

Şans Başarısının Üç-Parametrelili Lojistik Modelin Psikometrik Özellikleri Üzerindeki Etkilerinin İncelenmesi

İrem ÇELİK¹

¹ Ordu Üniversitesi, Ordu, Türkiye, iremcelik@odu.edu.tr, ORCID: 0009-0006-7465-7381

ÖZET

Üç parametrelili lojistik model, Madde Tepki Kuramı (MTK) çerçevesinde ayırt edicilik (a), güçlük (b) ve şans başarısı (c) parametrelerini içeren bir ölçme modelidir. Bu çalışma, şans başarısı parametresinin (c) üç parametrelili lojistik modelin psikometrik özellikleri üzerindeki etkilerini Madde Tepki Kuramı (MTK) ve Klasik Test Kuramı (KTK) perspektifleriyle karşılaştırmalı olarak incelemektedir. Araştırma kapsamında, c parametresinin farklı değerlerinin test güvenirliği, madde-toplam korelasyonu ve test bilgi fonksiyonu üzerindeki etkilerini izole etmek amacıyla bir Monte Carlo simülasyonu tasarlanmıştır. Simülasyon koşulları, modelin teorik gereklilikleri dikkate alınarak belirlenmiştir; a ve b parametreleri rastgele seçilip tüm tekrarlar da sabit tutularak c'nin saf etkisi analiz edilmiştir. Veri üretimi ve analizleri R programlama dili kullanılarak gerçekleştirilmiştir. Elde edilen bulgular, c parametresindeki artışın test güvenirliğini önemli ölçüde düşürdüğünü göstermektedir. Örneğin, c=0.10 iken Cronbach alfa katsayısı 0.72 olarak ölçülürken, c=0.40'ta bu değer 0.57'ye gerilemiştir. Madde-toplam korelasyonları da benzer şekilde c arttıkça zayıflamış, ortalama korelasyon 0.28'den 0.19'a düşmüştür. Bu durum, şans etkisinin hem iç tutarlılığı hem de madde geçerliliğini tehdit ettiğini ortaya koymaktadır. Test bilgi fonksiyonu analizleri ise, özellikle düşük yetenek düzeylerinde ($\theta < 0$) bilgi miktarının belirgin şekilde azaldığını ve ölçme hassasiyetinin bozulduğunu işaret etmektedir. Bu bulgular, madde geliştirme süreçlerinde şans başarısının minimize edilmesi ve test tasarımında MTK tabanlı bilgi fonksiyonu analizlerinin kullanılmasının gerekliliğini vurgulamaktadır. Bu çalışmanın, çoktan seçmeli testlerde sıklıkla göz ardı edilen şans başarısı parametresinin ölçme kalitesine etkilerini sistematik biçimde incelemesi bakımından önemli olduğu düşünülmektedir. Bu bulgular, madde geliştirme sürecinde şans başarısının minimize edilmesi ve test tasarımında MTK tabanlı bilgi fonksiyonu analizlerinin kullanılmasının gerekliliğini vurgulamaktadır.

ANAHTAR KELİMELE

Madde tepki kuramı, Şans başarısı, Üç-parametrelili lojistik model

Giriş

Madde Tepki Kuramı (MTK), bireylerin gözlemlenebilir test performansı ile bu performansın temelinde yatan ve doğrudan ölçülemeyen yetenekler arasındaki ilişkiyi modelleyen bir ölçme kuramı ve istatistiksel çerçevedir (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985). Klasik Test Kuramı'nın (KTK) madde ve birey parametre değişmezliğini sağlayamaması, ölçme hatasının tüm bireyler için sabit kabul edilmesi gibi kısıtlılıklarına çözüm amacıyla geliştirilen MTK, test ve ölçek geliştirme, test eşitleme, yanlılık analizleri gibi çeşitli ölçme problemlerinde önemli avantajlar sunmaktadır (DeMars, 2010). Özellikle birey düzeyinde hata değişiminin modellenenebilmesi ve her bir maddeye ait parametrelerin bağımsız olarak tahmin edilebilmesi, MTK'nın psikometrik değerlendirme süreçlerinde tercih edilmesini sağlamaktadır.

MTK'nın en ayırt edici özelliği, bireylerin yetenek düzeyleri ile test maddelerine verdikleri yanıtlar arasındaki ilişkiyi açıkça matematiksel olarak tanımlayabilmesidir. Bu tanım sayesinde, bireylerin farklı yetenek düzeylerinde belirli bir maddeye ilişkin sağladığı bilgi miktarı madde bilgi fonksiyonu ile ifade edilebilmekte; bu fonksiyonların toplamı ile elde edilen test bilgi fonksiyonu, testin farklı yetenek seviyelerindeki ölçme hassasiyetini değerlendirmeye imkân tanımaktadır (Baker, 2001). MTK'nın bu yönü, testin her birey için farklı düzeylerde ne derece

güvenilir ve geçerli ölçüm sunduğunu açıklamada klasik modellerin ötesine geçmesini sağlamaktadır. Nitekim MTK, günümüzde bilgisayar destekli adaptif testlerin temelini oluşturarak bireyselleştirilmiş ölçme sistemlerinin geliştirilmesine öncülük etmektedir.

MTK çerçevesinde yaygın biçimde kullanılan üç parametrelili lojistik (3PL) model; madde güçlüğü (b), madde ayırt ediciliği (a) ve şans başarısı (c) olmak üzere üç temel parametreye dayanmaktadır. Üç parametrelili lojistik (3PL) modele ilişkin eşitlik aşağıdaki gibidir (DeMars, 2010; Hutchinson, 1991):

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta-b)}}$$

3PL modeli, özellikle çoktan seçmeli testlerde önemli bir faktör olan rastgele yanıt davranışını dikkate alması bakımından diğer lojistik modellere göre daha gerçekçi ölçme imkânı sunmaktadır (Turgut, 2017). Bu modelde yer alan şans başarısı parametresi (c), bireylerin herhangi bir bilgiye sahip olmadan doğru yanıt verme olasılığını temsil eder. Teorik olarak 0 ile 1 arasında değer alabilen bu parametre, özellikle düşük yetenek düzeyindeki bireylerin performansına etkide bulunmakta ve testin genel geçerlik ve güvenilirlik yapısını değiştirebilmektedir (DeMars, 2010).

