Rulon formülünü kullanmaktır. Magnusson (1968), bu formülün testin iki yarı formuna ilişkin varyansların eşit olması sayılıtısını gerektirmediğini belirtmiştir. Bu formül ile testin tamamına ilişkin güvenilirlik katsayısı hesaplanmaktadır. Rulon formülü 2 no'lu eşitlikteki gibidir.

$$r_{tt} = 1 - \frac{S_d^2}{S_t^2} \quad (2)$$

r_{tt} : Güvenirlik katsayısı

S_d^2 : Testin iki yarı formundan alınan puanların farklarının varyansı

S_t^2 : Test puanlarının varyansı

Testin tamamına ilişkin eşdeğer yarılar güvenilirlik katsayısı, Guttman formülü ile de hesaplanabilir. Gulliksen (1967), testin iki yarı formuna ilişkin varyansların eşit olması sayılıtısı altında, Guttman ve Rulon formülleriyle elde edilecek eşdeğer yarılar güvenilirlik katsayılarının birbirine eşit olacağını ifade etmiştir. Guttman formülü 3 no'lu eşitlikteki gibidir.

$$r_{tt} = 2\left[1 - \frac{S_1^2 + S_2^2}{S_t^2}\right] \quad (3)$$

r_{tt} : Güvenirlik katsayısı

S_1^2 : Birinci formun varyansı

S_2^2 : İkinci formun varyansı

S_t^2 : Testin tamamının varyansı

Eşdeğer yarılar güvenilirliğini hesaplamada kullanılan diğer bir formül Flanagan formülüdür. Flanagan formülü de Rulon formülüne benzerlik göstermektedir. Testin iki yarı formuna ilişkin varyansların toplamı alınarak, hata varyanslarının kestiriminde bulunmaktadır (Guilford, 1954). Flanagan formülü 4 no'lu eşitlikteki gibidir.

$$r_{tt} = \frac{4r_{12}S_1S_2}{S_1^2 + S_2^2 + 2r_{12}S_1S_2} \quad (4)$$

r_{tt} : Güvenirlik katsayısı

S_1^2 : Birinci formun varyansı

S_2^2 : İkinci formun varyansı

r_{12} : Pearson Momentler Çarpımı korelasyon katsayısı

Alanyazında, testlerin iki yarıya ayrılması ve iki yarı güvenirliğinin farklı yöntemlerle elde edilmesine ilişkin az sayıda çalışma bulunmaktadır. Charter (2001), testin iki yarı formunun eşdeğer olmadığı durumlarda, Spearman Brown düzeltme formülü ile hesaplanan eşdeğer yarılar güvenirlilik katsayısının olduğundan daha yüksek çıktığını, dolayısıyla Flanagan formülünün kullanılmasının daha uygun olacağını belirtmiştir.

Callender ve Osburn (1977), yaptıkları çalışmada, en yüksek eşdeğer yarılar güvenirlilik katsayılarının elde edilebileceği MSPLIT adında bir algoritmadan söz etmiştir. MSPLIT algoritmasıyla, en yüksek toplam kovaryansın elde edilebileceği şekilde, testin iki yarı formuna maddeler atanmıştır. Sonuç olarak MSPLIT ve tekler-çiftler yöntemleri ile eşdeğer yarılar güvenirliliklerini ve KR-20 katsayılarını hesaplamışlardır. Araştırmanın sonucunda, en yüksek güvenirlilik katsayılarının MSPLIT yöntemi ile kestirildiği ortaya konmuştur.

Charter (1999), farklı örneklem büyüklüklerini kullanarak test-tekrar test, alternatif formlar, eşdeğer yarılar, alfa, sınıfıçı, puanlayıcı güvenirlilikleri ile geçerlik katsayılarının %95 güven aralığında ne derece kesin olarak kestirileceğini araştırmış ve bütün güvenirlilik kestirme yöntemlerinde, örneklem büyüklüğü arttıkça güven aralıklarının daraldığı gözlenmiştir. Güvenirlilik kestiriminin 0,90 ve üzeri olması durumunda, örneklem büyüklüğünün en az 400 ve üzeri olması gerektiği bulunmuştur.

Charter (2001), çalışmasında testin yarı formlarının standart sapmalarının birbirine eşit olmadığı durumlarda Spearman Brown düzeltme formülü ile yapılan hesaplamaların ne kadar hata oluşturduğunu belirlemeye çalışmıştır. Sonuç olarak, eşdeğer yarılar güvenirliğinin hesaplanmasında, Spearman Brown düzeltme formülü yerine, Flanagan formülünün kullanılmasının daha uygun olduğu, testin iki yarı formu eşdeğerlik sayılınsını karşılamadığında, Spearman Brown düzeltme formülü ile hesaplanan güvenirlilik katsayılarının olduğundan daha yüksek çıktığı belirlenmiştir.

Charter (2003), dergilerde, test kılavuzlarında ve test eleştiri kitaplarında yer alan güvenirlilik çalışmalarını ele alarak, araştırmacıların yaptıkları çalışmalarda hangi örneklem büyüklükleriyle çalıştıklarını ve hangi çalışmalarla yeterli kesinlikte güvenirlilik kestirimleri yapıldığını incelemiştir. Sonuç olarak %25, %50 ve %75'lik dilimlere bakıldığında iç tutarlılıkla ilgili çalışmaların çoğunda örneklem genişliğinin düşük olduğu ve yeterli kesinlikte güvenirlilik kestirimleri yapılamadığı görülmüştür. Bunun yanı sıra çalışmaların sadece %25'lik bir kısmında, yeterli kesinlikte güvenirlilik kestirimleri yapılabilecek büyüklükte örneklem genişlikleri kullanıldığı görülmüştür.

Walker (2006), testin iki yarısının farklı standart sapmalara sahip olduğunda, eşdeğer yarılar güvenirliğinin Flanagan formülüyle daha iyi kestirileceğinden yola çıkarak, Cronbach (1951) ve Charter (1996)'ın çalışmalarını ayrıntılı hale getirmiştir. Sonuç olarak yarı testlere ilişkin varyansların farklılaştığı ve yarı testler arasındaki korelasyonun 0,30 ve 0,70 arasında ve 0,30'dan küçük olduğu durumlarda Spearman-Brown düzeltme formülü ile elde edilen eşdeğer yarılar güvenirliğinin olduğundan daha yüksek hesaplandığı görülmüştür. Dolayısıyla bu gibi durumlarda Flanagan formülü ile eşdeğer yarılar güvenirliğinin daha doğru kestirileceğinin altı çizilmiştir.

