



Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi
Mustafa Kemal University Journal of the Faculty of Education
Yıl/Year: 2021 ♦ Cilt/Volume: 5 ♦ Sayı/Issue: 8, s. 32-46

ALT TESTLERİN ARTI DEĞER ÖZELLİĞİNE SAHİP OLMA DURUMLARININ FARKLI KOŞULLAR ALTINDA İNCELENMESİ*

Doç. Dr. Önder SÜNBÜL

Mersin Üniversitesi, Eğitim Fakültesi, ondersunbul@gmail.com

Orcid: 0000-0002-1775-1404

Dr. Öğr. Üyesi. Arzu UÇAR

Hakkari Üniversitesi, Eğitim Fakültesi, arzukapcik@gmail.com

Orcid: 0000-0002-0099-1348

Arş. Gör. Dr. Ebru BALTA

Ağrı İbrahim Çeçen Üniversitesi, Eğitim Fakültesi, ebrubalta2@gmail.com

Orcid: 0000-0002-2173-7189

Özet

Bu çalışmada, farklı alt alanlardan oluşan testlerin alt testlerinin artı değer gösterme sıklıklarının, çeşitli koşullar (alt test örneklem büyüklüğü, alt test uzunluğu, alt testler arasındaki korelasyon ve alt testlerin faktör kuvvetleri) altında incelenmesi amaçlanmıştır. Araştırmanın amacı doğrultusunda, birinci boyutta yer alan 15 maddenin özellikleri sabit tutularak, ikinci boyutta test uzunluğu (5, 10, 15), maddelerin faktör kuvveti (0.65, 0.70, 0.75, 0.80, 0.85, 1.0) koşulları altında iki boyutlu basit yapı ve iki kategorili (1-0) tamamlayıcı madde tepki kuramına dayalı 100 yinleme ile 1800 veri seti oluşturulmuştur. Madde vektörlerine açı verebilmek amacıyla boyutlar arası korelasyon (0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95) ve alt test grup örneklem büyüklüğü (500, 1000, 3000) koşulları değiştirilerek 432 farklı deneysel yapı oluşturulmuştur. Artı değer özelliği analizi, Haberman'ın yöntemi kullanılarak, alt test puanlarının, bu alt test puanlarının toplamıyla elde edilen toplam test puanlarının ve alt test puanlarıyla toplam test puanları kullanılarak elde edilen genişletilmiş puanların hata kareleri ortalamasında oransal azaltma (PRMSE) değerleri karşılaştırılarak gerçekleştirilmiştir. Veri üretiminde ve analizinde R.4.0.1. programı kullanılmıştır. Çalışmanın sonucunda, test uzunluğu ve alt testlerin faktör kuvveti miktarı artışı ile birlikte alt testlerin daha çok artı değer gösterdiği ancak alt testler arası korelasyon düzeyi artışı ile birlikte alt testlerin artı değer özelliğine sahip olma sıklıklarında azalma olduğu gözlemlenmiştir. Alt test grup örneklem büyüklüğünün farklılaşmasının ise alt testlerde artı değer özelliği taşıma sıklığında bir farklılık göstermediği sonucuna ulaşılmıştır. Alt testin psikometrik kalitesinin göstergelerinden artı değer özelliği taşıyıp taşıyamama durumu faktör analitik yöntemler ya da Yen OPI gibi analizlerle incelenebilir.

Anahtar Kelimeler: Artı Değer, Alt Test, Çok Boyutlu Madde Tepki Kuramı, Haberman

INVESTIGATION OF SUBTESTS WHICH HAVE ADDED VALUE FEATURE UNDER DIFFERENT CONDITIONS

Abstract

In this study, under various factors (group sample size of subtest, subtest length, correlation between subtests, and factor strengths of subtests (dimensions)) the frequency of subtests that are added value were examined. In line with the purpose of the study, by keeping the parameters of the 15 items in the first dimension constant for all generation conditions, in the second dimension, under the conditions of test length (5, 10, 15), factor strength of the items (0.65, 0.70, 0.75, 0.80, 0.85, 1.0), two-dimensional simple structured and two-dimensional 1800 datasets were generated with 100 replication based on two categorical (1-0) complementary item response theory. To give an angle to the item vectors, 432 different experimental structures were created by changing the conditions of correlation between subtests (0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95) and group sample size of subtest (500, 1000, 3000). The analysis of whether the subtests have added

* Bu araştırma makalesi, 02-05 Mayıs 2018 tarih aralığında Akdeniz Üniversitesi'nde düzenlenen Vth International Eurasian Educational Research Congress'de sözlü bildiri olarak sunulmuştur.

value features was carried out according to Haberman's method which the subtest scores, the total test scores to be obtained by the total of these subtest scores, and the extended scores obtained by using the subtest scores and the total test scores were produced and Proportional Reduction in MSE (PRMSE) values of these scores were calculated and PRMSE values of the three scores generated separately were compared. R.4.0.1 software was used for data generation and analyses. As a result of the study, it was observed that the subtests showed more added value with the increasing the subtest length and the amount of subtest factor strength, but with the increasing the level of correlation between the subtests, the frequency of having added value feature of the subtests decreased. It was concluded that the variation of the group sample size of subtest did not show a difference in the frequency of carrying the added value feature in the subtests. Whether or not the subtest has a added value feature, which is one of the indicators of psychometric quality, can be examined by factor analytical methods or analyzes such as Yen OPI.

Key Words: Added Value, Subtest, Multidimensional Item Response Theory, Haberman

Giriş

Eğitim alanında bireylerin istedik davranışları kazanıp kazanmadığını anlamak için ulusal ve uluslararası boyutta geleneksel kağıt-kalem testlerine sıklıkla başvurulmaktadır. Kullanılan testler alt alanlardan (matematik testinin, cebir ve geometri alt alanlarından, genel yetenek testinin, matematik, okuma ve yazma gibi alt alanlarından oluşması vb.) oluşabilmektedir. Alt test puanları, bireylerin öğrenme eksikliklerinin belirlenmesinde, daha başarılı olunan konuların analiz edilmesinde ya da okulların profillerinin ortaya çıkarılması amacıyla kullanılmaktadır. Ayrıca alt test puanları ile ölçülmesi istenilen özellik için kullanılan maddelerin başka bir özelliğin ölçümünde kullanılabilmesi ve böylelikle bu maddelerin bir araya gelmesi ile yeni bir alt özellik ortaya çıkarak testin psikometrik özelliklerine ilişkin daha detaylı bilgiler elde edilebilmektedir (Sinharay, Haberman ve Puan, 2007). Böylece, testi oluşturan alt alanların sağladığı bu avantajlarla alt test puanlarına olan ilginin arttığı görülmektedir. Bilgisayar ve teknolojinin hızlı bir şekilde günlük hayata girmesi ile birlikte test geliştirme sürecinde farklı teknikler geliştirilmiştir. Faktör analitik teknikler, test geliştirme sürecinde kullanılmasına rağmen bu teknikler ile incelenen özelliğin alt yapılarına ayrıştırılması sürecinde sıkıntılar oluşabilmektedir. Uluslararası ve ulusal düzeyde uygulanan testlerde, konu alanlarının her biri ayrı alt psikolojik boyut ve kendi içinde tek boyutlu varsayılarak toplam puanlar elde edilmekte ve önsel ağırlıklar ile bu alt testler ağırlıklandırılarak bireyler hakkında kararlar verilmektedir. Böylelikle, her maddeye eşit ağırlıklı puan atfedilerek yapay alt testlerin oluşturulması ve ağırlıkların kullanılması, ölçme puanlarına karışan hata miktarının artması sonucuna yol açmaktadır. Ölçme aracı geliştirmede, ilk ve temel amaç, ölçülecek olan psikolojik değişkenin yapısının ortaya çıkarılıp ölçülmesi öngörülen örtük özelliğin işevuruklaştırılmasıdır. İşevuruklaştırma için başvurulan yöntem ise faktör analitik yöntemlerdir. Ölçme araçları açısından bakıldığında, faktör, ortak özelliğe sahip maddeleri bir arada tutan bir bağ ve gerçek puan gibi doğrudan gözlenemeyen örtük bir değişken ifade etmektedir. Maddelerin faktör ile göstermiş olduğu ilişki de faktör yükü olarak tanımlanabilir (Crocker ve Algina, 1986). Yapı geçerliği çalışmalarında önemli bir yeri olan faktör analizi, açımlayıcı (exploratory) ve doğrulayıcı (confirmatory) olmak üzere ikiye ayrılmaktadır (McDonald, 1999). Açımlayıcı faktör analizi, psikolojik uzaydaki yapısı bilinmeyen verilerin yapılarının keşfedilmek istendiği durumlarda, doğrulayıcı faktör analizi ise psikolojik uzaydaki yerleşimleri öngörülen verilerin öngörünün doğruluğunun test edilmek istendiği durumlarda uygulanmaktadır.