Şans başarısı parametresinin model uyumu, madde parametre tahmini, ölçme değişmezliği, yanıluluk analizi ve test güvenirliliği üzerindeki etkileri, son yıllarda psikometrik literatürde yoğun olarak ele alınmaktadır. Örneğin, Cuhadar, Yang ve Paek (2021), c parametresinin modele dahil edilmemesinin a ve b parametre tahminlerini olumsuz etkilediğini ve model-data uyum indekslerini zayıflattığını göstermiştir. DeMars ve Jurich (2015), bu parametrenin ihmal edilmesinin değişen madde fonksiyonu (DMF) analizlerinde hatalı kararlar alınmasına neden olduğunu ortaya koymuştur. Finch ve French (2014) ise c parametresindeki küçük farklılıkların dahi madde parametre tahminlerini sistematik olarak saptırabileceğini göstermiştir. Bu çalışmalar, özellikle düşük veya yüksek c değerlerinin testin genel psikometrik özellikleri üzerinde ciddi etkiler yaratabileceğini vurgulamaktadır.

Ancak dikkat çekici bir eksiklik, şans başarısı parametresinin yalnızca MTK bağlamında değerlendirilmesi; KTK ile karşılaştırmalı olarak incelenmesine nadiren yer verilmesidir. Oysa KTK'da şans davranışı genellikle hata varyansı kapsamında değerlendirilmekte ve bu durum, test güvenirliliğini yapay olarak artırabilmektedir (Wise et al., 2006; Paek, 2014). Bu bağlamda, KTK ve MTK'nın şansa yaklaşım biçimlerinin karşılaştırmalı olarak analiz edilmesi, testin geçerlik ve güvenirliliğine ilişkin daha derinlemesine yorumlar yapılmasını sağlayabilir. Bu karşılaştırma, ölçme alanında hangi modelin hangi durumlarda tercih edilmesi gerektiği sorusuna da ışık tutabilir.

Ayrıca literatürde, şans başarısı parametresinin etkilerinin izole biçimde değerlendirildiği çalışmaların sınırlı sayıda olduğu dikkat çekmektedir. Çoğu araştırmada c parametresi, madde güçlüğü (b) ve ayırt edicilik (a) ile birlikte değerlendirilmekte, bu durum c parametresine özgü etkilerin net biçimde analiz edilmesini zorlaştırmaktadır. c parametresinin model üzerindeki etkilerinin daha sağlıklı analiz edilebilmesi için, diğer parametrelerin sabit tutulduğu simülasyon temelli tasarımlara ihtiyaç duyulmaktadır.

Bu bağlamda, mevcut çalışmanın amacı, üç parametrelili lojistik (3PL) modelde yer alan şans başarısı parametresinin (c) farklı düzeylerdeki etkilerinin, 1) testin güvenirliliği, 2) madde-toplam korelasyonu ve 3) test bilgi fonksiyonu üzerindeki yansımalarını analiz etmektir. Çalışmada ayrıca, elde edilen bulguların Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) perspektiflerinden yorumlanarak, iki yaklaşımın model seçimi ve sonuçların geçerliği üzerindeki etkileri karşılaştırmalı olarak değerlendirmektedir. Bu analiz, c parametresinin ölçme süreçlerindeki rolünü daha açık bir biçimde ortaya koyarak, model seçimine yönelik kuramsal ve pratik katkılar sağlamayı amaçlamaktadır.

Yöntem

Bu çalışma, çoktan seçmeli testlerde katılımcıların şans başarısını modelleyen üç parametrelili lojistik model (3 PL) kapsamında, şans parametresinin (c) farklı değerlerinin test güvenilirliği, madde fonksiyonları ve ölçme hata varyansı üzerindeki etkilerini ölçmeyi amaçlayan bir Monte Carlo simülasyon çalışmasıdır. Simülasyon çalışmaları, gerçek dünya sistemlerinin kontrollü koşullar altında dinamik davranışının modellenerek gözlemlenmesine izin verir (Banks vd., 2005; Kothari, 2004). Monte Carlo simülasyonu, verinin rastgele üretildiği ve özellikle karmaşık sistemlerin incelenmesinde veya gerçek dünyada elde edilmesi zor koşullarda sistemlerin sınırlarının test edildiği durumlarda tercih edilen etkili bir yaklaşımdır (Law ve Kelton, 2000). Bu araştırma şans başarısının test sonuçlarına yansımalarını izole ederek, bu durumun özellikle şans başarısının dikkate alındığı 3PL modeli üzerindeki etkilerini simülasyon yoluyla incelemeyi hedeflemektedir. Simülasyon sürecinde aşağıdaki adımlar izlenmiştir.

Simülasyon Koşulları

Simülasyonlar, R 4.3.1 programı ortamında, çok çekirdekli işlemci desteği ile paralel hesaplama yöntemleri uygulanarak gerçekleştirilmiştir. Veri üretimi ve analizleri için *foreach* (Revolution Analytics, 2022), *psych* (Revelle, 2023) ve *doParallel* (Weston ve Microsoft Corporation, 2022) paketleri kullanılmıştır (R Core Team, 2023). Simülasyon koşulları Tablo 1’de verilmiştir.

