

PIRLS 2001 TESTİNİN YAPI GEÇERLİLİĞİNİN ÜLKELERARASI KARŞILAŞTIRILMASI*

Araş. Gör. Dr. Murat AKYILDIZ
Ankara Üniversitesi Eğitim Bilimleri Fakültesi
Ölçme ve Değerlendirme ABD

ÖZET

Bu çalışmada, PIRLS 2001 başarı testinin, 35 ülkede yapılmış olan uygulamasından elde edilen faktör yapılarının birbirine denkliği çok gruplu doğrulayıcı faktör analizi ile incelenmiştir. PIRLS 2001 başarı testinin faktör yapısının 35 ülkede değişmezliği tüm ülkelerde uygulanmış olan dört ayrı alt ölçek bağlamında ayrı ayrı test edilmiştir. Elde edilen bulgular PIRLS 2001 testinin faktör yapısının uygulamaya katılan tüm ülkelerde birbirine orta düzeyde denk sayılabileceğini göstermiştir.

Anahtar Kelimeler: Yapı geçerliliği, Ölçme Değişmezliği, PIRLS

THE COMPARISON OF CONSTRUCT VALIDITIES OF THE PIRLS 2001 TEST BETWEEN COUNTRIES

ABSTRACT

In this work, the equivalence of factor structures obtained from the application of PIRLS 2001 test in 35 countries is analyzed via multigroup confirmatory factor analysis. The invariance of the factor structures of PIRLS 2001 test in 35 countries is tested in the level of the four subscales which are applied in all countries. The findings have shown that, the factor structure of PIRLS 2001 test is equivalent in all participant countries

Key Words: Construct Validity, Measurement Equivalence, PIRLS

* Bu makale, Araştırmacı tarafından Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Bilim Dalında Temmuz 2009 tarihinde kabul edilen Doktora tezinden uyarlanmıştır.

GİRİŞ

Uluslararası düzeyde 4. sınıf öğrencilerinin okuma becerileri ne düzeydedir? Bu öğrencilerin okuma alışkanlıkları ve okumaya yönelik tutumları bir ülkeden diğer ülkeye nasıl farklılıklar göstermektedir? Bu sorulara yanıt bulabilmek amacıyla, eğitim başarılarını karşılaştırmalı olarak incelemek için kurulmuş olan Boston College Uluslararası Çalışma Merkezi'nin desenlediği bir proje doğrultusunda uygulamaları yürüten ve merkezi Amsterdam'da bulunan Uluslararası Eğitim Başarısını Değerlendirme Kuruluşu (International Association For The Evaluation Of Educational Achievement, IEA) tarafından geliştirilen bir ölçme aracı, 2001 yılında Uluslararası Okuma Becerilerinde Gelişim Çalışması (Progress in International Reading Literacy Study, PIRLS) adıyla 35 ülkede uygulanmıştır (Mullis, Martin, Gonzalez, Ann, 2003). Uygulama çerçevesinde amaçlanan,

- a) 9 yaş grubu (4. sınıf) öğrencilerinin okuma becerileri ve okuma alışkanlıklarını,
 - b) Öğretmenlerin öğrencilere okuma becerisini kazandırmak için uyguladıkları öğretim metotları ve öğretim materyallerinin yeterli olup olmadığını,
 - c) Öğrencilerin okuma becerisini kazanmalarında ailelerinin katkılarını,
- uluslararası standart test ve anketlerle belirlemektir.

Uygulamaya Türkiye de dahil olmak üzere Almanya, Amerika Birleşik Devletleri, Arjantin, Belize, Bulgaristan, Çek Cumhuriyeti, Fas, Fransa, Hollanda, Hong Kong, İngiltere, İran, İskoçya, İsrail, İsveç, İtalya, İzlanda, Kanada, Kıbrıs Rum Kesimi, Kolombiya, Kuveyt, Letonya, Litvanya, Macaristan, Makedonya, Moldova, Norveç, Romanya, Rusya Federasyonu, Singapur, Slovakya, Slovenya, Yeni Zelanda, ve Yunanistan, katılmıştır.

PIRLS çalışması içinde okuma becerisi aşağıda belirtilen üç boyutta ele alınmıştır.

- a) Kavrama süreçleri
- b) Okuma amaçları
- c) Okuma alışkanlıkları ve okumaya yönelik tutumlar

Bu boyutlar çerçevesinde bilgi toplamaya çalışan PIRLS çalışmasının temel odak noktası kavrama süreçleri ve okuma amaçlarıdır. Okuma alışkanlıkları ve

okumaya yönelik tutumlar boyutu okuma becerisini belirlediği düşünülen dışsal değişkenler olarak değerlendirilmiştir.

PIRLS'ün okuma becerisini belirlemek için geliştirilmiş olan testi içindeki kavrama süreçleri ve okuma amaçları bölümlerinin dağılımı Tablo 1'de gösterilmiştir.

Tablo 1. PIRLS Okuma Becerileri Test Planı

Kavrama Süreçleri	Süreç İçindeki Yüzde	Toplam İçindeki Yüzde
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	%20	%50
Doğrudan çıkarımlar yapma	%30	
Fikir ve bilgileri birleştirme ve Yorumlama	%30	
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	%20	
Okuma Amaçları		
Okuma deneyimi	%50	%50
Bilgiyi elde etme ve kullanabilme	%50	

PIRLS maddelerinin okuma amaçlarına göre bölünmüş test içindeki dağılımı ise Tablo 2'de gösterilmiştir.

Tablo 2. Okuma Amaçlarına ve Süreçlerine Göre Bölünmüş PIRLS Değerlendirme Yüzdeleri

	Okuma Amaçları	
	Okuma Deneyimi	Bilgi Edinme
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	%9	%13
Doğrudan çıkarımlar yapma	%14	%9
Fikir ve bilgileri birleştirme ve Yorumlama	%20	%20
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	%6	%8

Tablo 2'de gösterilmiş dağılıma uygun olarak test geliştiriciler, geçerli ve güvenilir bir ölçme yapabilmek için en az sekiz okuma parçasına ihtiyaç olduğuna karar vermişler ve bu doğrultuda her bir okuma amacına dört okuma parçası düşecek şekilde sekiz metin yazmışlardır.

Geliştirilen sekiz uzun metnin okunması ve bu metinlerin sorularına yanıt verilmesi tek bir öğrenci için mümkün olmadığından uygulama matris örnekleme tekniği denen bir yöntemle yapılmıştır. Bu yöntemle göre, geliştirilen test materyali kırk dakikada yanıtlanabilecek ve maddeleri en az 15 puanlık bilgi sağlayabilecek sekiz blok oluşturacak şekilde bölünmüştür. Her bir okuma amacı için dört blok bulunmaktadır. İlk blok, okuma deneyimi elde etmek için okuma (Literary experience) bloğu olduğundan bu bloğa L bloğu denmiş ve dört blok L1'den L4'e doğru isimlendirilmiştir. Bilgi edinmek için okuma amacı (Information) bloğu da benzer şekilde I1'den I4'e kadar isimlendirilmiştir. Bu bloklar daha sonra on farklı kitapçık oluşturacak şekilde birleştirilmiştir. Böylece seksen dakikada tamamlanabilecek on farklı test kitapçığı (alt test) elde edilmiştir. Tablo 3, anlatılan düzenlemeyi göstermektedir. Bu on farklı test kitapçığının hepsi tüm ülkelerde uygulanmamıştır. Tüm ülkelerde ortak olarak uygulanan kitapçıklar sıfır, üç, altı ve sekiz numaralı kitapçıklardır (Martin ve diğerleri, 2003).