Maddelerin faktörlere verdiği yüklerin büyüklüğüne ve faktör yük dağılımına göre, basit (simple), yaklaşık basit (approximate simple) ve karışık (complex) olmak üzere üç farklı yapı tanımlamak mümkündür. Çok boyutlu bir testte, her madde sadece bir boyutu ölçüyorsa böyle bir çokboyutlu yapı "basit yapı" olarak adlandırılabilir (McDonald, 1999). Gerçek verilerde, basit yapının nadiren ortaya çıkmasından dolayı "yaklaşık basit" adıyla başka bir yapı tanımlanmıştır (Walker, Azen ve Schmitt, 2006). Çok boyutlu bir yapıda, madde birden fazla faktöre yük veriyor, ancak birincil olarak sadece bir faktörü ölçüyorsa, bu şekildeki yapılar "yaklaşık basit yapı" olarak adlandırılmaktadır. Çok boyutlu bir yapıda, maddenin birden fazla faktöre yük verdiği ve maddenin ait olduğu faktör ayrımının yapılamadığı faktör yapıları ise "karmaşık yapı" olarak adlandırılabilir (McDonald, 1999). İnsan özellikleri doğası gereği birbirleriyle ilişkilidir ve böylelikle basit bir yapı olmaları hemen hemen mümkün değildir. Bu nedenle, belirli bir özellik, diğerlerinden ayrı olmak yerine daha çok bir bileşen yapısındadır; maddelerin büyük bölümü de çokboyutlu yapıdadır. Bir özelliğin bileşen olması, özelliklerin bir kısmının kendi içinde tekboyutlu olmasına engel değildir; önemli olan bileşen yapısındaki kesişim alanlarının yok sayılmamasıdır (Erkuş, 2012). Ölçme aracının maddesi bir birim değildir; birim örtük özelliği temsil edecek en küçük parçadır ve bunun da her maddede farklı büyüklükte olma olasılığı yüksektir. Böylelikle, maddelerin ait olduğu örtük özelliğin parçası ve örtük özelliğe olan toplam katkısında farklılık görülebilmektedir. Dolayısıyla, psikolojik değişkenlere ilişkin maddelerin, büyük ölçüde çok boyutlu olduğu söylenebilir. Testlerde

yer alan maddelerin ölçtüğü özelliğin farklı olması, maddelerin toplanabilirliğini engellemez. Fakat ölçülen özellikle ilgili testten elde edilecek toplam puan ancak testin tek boyutluluğu ile mümkündür (Erkuş, 2012). Çok boyutlu ve bileşik yapılarda ise her alt testin ve toplam testin tek boyutluluğu sağlayabilmesinden dolayı hem alt testlerin maddeleri toplanarak ayrı ayrı, hem de tüm maddeler toplanarak toplam test bazında kararlar verilebilmektedir.

Uluslararası ve ulusal çapta geniş ölçekli ve yüksek riskli sınav uygulamalarında, bireylerin test maddelerine verdikleri tepkiler sonucunda bireyler hakkında bilgi edinilmektedir. Özellikle çoktan seçmeli testlerde, bireylerin maddelere verdikleri tepkiler doğru ise 1, yanlış ise 0 şeklinde iki kategorili olarak puanlanmaktadır. 1-0 şeklindeki puanlama sonucunda, bireyin, madde kökünde yoklanan davranış hakkında yanlış, kısmi bilgi, ya da hiç bilmeme gibi durumlar dikkate alınmamakta; bu durum da puanlama yetersizliği olarak ileri sürülmektedir (Akkuş ve Baykul, 2001). Böylelikle, uzman kanısına dayalı puanlama, Zinger Z₁ ve Z₂ puanlama gibi maddeleri farklı puanlama türleri geliştirilmiştir. Bu türlerin yanı sıra alt testlerin faktör yapıları dikkate alınarak, maddelerin faktör yükü, ortak varyans ve başka değerlere bağlı olarak ağırlıklarının belirlendiği ve puanlamanın bu durumlar dikkate alınarak yapıldığı yöntemler kullanılmaktadır (Yüce, 2012). Gerçekleştirilen bu çalışmaların alt testlerden daha çok bilgi elde etmeye yönelik olduğu görülmektedir. Son yıllarda bireylere uygulanan testlerden elde edilen sonuçların yanı sıra bu testlerin alt alanları gözetilerek bireylerin, alt alanlardaki profilleri oluşturulmaktadır. Başka bir ifadeyle, bireylerin, testlerin alt alanlarındaki konularda öğrenme eksikliklerinin ve yeterliklerinin belirlenmesinde, testlerin alt puanlarından faydalanılmaktadır. Alt alanlardan elde edilen bilgiler doğrultusunda, bireyler hakkında daha sağlıklı yönlendirme ve kararlara varılabilmektedir. Alt test puanlarının kullanımına olan ilginin artması ile birlikte toplam test ve alt test puanlarının amaca hizmet etme derecelerinin karşılaştırıldığı çalışmalar (Haberman, 2008; Sinharay ve diğer., 2007; Sinharay, 2010a; Sinharay, 2010b; Sinharay ve Haberman, 2011) incelendiğinde, her alt test puanının kullanımının uygun olmadığı ve böylelikle, alt test puanının kullanılabilme şartları arasında artı değer özelliğine sahip olması gerektiğinin belirtildiği görülmektedir. Artı değer özelliğine sahip olma durumu, alt test puanlarının yüksek güvenilirlik düzeyinde olması ve toplam testten elde edilen puanların alt testlerden elde edilen puanlara görece daha düşük güvenilirlik düzeyinde olması ve toplam test puanları ile alt testlerden elde edilen puanlar arasındaki korelasyon düzeyinin düşük olması gibi özellikler çerçevesinde belirlenmektedir. İlgili literatür (Sinharay ve diğer., 2007; Sinharay, 2010a, 2010b) incelendiğinde, alt testlerin hangi koşullar altında artı değer özelliği gösterdiğine ilişkin kesin, net bir çerçeve belirtilmemekle birlikte, alt testler arasında düşük korelasyon olduğu durumda alt test uzunluğunun artmasıyla birlikte alt testlerin artı değer özelliğine sahip olduğu görülmektedir. Bu araştırmada, farklı alt alanlardan oluşan testlerin alt testlerinin artı değer gösterme sıklıklarının, çeşitli koşullar (alt test grup örneklem büyüklüğü, alt test uzunluğu, alt testler arasındaki korelasyon ve alt testlerin (boyutların) faktör kuvvetleri) altında incelenmesi amaçlanmıştır. Araştırmanın amacı doğrultusunda, aşağıdaki sorulara cevap aranmıştır:

1. Çeşitli alt konu alanından oluşan testlerde, tek bir alt konu alanına yönelik maddelerden oluşan alt testlerin, çeşitli koşullar (alt test grup örneklem büyüklüğü, alt test uzunluğu, alt testler arasındaki korelasyon ve alt testlerin (boyutların) faktör kuvvetleri) altında artı değer özelliğine sahip olma durumuna etkisi nasıldır?
 - 1.1. Alt test grup örneklem büyüklüğünün, alt testlerde artı değer özelliğine sahip olma durumuna olan temel etkisi nasıldır?
 - 1.2. Alt test uzunluğunun, alt testlerde artı değer özelliğine sahip olma durumuna olan temel etkisi nasıldır?
 - 1.3. Alt testler arasındaki korelasyonun, alt testlerde artı değer özelliğine sahip olma durumuna olan temel etkisi nasıldır?

1.4. Alt testlerin faktör kuvvetinin, alt test grup örneklem büyüklüğü, alt test uzunluğu ve alt testler arası korelasyonun artı değer özelliğine sahip olma durumuna olan ortak etkisi nasıldır?

Ulusal ve uluslararası ilgili literatür incelendiğinde, çalışmada ele alınan değişkenlerin ve düzeylerinin, farklı alt konu alanlarından oluşan alt testlerinin artı değer gösterme sıklıklarının birlikte değerlendirildiği araştırmaya rastlanmamıştır. Böylelikle, araştırmadan elde edilen bulguların, ölçme ve değerlendirme alanına ilişkin özgün değer taşıyacağı ve alt konu alanlarından oluşan testlerden elde edilen test puanlarına yönelik olarak test geliştiricilere ve uygulayıcılara katkı sağlayacağı düşünülmektedir.

Yöntem

Bu bölümde araştırma modeli, simülasyon koşulları ve gerçekleştirilen analizlerle ilgili bilgiler yer almaktadır.

Araştırma Modeli

Bu araştırmada, simülasyon verileri kullanılarak farklı alt alanlardan oluşan testlerin alt testlerinin artı değer gösterme sıklıkları, çeşitli koşullar (alt test örneklem büyüklüğü, alt test uzunluğu, alt testler arasındaki korelasyon ve alt testlerin (boyutların) faktör kuvvetleri) altında incelenmiştir. Araştırmada, alt alanlardan oluşan testlerde, alt testlerin, çeşitli koşullar altında artı değer gösterme sıklıklarını belirlemek için Monte-Carlo simülasyon çalışması gerçekleştirilmiştir. Araştırmada, gerçek veri ile çalışmada ele alınan koşulların tümünün sağlanmasının mümkün olmamasından dolayı simülasyon verisi kullanılmıştır. Monte Carlo çalışmaları, verilerin bilgisayar aracılığıyla üretildiği deneysel çalışmaların görüntülerini yansıtan çalışmalardır (Harwell, Stone, Hsu ve Kirisci, 2016). Araştırmada, ayrıca, farklı alt alanlardan oluşan testlerin alt testlerinin artı değer gösterme sıklıklarının çeşitli koşullara göre ayrıntılı incelenerek ve karşılaştırılarak ortaya çıkarılması yönüyle araştırmanın betimsel araştırma özelliği taşıdığı söylenebilir. Betimsel araştırmalar, verilen bir durumu eksiksiz ve dikkatli bir şekilde tanımlandığı çalışmalardır (Fraenkel, Wallen ve Hyun, 2012). Bu çerçevede mevcut araştırma, betimsel araştırma niteliğinde bir simülasyon çalışması olarak değerlendirilebilir.

Simülasyon Deseni

Çalışmanın değişkenleri ve düzeyleri ilgili literatür temel alınarak belirlenmiştir. Alt testlerin, artı değer özelliği taşıması için gerekli belirgin şartlar olmamasına rağmen alt testler arasındaki korelasyon düzeyinin artı değer özelliği üzerinde etkiye sahip olduğu bazı araştırmacılar (Sinharay, 2010a; Sinharay, 2010b, Sinharay ve diğer., 2007) tarafından savunulmaktadır. Sinharay (2010a), alt alanlardan oluşan testlerin alt testleri arasındaki korelasyon değerlerini; 0.70, 0.75, 0.80, 0.85, 0.90 ve 0.95 olarak değişimleyerek, alt testler arasındaki korelasyonun alt test formlarının artı değer özelliklerine etkisini incelediği çalışmada, yüksek korelasyona sahip olan alt testlerin artı değer özelliği göstermediğini gözlemlemiştir. Bu çalışmada ise alt testlerin korelasyon düzeyleri 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 ve 0.95 olarak değişimlenmiştir.

Sünbül (2011), çeşitli boyutluluk özelliklerine sahip yapılarda, madde parametresinin değişmezliğini farklı modeller çerçevesinde incelediği çalışmada, ikinci boyutta yer alan maddelerin a² faktör kuvvetlerini, 0.25, 0.50, 0.75 ve 1.00 olarak değişimlemiştir. Çok boyutlu Madde Tepki Kuramında (ÇBMTK), faktör kuvveti düşük olduğunda madde parametrelerinin değişmezliğinin sağlanmadığı görülmüştür. Böylelikle, bu çalışmada, a² faktör kuvvetini 0.65, 0.70, 0.75, 0.80, 0.85 ve 1.00 olarak değişimlenmiştir. Sinharay (2010b) çalışmada, alt test grup örneklem büyüklüğünü 100, 1000 ve 4000, Sünbül (2011) ise çalışmada, 500, 1000 ve 3000

olarak belirlemiştir. Bu çalışmada, alt test grup örneklem büyüklüğü 500, 1000 ve 3000 olarak belirlenmiştir.

Sinharay (2010a, 2010b) alt test uzunluklarını 10, 20, 30 ve 50 olarak değiştirdiği çalışmalarında, test uzunluğu artışı ile birlikte alt testin artı değer özelliği gösterme sıklığının arttığını gözlemlemiştir. Bununla birlikte, Sünbül (2011) çalışmasında, ikinci boyut için madde sayılarını 4, 8, 16, 20 ve 24 olarak değiştirmiştir. Uçar (2016), alt test uzunluğunu 10, 15, 30, 50 ve 80 olarak değiştirdiği çalışmasında, alt test uzunluğunun 30, 40 ve 80 olduğu tüm koşullarda, alt testlerin artı değer özelliği gösterdiğini gözlemlenmiştir. Bununla birlikte, alt test uzunluğunun 15 olduğu durumda, alt testler arasında korelasyon düzeyinin yüksek olduğu (0.8, 0.9) ve alt test grup örneklem büyüklüğünün 20, 25 ve 50 olduğu koşullarda alt testlerin artı değer özelliği taşıma sıklığının azaldığı fakat alt test örneklem büyüklüğünün 100, 200 ve 500 olduğu durumda artı değer özelliği taşıma sıklığının arttığı gözlenmiştir. Bu çalışmada, birinci alt test uzunluğu 15 olarak sabitlenerek, ikinci alt testin madde uzunluğu 5, 10 ve 15 madde olacak şekilde değiştirilmiştir. Tablo 1'de ilgili literatürde yer alan çalışmalar göz önünde bulundurularak araştırmamanın değişkenleri ve değiştirilen düzeylerine yer verilmiştir.