Bu araştırmanın amacı, farklı homojenlik düzeyindeki örneklemlerde, farklı uzunluktaki testlerin, yarıya bölme yöntemlerine ve eşdeğer yarılar güvenirliliği kestirme tekniklerine göre hesaplanacak olan eşdeğer yarılar güvenirliliklerini, bir simülasyon çalışması yardımıyla incelemektir. Bu çalışmada kullanılan simülasyon tekniği ile çeşitli özellikteki evrenlerden basit seçkisiz örnekleme ile çekilen örneklemlerin eşdeğer yarılar güvenirlilikleri, birçok değişken göz önünde bulundurularak incelenmiş; örneklemler ve evrenler arasındaki uyuma bakılmıştır. Dolayısıyla bu çalışmanın araştırmacılara yol göstereceği düşünülmektedir. Ayrıca yapılan alanyazın taramasında simülasyon tekniği ile yapılmış çalışmaların az olması sebebiyle, çalışmanın ölçme ve değerlendirme alanına önemli bir katkı sağlaması beklenmektedir.

Araştırmanın amacına bağlı olarak aşağıdaki sorulara yanıt aranmıştır:

- a. Farklı homojenlik düzeyindeki örneklerde (düşük, orta, yüksek), farklı uzunluktaki testlerin (10, 20, 40, 80) yarıya bölme yöntemleriyle (tekler-çiftler, madde gücüne göre tekler-çiftler, seçkisiz atama) bölünerek, güvenilirlik kestirme formülleriyle (Spearman-Brown düzeltme formülü, Rulon, Flanagan) elde edilecek eşdeğer yarılar güvenilirliklerinin ortancaları ve standart hataları nasıldır?
- b. Farklı homojenlik düzeyindeki örneklerde (düşük, orta, yüksek), farklı uzunluktaki testlerin (10, 20, 40, 80) yarıya bölme yöntemleriyle (tekler-çiftler, madde gücüne göre tekler-çiftler, seçkisiz atama) bölünerek, güvenilirlik kestirme formülleriyle (Spearman-Brown düzeltme formülü, Rulon, Flanagan) elde edilecek eşdeğer yarılar güvenilirliklerinin, evren güvenilirlikleri ile uyumu nasıldır?

Yöntem

Veri toplama

Bu araştırma, simülatif bir çalışmaya dayalı olduğundan, öncelikle veri üretim çalışmaları gerçekleştirilmiştir. Veri üretim çalışmasında, Wingen3 bilgisayar programından yararlanılmıştır. Wingen 3 ile veri üretiminin birinci aşamasında, düşük, orta ve yüksek homojenlik düzeyindeki bireylerin gerçek puanları (theta), ortalaması 0,00 ve standart sapmaları sırasıyla 3,00; 2,00; 1,00 ve normal dağılıma uygun olacak şekilde belirlenmiştir.

İkinci aşamada, Madde Tepki Kuramlarından 2 Parametrelili Lojistik Modele ilişkin madde parametrelerinden a (ayırt edicilik) ve b (madde gücü) parametreleri için uniform dağılım kabul edilmiş, a parametresi 0,00 ve 2,00 arasında, b parametresi ise -3,00 ve 3,00 arasında değişecek şekilde belirlenmiştir. Bunun yanı sıra bu aşamada, madde sayısı 10, 20, 40 ve 80 olacak şekilde değişimlenmiştir.

Üçüncü aşamada, ilk iki aşamada girilen değerler birleştirilerek iki kategorili veri setleri üretilmiştir. Üç aşamanın da tamamlanması elde edilen 12 evrenin yapısı Tablo 1'de görülmektedir.

Tablo 1. Araştırmadaki Evrenlerin Yapısı

Homojenlik Düzeyi	Evren	Birey Sayısı (N)	Ortalama (μ)	Standart Sapma (σ)	Madde Sayısı (k)	a (ayırt edicilik)	b (madde gücü)
Düşük	1	10.000	0	3	10	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	2	10.000	0	3	20	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	3	10.000	0	3	40	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	4	10.000	0	3	80	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
Orta	5	10.000	0	2	10	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	6	10.000	0	2	20	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	7	10.000	0	2	40	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	8	10.000	0	2	80	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
Yüksek	9	10.000	0	1	10	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	10	10.000	0	1	20	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	11	10.000	0	1	40	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$
	12	10.000	0	1	80	$0,00 \leq a \leq 2,00$	$-3,00 \leq b \leq 3,00$

Örneklem/çalışma grubu

Örneklem aşamasında, araştırmacı tarafından MATLAB 6.5 programlama dilinde geliştirilen program yardımıyla her bir evrenden basit seçkisiz örneklem yöntemi ile 50 adet 50 birimlik; 50 adet 300 birimlik ve 50 adet 500 birimlik örneklem çekilmiştir. Bu işlemlerin sonucunda $12 \times 50 \times 3 = 1800$ adet örneklem elde edilmiştir.

İşlem

MATLAB 6.5 programı, elde edilen örneklemeleri üç farklı yarıya bölme yöntemi (tekler-çiftler, madde güçlüğüne göre tekler-çiftler, seçkisiz atama) ile yarıya bölmekte ve her bir bölme yönteminde üç farklı güvenilirlik kestirme tekniğiyle (Spearman-Brown düzeltme formülü, Rulon, Flanagan) güvenilirlik hesaplamalarını yapmaktadır. Dolayısıyla bu program sayesinde 12 evren için $12 \times 3 \times 3 = 108$ adet güvenilirlik hesaplaması elde edilmiştir.

Bir sonraki aşamada, 12 farklı evrenden basit seçkisiz örnekleme ile çeşitli büyüklüklerde ($n=50$, $n=300$ ve $n=500$) 50'şer örneklem çekilmiştir. Elde edilen 18000 adet örnekleme ait veriler farklı yarıya bölme yöntemleri (tekler-çiftler, madde güçlüğüne göre tekler-çiftler, seçkisiz atama) kullanılarak yarıya bölünmüştür. Her yarıya bölme yöntemi için güvenilirlik kestirimleri, 3 farklı güvenilirlik kestirme tekniği (Spearman-Brown düzeltme formülü, Rulon, Flanagan) ile gerçekleştirilmiştir. Dolayısıyla $12 \times 50 \times 3 \times 3 \times 3 = 16200$ adet güvenilirlik kestirimi elde edilmiştir. Gulliksen (1967), testin iki yarı formuna ilişkin varyansların eşit olması sayılısı altında, Guttman ve Rulon formülleriyle elde edilecek eşdeğer yarılar güvenilirlik katsayılarının birbirine eşit olacağını ifade etmiştir. Dolayısıyla Guttman formülü bu çalışmaya dahil edilmemiştir.