Tablo 3. PIRLS Blokları ve Kitapçık Düzenleri

Kitapçık Sırası	Kitapçık 1	Kitapçık 2	Kitapçık 3	Kitapçık 4	Kitapçık 5	Kitapçık 6	Kitapçık 7	Kitapçık 8	Kitapçık 9	Kitapçık 10
Kitapçık numarası	Sıfır	Bir	İki	Üç	Dört	Beş	Altı	Yedi	Sekiz	Dokuz
	L1	L2	L3	I1	I2	I3	L1	I2	I3	L4
	L2	L3	I1	I2	I3	L1	I1	L2	L3	I4

PIRSL 2001 testinin tümünde 98 madde bulunmaktadır. Bu 98 madde her bir kitapçığa ortalama 24 madde düşecek şekilde bölünmüştür. Aynı sorular birkaç kitapçıkta aynı anda yer alabilmiştir. Dolayısı ile her kitapçıkta ortak maddelerin bulunması sağlanmıştır. PIRLS 2001 uygulamasında kullanılan maddelerin bir kısmı çoktan seçmeli bir kısmı ise yapılandırılmış tepki (constructed-response) maddeleridir. Çoktan seçmeli maddelerin her birisi 1 puan değerindeyken yapılandırılmış tepki maddelerin puanları ise sorulan metnin hangi derecede kavrandığına bağlı olarak 1, 2 ya da 3 farklı puan değerinde olabilmektedir. PIRLS uygulamasındaki okuma becerisi sorularının dağılımı ve bu çalışmada tüm ülkelerde ortak olduğu için kullanılacak olan

sıfır, üç, altı ve sekiz numaralı alt testlerin sorularının tür olarak dağılımı Tablo 4'te gösterilmiştir.

Tablo 4. Madde Türüne ve Okuma Amacına Göre Maddelerin Test İçindeki Dağılımı

	Çoktan Seçmeli Maddeler	Yapılandırılmış Tepki Maddeleri			Toplam Madde Sayısı	Alınabilecek Toplam Puan
		1 Puan	2 Puan	3 Puan		
Tüm Test						
Okuma Amacıyla	25	14	9	3	51	66
Bilgi edinme Amacıyla	21	10	12	4	47	67
Toplam	46	24	21	7	98	133
Alt Test						
Sıfır	13	4	6	1	24	32
Üç	11	6	6	1	24	32
Altı	10	8	4	3	25	35
Sekiz	12	6	5	2	25	34

Tablo 4'te görüldüğü gibi, PIRLS 2001 uygulamasında kullanılan testin içinde yer alan çoktan seçmeli maddelerin sayısı (46) ile yapılandırılmış tepki maddelerinin sayısı (52) birbirine yakın olmasına rağmen yapılandırılmış tepki maddeleri daha çok sayıdadır. Tüm alt testlerde birden çok puanlamanın mümkün olduğu soruların sayısı yedidir. Fakat her alt testte özellikle üç farklı cevabın doğru kabul edildiği ve her bir doğruya farklı puan verildiği maddelerin sayısı her alt testte farklıdır. Farklı cevaplara üç farklı puanın verildiği maddelerin sayısı üç ile en çok altıncı testtedir. Ayrıca çoktan seçmeli madde sayısının oran olarak en az olduğu test de altı numaralı alt testtir.

Yapı Geçerliliği ve Ölçme Değişmezliği

Yapı kavramı, psikolojik özelliklerin ölçülmesi söz konusu olduğunda ölçülmek istenilen değişkeni ifade etmektedir (Cronbach, Meehl, 1955). Bu değişken kendi içinde tek boyutlu bir bütünlük taşıyabildiği gibi bir kaç alt boyuttan oluşabilmektedir (Brown, 2006). Örneğin umutsuzluk değişkeni kendi içinde tek boyutlu geleceğe dönük olumsuz beklentileri içeren bir yapıdır. Depresyon ise bireyin kendi iç dünyasında, kişilerarası ilişkilerinde ve dünya ile olan ilişkilerinde genel çöküntüyü ifade eden üç boyutlu bir yapıdır.

Yapı geçerliliğini test etmek amacıyla bazı yöntemler önerilmiştir (Cronbach ve Meehl, 1955; Baykul, 2000; Rodriguez, 2006). Yapı geçerliliği için düşük, yüksek şeklinde yorum yapmak yerine sayısal kanıtlarla çözümlene yapma arayışı yapı geçerliliği için istatistiksel bir teknik olan ve sayısal sonuçlar veren faktör analizinin kullanımını daha yaygın hale getirmiştir.

Yapı geçerliliği için elde edilen kanıtlar, uygulamanın yapıldığı grup ve benzer özellikleri taşıyan gruplar ile sınırlıdır. Birden fazla grup söz konusu olduğunda her grup için yapı geçerliliği için ayrıca kanıt toplanmalıdır. (Kline, 1998). Her grubun yapı geçerliliği kanıtları, bu grupların hepsinde geçerli olduğu düşünülen (beklenen) faktör yapısıyla uyum göstermelidir. Aynı ölçme aracının uygulandığı iki grupta faktör yapılarının birbirinden farklı olması, bu ölçme aracının her grupta farklı bir psikolojik özelliği ölçtüğünü gösterir. Böyle bir durumda aynı ölçme aracıyla ölçme yapıyor olmasına rağmen iki gruptan elde edilen sonuçların aynı anlama gelmediği yorumu yapılır.

Zaman içinde aynı gruptan alınmış farklı ölçümlerin ya da aynı zamanda farklı gruplardan alınmış ölçümlerin birbirine denkliğinden emin olunması, özellikle test geliştirme sürecinin önemli yönlerinden birisidir. Bir test heterojen bir gruba uygulandıysa, bu grup içinde yer alan alt gruplarda ölçme işlemine ait özelliklerin birbirine eşit olduğunun gösterilmesi gerekir. Bir ya da birkaç test maddesinin bir ya da birkaç alt grupta ölçmesi gereken psikolojik özellik yerine başka bir özelliği ölçmesi test yanlılığı olarak adlandırılır (Brown, 2006). Örneğin bir testi alan kadın ve erkek katılımcılardan elde edilen madde puanlarının aynı bilişsel özelliğe ilişkin olduğunun gösterilmesi gerekir. Bir maddenin kadınların ya da erkeklerin lehine ölçme yapması o test maddesinin yanlılığına ilişkin bilgi verir. Bu tür durumlarda testi alan grubun içindeki alt gruplarda ölçme modeli her grup için ayrıca test edilir. Örneğin tüm alt gruplarda maddelerin faktör yükleri birbirine eşit çıkarsa maddeler ile o maddelerin ölçmeye çalıştığı gizil değişken arasındaki ilişkinin tüm gruplarda birbirine denk olduğu söylenir. Bu ifade aynı zamanda o maddelerin tüm alt gruplarda aynı psikolojik özelliği aynı güçle ölçtüğünün de kanıtı olarak değerlendirilir. Bu belirlemeyi yapabilmek için önerilen teknik ise çok gruplu doğrulayıcı faktör analizidir (DFA). Özellikle psikolojik testlerin geliştirilmesi aşamasında testin sahip olması istenen psikometrik niteliklerinin testin uygulanacağı tüm gruplarda eşit olduğunun (değişmezliğinin) gösterilmesi için

