

Klasik Test Kuramına ve Faktör Analitik Yaklaşımına Göre Elde Edilen Farklı Güvenirlik Katsayılarının Karşılaştırılması*

Comparison of the Reliability Co-Efficients Obtained According to Classical Test Theory and Factor Analytic Approach

Ece BÜLBÜL **, Önder SÜNBÜL ***

Öz: Bu çalışmada paralel ve eş biçimli ölçme yapılarında güvenirliliği etkileyen faktörlerden, test uzunluğu, örneklem büyüklüğü ve faktör yükü dağılımları etkisi altında KR-20 ile faktör analitik yaklaşımda kullanılan Omega katsayısının incelenmesi amaçlanmıştır. Çeşitli faktörlere (örneklem büyüklüğü, test uzunluğu ve faktör yük düzeyleri) göre veri üretilip, üretilen verilerde her bir durum için 100 yinleme (replikasyon) yapılmıştır. Veriler, paralel ve Eş Biçimli ölçme yapılarında ayrı ayrı olarak, belirtilen koşulların çaprazlanması sonucu oluşan her bir duruma uygun olarak üretilmiştir. Veri üretimi R 3.0 programında Psych paketi ve araştırmacı tarafından yazılan kodlar kullanılarak gerçekleştirilmiştir. Paralel ve Eş Biçimli ölçme yapıları için farklı değişimleme ölçütlerine göre KR-20 ve Omega katsayılarının ortalama ve standart sapma değerleri hesaplanmıştır. Farklı ölçme yapılarında tek başına örneklem büyüklüğünün, katsayı ortalamaları ve standart sapmaları üzerinde anlamlı bir etkisi görülmemiştir. Ancak madde sayısı ve faktör yükü düzeylerinin katsayı ortalamaları ve standart sapmaları üzerinde anlamlı değişikliklere sebep olduğu gözlenmiştir. Paralel ölçme yapısında farklı madde sayılarında, örneklem büyüklüklerinde ve farklı faktör yüklerinde elde edilen Omega ve KR-20 katsayı ortalamaları, Eş Biçimli ve konjenerik ölçme yapılarında elde edilen katsayı ortalamalarına göre, beklenenin aksine, daha düşük çıkmıştır.

Anahtar Kelimeler: Paralel Ölçme Yapısı, Eş Biçimli Ölçme Yapısı, Omega Katsayısı, Güvenirlik

Abstract: The purpose of this study is the examination of the omega reliability coefficients used in KR-20 and factor analytic approach under the factors affecting reliability such as test length, sample size and factor loadings variance in parallel and tau-equivalent measurement structures. Data has been generated according to various factors (sample size, test length and factor loadings levels). 100 replications have been performed for each state in the generated data. The data have been generated separately in parallel and tau-equivalent measurement structures in compliance with each state consisted as a result of crossing the designated conditions. The data production has been carried out by using the codes written by the researcher and the psych package in R 3.0 program. The average and standart deviation values of KR-20 and omega coefficients have been calculated according to different alternation criterias for parallel and tau-equivalent measurement structures. No significant effect of sample size by itself on coefficient average and standart deviation in parallel and tau-equivalent measurement structures has been seen. However, it has been observed that the material number and factor loading levels have caused significant changes on the coefficient average and standart deviation. The omega and KR-20 coefficient averages obtained in different material numbers, different sample sizes and different factor loadings in parallel measurement structure has been, contrary to expectations, lower compared to the ones obtained in tau-equivalent and congeneric measurement structures.

Keywords: Parallel Measurement Structure, Tau-Equivalent Measurement Structure, Omega Coefficient, Reliability.

* Bu çalışma, birinci yazarın Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü'nde Temmuz 2016 tarihinde tamamlanan yüksek lisans tezinden üretilmiştir.

** Sorumlu yazar, Öğretim Görevlisi Ece BÜLBÜL, Toros Üniversitesi, İ.İ.S.B.F, Mersin-Türkiye, ORCID:0009-0000-0227-904X, e-posta: ece.bulbul@toros.edu.tr

*** Doç.Dr.Önder SÜNBÜL, Mersin Üniversitesi, Eğitim Bilimleri Fakültesi Mersin-Türkiye, ORCID: 0000-0002-1775-1404, e-posta: ondersunbul@gmail.com

Giriş

Ölçme ile hatalardan arınık, nesnel, standart ve geçerli ölçümler yapmak esas alınır. Bu sebepten nitelikli ölçme araçları geliştirilmesi büyük önem taşır. Bir ölçme aracının beklenen ve amaçlanan düzeyde nitelikli olabilmesi için güvenirlilik ve geçerliğinin sağlanmış olması gereklidir. Ölçme aracının geçerliği, test puanlarından ya da diğer değerlendirme biçimlerinden yapılan vardamaların uygunluğu olarak tanımlanır (Crocker ve Algina, 1986). Esas olarak bir ölçme aracının geçerliğini geliştirdiği amaca hizmet etme derecesi gösterir. Geçerlik kavramı güvenirlilikle birlikte ele alınmalıdır. Geçerlik ölçme araçları için ulaşılmak istenen nihai amaç olsa da geçerliğin sağlanabilmesi için güvenirlilik ön koşulunun sağlanmış olması gerekmektedir. Bu noktada geçerlik ve güvenirlilik arasındaki ilişkiler ele alınmalıdır. Bir ölçme aracının geçerli olabilmesi güvenirliliğine bağlı iken, güvenirliliğinin yüksek olması geçerli olduğu anlamına gelmeyebilir (Crocker ve Algina, 1986). Güvenirlilik birçok anlamda kullanılır ancak en temel anlamda güvenirlilik, ölçme aracının minimum hata ile ölçme yapabilme yeterliğidir (Turgut, 1972). Güvenirlilik kavramı, puanlayıcı güvenirliliği, kararlık anlamında güvenirlilik, iç tutarlık anlamında güvenirlilik olarak ayrı ayrı ele alınabilir. Bu çalışmanın kapsamı nedeniyle iç tutarlık anlamında güvenirlilik kavramı ele alınacaktır.

Klasik Test Kuramı ve Güvenirlilik Kavramı:

Güvenirlilik tanımı, Klasik Test Kuramı'nın etkisiyle genellikle, paralel testler arası korelasyon olarak tanımlanır (Gulliksen, 1967; Crocker ve Algina, 1986; Lord ve Novick, 1968). Magnusson (1967:61) a göre güvenirlilik, bir bireyin aldığı puanların kendi kendisiyle tutarlılığı ve belirli bir durumda ölçme aracından elde edilen sonucun yeniden üretilebilirliğidir.