Verilerin analizi

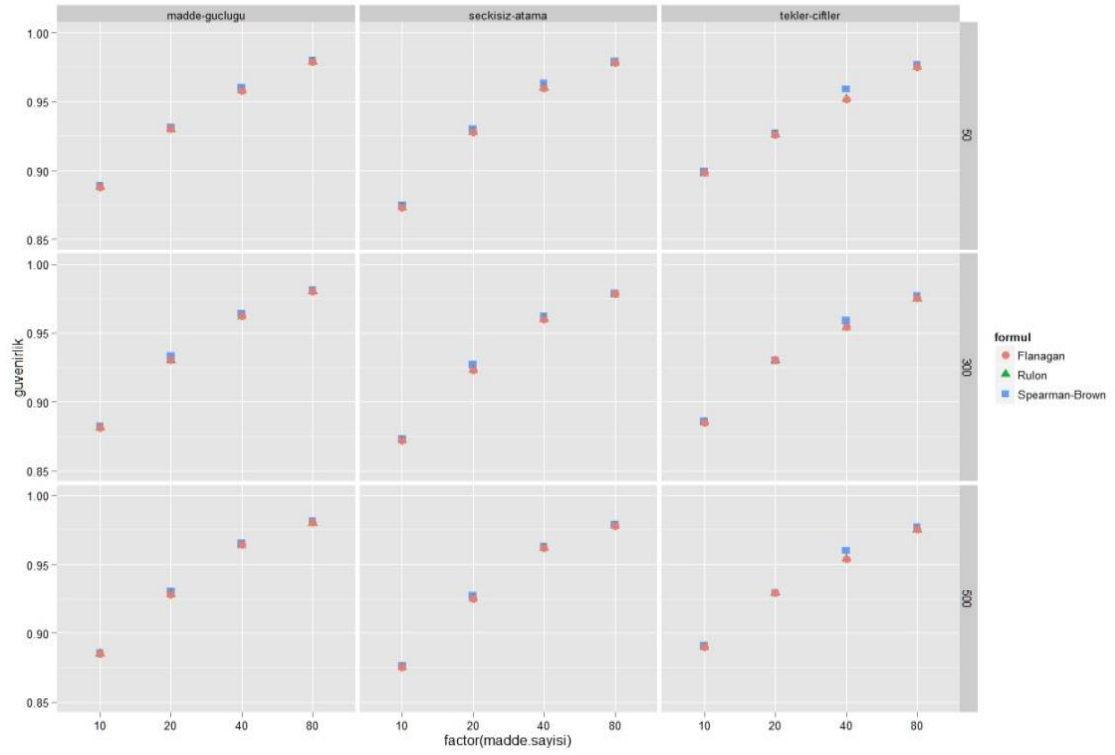
Farklı homojenlik düzeyindeki evrenlere ait güvenilirlik hesaplamaları ve örneklemelere ilişkin güvenilirlik kestirimlerinin ortancaları ve standart hataları; evrenler ve örneklemeler arası uyum madde sayısına, yarıya bölme yöntemlerine, örneklem büyüklükleri ve güvenilirlik kestirme tekniklerine göre tablolastırılmış ve R programlama dili ile hazırlanan çok değişkenli grafiksel gösterimlerle görsel hale getirilmiştir.

Bulgular

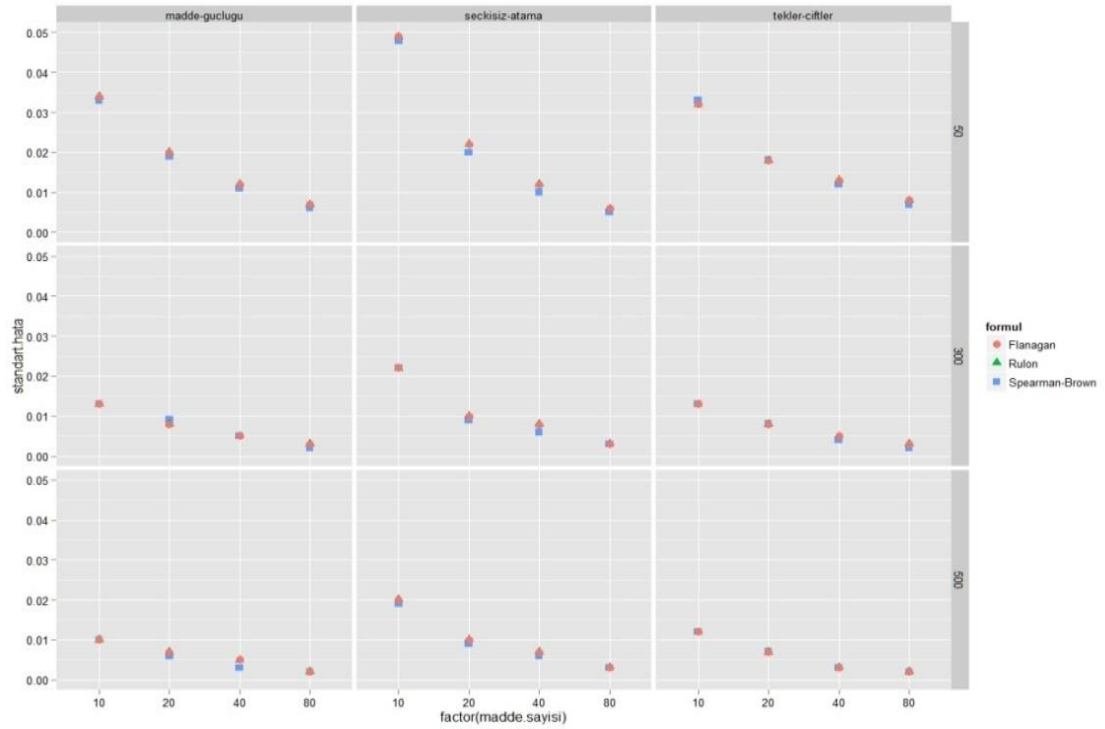
Birinci araştırma sorusuna ilişkin bulgular

Farklı homojenlik düzeyindeki (düşük, orta, yüksek) örneklemelerde, farklı uzunluktaki testlerin, yarıya bölme yöntemlerine ve güvenilirlik kestirme tekniklerine göre hesaplanan eşdeğer yarılar güvenilirliklerinin ortancalarına ve standart hatalarına ilişkin saçılım grafikleri Grafik 1, 2, 3, 4, 5 ve 6'da görülmektedir.

Grafiklerde görüldüğü üzere, örneklemeler homojenleştikçe kestirilen eşdeğer yarılar güvenilirlik ortancalarının düştüğü, standart hataların ise arttığı söylenebilir. En yüksek ortancalar düşük homojenlikteki örneklemelerde; en düşük ortancalar ise yüksek homojenlikteki örneklemelerde elde edilmiştir. Aynı zamanda güvenilirlik kestirimlerine ilişkin en yüksek standart hatalar yüksek homojenlikteki örneklemelerde, en düşük standart hatalar ise düşük homojenlik düzeyindeki örneklemelerde elde edilmiştir. Gulliksen (1967), hata puanlarının gruptan gruba değişmemesi sayılısı koşuluyla, bir testin heterojen bir gruba uygulanmasıyla elde edilecek güvenilirlik katsayılarının daha yüksek olacağını belirtmiştir.