çok gruplu DFA sıklıkla kullanılır (Thompson, 2004). Örneğin cinsiyete, etnik kökene, yaşa ya da olabilecek başka gruplayıcı değişkenlere bağlı olarak ölçme sonuçlarının birbirine denk olup olmadığı çok gruplu DFA ile test edilir. Ölçme değişmezliğinin sağlanması aynı zamanda ölçülen psikolojik değişkenin çeşitli gruplara genellenebilirliği açısından da önemlidir (Raykov ve Marcoulides, 2006). Örneğin bir psikiyatrik rahatsızlığın altında yatan psikolojik yapının farklı cinsiyetler, farklı ülkeler arasında aynı şekilde işlev gördüğünün kabul edilebilip edilemeyeceğine karar verirken ölçme değişmezliğinin sağlanmış olmasına dikkat edilir.

Çok gruplu DFA, bir DFA modelinin eş zamanlı olarak birden çok grupta analiz edilmesini içerir (Brown, 2006). Örneğin analiz iki grubu içeriyorsa iki ayrı kovaryans matrisi (her grubun kendisine ait kovaryans matrisi) kullanılır. Böylece iki grubun kovaryans matrislerinin birbirine denk sayılıp sayılmayacağı test edilir. Bu analiz aynı zamanda testle ölçülmüş olan psikolojik yapının testi alan tüm gruplarda birbirine denk olup olmadığını bir başka deyişle testin testi alan alt grupların hepsinde aynı psikolojik yapının ölçülebilme derecesini gösterir.

Uygulamaya katılan tüm ülkeler için büyük önem taşıyan PIRLS uygulamasının sonuçları, ülkelerarası karşılaştırmalar için çeşitli araştırmalarda farklı şekillerde yer almıştır. Bu araştırmaların tamamına yakını PIRLS'ten elde edilen aile, ebeveyn ya da öğrenci değişkenlerine ait verilerin öğrencilerin okuma ve okuduğunu anlama becerileriyle ilişkilerinin incelenmesini konu almaktadır. PIRLS uygulamasının yapı geçerliliği ve bu yapının uygulamaya katılan tüm ülkelerde birbirine denkliği daha önce test edilmemiştir. Uluslararası çalışmalarda ülkeler ortalamaları bakımından sıralanmaktadır. Sıralamada daha yukarıda olanlar daha başarılı kabul edilmektedirler. Oysa sıralama işleminde kullanılan ortalamaların farklı ülkelerde uygulanmış fakat birbirine denk sayılabilecek bir testten elde edildiğine dair kanıtlar çoğu kez sunulmamaktadır. Bu çalışma ile ülkelerin ortalamaları bakımından sıralanabileceği varsayımı sınınmıştır.

Eğitim politikalarında önemli değişikliklere yol açan uluslararası karşılaştırılmalı çalışmaların yapı geçerlilikleri uluslararası düzeyde büyük önem taşımaktadır. Genel geçerliliğin sıkıntılı olduğu durumlarda yapılacak karşılaştırmaların vereceği bilginin hiç bir bilimsel değeri ve anlamı olmayacaktır. Bu bağlamda, yapılan bu araştırma ile eğitim politikalarına yön veren uluslararası bir testin uluslararası geçerliliğinin tespiti,

eğitim politikalarına yön vermeden önce alınabilecek önlemlere yönelik katkılar getirecektir.

Araştırmanın amacı, PIRLS 2001 uygulamasının yapı geçerliliğini ortaya koyarak sonraki PIRLS uygulamalarının daha karşılaştırılabilir bilgiler sunabilmesine yardımcı olabilmektir. Buna ek olarak PIRLS 2001 uygulamasının yapı geçerliliği sonuçlarından yola çıkarak ülkelerarası karşılaştırmaların daha sağlıklı yapılabilmesini sağlamaktır.

Bu genel çerçevede aşağıdaki soruya yanıt aranmıştır.

PIRLS 2001 uygulaması 35 ülkenin her birinde okuma amaçlarına göre bölünmüş dört farklı kavrama sürecine uygun ölçme yapmakta mıdır?

YÖNTEM

Bu bölümde araştırmanın modeli, evreni, verilerin toplanması, verilerin analizi ve varsayımların test edilmesi sunulmuştur.

Araştırmanın Modeli

Bu araştırmada PIRLS 2001 testinin uluslararası uygulamasından elde edilmiş olan verilerden yola çıkılarak her ülkedeki yapı geçerliliği değerlerinin diğer ülkelerdeki yapı geçerliliği değerlerine denk sayılıp sayılmayacağı araştırılmıştır. Bu doğrultuda PIRLS 2001 uygulamasına ait olan veriler analize alınmıştır.

Yukarıdaki bilgiler doğrultusunda var olan bir durumun olduğu şekliyle araştırılması söz konusu olduğundan, bu araştırma tarama modelinde bir araştırmadır.

Evren

PIRLS 2001 uygulamasının evreni, uygulamayı yapan kuruluş olan International Association for the Evaluation of Educational Achievement tarafından “uluslararası istenen hedef evren” olarak isimlendirilmiştir. Bu evreni oluşturan bireylerin ise, test uygulaması sırasında büyük çoğunluğu dokuz yaşında bulunan öğrencilerin kayıtlı olduğu iki sınıftan üst sınıfta okuyan tüm öğrencileri kapsadığı belirtilmiştir (Mullis ve diğerleri, 2003). Örnekle açıklanacak olursa, Türkiye’de dokuz yaşında bulunan öğrenciler yoğunlukla üçüncü ve dördüncü sınıfta kayıtlı bulunmaktadır. Türkiye için