Ölçmede, gözlenen bir özelliğe ait olan gerçek değer bulunması amaçlanır. Ancak ölçmeye karışan çeşitli hatalar gerçek değer gözlenen değerden uzaklaşmasına neden olur. Gerçek değer, gözlenen puanlar ile kestirilmeye çalışılır (Baykul, 2000, s: 97). Klasik Test Kuramı da bu kestirimi yapabilmek için kurulan teorilerden biridir. Klasik Test Kuramına, kuram gerçek puanın kestirilmesi üzerine kurulduğu için gerçek puan Kuramı (True Score Theory) de denmektedir.

Ölçme sonucunda elde edilen sayıya gözlenen puan (X); ölçülmek istenilen değişkene ait gerçek değere ise gerçek puan (T) denir. Gerçek puan testin ölçtüğü özellik bireylerde pozitif veya negatif yönde değişmedikçe, ölçümden ölçüme değişmeyen değerdir (Gulliksen, 1950) ve ölçmeye karışan çeşitli hatalar nedeniyle doğrudan ölçülemez. Bu nedenle gerçek puan teoriktir ve gerçek puan ile gözlenen puan aynı olamayacağı için gözlenen puan, içinde hata barındırır.

Bireyin gerçek puanının hipotetik olmasından dolayı, gerçek puan gözlenirken ölçme araçlarından faydalanılır. Ölçme sonuçlarına çeşitli hatalar karışır. Bu hatalardan özellikle seçkisiz hatalar ölçme sonuçlarının güvenirliliğini doğrudan etkiler (Baykul, 2000). Eşitlik 1'de de görüleceği üzere bireyin gözlenen puanı (X), gerçek puanı (T) ile hata puanı (e) toplamıdır.

$$X=T+e \quad (1)$$

Klasik test kuramının temellerinin dayandığı bazı sayıtlılar vardır. Bunlar Gulliksen'e (1967) göre;

- Hata puanlarının ortalaması sıfırdır, $\mu_e = 0$
- Hata puanı ile gerçek puan arasındaki korelasyon sıfırdır, $r_{eT} = 0$
- Farklı testlerin hataları arasındaki korelasyon sıfırdır.

X, T ve E'nin ortalama, standart sapma ve korelasyonlarına ilişkin dağılımları hakkında sayıtlıların üretiminde bazı sorunlarla karşılaşmaktadır:

- Hata tanımının merkeze alınması, ölçmede önemli olan gerçek puanın göz ardı edilmesine neden olmaktadır. Ancak basit eşitlikler hataya ilişkin sayıtlılardan türetilmiştir.

- Gerçek puanın tanımının merkeze alınması ile gerçek puan önem verilmiş, hata puanı arka planda kalmış olacaktır. Ancak bu da zor eşitlikleri doğuracaktır. Bu nedenle birinci yaklaşım tercih edilmektedir (Gulliksen, 1950, s.5).

Bu sayıtların içerisinde hata terimlerinin ilişkisiz olduğuna ilişkin sayıtların bozulması modelin geçerliğini azaltan bir etkiye sahiptir (Novick ve Lewis, 1967). Bu sayıtlı, teste yer alan maddelerin paralel, eşdeğer ya da Eş Biçimli yapıya sahip olduklarında sağlanır.

Birleşik bir testte yer alan maddeler ölçülmek istenilen yapıyı eşit büyüklükte ve eşit duyarlılıkta ölçüyor ise paralel maddeler olarak tanımlanır. Maddeler ölçülmek istenilen yapıyı eşit büyüklükte ancak eşit olmayan duyarlılıkta ölçüyor ise bu tür maddeler için eşdeğer maddeler tanımlaması yapılır (Raykov, 1997).

Klasik Test Kuramında Güvenirlik:

KTK'nda gerçek puanlar varyansının maddeler arası kovaryanslar ile açıklanması, güvenilirlik kavramına bir zemin oluşturmaktadır. Güvenirlik kavramı gerçek puanlar varyansının gözlenen puanlar varyansına oranı olarak ifade edilmektedir.

$$X_i = T_i + e_i \quad \text{ve} \quad X_j = T_j + e_j \quad \text{olmak} \quad (2)$$

üzere

$$r_{X_i X_j} = \frac{\sum(T_i + e_i) \cdot (T_j + e_j)}{N \cdot s_{X_i} s_{X_j}} \quad (3)$$

$$s_{X_i} = s_{X_j} = s_X \quad (4)$$

$$r_{X_i X_j} = \frac{s_T^2}{s_X^2} \quad (5)$$

$$s_X^2 = s_T^2 + s_e^2 \quad \text{olduğundan dolayı} \quad (6)$$
$$r_{X_i X_j} = 1 - \frac{s_e^2}{s_X^2}$$

Klasik Test Kuramında Ölçmelerin Yapısı:

Gerçek puan ve maddeler arasındaki ilişkilere dayanan eşitlik modellemesi perspektifinden bakıldığında ölçme yapılarını aşağıdaki gibi tanımlamak mümkündür:

Eşitlik 1 ile ifade edilen Klasik Test Kuramının temel denklemi ve bu kurama ait bütün eşitlikler ölçmelerin paralel olduğu sayıtlı üzerine kurulur. Paralel ölçmelerde i ve j gibi iki ölçme maddesinin gözlenen puanları ile gerçek puanları arasındaki ilişki Eşitlik 7 ve Eşitlik 8'deki gibi gösterilir.

$$X_i = T_i + E_i \quad (7)$$

$$X_j = T_j + E_j \quad (8)$$

Eşitlik 10 ve Eşitlik 11 ile verilen durumlarda $T_i = T_j$ ve $E_i = E_j$ ise bu iki ölçme için *paralel ölçme* ifadesi kullanılır. Buna göre her iki madde; aynı özelliği eşit büyüklükte ve hata varyansları eşit şekilde ölçmektedirler.

Eğer ölçülmek istenilen gerçek puanlar arasındaki sabit bir bağıntı var ise;

$$T_i - T_j = a_{ij} \quad (9)$$

bu durumda her iki ölçme yapısı için *Eş Biçimli ölçme* (essentially tau-equivalent) adlandırılması yapılmaktadır (Traub, 1994).

Güvenirlik Belirlemede Madde Kovaryanslarına Dayanan Yöntemler

Kuder-Richardson Yöntemi

Test birçok paralel testin bileşkesi gibi görülür. Her madde diğer maddeyle paralelmiş gibi davranılır. Bu durumda n maddeli bileşik testin güvenilirliği Spearman-Brown formülüyle kestirilebilir. 1-0 ile puanlanan testlerde testin bütününe ait güvenilirlik KR-20 ve KR-21 formülleri ile hesaplanabilir.