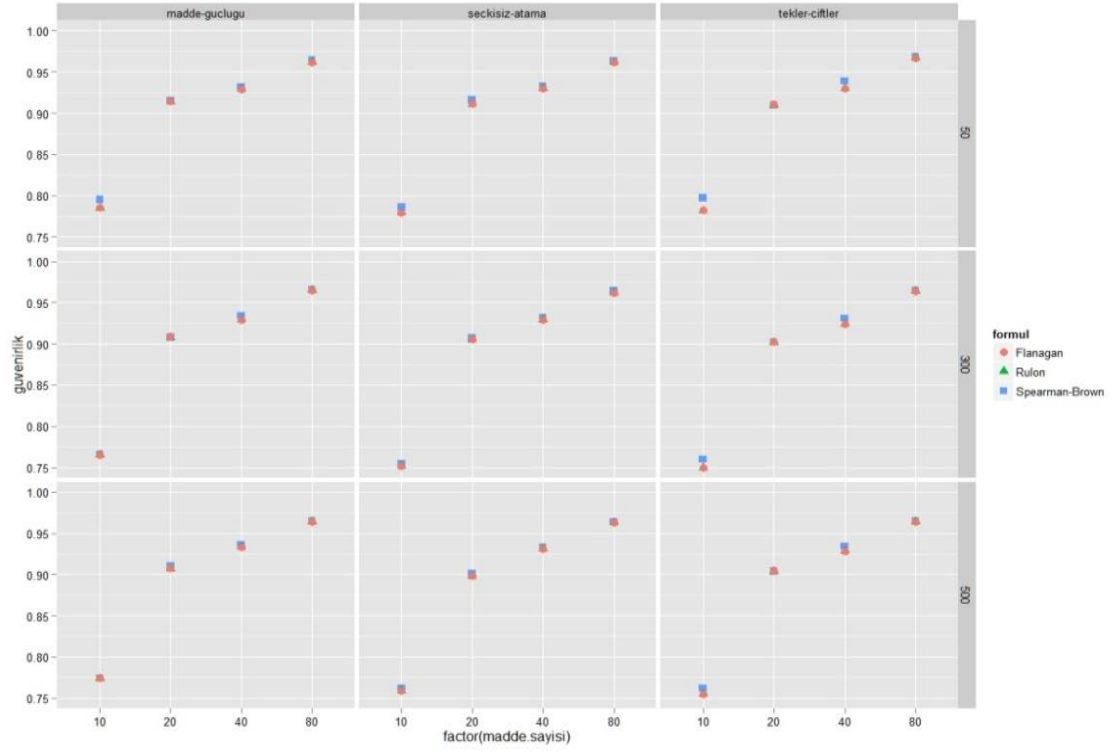


Grafik 1. Düşük Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlilik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliliklerinin Ortancalarına İlişkin Dağılım

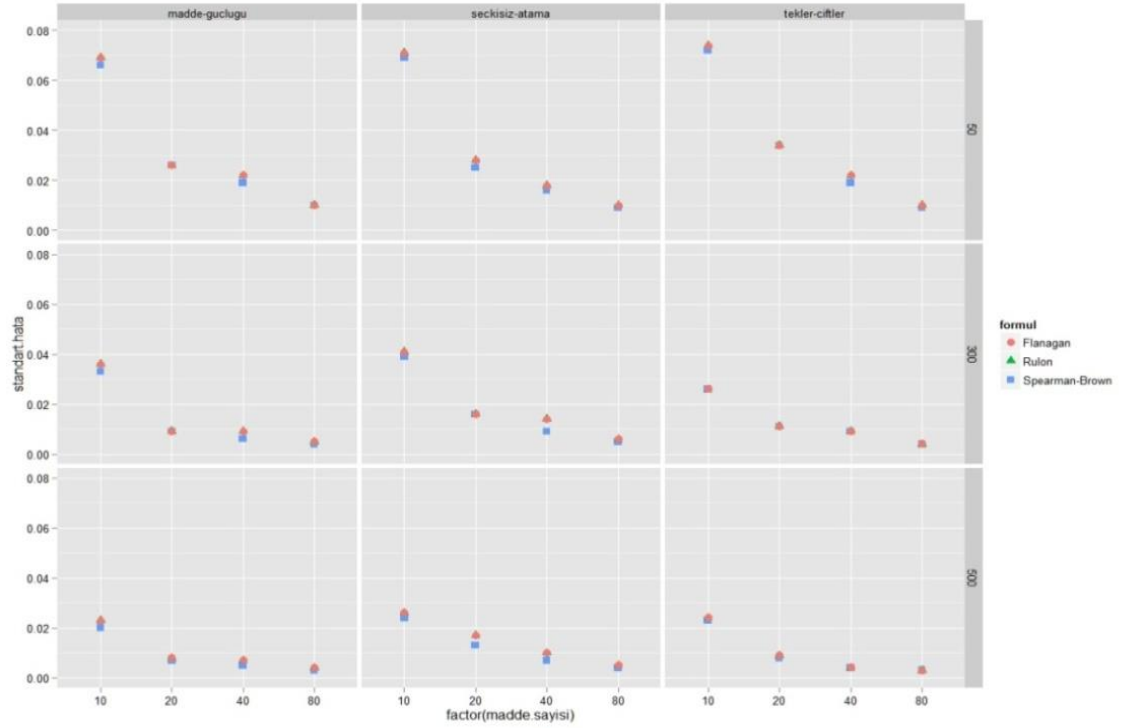


Grafik 2. Düşük Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlilik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliliklerinin Standart Hatalarına İlişkin Dağılım

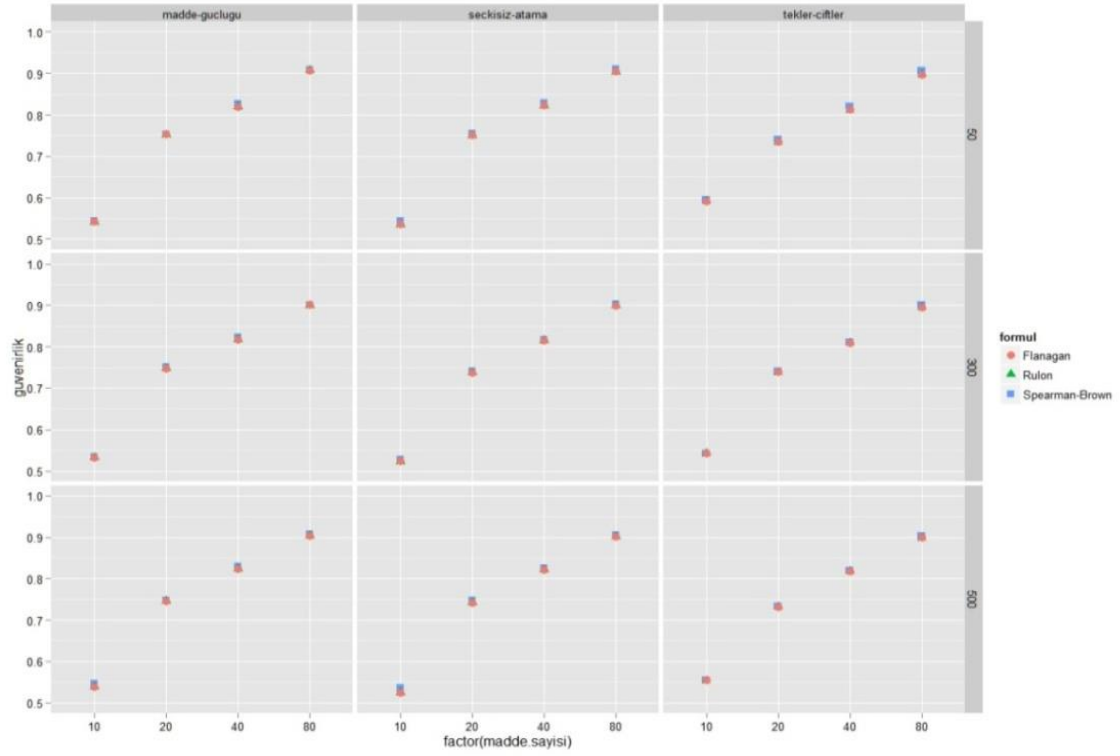
Eşdeğer Yarılar Güvenirliğinin Farklı Homojenlik Düzeyleri