hedef evren dördüncü sınıfta okuyan tüm öğrenciler olarak belirlenmiştir. Benzer durum araştırmaya katılan tüm ülkeler için geçerlidir. Bu durumda PIRLS 2001 uygulamasının tüm dünyadaki öğrencileri bir evren olarak görmek yerine araştırmaya katılan ülkelerdeki dördüncü sınıf öğrencilerini hedef evren olarak tanımladığı ortaya çıkmaktadır. Uygulamaya Almanya, Amerika Birleşik Devletleri, Arjantin, Belize, Bulgaristan, Çek Cumhuriyeti, Fas, Fransa, Hollanda, Hong Kong, İngiltere, İran, İskoçya, İsrail, İsveç, İtalya, İzlanda, Kanada, Kıbrıs Rum Kesimi, Kolombiya, Kuveyt, Letonya, Litvanya, Macaristan, Makedonya, Moldova, Norveç, Romanya, Rusya Federasyonu, Singapur, Slovakya, Slovenya, Türkiye, Yeni Zelanda ve Yunanistan olmak üzere 35 ülke katılmıştır. Her bir ülkedeki dördüncü sınıf öğrencilerinin örneklendirilmesi için iki aşamalı tabakalı küme örnekleme yöntemi kullanılmıştır. Her ülkedeki tabakalar o ülkenin okullarının dağılımlarıyla ilgili olduğu düşünülen önemli değişkenler göz önünde bulundurularak belirlenmiştir. Ülkelerin çoğunda tabakalar okulların büyüklükleri ile kırsal ve merkez (rural-urban) okul olma durumudur. Bazı ülkelerde okulların büyüklükleri bir tabaka olarak belirlenememiş bu durumda okulların resmi ve özel olması tabaka olarak belirlenmiştir. Örneklere, her ülke içinde yaklaşık 150 ilköğretim okulundan dördüncü sınıf düzeyindeki öğrenciler seçilmiştir. Örneklemlerin büyüklüğü 2488 ile 8253 arasında değişmiştir. Araştırmaya toplam 146490 öğrenci katılmıştır. Tablo 5, ülkelere göre örnekleme giren öğrencilerin sayılarının dağılımını göstermektedir.

Tablo 5. PIRLS 2001 Uygulamasına Katılan Ülkelerin Örnekleme Giren Öğrenci Sayıları

Ülke	N	Ülke	N	Ülke	N	Ülke	N
Almanya	7633	Kıbrıs	3001	Hong Kong	5050	Romanya	3625
Arjantin	3300	Kolombiya	5131	İngiltere	3156	Rusya Fed.	4093
ABD	3763	Kuveyt	7126	İran	7430	Singapur	7002
Belize	2909	Letonya	3019	İskoçya	2717	Slovakya	3807
Bulgaristan	3460	Litvanya	2567	İsrail	3973	Slovenya	2952
Çek Cumh.	3022	Macaristan	4666	İsveç	6044	Türkiye	5125
Fas	3153	Makedonya	3711	İtalya	3502	Yeni Zelanda	2488
Fransa	3538	Moldova	3533	İzlanda	3676	Yunanistan	2494
Hollanda	4112	Norveç	3459	Kanada	8253	TOPLAM	146490

Bu araştırmada örnekleme giren tüm ülkelerden elde edilen verilerin tamamı kullanılmıştır. Bu haliyle bu araştırmanın evreni PIRLS 2001 uygulamasından elde edilen verilerdir. Bu çalışmada tüm veriler kullanıldığı için araştırmada evrenin tamamına ulaşılmış, örneklem kullanılmamıştır.

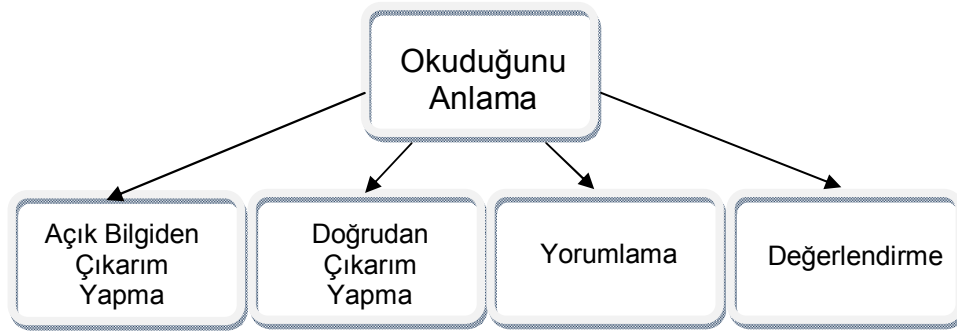
Verilerin Toplanması

PIRLS 2001 uygulamasından elde edilen veriler IEA'nın veritabanından tüm araştırmacılara serbest erişimle sunulmaktadır. Veriler IEA'nın veritabanından sağlanmıştır.

Verilerin Analizi

Elde edilen verilerin analizi aşağıdaki şekilde yapılmıştır.

PIRLS 2001 uygulamasında, okuma amaçlarına göre bölünmüş dört farklı kavrama sürecine uygun PIRLS faktör yapısının doğrulayıcı faktör analizi ile değerlendirilmesi için PIRLS faktör yapısı belirlenmiş ve bu çerçevede aşağıdaki faktör modeli kullanılmıştır.



PIRLS 2001 uygulamasına katılan öğrenciler teste yer alan soruların hepsini almamışlardır. Her bir öğrenci PIRLS kapsamında yer alan sekiz okuma parçasının sadece iki tanesini almıştır. On adet uygulama kitapçığının her birinde sadece iki tane okuma parçası bulunmaktadır. Sıfır, üç, altı ve sekiz numaralı kitapçıklarda yer alan okuma parçaları PIRLS kapsamındaki bütün okuma parçalarını kapsadığından ve diğer kitapçıklarda yer alan parçalar farklı okuma amaçlarını da içerdiğinden bu araştırma kapsamına dahil edilmemiş ve bu araştırmada sıfır, üç, altı ve sekiz numaralı kitapçıklar

dikkate alınmıştır. PIRLS kapsamında yer alan maddelerin büyük kısmı 0 ve 1 şeklinde puanlanırken, küçük bir kısmı 0 ve 1 şeklinde değil, kısmen doğru, oldukça doğru ve tamamen doğru gibi puanlanmaktadır. Bu maddelerin de tamamen doğru olan seçenek 1 ve diğer seçenekler 0 olacak şekilde yeniden kodlanarak 0 ve 1 şeklinde puanlanabileceği bildirilmiştir (Martin, Mullis, Kennedy, 2003). Bu doğrultuda tüm maddeler 0 ve 1 puan alınacak biçimde yeniden kodlanmıştır.

Çalışmada verileri analiz etmek için Lisrel 8.80 for Windows programı kullanılmıştır. Önce analizi yapılacak verilerin dağılımları betimsel istatistikleri aracılığı ile incelenmiş, daha sonra çok gruplu doğrulayıcı faktör analizi uygulanmıştır. Model veri uyumu için geliştirilmiş olan χ^2/sd oranı ve kestirimin hata kareleri ortalaması (RMSEA) indeksi kullanılmıştır. Büyük örneklerde, özellikle de kişi sayısının 250'den fazla olduğu örneklerde RMSEA değerlerinin uyum iyiliğinin değerlendirilmesinde diğer uyum indekslerine oranla daha uygun olduğuna dair öneriler bulunmaktadır (Hu, Bentler, 1999; Brown, 2006). Yapısal eşitlik modelleme ailesinden analizlerde kay kare değerlerinin büyük örneklerden etkilenerek her koşulda anlamlı farklılık bildirme eğiliminde olduğu bilinmektedir (Brown, 2006). İyi bir model veri uyumu için örneklem büyüklüğünden kaynaklanan istatistiksel zayıflıklara karşı daha güçlü olan ve büyük örneklerde diğer uyum indekslerine oranla daha uygun olduğu bildirilen RMSEA değerleri (Cheung & Rensvold, 2002) özellikle dikkate alınmıştır. RMSEA değerlerinin 0.05'ten küçük olmasının iyi bir model-veri uyumuna denk geldiği kabul edilmiştir.