Tablo 1.

Simülasyon Deseni

Değişkenler	Düzeyleri	Toplam Düzeyleri
Alt Test Grup Örneklem Büyüklüğü	500, 1000 ve 3000	3
Alt Test Uzunluğu	5, 10, 15	3
Alt Testler Arasındaki Korelasyon	0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 ve 0.95	8
Alt Testlerin Faktör kuvveti	0.65, 0.70, 0.75, 0.80, 0.85 ve 1.00	6
Toplam		3x3x8x3x6=432

Verilerin Üretilmesi

Çalışmada, alt test grup örneklem büyüklüğü (3 koşul), alt test uzunluğu (3 koşul), alt testler arasındaki korelasyon (8 koşul), alt testlerin faktör kuvveti (6 koşul) koşullarını içermek üzere toplamda 432 (3x3x8x6) koşul incelenmiştir. Çalışmada birinci boyutta yer alan 15 maddenin özellikleri sabit tutularak, ikinci boyutta test uzunluğu (madde sayısı (5, 10, 15), maddelerin faktör kuvveti (0.65, 0.70, 0.75, 0.80, 0.85, 1.0) koşulları altında iki boyutlu basit yapılı ve iki kategorili (1-0) tamamlayıcı madde tepki kuramına dayalı 100 yineleme ile 1800 veri seti oluşturulmuştur. Çalışmada, her bir koşulun düzeyi için 100 yineleme yapılarak veriler üretilmiştir. Araştırma sonuçlarının tutarlı ve genellenebilir olabilmesi için her veri seti üzerinde Sünbül (2011) 'ün çalışmasında 100 yineleme yaptığı gözlemlenmiştir. Veri setleri, R 3.4.0 programı kullanılarak üretilmiştir. Birinci boyutta yer alan 15 maddenin özellikleri bütün üretim koşulları için sabit tutulmuştur. Değişimler ise ikinci boyuttaki madde sayısı ve madde özellikleri ile boyutlar arası korelasyonlar üzerinde yapılmıştır. Üretilen bütün yapılarda basit yapı kullanılmış olup birinci boyutta, maddeler, sadece a1 ve d değerine, ikinci boyutta ise sadece a2 ve d değerlerine sahip olmuştur. Bu amaç doğrultusunda, ikinci boyutta yer alan madde sayıları 5, 10, 15 olarak değiştirilirken, bu maddelerin a2 faktör kuvveti ise 0.65, 0.70, 0.75, 0.80, 0.85 ve 1.00 olacak şekilde değiştirilmiştir. Böylelikle, 3x6=18 farklı deneysel hücre oluşmuştur. Madde vektörlerine açı verebilmek amacıyla, yukarıda belirtilen koşullara boyutlar arası korelasyonlar da eklenmiştir. Boyutlar arası korelasyonlar 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 ve 0.95 olarak değiştirilmiştir.

Ayrıca alt testlerde örneklem büyüklükleri 500, 1000 ve 3000 olan gruplar oluşturulmuştur. Böylelikle $6 \times 3 \times 8 \times 3 = 432$ farklı deneysel yapı oluşturulmuştur .

Verilerin Analizi

Psikolojik ve eğitimsel testler, genellikle değişik derecelerdeki çoklu özelliği veya özellik bileşkesini ölçmektedir (Ackerman, 1996). Alt testlerden elde edilen puanların, psikometrik kaliteleri test edildikten sonra bireylere rapor edilmektedir (Haberman ve Sinharay, 2010; Reckase ve Xu, 2015; Sawaki ve Sinharay, 2017; Wedman ve Lyren, 2015).

Bu çalışmada, ikili puanlanan maddeler için ÇBMTK modellerinden biri olan tamamlayıcı (compensatory) model kullanılmıştır. Reckase (2009), iki boyutlu ve iki parametrelili tamamlayıcı modelin lojistik olasılık fonksiyonunu,

$$P(X = 1 | \theta_1, \theta_2) = \frac{1}{1 + e^{-1.7(a_1\theta_1 + a_2\theta_2 + d)}} \quad (1)$$

Eşitlik 1 ile ifade etmektedir. Modelde yer alan parametreler incelendiğinde, her iki yeteneğe ilişkin iki ayırt edicilik parametresi (a_1 ve a_2) yer alırken, güçlük parametresine (d) ilişkin sadece bir parametrenin yer aldığı görülebilir. Farklı boyutlardaki yetenek düzeylerinin birbirini tamamlaması nedeniyle, madde güçlüğü her iki boyutta yer alan yetenek düzeylerinin ortak bir ürünü olarak ele alınmaktadır. Bununla birlikte, bireyin birinci ve ikinci boyuttaki yetenek seviyesi ise θ_1 ve θ_2 temsil edilmektedir.

Alt test puanının toplam testin puanından daha çok bilgi verip vermediğine karar vermek için üç farklı hesaplamayla alt test puanları elde edilebilmektedir (Haberman, 2008). Gerçek alt test puanı ve gözlenen alt test puanının, gözlenen toplam test puanı, gözlenen alt test puanı ve toplam test puanı üzerine olan regresyonu üç farklı hesaplama dayanan alt test puanları, alt test puan yaklaşımı, toplam test puanı yaklaşımı, toplam test puanı ve gözlenen alt test puan yaklaşımı olarak sırasıyla ifade edilen (2), (3) ve (4) numaralı eşitlikte verilen denklemlerde ifade edilmiştir.

$$S_x = L(\tau_x | S_x) = E(S_x) + \rho^2(S_x, \tau_x)[S_x - E(S_x)] \quad (2)$$

$$S_x = L(\tau_x | S_z) = E(\tau_x) + \beta(\tau_x | S_z)[S_z - E(S_z)] \quad (3)$$

$$S_x = L(\tau_x | S_x, S_z) = E(S_x) + \beta(\tau_x | S_x, S_z)[S_x - E(S_x)] + \beta(\tau_x | S_z, S_x)[S_z - E(S_z)] \quad (4)$$

Eşitlik 2, 3 ve 4'de $E(S_x)$ gözlenen alt test puan ortalamasını, $\rho^2(S_x, \tau_x)$ alt test puanının güvenilirlik katsayısını, $\beta(\tau_x | S_z)$ toplam test puanından yararlanarak elde edilen alt test puan denklemindeki regresyon katsayısını, $\beta(\tau_x | S_x, S_z)$ ve $\beta(\tau_x | S_z, S_x)$ ise toplam test puanı ve gözlenen alt test puanı kullanılarak elde edilen alt test puan denklemindeki regresyon katsayılarını ifade etmektedir. Eşitlik 2, 3 ve 4 kullanılarak elde edilen alt test puanları, gerçek alt test puan değerleri ile olan farkları yaklaşımının hatalarını ifade eder (Haberman, 2008). Bu hataların varyansları ise gerçek alt test puanı için hata kareleri ortalaması (Mean Square Error; MSE) değerini, bu değerlerin karekökü ise hata kareleri ortalamasının karekökü (Root Mean Square Error; RMSE) değerini ifade etmektedir. Üç farklı yaklaşımla elde edilen alt test puanlarından hangilerinin kullanımının uygun olacağı ya da toplam test puanının kullanılması gerektiği gibi kararlar verebilmek için gözlenen alt test puanı; toplam test puanı; hem gözlenen alt test puan hem de

toplam test puanı kullanılarak elde edilen alt puanların RMSE değerleri karşılaştırılmaktadır. Alt test puanında artı değer özelliği, en az hataya sahip olan yaklaşımda gözleendiği için RMSE değerleri küçük olan alt testlerin kullanılması önerilmektedir. Alt test puanının artı değer özelliğine sahip olup olmadığını belirlemede kullanılan diğer katsayı ise alt test puanının güvenilirlik katsayılarını içerisinde barındıran hata kareleri ortalamasında oransal azaltma (Proportional Reduction in MSE; PRMSE) değerleridir. PRMSE değeri büyük olan yaklaşım için alt test puanı artı değer özelliği göstermektedir (Haberman, 2008). Tablo 2’de, alt puanının artı değer özelliğine sahip olup olmadığını kontrol etmek için kullanılan RMSE ve PRMSE ifadeleri gösterilmiştir.