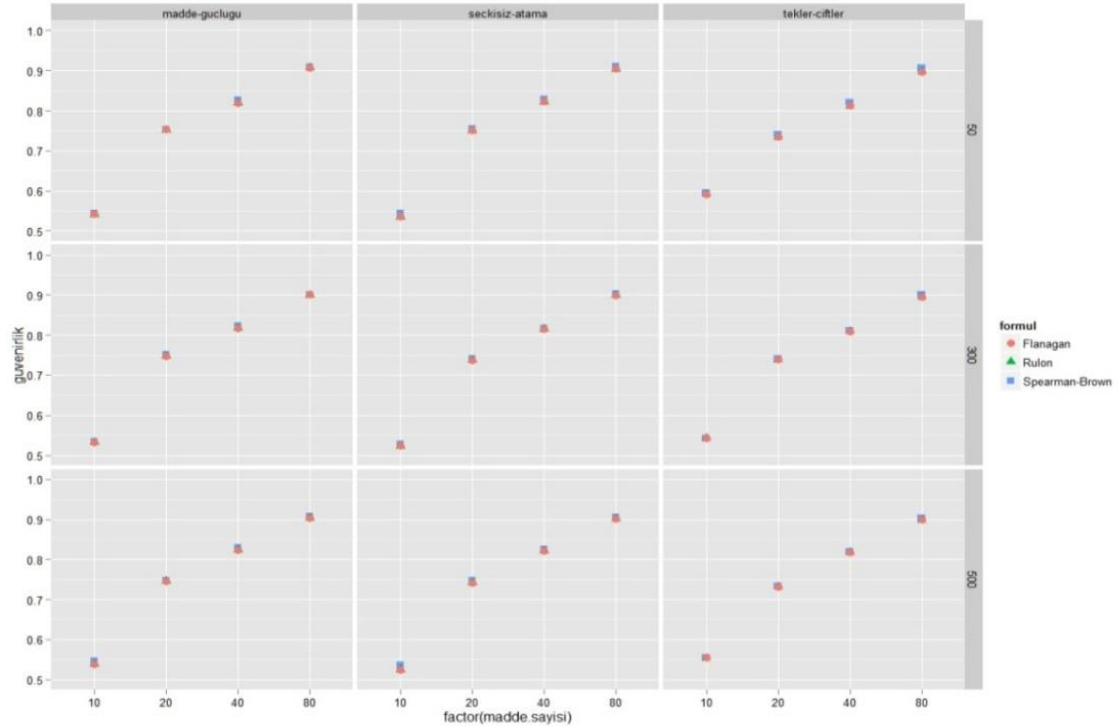
Grafik 3. Orta Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliklerinin Ortancalarına İlişkin Dağılım



Grafik 4. Orta Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliklerinin Standart Hatalarına İlişkin Dağılım



Grafik 5. Yüksek Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlilik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliliklerinin Ortancalarına İlişkin Dağılım



Grafik 6. Yüksek Homojenlik Düzeyindeki Örneklerde Farklı Uzunluktaki Testlerin, Yarıya Bölme Yöntemlerine ve Güvenirlilik Kestirme Tekniklerine Göre Hesaplanan Eşdeğer Yarılar Güvenirliliklerinin Standart Hatalarına İlişkin Dağılım

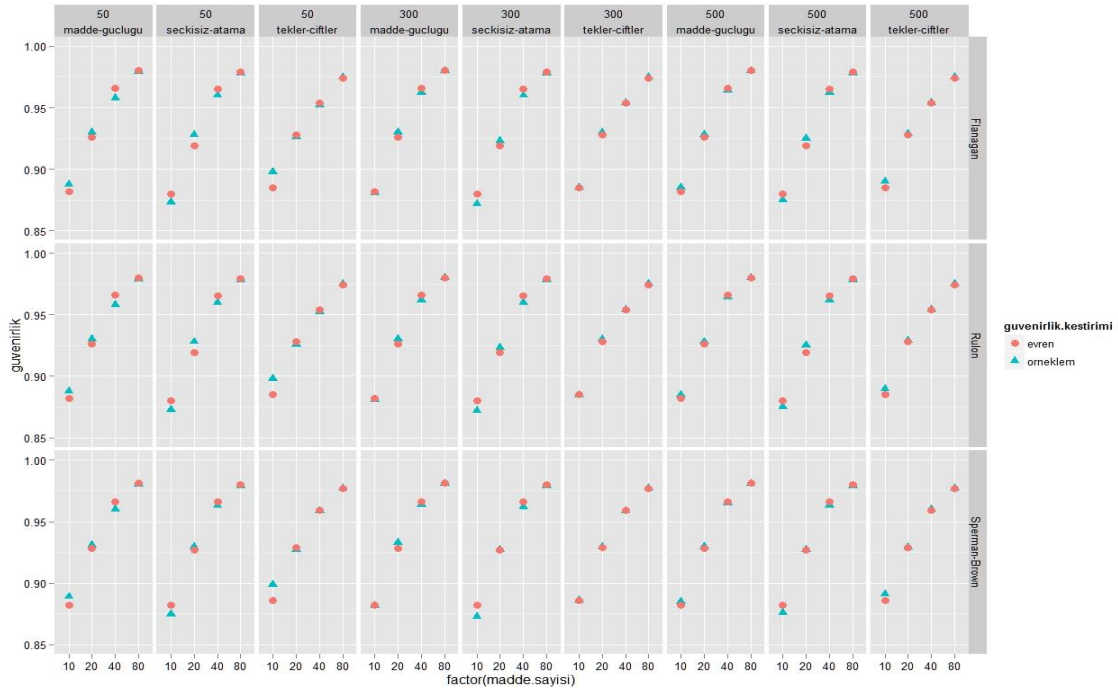
Elde edilen bulgular test uzunlukları açısından incelendiğinde, bütün homojenlik düzeylerinde, testteki madde sayısı arttığında eşdeğer yarılar güvenilirlik ortancalarının da arttığı, standart hataların ise azaldığı görülebilir. Özellikle yüksek homojenlik düzeyindeki örneklemlerde, testteki madde sayısı artışının diğer homojenlik düzeylerindeki örneklemlere göre çok daha büyük farklar oluşturduğu söylenebilir. Yüksek homojenlik düzeyindeki örneklemlerde, testteki madde sayısı 8 kat arttığında, eşdeğer yarılar güvenilirlik ortancaları 1,7 kat artarken, bu oran orta homojenlikteki örneklemlerde 1,3, düşük homojenlikteki örneklemlerde ise 1,1'dir.