Tablo 1. Simülasyon koşulları

Faktör	Düzeyler
Genel Parametreler	
Örneklem (N)	500
Test Uzunluğu	20
Tekrar Sayısı	1000 Monte Carlo tekrarı
Madde Parametreleri	
Yetenek Dağılımı (θ)	$\theta \sim N(0,1)$
Madde Güçlüğü (b)	$b \sim N(0,1)$
Madde Ayırt Ediciliği (a)	$a \sim \text{LogNormal}(\mu=0, \sigma=0.5)$
Şans Başarı Parametresi (c)	$c=0.10$ $c=0.15$ $c=0.20$ $c=0.25$ $c=0.30$ $c=0.35$ $c=0.40$
Yanıt Modeli	$P(Y = 1 \theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}}$
Analiz Parametreleri	
Theta Grid	$\theta \in [-4, 4]$ aralığında 100 eşit aralıklı nokta
Test Bilgi Fonksiyonu	$I(\theta) = \sum_{i=1}^{20} \frac{ai^2(1 - c)Pi(\theta)(1 - Pi(\theta))}{(c + (1 - c)Pi(\theta))^2}$
Ölçme Hata Varyansı	$\frac{1}{\sum I(\theta)}$

Simülasyon koşullarında belirtilen genel parametre sınırlandırmaları, modelin gerekliliklerini sağlayacak şekilde hem istatistiksel gücü sağlamak hem de hesaplama verimliliğini korumak amacıyla belirlenmiştir. Bu kapsamda, literatürdeki benzer simülasyon çalışmalarıyla tutarlı ve parametre tahminlerinin kararlılığını destekleyen orta büyüklükte örneklem tercih edilmiştir (DeMars, 2010; Finch ve French, 2014). Bununla birlikte hem pratik uygulanabilirliği yansıtmaları hem de madde bilgi fonksiyonlarının etkileşimini inceleyebilmek amacıyla kısa-orta uzunlukta bir test uzunluğu tercih edilmiştir (Baker, 2001). Tekrar sayısı ise Monte Carlo

simülasyonlarında istatistiksel kararlılık için standart kabul edilen tekrar sayısı olan 1000 olarak belirlenmiştir (Law ve Kelton, 2000).

Madde parametreleri, Tablo 1’de belirtilen dağılımlara göre üretilmiştir. Yetenek dağılımı, MTK modellerinde standart kabul edilen normal dağılımdan seçilmiştir. Bu seçim, parametrelerin yorumlanması kolaylaştırarak kuramsal tutarlılığın sağlanmasını sağlar (Hambleton ve Swaminathan, 1985). Madde güçlüğü için normal dağılım tercih edilirken, madde ayırt ediciliği için lognormal dağılım tercih edilmiştir. Lognormal dağılım, madde ayırt edicilik parametresi için pozitif değer üretme zorunluluğunu karşılar ve gerçek ayırt edicilik parametresini modellemeye uygundur (DeMars, 2010). Şans başarı parametresi (c) için ise .10 ile .40 aralığında .05’lik aralıklarla artan değerler tercih edilmiştir. Şans başarı parametresinin (c) pratikte .35’in üzerinde olmayacağı düşünülmekle birlikte model üzerindeki etkilerini incelemek amacıyla c = .40 değeri simülasyona dahil edilmiştir. Bu aralığın, parametrelerin test istatistikleri üzerindeki doğrusal ve doğrusal olmayan etkilerini incelemeyi mümkün kılmaktadır (Turgut, 2017). Şans başarı parametresinin (c) etkisinin izole edilebilmesi amacıyla a ve b parametreleri için belirlenen dağılımlardan rastgele seçim yapılmış ve tüm c değerleri bu değerler sabit tutulmuştur. Bu tasarım, c parametresinin izole edilmiş etkilerinin analizini mümkün kılmıştır.

Üç parametrelili lojistik modele (3PL) göre bir bireyin bir maddeyi doğru cevaplama olasılığı, $P(Y = 1 | \theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}}$ formülü ile tanımlanır. Bu formülde a madde ayırt ediciliğini, b madde güçlüğü, c ise şans başarı parametresini ifade eder. Test bilgi fonksiyonu ise tüm maddelerin bireysel bilgi fonksiyonlarının toplamı olarak tanımlanır ve ölçme duyarlılığını niceliksel olarak ifade eder: $I(\theta) = \sum_{i=1}^{20} \frac{ai^2(1-c)Pi(\theta)(1-Pi(\theta))}{(c+(1-c)Pi(\theta))^2}$ (Samejima, 1994). Ölçme hata varyansı ise bilgi fonksiyonunun tersi olarak tanımlanır ve MTK’nın birey düzeyine özgü hata tahminini temsil eder: $\frac{1}{\sum I(\theta)}$ (Hambleton ve Swaminathan, 1985). Bu yapılar, testin farklı yetenek düzeylerindeki hassasiyetini değerlendirmek için kullanılır.

Analiz sürecinde tüm yetenek seviyelerindeki psikometrik özelliklerin değerlendirilebilmesi amacıyla theta grid ($\theta \in [-4, 4]$) teorik bir aralıkta eşit aralıklı noktalar kullanılarak oluşturulmuştur (Hambleton ve Swaminathan, 1985). Test bilgi fonksiyonu, şans başarı parametresinin (c) bilgi miktarı üzerindeki etkisini açıkça yansıtmaması amacıyla Fisher bilgi kriterine dayalı olarak hesaplanmıştır (Samejima, 1994). Ölçme hata varyansı ise bilgi fonksiyonu ile ters orantılı olarak hesaplanmış ve MTK’nın yetenek düzeyine özgü ölçme hatası yaklaşımını temsil etmiştir (Hambleton ve Swaminathan, 1985).

Simülasyon koşulları yukarıda belirtilen değerlendirmeler göz önünde bulundurularak c parametresinin 3PL modelin psikometrik özellikleri üzerindeki etkilerinin incelenebilmesi amacıyla optimize edilmiştir.