X bir binom değişkeni

S_X^2 =bu değişkenin varyansı

p=maddelerin güçlük indeksleri olmak üzere;

n maddeli bileşik bir testin varyansı;

$$S_X^2 = \sum S_i^2 + 2 \sum r_{ik} S_i S_k \quad (i \neq k) \quad (10)$$

i ve k paralel maddeler olduğu için;

$$S_X^2 = \sum S_i^2 + n(n-1)\bar{r}_{ik}\bar{S}_i^2 \quad (11)$$

i ve k maddeleri arasındaki ortalama korelasyon Eşitlik 12 ile ifade edilir.

$$\bar{r}_{ik} = \frac{S_X^2 - \sum S_i^2}{(n-1) \sum S_i^2} \quad (12)$$

N maddelik bir testin güvenilirliği Spearman-Brown Formülü ile kestirildiğinde;

$$\begin{aligned} r_{XX} &= \frac{n\bar{r}_{ik}}{1 + (n-1)\bar{r}_{ik}} \\ &= n \cdot \frac{S_X^2 - \sum S_i^2}{(n-1) \sum S_i^2} \cdot \frac{1}{1 + (n-1) \left[\frac{S_X^2 - \sum S_i^2}{(n-1) \sum S_i^2} \right]} \\ &= \frac{n}{n-1} \left[\left(\frac{S_X^2 - \sum S_i^2}{\sum S_i^2} \right) \left(\frac{1}{1 + \left(\frac{S_X^2 - \sum S_i^2}{\sum S_i^2} \right) - 1} \right) \right] \\ &= \frac{n}{n-1} \left[\left(\frac{S_X^2 - \sum S_i^2}{\sum S_i^2} \right) \left(\frac{\sum S_i^2}{S_X^2} \right) \right] = \frac{n}{n-1} \left[\frac{S_X^2 - \sum S_i^2}{\sum S_X^2} \right] \end{aligned} \quad (13)$$

1- 0 şeklinde puanlanan maddeler için birleşik testin güvenilirliği yukarıdaki ispatlardan:

$$KR - 20 = r_{XX} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum pq}{S_X^2} \right) \quad (14)$$

$$KR - 21 = r_{XX} = \frac{n}{n-1} \cdot \left(n - \frac{\bar{s} \cdot (n-\bar{s})}{s_X^2} \right) \quad (15)$$

Guttman-Cronbach Alfa Katsayısı:

Klasik test kuramında paralel ölçmeler üzerine çalışılır ve paralel ölçme araçlarından elde edilen güvenilirlik katsayıları gerçek güvenilirliği verir. Ancak paralel olmayan ölçme yapısındaki ölçme araçlarından farklı güvenilirlik katsayıları elde edilebilir.

KTK'da maddelerin ölçme yapılarına göre iç tutarlık anlamındaki güvenilirlik katsayılarının elde edilmesi de farklılık göstermektedir. Bu durumda eğer birleşik testi oluşturan maddeler paralel ise maddeler arası korelasyonların güvenilirliğe karşılık geldiği bilinmektedir (Lord ve Novick; 1968).

Birleşik testi oluşturan maddeler eşdeğer ya da Eş Biçimli yapıda ise, maddelerin hata puanları varyansı eşit olmayacağından dolayı güvenilirlik katsayılarını elde edebilmek için korelasyon terimleri yerine maddeler arası kovaryans terimleri kullanılır.

α katsayısı, bileşenlere ait puanların, bileşik test puanlarıyla tutarlılığının bir ölçüsüdür. Çoklu puanlanan maddeler için kullanılır. Guttman-Cronbach alfa katsayısı eşitlik 16 ile verilir.

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum s_i^2}{s_X^2} \right) \quad (16)$$

Cronbach'ın α katsayısı aynı zamanda doğrulayıcı faktör analizi terimleri ile de ifade edilebilir (McDonald, 1985).

$$\alpha = \left(\frac{k}{k-1} \right) \left[\frac{k(\bar{\lambda})^2 - ((\bar{\lambda}^2))}{k(\bar{\lambda})^2 + \bar{\psi}^2} \right] \quad (17)$$

Cronbach'ın α katsayısı maddelere ilişkin kovaryansların eşit olduğu durumlarda güvenilirlik indeksini vermektedir. Ancak α katsayısı konjenerik ölçmelerde gerçek güvenilirliğin altında bir değer üretmektedir (Raykov, 1997).

Güvenirliği Etkileyen Faktörler

Bir ölçme aracının güvenilirliğinin sağlanması, ölçme hatalarının nereden kaynaklandığının bilinmesiyle yakından ilgilidir. Hata kaynaklarına bağlı olarak güvenilirliği saptama yöntemleri de değişecektir.

Crocker ve Algina (1986) güvenilirliği etkileyen faktörlerden; grubun homojenliği, zaman sınırı ve test uzunluğunu ele almakta; Gulliksen (1967) ise, test uzunluğu ve grup heterojenliğinin yanında Crocker ve Algina'nın zaman sınırına erişilemeyen maddelerin etkisi olarak değinmektedir. Lord ve Novick (1968) ise, güvenilirliği etkileyen ölçme koşulları altında; yorgunluk, bellek etkisi gibi kişinin yetenek düzeyindeki değişiklikler, ölçülen kişi ya da gözlem sayısı faktörlerini ele almaktadır. Daha geniş bir sınıflama içinde; 1) testi alan bireylerin özellikleri, 2) testin kendisinin özellikleri, 3) test puanlarının kullanılma amacı ve 4) güvenilirlik kestiriminde kullanılacak yöntem biçiminde konuyu ele almaktadır (Erkuş,2003).

İlgili Çalışmalar

Raykov (1997) çalışmasında, bileşik güvenirligi kestirirken konjenerik ölçümlerden faydalanmıştır. Raykov bu çalışmada konjenerik ölçme modellerinde faktör analitik yaklaşımla elde edilen katsayıların, alfa katsayısına göre daha gerçekçi sonuçlar verdiği belirtmiştir. Alfa katsayısının konjenerik ölçümlerde zayıf kalmasının sebebinin ise, alfa katsayısının paralel ölçümlerin varsayımları altında çıkarıldığı sonucuna bağlamıştır.

Graham (2006) çalışmasında güvenirligi için konjenerik ve Eş Biçimli ölçme yapılarından kestirmiştir. Bu çalışmada oluşturulan konjenerik ve Eş Biçimli modellerde faktör analitik yaklaşımlarda kullanılan güvenirlilik katsayıları ve alfa katsayıları hesaplanmıştır. Alfa katsayısının konjenerik ve Eş Biçimli ölçme yapılarında zayıf kaldığı görülmüştür.