Elde edilen bulgular yarıya bölme yöntemleri açısından incelendiğinde, tüm homojenlik düzeyindeki örneklemlerde, en yüksek eşdeğer yarılar güvenilirlik ortancaları ve en düşük standart hatalar tekler-çiftler ve madde güçlüğüne göre tekler-çiftler yöntemleriyle yarıya bölünen testlerden elde edildiği söylenebilir. Guilford (1954), testteki maddelerin kolaydan zora doğru sıralanarak tekler-çiftler yöntemiyle yarıya bölünmesiyle, iki yarı testin eşit güçlüklerde ve paralel olacağını belirtmiştir.

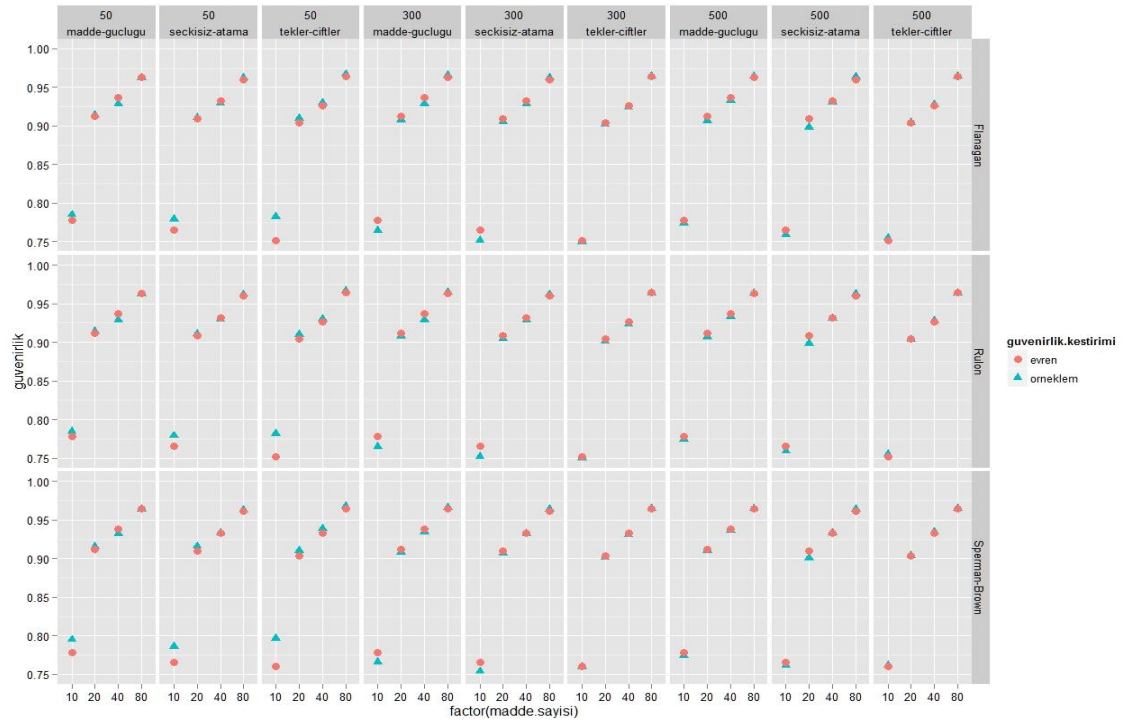
Elde edilen bulgular güvenilirlik kestirme teknikleri açısından incelendiğinde, tüm homojenlik düzeylerinde kullanılan güvenilirlik kestirme teknikleri ile elde edilen eşdeğer yarılar güvenilirlik ortancalarının ve standart hatalarının birbirine çok yakın olduğu söylenebilir. Özellikle Rulon ve Flanagan formülleriyle kestirilen eşdeğer yarılar güvenilirliklerinin ortancalarının, üç homojenlik düzeyinde de birbirinin aynı olduğu görülebilir. Gulliksen (1957), iki yarı testin varyanslarının eşit olması durumunda, kullanılan formüllerin benzer sonuçlar vereceğini ifade etmiştir.

İkinci araştırma sorusuna ilişkin bulgular

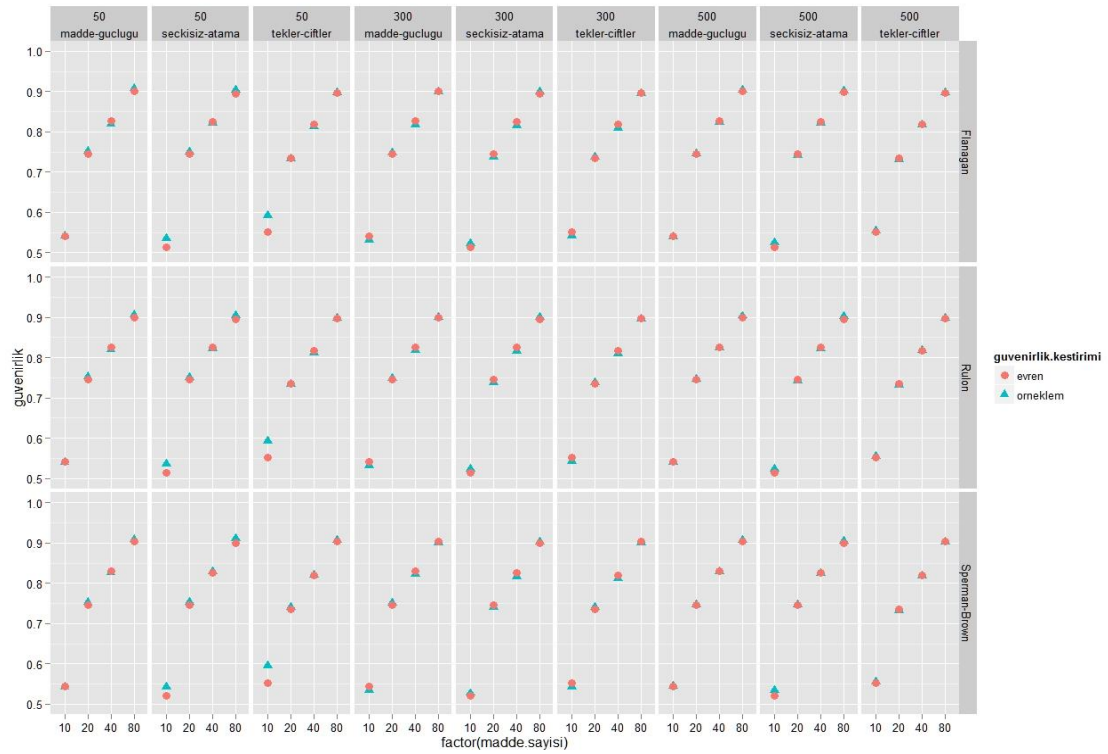
Farklı homojenlik düzeyindeki örneklemlerde (düşük, orta, yüksek), farklı uzunluktaki testlerin, yarıya bölme yöntemlerine ve güvenilirlik kestirme tekniklerine göre hesaplanan eşdeğer yarılar güvenilirliklerinin ortancaları ve evren güvenilirlikleri arasındaki uyuma ilişkin saçılım grafikleri Grafik 7, 8 ve 9'da görülmektedir.