Bulgular

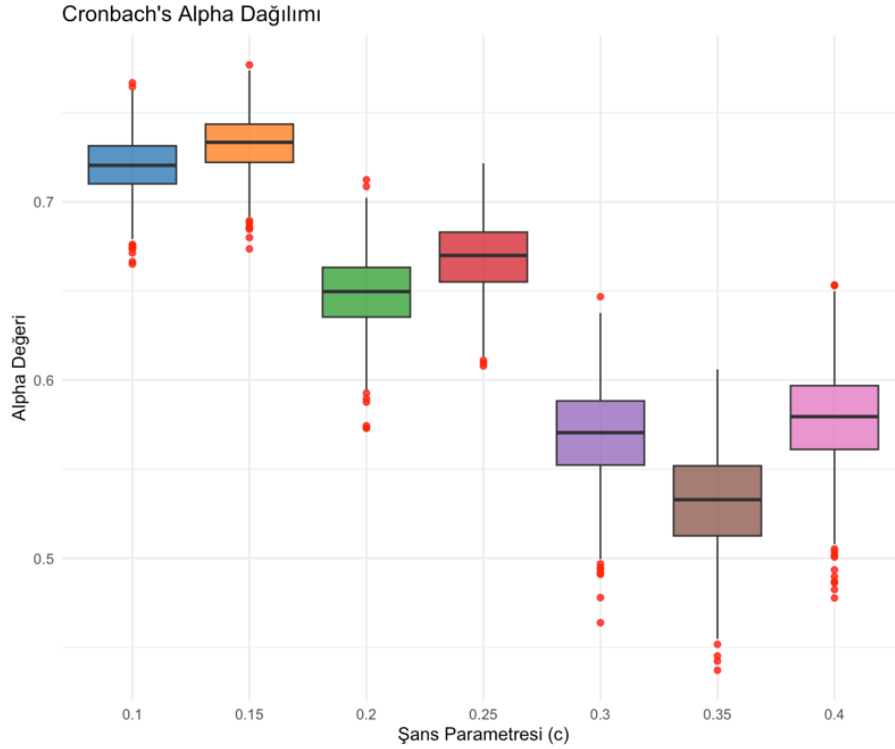
Bu çalışmada 3PL kapsamında şans parametresinin (c) ikili puanlanan (0-1) çoktan seçmeli testlerin psikometrik özellikleri üzerindeki etkilerini incelemek amacıyla Cronbach’s alfa katsayısı, madde-toplam korelasyonu, test bilgi fonksiyonu analizleri gerçekleştirilmiştir. Tablo 2, c parametresinin değişimine göre Cronbach alfa katsayısı ve ortalama madde-toplam korelasyonlarını sunmaktadır.

Tablo 2. Güvenirlik ve Ortalama Madde-Toplam Korelasyonları

c	α	$\bar{r}$	SD _r
0.10	0.72	0.28	0.08
0.15	0.73	0.30	0.10
0.20	0.64	0.23	0.08
0.25	0.66	0.25	0.10
0.30	0.56	0.19	0.07

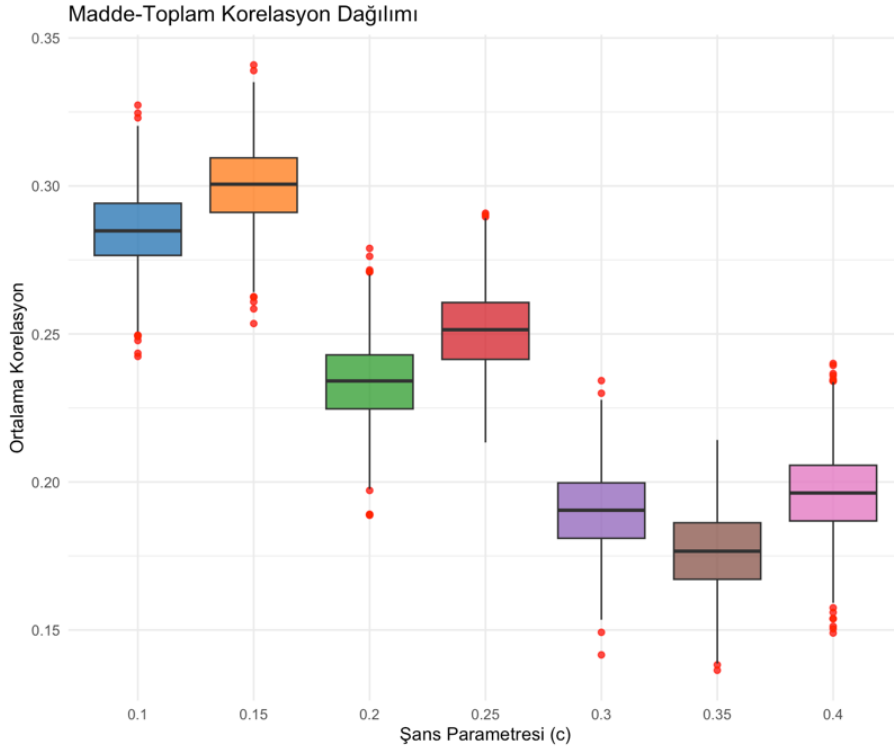
0.35	0.53	0.17	0.07
0.40	0.57	0.19	0.07

Tablo 2 incelendiğinde, c parametresi arttıkça Cronbach alfa katsayısında düşüş gözlemlenmektedir. Şans başarısı parametresi (c) .10 iken .72 olan Cronbach alfa katsayısı, bu değer .40 olduğunda .57'ye kadar gerilemiştir. Benzer şekilde, ortalama madde-toplam korelasyon değerleri c parametresindeki artış ile düşme eğilimi göstermiştir ve .28'den .19'a kadar gerilemiştir. Madde-toplam korelasyonlarının standart sapma değerleri ise .007 ile .10 arasında değişim göstermiştir. Şekil 1 ve 2, c parametresindeki değişime göre Cronbach alfa ve ortalama madde-toplam korelasyon değerlerine ilişkin bu dağılımları göstermektedir.