Yurdugül(2006) çalışmasında ölçme yapıları ve güvenirlilik için temel kavramları ele almış ve değişik ölçme kümelerinde Cronbach'ın α , Armor'un θ , Heise ve Bohrnstedt'in Ω , Revelle'nin β ve McDonald'ın ω güvenirlilik katsayılarını karşılaştırmıştır. Testteki maddeler paralel ölçmeler olduğu sürece güvenirlilik katsayıları aynı değerler üretirken ($\alpha=\theta=\Omega=\omega=\beta$) konjenerik ölçmelerde ($\alpha=\theta=\Omega=\beta<\omega$) eşitsizliği ortaya çıkmaktadır. Buna göre, çoklu derecelenmiş testlerde maddelere ilişkin kovaryans terimleri ya da faktör yükleri eşit olduğunda α , β , Ω ya da ω aksi takdirde ise yalnızca ω katsayısının kullanımını önermiştir.

Sijtsma (2009) çalışmasında boyutluluk arttıkça α 'nın değişmediğine ancak diğer gibi katsayılarının arttığına dikkat çekmiştir. Bu durumda güvenirlilik için α katsayısının her zaman uygun bir tercih olmadığı sonucuna ulaşılabilir

Araştırmanın Amacı ve Önemi

Bu araştırmada, klasik test kuramına ve faktör analitik yaklaşımlara göre çeşitli etmenler altında güvenirlilik katsayılarının karşılaştırılması amaçlanmıştır. Bu amaçla, paralel, eş biçimli ve konjenerik ölçme yapılarında, test uzunluğu, örneklem büyüklüğü ve faktör yükü düzeylerine göre alfa, KR-20 ve faktör analitik yaklaşım dikkate alınarak hesaplanan katsayılardan Omega incelenmiştir.

Alan yazın incelendiğinde güvenirlilik katsayılarının karşılaştırılmasına yönelik çalışmaların bileşik testlerin güvenirligi üzerine yapıldığı gözlenmiştir.

Bu araştırmanın, Klasik Test Kuramı'ndaki varsayımların karşılanmadığı durumlarda faktör analitik yaklaşımlarla güvenirlilik hesaplamalarının yapılabileceğini ortaya koyması ve paralel ölçmelerin yanı sıra konjenerik ve eş biçimli ölçme yapılarının güvenirliklerinin çeşitli faktörler altında incelenmesi açısından önemli olduğu ve alanyazına dikkate değer bir katkı sağlayacağı düşünülmektedir.

Problem Cümlesi

Paralel, eş biçimli ve konjenerik ölçme yapılarında, madde sayısı, örneklem büyüklüğü ve faktör yükü dağılımı gibi güvenirligi etkileyen etmenler altında, KR-20 ve faktör analitik yaklaşımla elde edilen Omega güvenirlilik katsayıları nasıldır?

Alt Problemler

1. Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının ortalamalarına ortak etkileri nasıldır?

2. Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının standart sapmalarına ortak etkileri nasıldır?

Sınırlılıklar

Bu araştırma;

- 1) Paralel ve Eş Biçimli ölçme yapılarında üretilen 1000, 2000 ve 3000 örneklem büyüklükleri ile,
- 2) Paralel ve Eş Biçimli ölçme yapılarında üretilen 10,20 ve 50 maddelik test uzunluğu ile,
- 3) Paralel ve Eş Biçimli ölçme yapılarında üretilen düşük, orta ve yüksek boyutluluk düzeyleri ile,
- 4) Kullanılan bilgisayar programıyla üretilen simülasyon veriler ile,
- 5) Araştırmaya dahil edilen güvenilirlik katsayıları ile sınırlıdır.

Yöntem

Bu çalışmada, örneklem verilerinden hareketle evrene genelleme yapmak yerine, paralel ve eş biçimli ölçme yapılarında, güvenilirliği etkileyen çeşitli etmenler altında, KR-20 ve faktör analitik yaklaşımda kullanılan güvenilirlik katsayılarının karşılıklı olarak incelenmesi amaçlanmıştır. Elde edilen güvenilirlik katsayılarının çeşitli koşullar altında benzerlik ve farklılıklar inceleneceğinden araştırma, temel araştırma niteliğindedir.

Veri Üretimi ve Veri Analizi

Veri üretimi için yapılan çalışmada, örneklem büyüklüğü, test uzunluğu ve faktör yük düzeyleri değiştirilerek, her bir durum için 100 yinleme (replikasyon) yapılmıştır. Veri üretimi R 3.0 programında Psych paketi ve yazılan kodlar kullanılarak gerçekleştirilmiştir.

Araştırma kapsamında değiştirilen faktörler ve bu faktör düzeyleri ile ilişkili değerler belirlenirken, ilgili literatürde yer alan benzer çalışmalardaki bulgular dikkate alınmıştır. Değiştirilen faktörler ve bu faktörlere ilişkin düzeyler aşağıda belirtilmiştir.

Örneklem Büyüklüğü: İlgili literatür incelendiğinde, örneklem büyüklüğünün faktör yükü kestiriminde etkili olduğu görülmektedir. Örneğin; Sünbül (2013) çalışmasında Klasik Test Kuramı için 500'den büyük örneklemelerde stabil kestirim sağlanabiliyorken, çok boyutlu yapılarda MIRT modellemeleri için en sağlıklı sonuçların örneklem büyüklüğünün 3000 olduğu durumlarda sağlandığını belirtmiştir.

Bu çalışma kapsamında üretilen verilerde Omega katsayısı kestirimi MIRT modellemesiyle yapıldığı için, örneklem büyüklükleri, parametre kestiriminin farklı örneklem büyüklüklerinde nasıl değişeceğini belirlemek amacıyla 1000, 2000 ve 3000 olmak üzere üç düzey olarak değiştirilmiştir.

Madde Sayısı: Testin uzunluğu, hem gözlenen puan varyanslarını hem de gerçek puan varyanslarını etkiler. Bir testteki madde sayısı artırıldığında, gerçek puan varyansı ve hata puan varyansı da artacaktır (Magnusson, 1968). Bu çalışmada test uzunluğunun güvenilirlik katsayılarını etkilediği göz önüne alınarak farklı madde sayılarında kestirilen parametrelerin nasıl değişeceğini belirlemek amacıyla, madde sayıları 10,20 ve 50 olarak belirlenmiştir.

Faktör yükü düzeyleri: Bu çalışmada farklı ölçme yapıları için farklı faktör yük düzeylerinde veriler üretilmiştir. Eş Biçimli ve paralel ölçme yapıları için faktör yük düzeyleri 0.5,0.6,0.7,0.8 ve 0.9 olmak üzere beş düzey olarak belirlenmiştir. Konjenerik ölçme yapısı için faktör yükleri, 0.5-0.7 aralığında ve 0.5-0.9 aralığında olmak üzere seçkisiz olarak belirlenmiştir.