Son yıllarda çok gruplu faktör analizi uygulamalarında modellerin uyum iyiliği katsayılarının birbirinden istatistiksel olarak anlamlı bir farklılık gösterip göstermedikleri incelenmiş, RMSEA değerleri arasındaki farkın 0,01'ten büyük olması durumunda iki modelin birbirinden farklı olduğu yönünde yorum yapılabileceği bildirilmiştir (Cheung, Rensvold, 2000; Cheung, Rensvold, 2002). Cheung ve Rensvold yaptıkları bu simülasyon çalışmasında, 768000 farklı veri seti üretip, bu setlerden 96 farklı faktör koşulu yaratmışlar, çeşitli uyum indekslerinin %99 güven aralıklarını hesaplamışlardır. RMSEA için belirledikleri güven aralığı iki RMSEA değeri arasındaki farkın anlamlı olabilmesi için 0.01 kadar farkın olması gerektiğini ortaya koymuştur. Bu çalışmalardan elde edilen bilgiler doğrultusunda, modellerin RMSEA değerlerinin farkına bakılmış (Δ RMSEA), farkın bildirilen bu kritik değeri aştığı durumlarda daha

küçük RMSEA değerine sahip olan model daha uyumlu model olarak seçilmiştir. Farkın bu kritik değeri aşmadığı durumlarda Model A dışındaki modellerin Model A'dan daha iyi bir açıklama getiremedikleri kabul edilmiş, faktör yüklerinin ve hata varyanslarının gruplarda birbirine denk olduğu üzerine kurulu olan Model A, yani test edilen ölçme aracının tüm ülkelerde aynı faktör yapısını ölçtüğü kabul edilmiştir (Kline, 1998; Brown, 2006).

BULGULAR ve YORUMLAR

Bu bölümde, araştırma sonucunda elde edilen bulgular sunulmuştur. Araştırma sorusuna cevap oluşturacak analizlerin bitiminin ardından elde edilen bulgulara dair yorumlar ve tartışmalar verilmiştir.

Araştırma sorusuna cevap bulabilmek için, PIRLS projesi kapsamında 35 ülkede uygulanan 0, 3, 6 ve 8 numaralı alt testlerin faktör yapılarının kültürler arası eş değerliği yapısal eşitlik modelleri yöntemlerinden çok gruplu doğrulayıcı (confirmatory) faktör analizi yöntemi ile incelenmiştir. Her alt test için analizler ayrı başlıklar altında aşağıda sunulmuştur.

Hare ve Pufflings okuma parçalarından oluşan sıfır numaralı alt teste ilişkin bulgular

Hare ve Pufflings okuma parçalarından oluşan sıfır numaralı alt testten elde edilen puanların dağılımına ait betimleyici istatistikler ülkelere göre Tablo 6'da verilmektedir.

Tablo 6. Hare ve Pufflings Okuma Parçalarından Oluşan Sıfır Numaralı Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel istatistikler	N	Art. Ort.	Ortanca	SS	Varyans	Çarpıklık	Basklık	Ranj	Min.	Maks.	Testteki Madde Sayısı	Testin KR20 Güv. Kat.	Testin iki yarı Güv. Kat.
Türkiye	1283	10.68	11	4.61	21.29	0.12	-0.64	24	0	24	24	0.81	0.74
Almanya	1903	14.20	15	4.36	19.04	-0.44	-0.24	23	1	24	24	0.75	0.67
ABD	926	14.35	15	4.79	22.99	-0.56	-0.25	23	1	24	24	0.82	0.76
Arjantin	794	9.05	9	4.79	22.98	0.27	-0.62	23	0	23	24	0.80	0.75
Belize	290	5.10	4	3.36	11.28	1.24	1.93	19	0	19	24	0.71	0.66
Bulgaristan	879	15.16	16	4.54	20.63	-0.77	0.33	24	0	24	24	0.75	0.71
Çek Cumh.	748	14.09	15	4.25	18.02	-0.45	-0.39	21	2	23	24	0.77	0.72

Tablo 6. Devam

Fas	774	6.31	6	4.13	17.05	0.97	1.05	22	0	22	24	0.79	0.77
Fransa	883	13.97	14	4.26	18.11	-0.22	-0.53	22	2	24	24	0.75	0.68
Hollanda	1018	14.85	15	3.69	13.60	-0.36	-0.19	20	4	24	24	0.75	0.71
Hong Kong	1259	14.18	14	4.23	17.91	-0.45	0.04	24	0	24	24	0.73	0.69
İngiltere	777	14.71	16	4.90	24.06	-0.62	-0.21	23	1	24	24	0.81	0.75
İran	1896	9.60	10	4.08	16.66	0.08	-0.57	21	0	21	24	0.74	0.68
İskoçya	659	13.70	14	4.97	24.67	-0.45	-0.41	23	0	23	24	0.80	0.76
İsrail	985	12.98	13	5.14	26.44	-0.18	-0.73	24	0	24	24	0.83	0.80
İsveç	1489	15.57	16	3.89	15.12	-0.68	0.42	23	1	24	24	0.73	0.68
İtalya	874	14.57	15	4.43	19.65	-0.5	-0.27	23	1	24	24	0.77	0.70
İzlanda	353	12.10	13	4.22	17.81	-0.29	-0.49	20	2	22	24	0.76	0.67
Kanada	2111	14.06	14	4.37	19.07	-0.37	-0.30	24	0	24	24	0.78	0.73
Kıbrıs	749	12.36	13	4.37	19.08	-0.23	-0.46	23	0	23	24	0.78	0.74
Kolombiya	1284	9.20	9	4.26	18.18	0.21	-0.51	21	0	21	24	0.78	0.72
Kuveyt	656	8.78	8	4.95	24.52	0.22	-0.89	21	0	21	24	0.81	0.75
Letonya	747	14.62	15	3.70	13.71	-0.58	0.45	24	0	24	24	0.69	0.65
Litvanya	702	15.08	16	4.26	18.13	-0.52	-0.27	21	2	23	24	0.76	0.76
Macaristan	1164	13.79	14	4.07	16.60	-0.32	-0.37	21	2	23	24	0.75	0.67
Makedonya	935	10.43	10	5.37	28.82	0.04	-1.03	22	0	22	24	0.82	0.79
Moldova	859	11.94	12	4.50	20.26	-0.16	-0.29	24	0	24	24	0.79	0.70
Norveç	863	11.82	12	4.79	22.98	-0.19	-0.78	21	1	22	24	0.80	0.77
Romanya	919	13.80	14	4.90	23.98	-0.44	-0.47	23	1	24	24	0.80	0.76
Rusya Fed.	1022	13.73	14	4.10	16.84	-0.26	-0.32	23	1	24	24	0.73	0.67
Singapur	1747	13.59	14	4.83	23.29	-0.44	-0.47	23	1	24	24	0.84	0.79
Slovakya	944	13.14	13	4.07	16.58	-0.36	-0.3	21	1	22	24	0.74	0.71
Slovenya	736	11.21	11	4.10	16.82	0.01	-0.31	22	1	23	24	0.72	0.66
Yeni Zelanda	608	13.71	15	5.28	27.83	-0.49	-0.60	24	0	24	24	0.81	0.81
Yunanistan	633	14.28	15	4.29	18.40	-0.46	-0.20	22	1	23	24	0.76	0.69