Şekil 1. c parametresine göre Cronbach alfa dağılımı

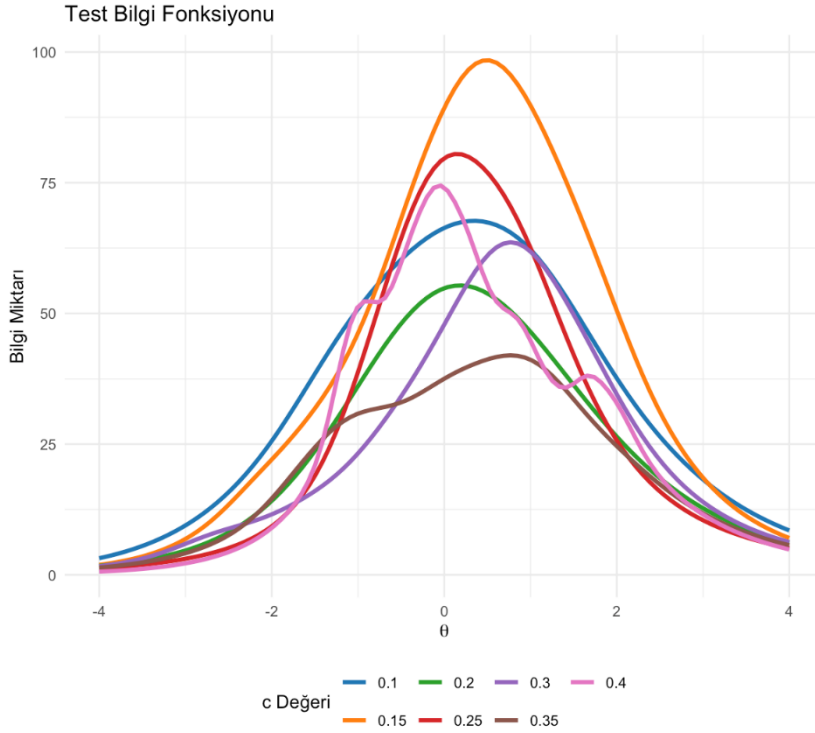
Şekil 1, farklı şans başarı (c) parametreleri altında hesaplanan Cronbach alfa katsayılarının dağılımını göstermektedir. Grafik, c parametresindeki artışın testin iç tutarlılığı üzerindeki etkilerini açıkça ortaya koymaktadır. Elde edilen sonuçlara göre, c parametresi arttıkça Cronbach alfa katsayısında genel bir düşüş eğilimi gözlenmektedir. Bu durum, yüksek c değerlerinin testin güvenilirliğini olumsuz etkilediğini göstermektedir. Özellikle .35 ve .40 gibi yüksek c değerlerinde alfa katsayılarının belirgin şekilde azaldığı görülmektedir. Bu bulgu, rastgele yanıtlamaların testin genel iç tutarlılığına zarar verdiğini ve güvenilirlik katsayılarının yapay biçimde düşmesine neden olabileceğini göstermektedir.



Şekil 2. c parametresine göre ortalama madde-toplam korelasyon dağılımı

Şekil 2’de c parametresine göre ortalama madde-toplam korelasyonlarının dağılımı sunulmaktadır. Grafik, madde düzeyinde iç tutarlılık ölçütü olarak kabul edilen madde-toplam korelasyonlarının şans başarısından nasıl etkilendiğini göstermektedir. c parametresindeki artışla birlikte korelasyon değerlerinde düşüş eğilimi belirginleşmektedir. Bu düşüş, özellikle .30 ve üzeri c değerlerinde daha belirgin hale gelmektedir. Bulgular, şans başarısı etkisinin sadece test düzeyinde değil, madde düzeyinde de psikometrik kaliteyi azalttığını ve madde ile toplam puan arasındaki ilişkiyi zayıflatıldığını ortaya koymaktadır. Bu nedenle, test geliştirme süreçlerinde madde-toplam korelasyonlarının yorumlanmasında c parametresinin potansiyel etkisi göz önünde bulundurulmalıdır.

Buna ek olarak test bilgi fonksiyonuna ilişkin bulgular, Şekil 3’te sunulmuştur. Farklı c parametrelerine göre test bilgi fonksiyonlarının değişim gösterdiği ve c değeri arttıkça bilgi fonksiyonundaki dalgalanmaların arttığı görülmektedir. Test bilgi fonksiyonlarının en tutarlı olduğu c değeri .15 iken, en fazla dalgalanmanın .35 ve .40 değerlerinde olduğu belirlenmiştir.



Şekil 3. c parametresine göre test bilgi fonksiyonlarının dağılımı

Şekil 3'te farklı c değerlerine göre elde edilen test bilgi fonksiyonları karşılaştırmalı olarak sunulmuştur. Grafik, testin yetenek düzeylerine göre sağladığı bilgi miktarının, c parametresine bağlı olarak nasıl değiştiğini yansıtmaktadır. Bulgular, düşük c değerlerinde (özellikle .10 – .15 aralığında) bilgi fonksiyonunun daha tutarlı ve belirgin olduğunu göstermektedir. Ancak c değeri arttıkça, test bilgi fonksiyonlarında dalgalanmaların belirgin biçimde arttığı gözlenmiştir. Bu durum, yüksek şans başarısı değerlerinin testin farklı yetenek düzeylerindeki ölçme hassasiyetini azalttığını ve bilgi miktarının homojenliğini bozduğunu göstermektedir. En fazla dalgalanmanın c = .35 ve c = .40 değerlerinde ortaya çıkması, bu aralıkların ölçme doğruluğu açısından kritik eşikler olabileceğini düşündürmektedir. Bu sonuçlar, test bilgi fonksiyonunun yorumlanmasında c parametresinin göz ardı edilmemesi gerektiğini açıkça ortaya koymaktadır.

Sonuç ve Tartışma

Bu çalışma, 3PL modelinde şans başarısı parametresinin (c) farklı düzeylerinin test güvenilirliği üzerindeki etkilerini hem Klasik Test Kuramı (KTK) hem de Madde Tepki Kuramı (MTK) çerçevesinde sistematik biçimde ele almıştır. Elde edilen bulgular, c parametresindeki artışın Cronbach alfa katsayısında azalmaya, madde-toplam korelasyonlarında düşüşe ve test bilgi fonksiyonlarında dalgalanmalara yol açtığını ortaya koymaktadır. Bu durum, şans başarısının testin iç tutarlılığı ve ölçme doğruluğu açısından doğrudan bir tehdit unsuru olduğunu göstermektedir (Paek, 2015; Zimmerman ve Williams, 2003).