Tablo 1
Yapılan Değişimlere İlişkin Düzenek

Faktör	Düzy Sayısı	Düzy Değerleri
Örnekleme Büyüklüğü	3	1000 2000 3000
Madde Sayısı	3	10 20 50
Faktör Yüky Düzeyleri	5	0.5 0.6 0.7 0.8 0.9
Ölçme Yapıları	3	Paralel Eş Biçimli
Replikasyon	100	

İşlem ve Verilerin Analizi

Veriler, paralel ve Eş Biçimli ölçme yapılarında ayrı ayrı olarak, belirtilen koşulların çaprazlanması sonucu oluşan her bir duruma uygun olarak üretilmiştir. Veri üretimi R 3.0 programında psych paketi ve araştırmacı tarafından yazılan kodlar kullanılarak gerçekleştirilmiştir.

Üretilen verilerden parametre kestirimleri sirt paket programı ve araştırmacı tarafından R 3.0 programı kullanılarak yazılan kodlar ile gerçekleştirilmiştir. Paralel, Eş Biçimli ve konjenirik ölçme yapıları için farklı değişimleme ölçütlerine göre KR-20 ve omega katsayıları hesaplanmıştır.

Paralel, Eş Biçimli ve konjenirik ölçme yapıları için farklı değişimleme ölçütlerine göre farklı değişimleme ölçütlerinde hesaplana güvenilirlik katsayılarının ortalamalarının ve standart sapmalarının yer aldığı bir matris oluşturulmuştur. Elde edilen ortalama ve standart sapma değerleri göz önüne alınarak değişimlenen tüm faktörlerin temel etkileri hesaplanmıştır. Temel etkiler her bir faktörün her bir düzeyi için elde edilen değerlerin ortalamaları alınarak hesaplanmıştır. Elde edilen bulguların ortak etkiler açısından yorumlanabilmesi için çok değişkenli Grafikselle yöntemlerden yararlanılmıştır. Grafik çizimleri, R 3.0 programı kullanılarak gerçekleştirilmiştir.

Bulgular

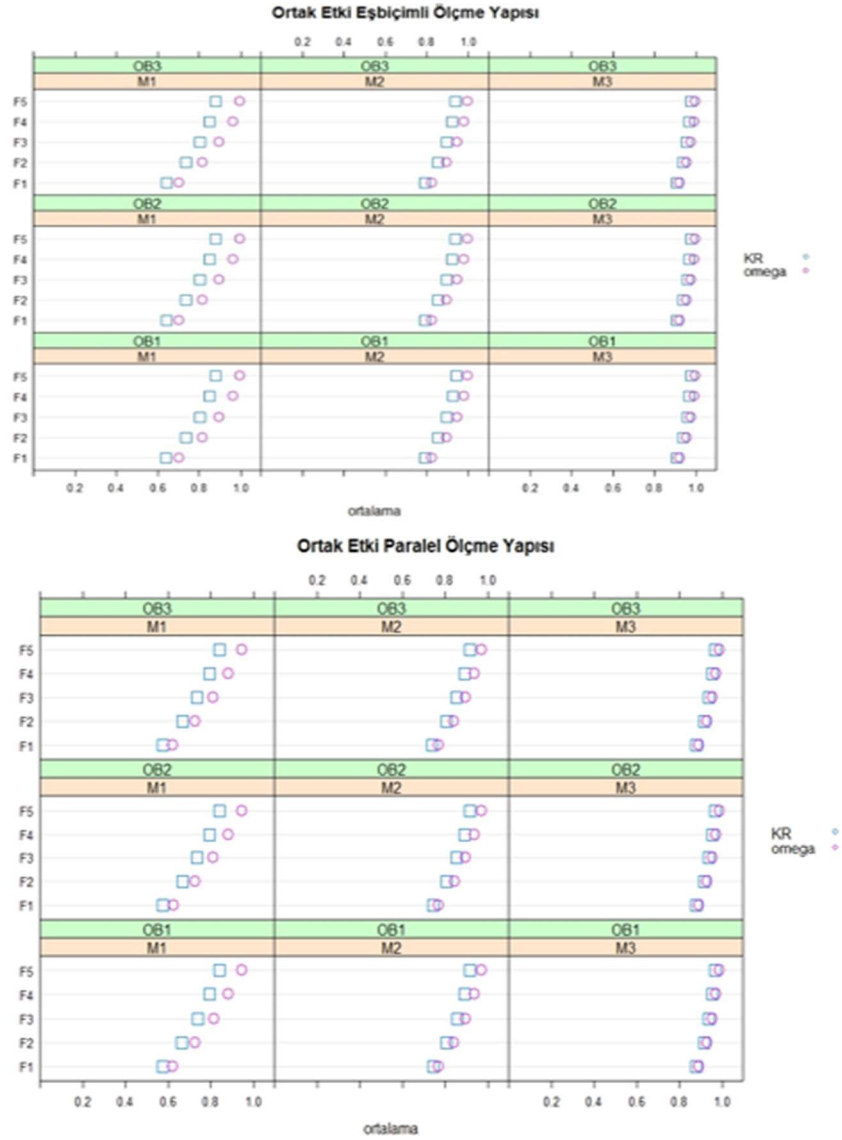
Bu bölümde, alt problemlere ilişkin bulgulara yer verilmiştir. Çeşitli faktörlerin temel etkilerinin ve temel etkilere ilişkin ortalamaların gösteriminde 2 boyutlu tablo ve grafiklerden, ortak etkilerin gösteriminde ise, çok boyutlu veri grafiklerinden yararlanılmıştır. Grafikler arası standardın korunabilmesi için temel etki grafiklerinde standart sapma bulgularının yer aldığı grafikler minimum 0.00, maksimum 0.05; ortalamaların yer aldığı grafikler ise minimum 0.50, maksimum 1.00 olacak şekilde ölçeklenmiştir.

Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının ortalamalarına ortak etkileri nasıldır?

Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının ortalamalarına ortak etkileşimlerine ilişkin sonuçları Grafik 20’de gösterilmektedir.

Grafiklerde yer alan koşulların ortak etkileri açısından KR-20 ve Omega ortalamalarının dağılımları incelendiğinde, Omega ortalamasının her koşulda KR-20 ortalamasından büyük olduğu görülmektedir. Paralel ve Eş Biçimli ölçme yapılarının ikisinde de madde sayısı 10 iken Omega ve KR-20 arasındaki fark fazla iken, madde sayısı arttıkça katsayılar birbirine yaklaşmaktadır. Örneklem büyüklüğü tek başına KR-20 ve Omega ortalamalarında ciddi bir artış sağlamamakta, ancak madde sayısı ve örneklem büyüklüğü birlikte değerlendirildiğinde ikisindeki artışla birlikte KR-20 ve Omega ortalamaları artarak birbirlerine yaklaşmaktadırlar.