Tablo 2.
Kestirilen Alt Puanların Artı Değerini Veren İstatistikler

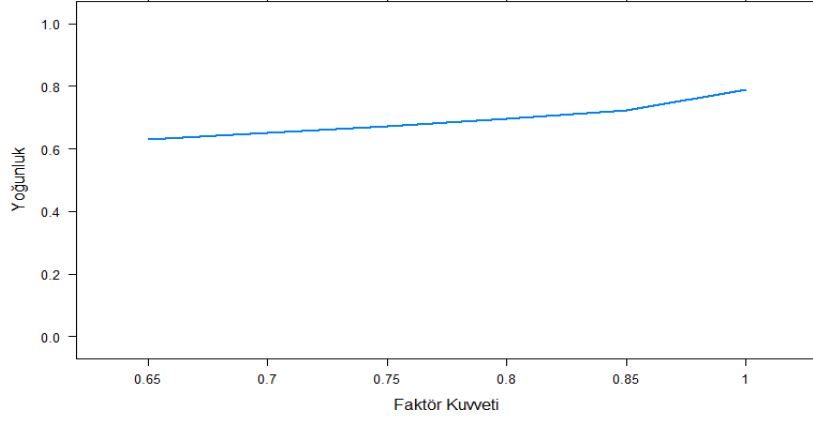
Kestirim Yöntemi	RMSE	PRMSE
Alt Puan	$\sigma_{\tau_x} [1 - \rho^2(S_x, \tau_x)]^{1/2}$	$\rho^2(S_x, \tau_x)$
Toplam Puan	$\sigma_{\tau_x} [1 - \rho^2(S_z, \tau_x)]^{1/2}$	$\rho^2(S_z, \tau_x)$
Toplam Puan ve Alt Puan	$\sigma_{\tau_x} [1 - \rho^2(S_x, \tau_x)]^{1/2} [1 - \rho^2(S_z, \tau_x \cdot S_x)]^{1/2}$	$1 - [1 - \rho^2(S_x, \tau_x)] [1 - \rho^2(S_z, \tau_x \cdot S_x)]$

Tablo 2 incelendiğinde, gözlenen alt test puanı kullanılarak elde edilen alt test puanının S_x , toplam test puanı kullanılarak elde edilen alt test puanının S_z , hem gözlenen alt test puanı hem de toplam test puanı kullanılarak elde edilen alt test puanının ise S_{xz} ile ifade edildiği görülmektedir. Bu puanların PRMSE değerleri ise sırasıyla $PRMSE_x$, $PRMSE_z$ ve $PRMSE_{xz}$ ile gösterilmek üzere; $PRMSE_x > PRMSE_z$ olduğu durumda alt test puanının artı değere sahip olduğu görülmektedir. Böylelikle, bu durum, alt test puanının kullanılabilir olduğunu ve alt test puanının toplam test puanına göre daha çok bilgi verdiğini ve alt test puanı ile profil oluşturulabileceğini göstermektedir. $PRMSE_{xz} > PRMSE_x$ ve $PRMSE_{xz} > PRMSE_z$ olduğu durumda ise toplam test puanı ve gözlenen alt test puanından elde edilen puan, artı değer özelliği taşımaktadır. Bu puana, özel olarak genişletilmiş puan adı verilmektedir (Haberman, 2008). Genişletilmiş puan, alt test puanından daha çok bilgi içermekle birlikte test formları arasındaki karşılaştırmayı kolaylaştırmaktadır (Sinharay ve Haberman, 2011).

Çalışmada, alt testlerin artı değer özelliğine sahip olma durumlarının çeşitli koşullara göre incelenmesi için alt test puanlarının, bu alt test puanlarının toplamıyla elde edilen toplam test puanlarının ve alt test puanlarıyla toplam test puanlarının kullanılarak elde edilen genişletilmiş puanların PRMSE değerleri hesaplanmıştır. Elde edilen üç ayrı üretilmiş puan için PRMSE değerleri karşılaştırılmıştır. Alt testin artı değer özelliği göstermesi için, alt testten elde edilen alt test puanının PRMSE değerinin toplam test puanı kullanılarak elde edilen puanın PRMSE değerinden büyük olması gerekmektedir. Bu işlem, her bir test formu için 100 kez tekrarlanarak artı değer özelliği ile karşılaştırma frekansı elde edilmiştir. Verilerin analizinde, R 3.4.0 programı kullanılmıştır. Alt test grup örneklem büyüklüğü, alt testlerin faktör kuvveti, alt test uzunluğu ve alt testler arasındaki korelasyon düzeyleri koşullarının artı değer özelliğine etkisini incelemek için temel etki grafikleri ve koşulların ortak etkisini gözlemek için ortak etki grafiklerinden yararlanılmıştır.

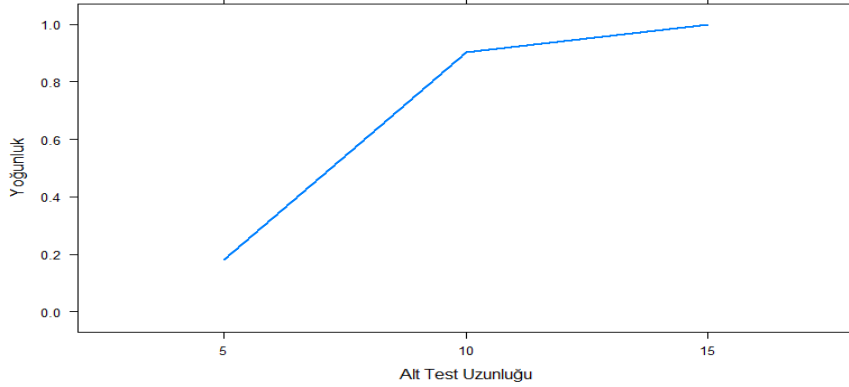
Bulgular

Çalışmada, alt testlerin faktör kuvveti 0.65, 0.70, 0.75, 0.80, 0.85 ve 1 olacak şekilde altı farklı düzeyde değişimlenmiştir. Şekil 1’de, alt testlerin faktör kuvvetinin alt testlerin artı değer özelliğine olan temel etkilerine yönelik grafik gösterilmektedir.