Sıfır numaralı alt testin dağılımı hakkında bilgi edinebilmek amacıyla çarpıklık ve basıklık değerleri incelendiğinde bu değerlerin çarpıklık için -0.77 ile 1.24 arasında değiştiği görülmüştür. Basıklık değerlerinin ise -1.03 ile 1.93 arasında değiştiği görülmüştür. Basıklık ve çarpıklık değerlerinin ± 1.00 arasında değer alması dağılımların normal olduğuna işaret etmektedir (George ve Mallery, 2001). Belize'nin basıklık ve çarpıklık değerleri dışındaki tüm ülkelerin basıklık ve çarpıklık değerleri bu ülkeler için dağılımların normal olduğunu göstermektedir. Belize'den elde edilen puanlar, sağa çarpık ve ve sivri bir dağılım göstermektedir. KR20 güvenilirlik katsayılarının 0.69 ile 0.84 arasında değiştiği, iki yarı güvenilirlik katsayılarının ise 0.65 ile 0.81 arasında değiştiği görülmüştür. Bu değerler sıfır numaralı alt ölçeğin tüm ülkelerde yeterli derecede güvenilirliğe sahip olduğunu göstermektedir. Yukarıda betimsel istatistikleri verilmiş olan ülkelerin sıfır numaralı alt testten elde edilen puanlarının çok gruplu doğrulayıcı faktör analizi yapılmıştır.

Hare ve Pufflings okuma parçalarından oluşan sıfır numaralı alt test için çok gruplu doğrulayıcı faktör analizi sonuçları:

Hare ve Pufflings okuma parçalarına ait değerlendirme sorularından oluşan sıfır numaralı alt testteki maddelerin ölçtüğü varsayılan kavrama süreçlerine (boyutlar) göre dağılımı Tablo 7’de verilmektedir (Gonzalez ve Kennedy, 2003; Mullis, Martin, Gonzalez ve Ann, 2003).

Tablo 7. Hare ve Pufflings Okuma Parçalarından Oluşan Sıfır Numaralı Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımları

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Hare	Pufflings
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar	1, 2, 4	1, 2, 3, 5, 9
Doğrudan çıkarımlar yapma	5, 7	10
Fikir ve bilgileri birleştirme ve yorumlama	6, 8, 9, 10, 11	4, 6, 7, 8, 11, 12
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	3	13

Tablo 7’de görüldüğü gibi Hare okuma parçasında 11 soru Pufflings okuma parçasında ise 13 soru vardır. Teste ait bu yapının test uygulamasına katılan ülkeler arasında eşdeğer olup olmadığını incelemek için 35 ülkeden elde edilen veriler üzerinden çok gruplu doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 8 hazırlanmıştır.

Tablo 8. Hare ve Pufflings Okuma Parçalarından Oluşan Sıfır Numaralı Alt Test İçin Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2/sd	RMSEA	$\Delta\chi^2_{sd}$	Δ RMSEA
Model A	57514.24 ₁₀₄₄₆	5.50	0.068		
Model B	50259.32 ₉₆₃₀	5.22	0.065	7254.92 ₈₁₆ *	0.003
Model C	49844.28 ₉₆₀₆	5.19	0.065	7969.96 ₈₄₀ *	0.003
Model D	57070.61 ₁₀₄₂₂	5.48	0.067	443.63 ₂₄ *	0.001

Model A: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

Model B: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

Model C: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

Model D: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

Model B, Model C ve Model D'nin temel model olan Model A ile yapılan ikili karşılaştırmalarında şu bulgular elde edilmiştir. Ülkelerin kendi faktör yüklerine sahip olduğu Model B'nin, faktör yükleri ile hata varyanslarının serbest olduğu Model C'nin ve her ülkede hata varyanslarının birbirinden farklı olduğu Model D'nin uyum katsayılarının (RMSEA değerleri), faktör yüklerinin tüm ülkelerde eşit olduğunu varsayan A modelinden istatistiksel olarak anlamlı derecede daha iyi uyum gösteremedikleri görülmüştür. χ^2/sd oranları da incelendiğinde bu oranların birbirlerine çok yakın oldukları görülmektedir. Tüm modellerin χ^2/sd oranları 5 civarındadır. Model A'nın χ^2/sd oranı 5.50, RMSEA değeri ise 0.068'dir. Tüm ülkelerde faktör yapısının farklı olduğu üzerine kurulu olan Model B (RMSEA= 0.065; $\chi^2/sd=5.22$; $\Delta RMSEA=0.003$), Model C (RMSEA=0.065; $\chi^2/sd=5.19$; $\Delta RMSEA=0.003$), ve Model D (RMSEA=0.067; $\chi^2/sd=5.48$; $\Delta RMSEA=0.001$), faktör yapısının tüm ülkelerde birbirine denk olduğunu kabul eden Model A'dan daha iyi bir uyum gösteremediği için reddedilmiştir. Buna göre uygulamaların her ülkede aynı faktör yapısına sahip olduğu üzerine kurulu model olan Model A'dan daha uyumlu bir model bulunmadığı için Model A kabul edilmiştir. Model A, sıfır numaralı alt testin tüm ülkelerde benzer faktör yapısına uygun ölçme yaptığı üzerine kuruludur. Sıfır numaralı alt testin tüm ülkelerde birbirine benzer faktör yapısıyla ölçme yaptığı kabul edilmiştir. Ölçeğin faktör yapısı tüm ülkeler için evrensel bir yapı göstermektedir.

Uygulamanın tüm ülkelerde eşit olduğunu kabul eden Model A'nın RMSEA değeri (0.068) ve χ^2/sd oranı (5.50) incelendiğinde Hare ve Pufflings okuma parçalarından oluşan sıfır numaralı alt testin tüm ülkelerde orta derecede bir uyumla benzer faktör yapısında ölçme yaptığını görülmektedir.

Flowers ve Antarctica okuma parçalarından oluşan üç numaralı alt testte ilişkin bulgular:

Flowers ve Antarctica okuma parçalarından oluşan üç numaralı alt testten elde edilen puanların dağılımına ait betimleyici istatistikler ülkelere göre Tablo 8'de verilmektedir.