Eş Biçimli ölçme yapısında KR-20 ve Omega ortalamaları her koşulda paralel ölçme yapısından bir miktar daha yüksek çıkmıştır. Bu sonuca Eş Biçimli ölçme yapısında hata varyanslarının, farklı faktör yükü düzeylerinde o faktör yükünün karesi ile toplandığında 1’i geçmeyecek şekilde 0.20 ile 0.70 arasında hata varyansı değerlerini içeren matristen seçkisiz olarak alınması ancak, paralel ölçme yapısında hesaplanan katsayılardaki hata varyanslarının sabit olması sebep olmaktadır.

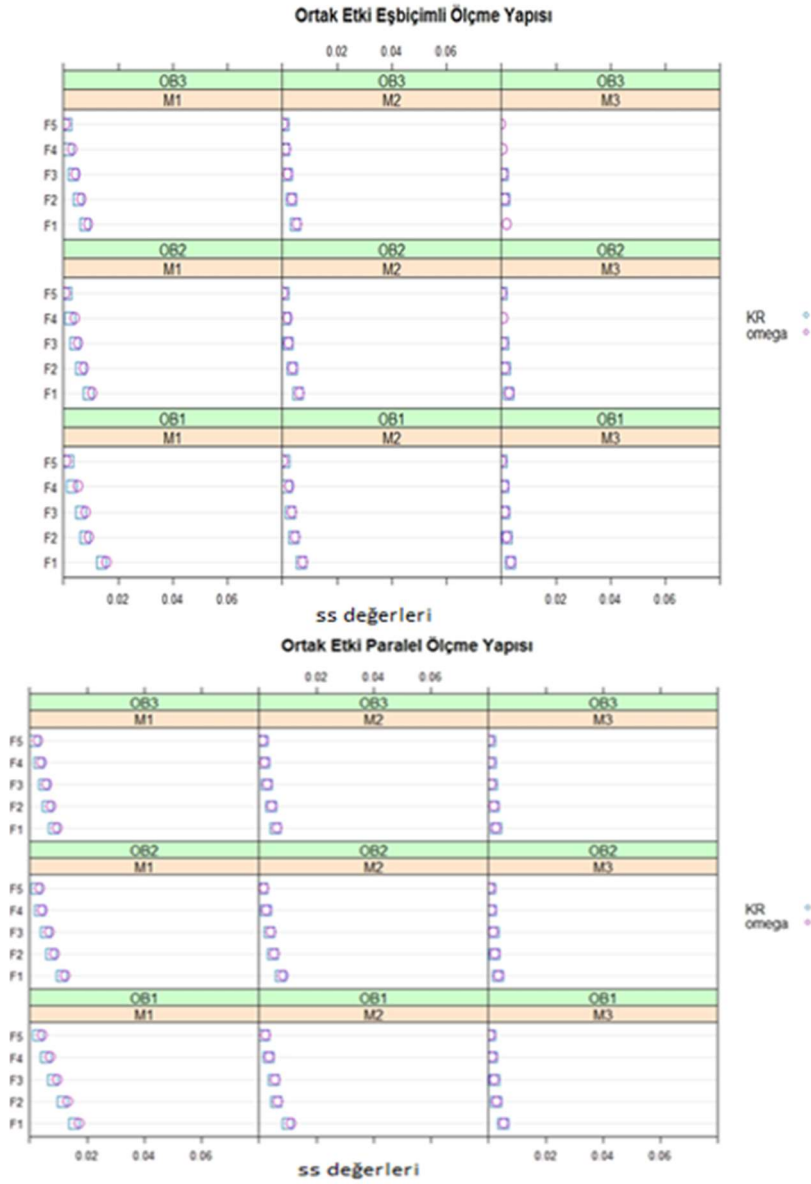


Grafik 1. Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, beş farklı faktör yükü düzeyi altında, KR-20 ve Omega katsayılarının ortalamalarına ortak etki değerleri

Grafik 1 incelendiğinde faktör yükü düzeyleri arttıkça Eş Biçimli ve paralel ölçme yapılarında tüm koşullarda KR-20 ve Omega ortalamaları artmaktadır. Grafiklerdeki saçılımlar göz önüne alındığında faktör yükü düzeyinin ve madde sayısı artışının KR-20 ve Omega ortalamalarının artması sağladığı, ancak örneklem büyüklüğünün katsayı ortalamalarında ciddi bir değişikliğe neden olmadığı söylenebilir.

Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının standart sapmalarına ortak etkileri nasıldır?

Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, farklı faktör yükü düzeyleri altında, KR-20 ve Omega katsayılarının standart sapmalarına ortak etkileri sonuçları Grafik 2’de gösterilmektedir.



Grafik 2. Paralel ve eş biçimli ölçme yapılarında; madde sayısının ve örneklem büyüklüğünün, beş farklı faktör yükü düzeyi altında, KR-20 ve Omega katsayılarının standart sapmalarına ortak etki değerleri

Grafiklerde yer alan koşulların ortak etkileri açısından KR-20 ve Omega standart sapmalarının dağılımları incelendiğinde, Omega standart sapmasının her koşulda KR-20 standart sapmasından bir miktar daha büyük olduğu görülmektedir. Paralel ve Eş Biçimli ölçme yapılarının ikisinde de madde sayısı 10 iken Omega ve KR-20 arasındaki fark fazla iken, madde sayısı arttıkça katsayılar birbirine yaklaşmaktadır. Örneklem büyüklüğü tek başına KR-20 ve Omega standart sapmalarında ciddi bir değişiklik yaratmamakta, ancak madde sayısı ve örneklem büyüklüğü birlikte değerlendirildiğinde ikisindeki artışla birlikte KR-20 ve Omega standart sapmaları azalarak birbirlerine yaklaşmaktadırlar.

Eş Biçimli ölçme yapısında KR-20 ve Omega standart sapmaları, her koşulda, paralel ölçme yapısından bir miktar daha düşük çıkmıştır. Bu sonuca Eş Biçimli ölçme yapısında hata varyanslarının, farklı faktör yükü düzeylerinde o faktör yükünün karesi ile toplandığında 1'i geçmeyecek şekilde 0.20 ile 0.70 arasında hata varyansı değerlerini içeren matristen seçkisiz

olarak alınması ancak, paralel ölçme yapısında hesaplanan katsayılardaki hata varyanslarının sabit olması sebep olmaktadır.

Grafik 2 incelendiğinde faktör yükü düzeyleri arttıkça Eş Biçimli ve paralel ölçme yapılarında tüm koşullarda KR-20 ve Omega standart sapmaları azalmaktadır. Grafiklerdeki saçılımlar göz önüne alındığında faktör yükü düzeyinin ve madde sayısı artışının KR-20 ve Omega standart sapmalarının azalmasını sağladığı, ancak örneklem büyüklüğünün katsayı standart sapmalarında dikkat çeken bir değişikliğe neden olmadığı söylenebilir.