Şekil 1. Alt Testlerin Faktör Kuvvetinin Alt Testlerin Artı Değer Özelliğine Olan Temel Etki Grafiği

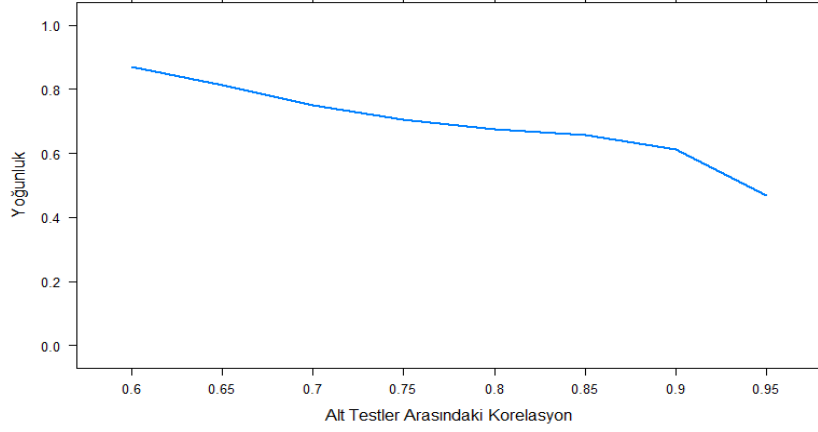
Şekil 1’de, alt testlerin faktör kuvvetinin değişimlenmesi sonucunda alt test çiftlerinin artı değer özelliği taşıma sıklığına ilişkin elde edilen bulgular görülmektedir. Şekil 1 incelendiğinde, alt testlerin faktör kuvveti 0.65 olduğu durumda, her iki alt testte de artı değer gözlenme sıklığının 63 olduğu, faktör yükü 1 olduğunda ise sıklığın 78’e çıktığı gözlenmektedir. Böylelikle, alt testlerin faktör kuvveti arttıkça her iki alt testinde artı değer özelliği taşıma sıklıklarının arttığı görülmektedir. Şekil 2’de, alt testte yer alan madde sayılarındaki değişimin artı değer özelliğine olan temel etkisine yönelik grafik gösterilmektedir.



Şekil 2. Test Uzunluğunun Alt Testlerin Artı Değer Özelliğine Olan Temel Etki Grafiği

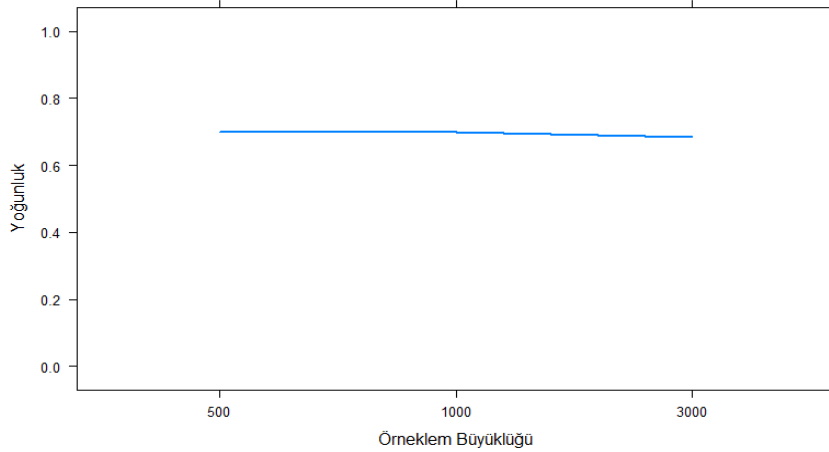
Şekil 2’de, alt test uzunluğunun değişmesi ile, alt testlerin artı değer özelliği taşıma sıklığına ilişkin bulgular görülmektedir. Şekil 2 incelendiğinde, ikinci faktörde yer alan madde sayısının 5 olduğu durumda, her iki alt testte de artı değer gözlenme sıklığı 18 iken madde sayısının

15 olduğu durumda bu sıklığın 99'a çıktığı görülmektedir. Böylelikle, alt testte yer alan madde sayısı arttıkça alt testlerin her ikisinde de artı değer özelliği gösterme sıklığının arttığı görülmektedir. Şekil 3'te alt testler arasındaki korelasyon düzeyinin alt testlerin artı değer özelliği gösterme durumuna olan temel etkisine yönelik grafik gösterilmektedir.



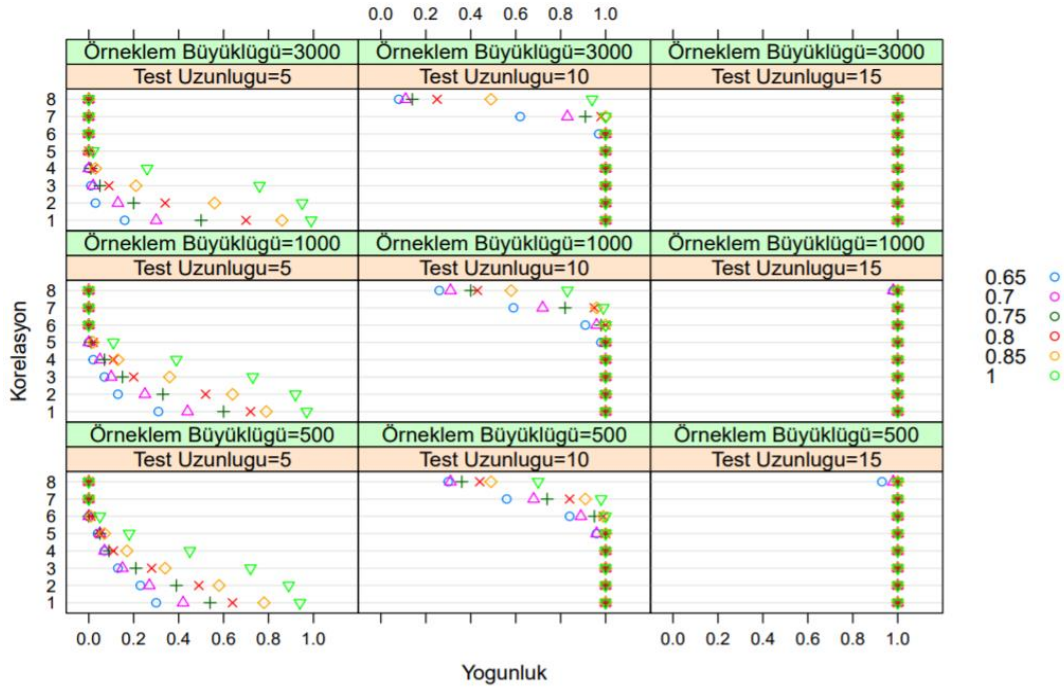
Şekil 3. Alt Testler Arasındaki Korelasyonun Artı Değer Özelliğine Olan Temel Etki Grafiği

Şekil 3'de, alt testler arasındaki farklı korelasyon düzeylerinde, alt testlerin artı değer özelliği gösterme sıklığına ilişkin bulgular görülmektedir. Şekil 3 incelendiğinde, alt testler arasındaki korelasyon düzeyinin 0.60 olduğu durumda, artı değer özelliğine sahip alt test çiftlerinin sayısının 86 olduğu, alt testler arasında korelasyon düzeyinin 0.95 olduğu durumda ise alt test çiftlerinin artı değer özelliği gösterme sıklığının 46 olduğu görülmektedir. Böylelikle, alt testler arasındaki korelasyon düzeyi arttıkça alt testlerin artı değer özelliği gösterme sıklıklarının azaldığı görülmektedir. Şekil 4'de, alt test grup örneklem büyüklüğünün alt testlerin artı değer özelliğine sahip olma durumlarına olan temel etkisine yönelik grafik gösterilmektedir.



Şekil 4. Alt Test Grup Örneklem Büyüklüğünün Artı Değere Özelliğine Olan Temel Etki Grafiği

Şekil 4' de, değişimlenen alt test grup örneklem büyüklüklerine göre her iki alt testin de artı değer özelliği taşıma sıklığına yönelik bulgular görülmektedir. Şekil 4 incelendiğinde, alt testlerde yer alan grupların örneklem büyüklüğü değiştikçe alt testlerin artı değer özelliği taşıma durumları arasında bir farklılık oluşmadığı görülmektedir. Şekil 5'de, çalışmada değişimlenen koşulların, alt testlerin artı değer özelliği gösterme durumuna olan ortak etkisine yönelik grafik gösterilmektedir.