Tablo 9. Flowers ve Antartica Okuma Parçalarından Oluşan Üç Numaralı Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel İstatistikler	N	Arit. Ort.	Ortanca	SS	Varyans	Çarpıklık	Basıklık	Ranj	Minimum	Maksimum	Testteki Madde Sayısı	Testin KR20 Güv. Katsayısı	Testin İki Yarı Güv. Katsayısı
Türkiye	423	10.60	10	4.86	23.61	0.07	-0.83	21	0	21	24	0.83	0.80
Almanya	633	16.15	17	4.14	17.11	-0.67	0.19	22	2	24	24	0.74	0.72
ABD	315	16.24	17	5.04	25.38	-0.80	0.10	23	1	24	24	0.85	0.80
Arjantin	279	10.42	10	5.40	29.12	0.23	-0.92	23	0	23	24	0.81	0.80
Belize	290	5.87	5	4.54	20.58	1.13	0.73	20	0	20	24	0.87	0.82
Bulgaristan	281	17.23	19	4.76	22.62	-1.08	0.81	24	0	24	24	0.77	0.74
Çek Cumh.	252	15.76	16	4.30	18.45	-0.67	0.14	20	3	23	24	0.69	0.65
Fas	259	5.77	5	4.07	16.60	0.89	0.39	19	0	19	24	0.78	0.72
Fransa	296	15.53	16	4.42	19.57	-0.49	-0.16	21	3	24	24	0.74	0.64
Hollanda	335	17.39	18	3.80	14.41	-0.88	0.77	20	4	24	24	0.75	0.67
Hong Kong	420	14.73	15	3.96	15.66	-0.56	0.19	22	0	22	24	0.71	0.63
İngiltere	276	16.75	18	4.66	21.74	-1.03	0.90	24	0	24	24	0.79	0.68
İran	611	10.59	10	5.45	29.70	0.04	-0.99	22	0	22	24	0.80	0.73
İskoçya	236	15.69	17	4.76	22.68	-0.71	-0.21	21	2	23	24	0.78	0.74
İsrail	325	13.99	15	5.73	32.82	-0.55	-0.79	23	0	23	24	0.86	0.85
İsveç	508	18.21	19	3.81	14.53	-0.96	0.90	21	3	24	24	0.69	0.62
İtalya	301	16.47	17	4.04	16.34	-0.65	0.13	21	3	24	24	0.76	0.70
İzlanda	356	14.50	15	5.08	25.76	-0.48	-0.50	22	2	24	24	0.78	0.75
Kanada	683	16.20	17	4.38	19.14	-0.68	0.10	22	2	24	24	0.80	0.73
Kıbrıs	239	13.16	14	5.24	27.46	-0.38	-0.54	23	0	23	24	0.79	0.76
Kolombiya	425	10.47	10	4.74	22.43	0.26	-0.60	22	1	23	24	0.78	0.75
Kuveyt	707	7.17	6	4.37	19.09	0.55	-0.49	19	0	19	24	0.76	0.70
Letonya	256	16.70	17	4.37	19.08	-0.64	-0.07	21	3	24	24	0.77	0.74
Litvanya	229	15.85	16	4.29	18.38	-0.55	-0.19	20	3	23	24	0.79	0.72
Macaristan	392	15.83	17	4.29	18.38	-0.63	-0.01	24	0	24	24	0.75	0.71
Makedonya	307	11.06	12	5.81	33.71	-0.12	-1.14	22	0	22	24	0.81	0.71
Moldova	301	13.62	15	4.90	23.98	-0.53	-0.49	21	1	22	24	0.75	0.64
Norveç	279	14.24	15	4.88	23.78	-0.61	-0.20	21	2	23	24	0.75	0.72
Romanya	293	14.79	16	5.37	28.85	-0.54	-0.59	22	1	23	24	0.81	0.75
Rusya Fed.	335	15.26	16	4.33	18.78	-0.82	0.30	22	1	23	24	0.65	0.60
Singapur	583	15.38	16	5.15	26.54	-0.73	-0.22	22	1	23	24	0.85	0.82
Slovakya	314	14.77	15	4.36	18.97	-0.64	0.06	21	2	23	24	0.74	0.68
Slovenya	245	14.52	15	4.58	20.97	-0.42	-0.23	22	1	23	24	0.78	0.72
Yeni Zelanda	211	15.64	16	5.16	26.64	-0.68	-0.21	23	1	24	24	0.83	0.81
Yunanistan	212	15.66	17	4.60	21.16	-0.67	-0.14	22	2	24	24	0.68	0.57

Ülkelere ilişkin betimsel istatistikler, çarpıklık ve basıklık katsayıları bakımından incelendiğinde, çarpıklık katsayılarının -1.08 ile 1.13 arasında değer aldığı, basıklık katsayılarının ise -1.14 ile 0.90 arasında bulunduğu görülmüştür. Bu değerlerin ± 1.00 sınırında ya da bu sınıra çok yakın olması nedeniyle tüm dağılımların normal ya da normale çok yakın olduğu görülmüştür. Sıfır numaralı alt testte olduğu gibi üç numaralı alt testte de Belize'den elde edilen puanların dağılımı normalden çok fazla olmasa da sağa çarpık bir dağılım göstermektedir. Üç numaralı alt testin farklı ülkelerdeki

uygulamalarından elde edilen KR20 güvenilirlik katsayıları incelendiğinde KR20 güvenilirlik değerlerinin 0.65 ile 0.87 arasında, iki yarı güvenilirlik katsayılarının ise 0.57 ile 0.85 arasında olduğu belirlenmiştir. Bu değerler üç numaralı alt testin tüm ülkelerde yeterli güvenilirliğe sahip olduklarını göstermektedir.

Flowers ve Antarctica okuma parçalarından oluşan üç numaralı alt test için çok gruplu doğrulayıcı faktör analizi sonuçları:

Flowers ve Antarctica okuma parçalarına ait değerlendirme sorularından oluşan üç numaralı alt testteki maddelerin ölçtüğü varsayılan kavrama süreçlerine (boyutlar) göre dağılımı Tablo 10'da verilmektedir (Gonzalez ve Kennedy, 2003; Mullis ve diğerleri, 2003).

Tablo 10. Flowers ve Antarctica Okuma Parçalarından Oluşan Üç Numaralı Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımları

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Flowers	Antarctica
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	4, 9	1, 2, 3, 6, 8
Doğrudan çıkarımlar yapma	2, 3, 5, 6, 10	5, 7
Fikir ve bilgileri birleştirme ve yorumlama	1, 7, 8, 12	4, 9
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	11, 13	10, 11

Tablo 10'da görüldüğü gibi Flowers okuma parçasında 11 soru Antarctica okuma parçasında ise 13 soru vardır. Üç numaralı alt teste ait bu yapının test uygulamasına katılan ülkeler arasında eşdeğer olup olmadığını incelemek için 35 ülkeden elde edilen veriler üzerinden çok gruplu doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 11 hazırlanmıştır.

Tablo 11. Flowers ve Antarctica Okuma Parçalarından Oluşan Üç Numaralı Alt Test İçin Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2/sd	RMSEA	$\Delta\chi^2_{sd}$	Δ RMSEA
Model A	29176.50 ₁₀₄₄₆	2.79	0.071		
Model B	26628.81 ₉₆₃₀	2.76	0.071	2547.69 ₈₁₆ *	0,000
Model C	26614.12 ₉₆₂₈	2.76	0.071	2562.38 ₈₁₈ *	0,000
Model D	29164.97 ₁₀₄₄₄	2.79	0.071	11.53 ₂	0,000

Model A: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

Model B: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

Model C: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

Model D: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

Tablo 11’de görüldüğü gibi Model B, Model C ve Model D’nin temel model olan Model A ile ikili karşılaştırmaları yapılmıştır. Tüm modellerin uyum katsayılarının birbirine eşit olduğu görülmektedir. Üç numaralı alt testin tüm ülkelerde aynı faktör yapısıyla ölçme yaptığı kabul eden Model A’nın RMSEA değeri 0.071 olarak bulunmuştur. Diğer modellerin RMSEA değerleri de Model A’nın RMSEA değerine eşittir. χ^2/sd oranları incelendiğinde tüm modellerin oranlarının birbirlerine benzer olduğu görülmektedir. Modeller arasındaki kay-kare değerlerinin farkı anlamlı olmasına rağmen bu farklılık kay-kare değerinin örneklem büyüklüğünden etkilenerek büyük örneklerde anlamlı sonuç verme eğilimine girmesi nedeniyle yorumlanmamıştır. Bu durumda faktör yapısının tüm ülkelerde birbirinden farklı olabileceği üzerine kurulu bulunan diğer modellerin Model A’dan daha iyi bir sonuç üretmedikleri görülmüştür. Bu nedenle üç numaralı alt testin faktör yapısının uygulamanın yapıldığı tüm ülkelerde birbirine denk olduğu görülmüştür. Sonuç olarak üç numaralı alt testin uygulamaya katılan 35 ülkede dört ayrı kavrama süreciyle ifade edilen boyutlarda ölçme yaptığı söylenebilir. Üç numaralı alt testin yapı geçerliliği evrenseldir.