Tartışma, Sonuç ve Öneriler

Paralel ve Eş Biçimli ölçme yapılarında katsayı ortalamaları için faktörlerin ortak etkileri incelendiğinde, örneklem büyüklüğü, madde sayısı ve faktör yükü düzeyleri arttıkça ortalamaların yükseldiği ve Omega ile KR-20 ortalamalarının giderek birbirlerine yaklaştığı söylenebilir. Örneklem büyüklüğü tek başına ortalamalar üzerinde ciddi bir artışa neden olmazken faktör yükü düzeyi ve madde sayısının artmasıyla birlikte incelendiğinde katsayı ortalamalarında önemli bir artış sağladığı sonucuna ulaşılabilir. Faktörlerin ortak olarak artırılması katsayı ortalamalarının daha yansız kestirimini sağlamaktadır.

Paralel ve Eş Biçimli ölçme yapılarında faktörlerin ortak etkilerinin katsayı standart sapmalarına olan etkileri incelendiğinde, örneklem büyüklüğü, madde sayısı ve faktör yükü düzeyleri arttıkça standart sapmaların azaldığı, Omega ve KR-20 standart sapmalarının giderek birbirlerine yaklaştığı söylenebilir. Örneklem büyüklüğü tek başına ortalamalar üzerinde ciddi bir azalmaya neden olmazken faktör yükü düzeyi ve madde sayısının artmasıyla birlikte incelendiğinde katsayı standart sapmalarında önemli bir azalma sağladığı sonucuna ulaşılabilir. Faktörlerin ortak olarak artırılması katsayı ortalamalarının daha yansız kestirimini sağlamaktadır.

Paralel ölçme yapısında farklı madde sayılarında, farklı örneklem büyüklüklerinde ve farklı faktör yüklerinde elde edilen Omega ve KR-20 katsayı ortalamaları, Eş Biçimli ve konjenerik ölçme yapılarında elde edilen katsayı ortalamalarına göre, beklenenin aksine, daha düşük çıkmıştır. Paralel ölçme yapısında üretilen verilerde hata varyansları, paralel ölçmelerin yapısı gereği, sabit tutulmuştur. Ancak Eş Biçimli ölçme yapısında hata varyansları, farklı faktör yükü düzeylerinde o faktör yükünün karesi ile toplandığında 1'i geçmeyecek şekilde 0.20 ile 0.70 arasında hata varyansı değerlerini içeren matristen seçkisiz olarak alınmıştır. Bu durumda ölçmelere karışan hata varyansları Eş Biçimli ölçme yapılarında paralel ölçme yapılarına göre daha düşük olabilmektedir. Örneğin 0.50 faktör yükü düzeyinde 10 madde uzunluğu ve 1000 örneklem büyüklüğünde, paralel ölçme yapısında Omega değeri ortalaması 0.62 iken, Eş Biçimli ölçme yapısında 0.70 olarak elde edilmiştir. Örnekteki koşullarda paralel ölçme yapısında hata varyansı sabit ve 0.75 iken Eş Biçimli ölçme yapısında 0.20 ile 0.70 arasında değişen değerlere göre seçkisiz olarak alınmıştır. Elde edilen bulgulardan yola çıkarak farklı ölçme yapılarında, madde uzunluğu örneklem büyüklüğü gibi faktörlerden daha çok, ölçmeye karışacak hata miktarları hesaplanan katsayıları etkilemektedir çıkarımı yapılabilir.

Benzer bir çalışma, madde sayıları, örneklem büyüklüğü ve faktör yükü düzeyleri değişimlenmiştir. Aynı çalışma veri üretimi aşamasında grup heterojenliği ve hata varyansları değişimlenerek yapılabilir. Bu çalışmada sadece ikili derecelenmiş (dichotomous) maddelerde veri üretildiği için KR-20 ve Omega katsayıları kullanılmıştır. Benzer bir çalışma çoklu derecelenmiş (polythomous) maddelerde α katsayısı ve Omega katsayısı kullanılarak yapılabilir. Simülasyon ile yapılan bu çalışma gerçek koşullarda tekrarlanabilir.

Kaynaklar

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Company.
- Alwin, D. F. (1976). Attitude scales as congeneric tests: A re-examination of an attitude-behavior model. *Sociometry*, 39, 377-383.

- Atılğan, H., Kan A., ve Doğan N. (2009). *Eğitimde ölçme ve değerlendirme*, Anı Yayıncılık,
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik Test Teorisi ve uygulaması*. ÖSYM Yayınları.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and psychological measurement*, 66(6), 930-944.
- Gulliksen, H. (1950). *Theory of mental tests*. John Wiley & Sons, Inc.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Erlbaum.
- McDonald, R. P. (1999). *Test theory: a unified treatment*. Erlbaum.
- Magnusson, D. (1968). *Test Theory*. Addison-Wesley Publishing Company
- Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32(1), 1-13.
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173-184.
- Raykov, T. (2001). Bias of coefficient alfa for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69– 76
- Raykov, T. (1997). Scale Reliability, Cronbach's Coefficient Alpha, and Violations of Essential Tau-Equivalence with Fixed Congeneric Components. *Multivariate Behavioral Research*, 32(4), 329-353
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Sünbül, Ö. ve Erkuş, A. (2013). Madde parametrelerinin değişmezliğinin çeşitli boyutluluk özelliği gösteren yapılarda madde tepki kuramına göre incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2).
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçme teknikleri*. ÖSYM yayınları.
- Turgut, M. F. (1980). *Eğitimde Ölçme ve Değerlendirme*. Gül Yayınevi
- Yurdugül, H. (2005). Konjenerik test kuramı ve konjenerik madde analizi: Tek boyutlu çoktan seçmeli testler üzerine bir uygulama. *A.Ü. Eğitim Bilimleri Fakültesi Dergisi*, 38(2).
- Yurdugül, H. (2006). Çoktan seçmeli testlerde klasik güvenilirlik terimlerinin madde-yanıt kuramından elde edilmesi. *A. Ü. Eğitim Bilimleri Fakültesi Dergisi*, 39(2), 27-44.
- Yurdugül, H. (2006). Paralel, eşdeğer ve konjenerik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *A.Ü. Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15-37.