Sonuçlar, test geliştirme süreçlerine yönelik çeşitli uygulamalı çıkarımları da beraberinde getirmektedir. Öncelikle, çoktan seçmeli test maddelerinin oluşturulmasında çeldirici seçeneklerin dikkatle yapılandırılması gerekmektedir. Özellikle düşük yetenek düzeyindeki bireylerin rastgele doğru yanıt verme olasılığını azaltmak amacıyla, işlevsel çeldiriciler kullanılarak c parametresi düşürülebilir (Hambleton ve Swaminathan, 1985). Bu bağlamda, madde yazımı sürecinde hem içerik geçerliği hem de çeldirici işlevselliği birlikte değerlendirilmelidir.

İkinci olarak, test analizlerinde yalnızca klasik güvenirlik ölçütlerinin (ör. Cronbach alfa) yanı sıra, MTK temelli test bilgi fonksiyonu gibi duyarlılığı yüksek metriklerin de dikkate alınması önerilmektedir. Test bilgi fonksiyonları, yetenek düzeylerine göre ölçme hatalarının daha ayrıntılı değerlendirilmesini mümkün kılmakta ve özellikle bilgisayar adaptif test uygulamaları için değerli veriler sunmaktadır (Samejima, 1994; Baker, 2001). Yüksek c değerlerinin gözlemlendiği maddelerin belirlenerek testten çıkarılması ya da revize edilmesi, testin genel güvenirliğini artırmak adına önemli bir adımdır.

Üçüncü olarak, sınav uygulama süreçleri de şans başarısı üzerinde etkili olabileceğinden, uygulama koşullarının dikkatle planlanması gerekmektedir. Sınav süresinin yeterli verilmesi, öğrencilerin motivasyonlarının artırılması ve hızlı yanıt davranışlarının önlenmesi, testin ölçme duyarlılığını olumlu yönde etkileyebilir. Wise, Bhola ve Yang (2006), hızlı yanıt davranışlarının testin iç tutarlılığını yapay olarak yükselttiğini ve test güvenirliğini tehdit ettiğini belirtmiştir. Bu nedenle, sınav sırasında yanıt süreleri ve yanıt kalitesi gibi verilerin birlikte izlenmesi önerilmektedir.

Bulgular, Paek'in (2014) MTK'ya dayalı simülasyon çalışmasıyla da örtüşmektedir. Paek, c parametresinin test uzunluğu ve madde güçlük dağılımına bağlı olarak Cronbach alfa katsayısında anlamlı düşüslere neden olduğunu belirtmiştir. Mevcut çalışmanın özgün katkısı, bu etkinin hem KTK hem de MTK bağlamında çok boyutlu olarak ele alınması ve madde düzeyindeki analizlerle desteklenmiş olmasıdır. Özellikle, c parametresindeki artışın madde-toplam korelasyonlarında sistematik azalmaya yol açtığı belirlenmiştir. Bu bulgu, maddelerin toplam puanla ilişkisini zayıflatarak geçerliliğini tehlikeye sokmakta ve testin içerik doğruluğunu olumsuz etkilemektedir (DeMars, 2010; Baker, 2001).

MTK çerçevesinde elde edilen bulgular da literatürle paralellik göstermektedir. Hambleton ve Swaminathan (1985), c parametresinin madde karakteristik eğrilerini yukarı kaydırarak düşük yetenek düzeylerinde doğru yanıt olasılığını yapay olarak artırdığını ifade etmişlerdir. Bu durum, test bilgi fonksiyonu üzerinden değerlendirildiğinde daha açık biçimde görünmektedir. Bu çalışmada, c parametresi arttıkça test bilgi fonksiyonunda düzensizliklerin arttığı, özellikle $c = .35$ ve $.40$ düzeylerinde ölçme duyarlılığının ciddi biçimde azaldığı görülmüştür. Samejima (1994) ve Finch ve French (2014), benzer şekilde, yüksek c değerlerinin test bilgi fonksiyonlarını olumsuz etkilediğini ve testin belirli yetenek düzeylerinde geçerli sonuçlar üretme kapasitesini azalttığını rapor etmişlerdir.

Bu bağlamda çalışma hem klasik hem de modern test kuramlarını aynı analitik çerçevede bir araya getirerek c parametresinin etkilerini çok boyutlu bir bakış açısıyla incelemektedir. Literatürde bu tür bütüncül analizlerin az sayıda olması, çalışmanın özgün katkısını vurgulamaktadır. Ayrıca, bu çalışmada elde edilen bulgular, yüksek c parametresine sahip maddelerin testten çıkarılması, revize edilmesi veya işlevsel çeldiricilerle desteklenmesi gerektiğini göstermektedir (Turgut, 2017; Zimmerman ve Williams, 2003). Formül puanlama ve cevapsız bırakma gibi stratejilerin de şans etkisini azaltıcı biçimde kullanılması, test güvenirliğini artırma noktasında önemli uygulamalardır. Çalışmanın pratik yansımaları da dikkate değerdir. Özellikle yüksek c parametresine sahip maddelerin testten çıkarılması, revize edilmesi veya çeldiricilerin işlevselliğinin artırılması önerilmektedir (Hambleton ve Swaminathan, 1985). Ayrıca, formül puanlama uygulamaları ve cevapsız bırakma stratejileri gibi tekniklerin kullanımı, şans başarısından kaynaklanan sistematik hataları azaltmada etkili olabilir (Zimmerman ve Williams, 2003; Turgut, 2017). Gelecekte yapılacak araştırmalar, farklı test uzunlukları, madde dağılımları ve popülasyon yapıları altında c parametresinin etkilerini daha ayrıntılı biçimde ele alarak, alandaki kuramsal ve pratik bilgi birikimini daha da geliştirebilir.