Şekil 5. Değişimlenen Koşulların Artı Değer Özelliğine Olan Ortak Etki Grafiği

Şekil 5 incelendiğinde, ikinci alt testte yer alan madde sayılarının artmasıyla birlikte her iki alt testinde değişimlenen tüm faktör ve faktör düzeylerinde artı değer özelliğine sahip olma sıklığının arttığı görülmektedir. Benzer olarak, a2 parametresinin değeri arttıkça da değişimlenen faktörler ve düzeyleri altında her iki alt testin artı değer özelliğine sahip olma sıklığının arttığı görülmektedir. Bununla birlikte, alt testler arasındaki korelasyon miktarı arttıkça özellikle ikinci alt test uzunluğunun 5 ve 10 olduğu durumlarda, alt testlerin artı değer özelliğine sahip olma sıklığının azaldığı gözlenmektedir. İkinci alt test uzunluğunun 15 olduğu durumda, genel olarak tüm şartlar altında, artı değer özelliğine sahip alt testlerle karşılaşma oranının 1'e yakın olduğu görülmektedir. Test uzunluğunun 15, alt test grup örneklem büyüklüğünün 500, alt testler arasındaki korelasyonun 0.95 ve a2 parametre değerinin 0.65 olduğu koşulda artı değer özelliğine sahip alt testlerle karşılaşma oranının diğer a2 parametre değerlerine göre daha az olduğu gözlenmektedir.

Tartışma ve Sonuç

Alt testler, bireylerin ya da okulların profillerinin oluşturulmasında, bireylerin öğrenim düzeylerinin takip edilmesinde, bireylerin bir kuruma seçilip yerleştirilmesinde ve bireylerin işe alımında kullanılmaktadır. Alt test puanlarının kullanımı ise, alt testlerin artı değer özelliği gösterdiği koşullarda uygundur (Sinharay ve diğer., 2007; Haberman, 2008; Sinharay, 2010a; Sinharay, 2010b, Sinharay ve Haberman, 2011). Dolayısıyla, alt test puanlarının rapor edilebilir ya da tek başlarına kullanılabilir olması için alt testlerin artı değer özelliği göstermesi gerekmektedir. Çalışmada, alt test grup örneklem büyüklüğü, alt test uzunluğu, alt testler arasındaki korelasyon düzeyi, ve alt testlerin faktör kuvveti değişkenlerinin artı değer özelliği üzerindeki temel etkileri ve ortak etkisi incelenmiştir.

Çalışmada değişimlenen koşulların temel etkileri bulgularına göre alt test uzunluğu ve alt testlerin faktör kuvveti değerlerinin artışı ile birlikte alt testlerin artı değer özelliği gösterme durumunun arttığı görülmüştür. Haberman (2008), Sinharay (2010a, 2010b), Sinharay ve diğer (2007), çalışmalarında, alt test uzunluğunun artışı ile birlikte artı değer özelliği gösteren testlerin sayısında artış olduğunu belirtmişlerdir. Bu durum, alt test uzunluğunun artışı alt testten elde edilen puanların güvenilirliğinin artmasına ve alt test faktör kuvvetinin artışının ise ilgili boyutu temsil eden maddelerin daha ayırt edici olmasına neden olmaktadır. Böylelikle, alt test puanının tek başına, alt testin ilgili özelliği ile ilgili daha çok bilgi verdiği görülmektedir. Bu durumdaki alt test puanının artı değer özelliği incelendiğinde, artı değer özelliğinin daha yaygın görüldüğüne ulaşılmaktadır. Alt testler arasındaki korelasyon düzeylerinin artışı ise, artı değer özelliği taşıyan alt test sayısında azalmaya neden olmuştur. Sinharay (2010a, 2010b) çalışmasında, alt testler arasındaki korelasyon düzeyinin artması ile birlikte testlerin artı değer özelliği gösterme durumlarında azalma meydana geldiğini belirtmiştir. Böylelikle, alt testlerin testin toplamı ile daha ilişkili yani ilgili özelliğın toplam test puanı ile rapor edilmesi gerektiği sonucu ortaya çıkmaktadır. Başka bir ifadeyle, alt testin alt alanını bir boyut olarak temsil etmediği, dolayısıyla, alt test puanının rapor edilebilir özellikte olmadığı söylenebilir. Alt test grup örneklem büyüklüğünün, testin artı değer özelliği taşıyıp taşıyamaması üzerine bir etkisi görülmemiştir.

İkinci alt test uzunluğunun 15 olduğu durumda, her iki alt testin de artı değer özelliği taşıma durumunun bütün koşullar altında genel olarak 1'e yakın değerler aldığı görülmektedir. Ayrıca, alt testlerin faktör kuvvetinin 1 ve alt testler arasındaki korelasyonun en çok 0.65 olduğu durumlar da ise tüm şartlar altında, her iki alt testin artı değer özelliği taşıma sıklıklarının % 80'in üzerinde olduğu bulunmuştur. Çalışmada ele alınan değişkenlerin ortak etkilerinde ise artı değer özelliği taşıyan alt test çiftleri sayısında, alt test uzunluğunun ve alt test faktör kuvvetinde meydana gelen artma ile alt testler arasındaki korelasyon miktarındaki azalma ile artış olduğu görülmektedir.

Öneriler

Bu çalışmada, diğer çalışmalardan farklı olarak gerçekleştirilen alt testlerin faktör kuvvetlerinin de değişimlenmesidir. Bu değişimleme sonucunda, alt test faktör kuvvetinin artması ile birlikte, artı değer özelliği taşıyan alt test sayısının sıklığının arttığı görülmüştür. Böylelikle, araştırmada simülasyon düzeneğinde yer alan değişkenlerin düzeyleri farklı değişimlenerek çalışma gerçekleştirilebilir. Ayrıca, alt test puanlarının rapor edilebilir olması ile ilgili farklı analizler kullanılarak karşılaştırma yapılabilir. Gerçekleştirilen bu çalışmada, artı değer özelliği analizi Haberman'ın (2008) yöntemine dayanmaktadır. Alt testin psikometrik kalitesinin göstergesi bir başka ifade ile artı değer özelliği taşıyıp taşıyamama durumu faktör analitik yöntemler ya da Yen OPI gibi analizlerle incelenebilir (Haberman ve Sinharay, 2010; Reckase ve Xu, 2015; Sawaki ve Sinharay, 2017; Wedman ve Lyren, 2015; Weeks, 2018).