Extended Abstract

Introduction

Measurement is based on making error-free, objective, standard and valid measurements. For this reason, it is of great importance to develop qualified measurement tools. In order for a measurement tool to be qualified at the expected and intended level, its reliability and validity must be ensured. Validity of a measurement tool is defined as the appropriateness of inferences made from test scores or other forms of evaluation (Crocker and Algina, 1986). Essentially, the validity of a measurement tool is determined by the degree to which it serves the purpose for which it was developed. The concept of validity should be considered together with reliability. Although validity is the ultimate goal to be achieved for measurement tools, the prerequisite of reliability must be met in order to ensure validity. At this point, the relationships between validity and reliability should be addressed. While the validity of a measurement tool depends on its

reliability, high reliability does not necessarily mean that it is valid (Crocker and Algina, 1986). Reliability is used in many senses, but in the most basic sense, reliability is the ability of the measurement tool to make measurements free from errors (Turgut, 1972). The concept of reliability can be discussed separately as rater reliability, reliability in the sense of stability, and reliability in the sense of internal consistency. Due to the scope of this study, the concept of reliability in the sense of internal consistency will be discussed.

In this research, it was aimed to compare reliability coefficients under various factors according to classical test theory and factor analytic approaches. For this purpose, in parallel, tau-equivalent and congeneric measurement structures, alpha, KR-20 and Omega, one of the coefficients calculated taking into account the factor analytic approach, will be examined according to test length, sample size and factor load levels.

When the literature was examined, it was observed that studies on the comparison of reliability coefficients were conducted on the reliability of composite tests. It is thought that this research is important in terms of revealing that reliability calculations can be made with factor analytic approaches in cases where the assumptions in Classical test theory are not met, and examining the reliabilities of congeneric and tau-equivalent measurement structures as well as parallel measurements under various factors, and will make a notable contribution to the literature.

The hypotheses of the research are as follows:

What are the omega reliability coefficients obtained with KR-20 and factor analytic approach under factors affecting reliability such as number of items, sample size and factor load distribution in parallel, tau-equivalent and congeneric measurement structures?

1. In parallel and tau-equivalent measurement structures,
 - a) Number of items,
 - b) Sample size,
 - c) Factor load distribution,

What are the common effects of KR-20 and omega coefficients on their averages?

2. In parallel and tau-equivalent measurement structures,
 - a) Number of items,
 - b) Sample size,
 - c) Factor load distribution,

What are the common effects of KR-20 and omega coefficients on standard deviations under different factor loading levels?

Method

In this research, instead of generalizing to the population based on sample data, it is aimed to mutually examine the reliability coefficients used in KR-20 and factor analytic approach under various factors affecting reliability in parallel and tau-equivalent measurement structures. The research is basic research since the similarities and differences of the reliability coefficients obtained will be examined under various conditions.

In the data generation study, sample size, test length and factor load levels were changed and 100 replications were made for each case. Data generation was carried out using the psych package and written codes in the R 3.0 program. When the related literature is examined, it is seen that sample size is effective in factor loading estimation. For example, Sünbül (2013) stated in his study that while stable estimation can be achieved in samples larger than 500 for Classical Test Theory, the healthiest results for MIRT modeling in multidimensional structures are provided when the sample size is 3000. Since the estimation of the Omega coefficient in the data produced within the scope of this study was performed by MIRT modeling, the sample sizes were varied in three levels as 1000, 2000 and 3000 in order to determine how the parameter estimation would change at different sample sizes. The length of the test affects both observed score variances and true score variances. In this study, considering that the test length affects the reliability

coefficients, the number of items was determined as 10, 20 and 50 in order to determine how the estimated parameters would change with different item numbers. In this study, data with different factor loading levels were generated for different measurement constructs. For the congruent and parallel measurement structures, factor loading levels were determined as 0.5, 0.6, 0.7, 0.8 and 0.9. For the congeneric measurement structure, factor loadings were randomly determined between 0.5-0.7 and 0.5-0.9. While determining the factors that changed within the scope of the research and the values related to these factor levels, the findings of similar studies in the relevant literature were taken into account. The changing factors and the levels related to these factors are stated below.

Factor	Number of Levels	Level Values
Sample Size	3	1000 2000 3000
Number of item	3	10 20 50
Level of Factor Loads	5	0.5 0.6 0.7 0.8 0.9
Measurement Structures	3	Parallel Tau-equivalent
Replication	100	

Parameter estimations from the data produced were made with the sirt package program and the codes written by the researcher using the R 3.0 program. KR-20 and omega coefficients were calculated according to different modification criteria for parallel and tau-equivalent measurement structures.

A matrix was created containing the averages and standard deviations of the reliability coefficients calculated at different modification criteria for parallel, tau-equivalent and congeneric measurement structures. Taking into account the obtained mean and standard deviation values, the main effects of all changing factors were calculated. Main effects were calculated by taking the average of the values obtained for each level of each factor. Multivariate graphical methods were used to interpret the findings in terms of common effects. Graphic drawings were made using the R 3.0 program.

Result and Discussion

When the joint effects of the factors for the coefficient means in parallel and congruent measurement structures are examined, it can be said that as the sample size, number of items and factor loading levels increase, the means increase and the Omega and KR-20 means gradually approach each other. While sample size alone does not cause a significant increase in the averages, it can be concluded that it provides a significant increase in the coefficient averages when examined together with the increase in the factor loading level and the number of items. Increasing the factors jointly provides a more unbiased estimation of the coefficient means.

When the effects of the common effects of the factors on the coefficient standard deviations in parallel and congruent measurement structures are examined, it can be said that as the sample size, number of items and factor loading levels increase, the standard deviations decrease and the Omega and KR-20 standard deviations gradually approach each other. While sample size alone does not cause a significant decrease in the averages, it can be concluded that it provides a

significant decrease in the coefficient standard deviations when examined together with the increase in the factor loading level and the number of items. Increasing the factors jointly provides a more unbiased estimation of the coefficient means.

Contrary to expectations, the Omega and KR-20 coefficient means obtained at different item numbers, different sample sizes and different factor loadings in the parallel measurement structure were lower than the coefficient means obtained in the congruent and congeneric measurement structures. The error variances in the data produced in the parallel measurement structure were kept constant due to the nature of parallel measurements. However, in the congruent measurement structure, error variances were randomly taken from the matrix containing error variance values between 0.20 and 0.70 so that when summed with the square of that factor loading at different factor loading levels, it would not exceed 1. In this case, the error variances involved in the measurements may be lower in congruent measurement structures than in parallel measurement structures. For example, at a factor loading level of 0.50, with a length of 10 items and a sample size of 1000, the average Omega value was 0.62 in the parallel measurement structure, while it was 0.70 in the congruent measurement structure. While the error variance was constant and 0.75 in the parallel measurement structure under the conditions in the sample, it was taken randomly according to values ranging between 0.20 and 0.70 in the uniform measurement structure. Based on the findings, it can be concluded that in different measurement structures, the amount of errors that will be involved in the measurement affects the calculated coefficients rather than factors such as item length and sample size.

In this study, the number of items, sample size and factor loading levels were varied. The same study can be conducted by varying group heterogeneity and error variances during the data generation phase.