Grafik 7. Düşük Homojenlik Düzeyindeki Örneklemlere İlişkin Eşdeğer Yarılar Güvenirlik Kestirimlerinin Ortancaları ve Evrenlere İlişkin Eşdeğer Yarılar Güvenirliklerinin Uyumu



Grafik 8. Orta Homojenlik Düzeyindeki Örneklemelere İlişkin Eşdeğer Yarılar Güvenirlik Kestirimlerinin Ortancaları ve Evrenlere İlişkin Eşdeğer Yarılar Güvenirliklerinin Uyumu



Grafik 9. Yüksek Homojenlik Düzeyindeki Örneklemelere İlişkin Eşdeğer Yarılar Güvenirlik Kestirimlerinin Ortancaları ve Evrenlere İlişkin Eşdeğer Yarılar Güvenirliklerinin Uyumu

Elde edilen bulgular evrenler ve örneklemeler arasındaki uyum açısından incelendiğinde, üç homojenlik düzeyinde de basit seçkisiz örnekleme ile elde edilen örneklemelerin çekildikleri evrenlerle uyum içerisinde olduğu söylenebilir.

Genel olarak bakıldığında evrenlerden elde edilen güvenirlilik katsayıları ve örneklemelerden elde edilen güvenirlilik ortancalarının virgülden sonra ikinci ya da üçüncü basamakta farklılaştığı görülebilir. Homojenlik düzeyleri açısından tek tek ele alındığında, düşük homojenlik düzeyindeki evrenler ve örneklemeler arasındaki uyumun en yüksek olduğu, homojenlik düzeyi arttıkça evren ve örneklemeler arasındaki uyumun da azaldığı söylenebilir. Yine aynı şekilde test uzunluğu ve örneklem büyüklükleri arttıkça tüm homojenlik düzeylerindeki evrenler ve örneklemeler arasındaki uyumun arttığı görülebilir. Yarıya bölme yöntemlerinden ağırlıklı olarak tekler-çiftler ve madde gücüne göre tekler-çiftler yöntemlerinin kullanıldığı durumlarda evren ve örneklemelerin tam bir uyum içerisinde olduğu durumlar gözlemlendiği, kullanılan güvenirlilik kestirme tekniklerinin ise evren ve örnekleme ilişkin güvenirlilik katsayıları arasındaki uyum üzerinde herhangi bir etkisi olmadığı söylenebilir.

Tartışma / Sonuç ve Öneriler

Bu çalışmada, farklı homojenlik düzeyindeki örneklemelerde, farklı uzunluktaki testlerin, yarıya bölme yöntemlerine ve eşdeğer yarılar güvenirliliği kestirme tekniklerine göre hesaplanacak olan eşdeğer yarılar güvenirliliklerinin, bir simülasyon çalışması yardımıyla incelenmesi amaçlanmıştır.

Yapılan çalışmalar sonucunda, çalışma grubu heterojenleştikçe eşdeğer yarılar güvenirliliklerinin arttığı, standart hataların ise azaldığı belirlenmiştir. Araştırmada yer alan üç farklı örneklem büyüklüğünden elde edilen eşdeğer yarılar güvenirlilikleri, tüm homojenlik düzeylerinde birbirine yakın çıkmıştır. Fakat elde edilen sonuçlar güvenirlilik kestirimlerinin standart hataları açısından ele alındığında, örneklem büyüklüğünün artmasıyla birlikte standart hataların da azaldığı ortaya çıkmıştır. Tüm homojenlik düzeyindeki gruplarda, test uzunluğu arttıkça eşdeğer yarılar güvenirlilik ortancaları artmış, güvenirlilik kestirimlerine ilişkin standart hataların azaldığı gözlenmiştir. Tüm homojenlik düzeyindeki gruplarda, genel olarak tekler-çiftler ve madde gücüne göre tekler-çiftler yöntemleriyle yarıya bölünen testlerden elde edilen eşdeğer yarılar güvenirlilik ortancaları, seçkisiz atama yöntemiyle yarıya bölünen testlerden elde edilen eşdeğer yarılar güvenirlilik ortancalarına göre daha yüksek elde edilirken, standart hatalarının ise daha düşük çıktığı görülmüştür. Tüm homojenlik düzeyindeki gruplarda, Spearman-Brown düzeltme formülü, Rulon ve Flanagan formüllerinin kullanılmasıyla elde edilen eşdeğer yarılar güvenirlilik ortancaları ve standart hataları arasındaki fark oldukça azdır. Özellikle Rulon ve Flanagan formülleri ile elde edilen eşdeğer yarılar güvenirlilik ortancalarının tüm koşullar göz önünde bulundurulduğunda birbirine eşit olduğu gözlenmektedir. Araştırmada yer alan grupların homojenlik düzeyleri arttıkça, evren ve örneklemeler arasındaki uyumun azaldığı ortaya çıkmıştır. Aynı zamanda örneklem büyüklüğü ve test uzunluğu arttıkça; tekler-çiftler ve madde gücüne göre tekler-çiftler yöntemlerinin kullanıldığı durumlarda evren ve örneklemeler arasındaki uyumun arttığı bulunmuştur.

Bu çalışmadan elde edilen bulgulara dayanarak, bir ölçme aracının güvenirliliğinin hesaplanmasında eşdeğer yarılar yönteminden faydalanılması durumunda; çalışma grubunun olabildiğince heterojen bir yapıya sahip olması; güvenirlilik kestirimlerinin olabildiğince hatasız olabilmesi için, en az 300 birimlik örneklem büyüklükleri ile çalışılması; özellikle çalışma grubunun homojen olduğu durumlarda, yüksek güvenirlilik kestirimleri için test uzunluğunun en az 40 olması; çalışma grubunun heterojen olduğu durumlarda, testin madde gücüne göre tekler-çiftler yöntemleriyle yarıya bölünmesi; dolayısıyla testte yer alan maddelerin güçlük indeksleri hesaplandıktan sonra yarıya bölme işlemine geçilmesi; teste ilişkin güvenirlilik kestirimleri yapılırken, Spearman-Brown formülünün yanı sıra Rulon veya Flanagan formüllerinden birinin de kullanılması; aradaki farkın yüksek olması durumunda, testin iki yarı formuna ilişkin varyansların incelenmesi önerilebilir.