Sınırlılıklar ve Öneriler

Bu araştırmanın bulguları, belirli varsayımlara dayalı olarak gerçekleştirilen simülasyon koşulları kapsamında değerlendirilmelidir. Çalışmada kullanılan madde parametreleri (a, b, c) belirli dağılımlardan rastgele seçilmiş olup, bu dağılımların gerçek test uygulamalarındaki

çeşitliliği ve bağlama özgü farklılıkları tam olarak temsil etmeyebileceği göz önünde bulundurulmalıdır. Ayrıca kullanılan madde sayısı 20 ile sınırlı tutulmuş, bu durum daha uzun testlerde ya da adaptif test ortamlarında parametrelerin farklılaşan etkilerinin kapsam dışı kalmasına neden olmuştur. Bununla birlikte, analizlerde kullanılan yetenek dağılımı, genellikle varsayılan normal dağılıma dayalı olarak oluşturulmuş ve theta grid $[-4, 4]$ aralığında eşit aralıklarla belirlenmiştir. Oysa gerçek sınav uygulamalarında yetenek düzeyleri bu kadar düzenli ve simetrik bir yapıya sahip olmayabilir; örneğin pozitif ya da negatif çarpıklık içeren örneklem ya da uç değerlerin yoğunlaştığı dağılımlar modelin tahmin gücünü farklı şekilde etkileyebilir. Buna ek olarak, 3PL modelinin varsayımları olan tek boyutluluk ve yerel bağımsızlık gibi koşullar simülasyon ortamında sağlanmış olsa da, bu varsayımlar gerçek testlerde sıklıkla ihlal edilebilmekte ve bu da modelin uyumu ile parametre tahminlerinin geçerliğini sınırlandırabilmektedir.

Bu sınırlılıklar göz önünde bulundurularak, gelecekte yapılacak araştırmalarda bazı genişletmelere yer verilmesi önerilmektedir. Öncelikle, simülasyon ortamında elde edilen bulguların geçerliliğini değerlendirmek amacıyla gerçek sınav verileriyle karşılaştırmalı analizlerin yapılması önem arz etmektedir. Gerçek veriler üzerinden elde edilecek sonuçlar, modelin uygulama gücünü ve pratik geçerliliğini daha doğru şekilde ortaya koyacaktır. Ayrıca, farklı test uzunlukları ve madde sayıları kullanılarak yapılacak analizler, özellikle test bilgi fonksiyonu ve ölçme hatası varyansı gibi ölçütlerin stabilitesine dair daha kapsamlı sonuçlar sağlayabilir. Normal dağılımın ötesine geçilerek, çarpık veya çok modlu yetenek dağılımları üzerinden yapılacak analizler, modelin farklı örneklem yapıları altındaki performansını ortaya koyacaktır. Bununla birlikte, şans başarı parametresi (c) için sabit ve artan değerler yerine, örneğin beta dağılımı gibi daha gerçekçi ve esnek dağılımlar kullanılarak modellenmesi, test davranışlarının daha doğal bir biçimde temsil edilmesine katkı sağlayabilir. Son olarak, çok boyutlu test yapıları ve IRT modelleriyle yapılacak karşılaştırmalı çalışmalar, şans başarısı etkisinin yalnızca tek bir yetenek boyutuyla sınırlı olmadığını ve daha karmaşık yapıların dikkate alınması gerektiğini gösterebilir. Bu doğrultuda yapılacak çalışmalar, özellikle düşük güvenilirlik düzeyine sahip testlerde şans başarısının etkisini daha hassas biçimde değerlendirmeye olanak sağlayacak ve 3PL modelinin psikometrik analizlerdeki işlevselliğini artıracaktır.

Etik Kurul Onay Bilgileri (The Ethical Committee Approval)

Bu çalışma, yapay olarak üretilmiş veriler kullanılarak gerçekleştirildiğinden etik kurul onayı gerektirmemektedir.

Çıkar Çatışması (Conflict of Interest)

“Şans Başarısının Üç-Parametrelili Lojistik Modelin Psikometrik Özellikleri Üzerindeki Etkilerinin İncelenmesi” başlıklı makalemin herhangi bir kurum, kuruluş, kişi ile mali çıkar çatışması yoktur.

Finansal Destek (Financial Support)

Çalışma için herhangi bir finansal destek alınmamıştır.

Kaynaklar

- Baharuddin, B., Handayani, L., & Rusli, R. (2024). *Enhancing biochemistry assessment quality in medical education through item response theory (IRT)*. Research Square. <https://doi.org/10.21203/rs.3.rs-3859753/v1>
- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2005). *Discrete-event system simulation* (4th ed.). Pearson.