Kaynakça

- Ackerman, T. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20(4), 311–329. doi:10.1177/014662169602000402
- Akkuş, O., & Baykul, Y. (2001). Çoktan seçmeli test maddelerini puanlamada, seçenekleri farklı biçimlerde ağırlıklandırmanın madde ve test istatistiklerine olan etkisinin incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 9 – 15. <https://dergipark.org.tr/tr/download/article-file/87948> . (Erişim Tarihi: 05.04.2018)
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Erkuş, A. (2012). Var olan ölçek geliştirme yöntemleri ve ölçme kuramları psikolojik ölçek geliştirmede ne kadar işlevsel: yeni bir öneri. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(2), 279–290. <https://dergipark.org.tr/tr/download/article-file/65963> .(Erişim Tarihi: 02.04.2018)
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* . New York: McGraw-Hill
- Haberman, S. J. (2008). When can subscores have value?. *Journal of Educational and Behavioral Statistics*, 33, 204–229. doi:10.3102/1076998607302636
- Haberman, S. J., & Sinharay, S. (2010). Reporting of subscores using multidimensional item response theory. *Psychometrika*, 75(2), 209–227. doi:10.1007/s11336-010-9158-4
- Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (2016). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101–125. doi:10.1177/014662169602000201
- McDonald, R. P. (1999). *Test theory: A unified treatment*. New Jersey: Lawrence Erlbaum Assoc.
- Reckase, M. D., & Xu, J.R. (2015). The evidence for a subscore structure in a test of english language competency for english language learners. *Educational and Psychological Measurement*, 75(5), 805– 825. doi:10.1177/0013164414554416
- Sawaki, Y., & Sinharay, S. (2017). Do the TOEFL iBT® section scores provide value-added information to stakeholders?. *Language Testing*, 1-28. doi:10.1177/0265532217716731
- Sinharay, S., Haberman, S., & Puhan, G. (2007). Subscores based on classical test theory: to report or not to report. *Educational Measurement: Issues and Practice*, 26(4), 21–28. doi:10.1111/j.1745-3992.2007.00105.x
- Sinharay, S., & Haberman, S. J. (2011). Equating of augmented subscores. *Journal of Educational Measurement*, 48(2), 122–145. doi:10.1111/j.1745-3984.2011.00137.x
- Sinharay, S. (2010a). *When can subscores be expected to have added value? Results From Operational and Simulated Data*. (ETS Research Report No RR-10-16, August). <https://www.ets.org/Media/Research/pdf/RR-10-16.pdf>
- Sinharay, S. (2010b). How often do subscores have added value? results from operational and simulated data. *Journal of Educational Measurement*, 47(2), 150–174. doi:10.1111/j.1745-3984.2010.00106.x
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, madde tepki kuramı ve çok boyutlu madde tepki kuramı çerçevesinde incelenmesi*. Yayınlanmamış Doktora Tezi, Mersin Üniversitesi, Mersin.
- Uçar, A. (2016). *Artı değer özelliğine sahip alt testlerde eşitleme yöntemlerinin karşılaştırılması*.

- Yayınlanmamış Yüksek Lisans Tezi, Mersin Üniversitesi, Mersin.
- Walker, C. M., Azen, R., & Schmitt, T. (2006). Statistical versus substantive dimensionality: the effect of distributional differences on dimensionality assessment using DIMTEST. *Educational and Psychological Measurement*, 66, 721-738.
- Wedman, J., & Lyren, P.E. (2015). Methods for examining the psychometric quality of subscores: A review and application. *Practical Assessment, Research and Evaluation*, 20(21), 1-14. <https://scholarworks.umass.edu/pare/vol20/iss1/21/>. (Erişim Tarihi: 05.04.2018)
- Weeks, J. P. (2018). An application of multidimensional vertical scaling. *Measurement: Interdisciplinary Research and Perspectives*, 16(3), 139–154. doi:10.1080/15366367.2018.1502005.
- Yüce, G. (2012). *Seviye belirleme alt testlerinin önsel (a priori) ve sonsal (a posteriori) olarak belirlenmesinin bireylerin sınıflama ve sıralamaları açısından incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Mersin Üniversitesi, Mersin.

EXTENDED ABSTRACT

Introduction:

Subtest scores are used to determine the learning deficiencies of the individual, to analyze the subjects in which she/he is more successful, or to reveal the profiles of the schools. Thus, with the combination of these items, a new subfield emerges and more detailed information about the structure of the test can be obtained (Sinharay, Haberman, & Puan, 2007). With the increasing interest in the use of subtest scores, studies were conducted on the quality of the subtest score (Haberman, 2008; Sinharay et al., 2007; Sinharay, 2010a; Sinharay, 2010b; Sinharay & Haberman, 2011). As a result of the researches on which of the total test and subtest scores would serve the purpose more, it was stated that each subtest score may not be suitable for use, and the conditions required for its use, albeit partially. In order for the subtest score to be usable, it was emphasized that the score should have an added value. For having added value, features such as high reliability of subtest scores, relatively lower reliability of the total test, and different subtests are considered (Sinharay, 2010a, 2010b). Although there is no clear framework in the literature about the conditions under which subtests have added value, it has been found that the subtests have added value with the low correlation between the subtests and the increase in the length of the subtests (Sinharay et al., 2007; Sinharay, 2010a, 2010b). In this study, the subtests of the tests consisting of different subdomains, test length, the correlation between the subtests, and the conditions under which the subtests change in different levels of the factor strengths have been investigated. It is thought that the findings obtained will have a unique value and will contribute to the field of measurement and evaluation.

Method:

In this study, under various factors (group sample size of subtest, subtest length, correlation between subtests, and factor strengths of subtests) the frequency of subtests that are added value was examined. Simulation study was conducted to investigate of subtests which have added value feature under different conditions. In this framework, this research can be evaluated as a simulation study in the nature of descriptive research. The variables and levels of the study were determined on the basis of the literature. In line with the purpose of the study, by keeping the parameters of the 15 items in the first dimension constant for all generation conditions, in the second dimension, under the conditions of test length (5, 10, 15), factor strength of the items (0.65, 0.70, 0.75, 0.80, 0.85, 1.0), two-dimensional simple structured and two-dimensional 1800 datasets were generated with 100 replication based on two categorical complementary item response theory. In order to give an angle to the item vectors, 432 different experimental structures were created by changing

the conditions of correlation between subtests (0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90 ,0.95) and group sample size of subtest (500, 1000, 3000). The analysis of whether the subtests have added value features was carried out according to Haberman's method which the subtest scores, the total test scores to be obtained by the total of these subtest scores, and the extended scores obtained by using the subtest scores and the total test scores were produced and Proportional Reduction in MSE (PRMSE) values of these scores were calculated and PRMSE values of the three scores generated separately were compared. R.4.0.1 software was used for data generation and analyses.

Results:

As a result of the study, it was observed that the subtests showed more added value with the increasing the subtest length and the amount of subtest factor strength, but with the increasing the level of correlation between the subtests, the frequency of having added value feature of the subtests decreased. It was concluded that the variation of the group sample size of subtest did not show a difference in the frequency of carrying the added value feature in the subtests.

Discussion and Conclusion:

According to the findings of the main effects of the factors, it was seen that the subtests showing added value feature increased as the subtest length and factor strength values of the subtest increased. Haberman (2008), Sinharay (2010a, 2010b), Sinharay et al. (2007) stated in their study that the number of tests showing the added value of the test increased with the increase in the length of the subtest. This situation causes the reliability of the subtest to increase as the length of the subtest increases also the increase in the dimension factor strength causes the items representing the relevant dimension to be more distinctive. Therefore, it is seen that the subtest score alone gives more information about the related trait of the subtest. When the subtest that has added value in this situation is examined, it is found that added value is more common in these subtests. The increase in the level of correlation between subtests caused a decrease in the number of tests with added value. Sinharay (2010a, 2010b) stated in his study that with the increase in the level of correlation, there was a decrease in the cases of the tests having added value. Therefore, the result is that the subtests are more related to the total of the test, that is, the related trait should be reported with the total test score. In other words, it can be said that the subtest score is not reportable because it does not represent the subdomain of the subtest as a dimension. Although the sample size does not seem to have an effect on whether the test has added value or not, the added value analysis performed is based on Haberman's (2008) Classical Test Theory. In other words, if the psychometric quality indicator of the subtest is examined with factor analytical methods or analyzes such as Yen OPI, it is necessary to examine the effect of the sample size variable on added value (Haberman & Sinharay, 2010; Reckase & Xu, 2015; Sawaki & Sinharay, 2017; Wedman & Lyren, 2015 Weeks; 2018).