Bu araştırmada yer alan veri üretimi, Tek Boyutlu Madde Tepki Kuramı modellerinden İki Parametrelili Lojistik Modelle sınırlıdır. Dolayısıyla çalışma, Tek Boyutlu Madde Tepki Kuramının diğer modelleriyle ya da çok boyutlu Madde Tepki Kuramı ile yeniden desenlenebilir. Araştırmanın veri üretimi aşamasında, bireylerin gerçek puanlarına ilişkin belirlenen ortalama ve standart sapma değerleri, test uzunlukları, örneklem büyüklükleri, yarıya bölme yöntemleri ve güvenilirlik kestirme teknikleri farklı şekillerde değişimlenebilir. Araştırma farklı güvenilirlik hesaplama yöntemleri ile yinelenir. Test süresi de simüle edilerek güvenilirlik kestirimleri hesaplanabilir.

Kaynaklar

- Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. New York: Macmillan Pub.Co.Inc.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Burnett, J. D. (1974). Parallel measurements and the Spearman-Brown formula. *Educational and Psychological Measurement, 34*, 785-788.
- Callender, J. C., & Osburn, H. G. (1977). A method for maximizing split-half reliability coefficients. *Educational and Psychological Measurement, 37*, 819.
- Charter, R. A. (2001). It's time to bury the Spearman-Brown "prophecy" formula for some common applications. *Educational and Psychological Measurement, 61* (4), 690-696.
- Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology, 130* (24), 117-129.
- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21* (4), 559-566.
- Cortina, M. J. (1993). What is coefficient Alpha? An Examination of theory and applications. *Journal of Applied Psychology, 78* (1), 98-104.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests, *Psychometrika, 16* (3), 297-334.
- Cureton, Edward E. (1958). The definition and estimation of test reliability. *Educational and Psychological Measurement, 18*, 715-738.
- Fiske, D. W. (1973). Homogeneity of sample and of test responses. *Educational and Psychological Measurement, 33*, 239-253.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill Book Co.
- Gulliksen, H. (1967). *Theory of mental tests*. John Wiley & Sons. Inc.
- Lord, M. F., & Novick, M. R. (1968). *Statistical theories of mental test scores*. New Jersey: Addison-Wesley Pub. Co. Inc.
- Magnusson, D. (1968). *Test theory*. Massachusetts. Addison-Wesley.
- Miller, M. B. (1995). Coefficient Alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling, 2* (3), 255-273.
- Walker, D. A. (2006). A comparison of the Spearman-Brown and Flanagan-Rulon formulas for split half reliability under variance parameter conditions, *Journal of Modern Applied Statistical Methods, 5* (2), 443-451.

Extended Abstract

Introduction

In education and psychology, measurement and evaluation has a great role for making decisions about individual's psychological constructs. With measurement process, decisions made about the psychological constructs which can't observed directly. In order to evaluate

these constructs properly, the measurement instruments must be reliable. Whenever a test administered, the test user would like some assurance that the results could be replicated if the same individuals were tested again under similar circumstances. The desired consistency of test scores is called reliability. There are some factors that affects the test reliability. These are test length, time limit and group homogeneity. If the items in a test increased n times, true score variances increase n^2 times and error scores variances increases n times. Because of this situation, when the test length increased, the reliability of the test increased. Time limits should be long enough to allow all or nearly all of the examinees to finish the test, otherwise the estimation of test reliability may be artificially increased. The reliability of a test will be higher when there is a larger genuine variation in the population of examinees.

There are some procedures to estimate the reliability of a test. These are procedures requiring two test administrations (Test-Retest Method, Alternate Form Method) and methods requiring a single test administration (Cronbach Alpha, Kuder-Richardson, Split-Half Methods). In split-half reliability which is the main subject of this thesis, after administrating a test to a group of examinees, the test developer divides the items into two subtests, each half the length of the original test. The most popular methods for dividing a test into halves are, odds-evens, odds-evens according to item difficulties, random assignment and content matching methods. After dividing a test with one of these methods, the split-half reliability of the test can be estimated by one of the techniques Spearman-Brown prophecy formula, Rulon, Guttman or Flanagan formulas.

The purpose of this study is to examine split-half reliabilities of tests in different lengths for low, medium and high homogeneity level groups based on different split-half methods and reliability estimation techniques.

Method

For the purpose of this study, unidimensional data simulation was realized depending on 2 Parametric Logistic Model of Item Response Theory. The data simulation of the study was generated on Wingen 3 software. Wingen was developed to generate dichotomous and polytomous item response data for several IRT models and for many conditions that arise in practice. At the first step of the data generation with Wingen, to simulate 0-1 dichotomous data for 10000 examinees, the mean of ability parameters were set to 0.00; standard deviations were set to 1.00, 2.00, 3.00. At the second step, uniform distribution was chosen for item parameters a and b . The parameter a was set between 0.00 and 2.00 and b was set between -3.00 and 3.00. And at this step, the test lengths were set to 10, 20, 40 and 80. As a result of these steps, 12 different populations were achieved as a result of the data simulation. In order to determine how split-half reliabilities is affected by sample size, 50 samples for 50 sample size, 50 samples for 300 sample size and 50 samples for 500 sample size were drawn from populations by using simple random sampling. Thus 1800 samples were displayed. The tests about populations and samples were divided into halves with odds-evens, odds-evens according to item difficulties and random assignment methods; the split-half reliabilities were computed by Spearman-Brown prophecy formula, Rulon formula and Flanagan formula. This process was realized by programming codes written in MATLAB 6.5 programming language.

Besides, medians and standard errors of sample reliabilites were computed. Also the consistency between the population reliabilities and the sample reliabilities were investigated. The medians and the standard errors of the estimated sample reliabilities and the consistency between the reliabilities of the populations and samples were visualized by using multivariate graphs. This process was realized by the codes written in R programming language.

Result and Discussion

On examining the results, it was found that the more heterogeneous is the sample, the higher is the split-half reliabilities and the standard errors of reliabilities were decreased. Although the split-half reliabilities in three different sample sizes for all homogeneity levels were similar, when the sample sizes were high, the standard errors for split-half reliabilities were low. In all

homogeneity levels, as the test length increased split-half reliabilities were increased; standard errors of the reliabilities were decreased. When the tests split into halves with odds-evens method and odds-evens according to item difficulties method, the split-half reliabilities were higher; the standard errors were lower. The difference between the standard errors and the medians of split-half reliabilites attained by Spearman-Brown prophecy formula, Rulon formula and Flanagan formula in all homogeneity groups was scarcely any. It was found that the consistency between populations and samples was higher in the cases where the groups were heterogenous; the test length and the sample sizes were high and odds-evens method and odds-evens according to item difficulties method were used.