- Birnbaum, A. (2008). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Addison Wesley.
- Bolt, D., & Liao, X. (2021). On the positive correlation between DIF and difficulty: A new theory on the correlation as methodological artifact. *Journal of Educational Measurement*, 58(4), 465–491. <https://doi.org/10.1111/jedm.12302>
- Brown, W. M., Templin, J., & Cohen, A. S. (2015). Model selection issues in the three parameter logistic model. *Applied Psychological Measurement*, 39(2), 123–138.
- Bulut, O. (2015). Applying item response theory models to entrance examination for graduate studies: Practical issues and insights. *Journal of Measurement and Education in Education and Psychology*, (2). <https://doi.org/10.21031/epod.17523>
- Camilli, G. (2018). IRT scoring and test blueprint fidelity. *Applied Psychological Measurement*, 42(5), 393–400. <https://doi.org/10.1177/0146621618754897>
- Carroll, J. B. (1945). The effect of difficulty and chance success on test reliability. *Psychometrika*, 10(1), 1–9.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Cuhadar, C., Yang, Y., & Paek, I. (2021). The impact of ignoring guessing parameter on measurement invariance testing. *Educational and Psychological Measurement*, 81(4), 601–623.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- DeMars, C. E., & Jurich, D. P. (2015). Guessing and differential item functioning. *Educational Measurement: Issues and Practice*, 34(1), 23–31.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Erlbaum.
- Finch, H., & French, B. F. (2014). The impact of differential guessing on item parameter estimation. *Journal of Educational Measurement*, 51(3), 243–260.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer.
- Hutchinson, T. P. (1991). *Ability, partial information, guessing: Statistical modelling applied to multiple-choice tests*. Rumsby Scientific Publishing.
- Ihda, M., & Retnawati, H. (2023). Analysis of the distractor of the multiple-choice test using classical test theory (CTT) and item response theory (IRT). In *Proceedings of the 4th International Conference on Research of Educational Administration and Management (ICREAM 2023)* (pp. 196–203).
- Kothari, C. R. (2004). *Research methodology: Methods and techniques* (2nd ed.). New Age International Publishers.
- Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis* (3rd ed.). McGraw-Hill.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, 7.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Paek, I. (2014). An investigation of the impact of guessing on coefficient alpha and reliability. *Applied Psychological Measurement*, 39(4), 264–277.
- Paek, I. (2015). The effect of guessing parameter on reliability estimation in IRT. *Educational Assessment*, 20(2), 154–171.

- R Core Team. (2023). R: A language and environment for statistical computing (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. <https://www.Rproject.org/>
- Revelle, W. (2023). psych: Procedures for Psychological, Psychometric, and Personality Research (Version 2.3.9) [R package]. Northwestern University. <https://CRAN.Rproject.org/package=psych>
- Revolution Analytics. (2022). foreach: Provides Foreach Looping Construct for R (Version 1.5.2) [R package]. <https://CRAN.R-project.org/package=foreach>
- Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement*, 18(3), 229–244.
- Turgut, M. F. (2017). *Şans başarısının test puanlarına etkisi*. Ankara Üniversitesi Yayınları.
- Wang, B., Huggins-Manley, A., Kuang, H., & Xiong, J. (2024). Enhancing effort-moderated item response theory models by evaluating a two-step estimation method and multidimensional variations on the model. *Educational and Psychological Measurement*, 85(2), 401–423.
- Weston, S., & Microsoft Corporation. (2022). doParallel: Foreach Parallel Adaptor for the 'parallel' Package (Version 1.0.17) [R package]. <https://CRAN.R-project.org/package=doParallel>
- Wise, S. L., Bholá, D. S., & Yang, S. (2006). Taking the time to improve the validity of low stakes tests: The effort-monitoring CBT. *Educational Measurement: Issues and Practice*, 25(2), 21–30. <https://doi.org/10.1111/j.1745-3992.2006.00054.x>
- Zimmerman, D. W., & Williams, R. H. (2003). A new look at the influence of guessing on the reliability of multiple-choice tests. *Applied Psychological Measurement*, 27(5), 357–371.

EXTENDED ABSTRACT

Examining the Effects of Guessing Parameter on the Psychometric Properties of the Three-Parameter Logistic Model

Introduction

Item Response Theory (IRT) is a measurement theory that models the relationship between individuals' test performance and their basic abilities. It was developed to overcome the limitations of Classical Test Theory (CTT) and contributes to the solution of problems such as test development and equating.

The advantages of IRT include its ability to model the discrimination of test items and its individual-based approach to measurement error. Whereas CTT assumes constant error, IRT provides individualized assessment which is important in computer-assisted and adaptive testing.

One of the main components of the IRT is the item characteristic curve (ICC). This curve shows the probability of individuals answering test items correctly depending on their ability level. Different measurement scenarios can be created with 1PL, 2PL and 3PL models based on logistic functions. The 3PL model includes guessing parameter (c) in addition to item difficulty (b) and item discrimination (a) parameters. The c parameter expresses the probability that individuals respond correctly by chance and is a critical factor for measurement reliability.

Research shows that omitting the c parameter in the 3PL model has negative effects on test statistics. When the c parameter is not included in the model, ability estimates may be inaccurate and systematic errors may occur in item analyses. Moreover, test reliability decreases as the c parameter increases. Especially in short tests, chance success was found to have significant effects on reliability.

Method

This study used Monte Carlo simulation to examine the effects of the guessing parameter (c) on test reliability and measurement error variance within a three-parameter logistic model (3PL). Simulations were performed in R 4.3.1 environment and various packages were used for data generation and analysis.

Sample size was limited to 500, test length to 20 items and 1000 Monte Carlo replications. The ability distribution was normal, item difficulty was normal, and item discrimination was lognormal. The chance success parameter (c) took values ranging from 0.10 to 0.40.

Theta grid was set in the range $[-4, 4]$ and the test information function was calculated according to the Fisher information criterion. The measurement error variance is inversely proportional to the information function and reflects the individual measurement accuracy. Simulation results were optimized to analyze the effect of parameter c on the test statistics.

Results

In this study, the effect of the chance parameter (c) in the 3PL model on test reliability and information function is examined. The results show that as c increases, Cronbach's alpha coefficient decreases from 0.72 to 0.57 and item-total correlation decreases from 0.28 to 0.19. Standard deviations ranged between 0.007-0.10.

Test information functions were found to be most stable at $c = 0.15$ and most volatile at $c = 0.35$ and 0.40 . The findings suggest that chance achievement has negative effects on test statistics.

Discussion

This study examines the effect of the c parameter on test reliability in the 3PL model. It has been shown that chance achievement decreases reliability by increasing the error variance in the test results. As the c parameter increased, there was a significant decrease in Cronbach's alpha coefficient and item-total correlations weakened.

In the context of ITC, the increase in the c parameter negatively affected the item information function and decreased the measurement accuracy of the test. It was observed that items with high c values reduced the overall sensitivity of the test. It is recommended that such items be removed or revised during the test development process.

In conclusion, the negative effects of chance achievement on test reliability and measurement precision were confirmed. It is recommended that tests be designed to minimize the effect of chance through formula scoring or alternative strategies.