3.3. Mice ve River okuma parçalarından oluşan altı numaralı alt teste ilişkin bulgular

Mice ve River okuma parçalarından oluşan altı numaralı alt testten elde edilen puanların dağılımına ait betimleyici istatistikler ülkelere göre Tablo 12’de verilmektedir.

Tablo 12. Mice ve River Okuma Parçalarından Oluşan Altı Numaralı Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel istatistikler	N	Aritmetik Ortalama	Ortanca	Standart Sapma	Varyans	Çarpıklık	Basıklık	Ranj	Minimum	Maksimum	Testteki Madde Sayısı	Testin KR20 Güvenilirlik Katsayısı	Testin iki yarı güvenilirlik katsayısı
Türkiye	425	10.31	10	5.25	27.64	0.21	-0.70	24	0	24	25	0.82	0.75
Almanya	658	14.87	16	5.08	25.84	-0.60	-0.23	24	0	24	25	0.82	0.80
ABD	309	15.88	17	5.12	26.25	-0.83	0.30	24	1	25	25	0.83	0.78
Arjantin	283	9.05	8	5.08	25.88	0.31	-0.56	23	0	23	25	0.80	0.71
Belize	299	5.14	4	3.91	15.32	1.28	1.75	21	0	21	25	0.85	0.87
Bulgaristan	289	15.95	17	4.77	22.78	-0.64	0.18	25	0	25	25	0.78	0.70
Kanada	687	15.06	16	5.21	27.23	-0.59	-0.28	24	0	24	25	0.82	0.76
Çek Cumh.	252	15.84	17	4.27	18.25	-0.69	-0.09	20	4	24	25	0.69	0.65
Fas	254	5.85	5	3.89	15.19	1.00	1.13	21	0	21	25	0.79	0.94
Fransa	290	15.71	16	4.48	20.14	-0.70	0.07	21	3	24	25	0.77	0.70
Hollanda	359	16.94	18	3.99	15.96	-0.61	0.27	22	3	25	25	0.77	0.73
Hong Kong	425	15.65	16	4.64	21.61	-0.82	0.34	24	0	24	25	0.76	0.66
İngiltere	270	16.21	17	5.46	29.89	-0.71	-0.20	24	0	24	25	0.84	0.79
İran	634	9.44	9	4.56	20.81	0.22	-0.73	21	0	21	25	0.77	0.72
İskoçya	226	14.67	15.5	5.06	25.64	-0.57	-0.22	23	1	24	25	0.78	0.73
İsrail	325	13.79	14	5.46	29.82	-0.36	-0.67	24	0	24	25	0.82	0.77
İsveç	496	17.69	18	4.42	19.61	-0.95	0.97	23	2	25	25	0.75	0.71
İtalya	278	15.07	16	4.48	20.10	-0.73	0.55	24	0	24	25	0.75	0.69
İzlanda	369	13.59	14	5.07	25.75	-0.28	-0.46	24	0	24	25	0.78	0.72
Kıbrıs	259	12.74	13	5.32	28.30	-0.14	-0.68	22	1	23	25	0.83	0.79
Kolombiya	430	9.40	9	4.96	24.63	0.21	-0.82	20	0	20	25	0.82	0.71
Kuveyt	686	6.40	6	3.71	13.80	0.51	-0.20	19	0	19	25	0.73	0.73
Letonya	256	15.76	16	4.18	17.53	-0.67	-0.01	19	4	23	25	0.69	0.66
Litvanya	227	15.78	17	4.14	17.19	-0.60	0.22	22	2	24	25	0.72	0.69
Macaristan	389	15.34	16	4.72	22.31	-0.62	-0.16	23	1	24	25	0.77	0.69
Makedonya	305	9.38	9	5.3	28.11	0.18	-0.90	22	0	22	25	0.75	0.70
Moldova	305	12.32	12	4.64	21.54	-0.19	-0.61	22	0	22	25	0.72	0.59
Norveç	288	13.41	14	5.20	27.14	-0.32	-0.42	23	1	24	25	0.79	0.78
Romanya	293	13.13	14	5.26	27.71	-0.22	-0.66	24	0	24	25	0.83	0.78
Rusya Fed.	347	15.27	16	4.47	20.03	-0.71	0.02	21	2	23	25	0.72	0.62
Singapur	582	14.65	15	5.41	29.31	-0.63	-0.24	25	0	25	25	0.85	0.85
Slovakya	311	14.17	15	4.47	20.02	-0.64	0.04	22	1	23	25	0.69	0.65
Slovenya	240	13.39	14	4.80	23.05	-0.35	-0.46	24	1	25	25	0.75	0.71
Yeni Zelanda	201	15.04	16	6.09	37.13	-0.58	-0.55	24	1	25	25	0.87	0.84
Yunanistan	209	14.92	15	4.50	20.25	-0.29	-0.71	20	4	24	25	0.76	0.71

Altı numaralı alt testin uygulandığı ülkelerin betimsel istatistikleri, dağılım hakkında fikir elde edinebilmek amacıyla incelendiğinde dağılımların çarpıklık değerlerinin -0.96 ile 1.28 arasında değiştiği görülmüştür. Çarpıklık değeri 1.00'in

üzerinde olan tek ülke diğer alt test uygulamalarında da olduğu gibi Belize'dir. Belize dışındaki tüm ülkelerin çarpıklıklarının normalden büyük bir sapmayı göstermediği belirlenmişken Belize'den elde edilen puanların sağa çarpık olduğu görülmektedir. Basıklık katsayıları incelendiğinde, bu değerlerin -0.90 ile 1.75 arasında değiştiği görülmüştür. Basıklık değerleri 1.00'in üzerinde bulunan iki ülke Fas ve Belize'dir. Bu ülkelerin basıklık katsayılarından dağılımların sivri bir yapı gösterdikleri anlaşılmaktadır. Diğer ülkelerde dağılımların normal ya da normale çok yakın olduğu görülmektedir. Ölçeğin tüm ülkelerde güvenilir ölçme yapıp yapmadığını anlamak amacıyla uygulamalardan elde edilen KR20 güvenilirlik katsayıları incelenmiş bu değerlerin 0.69 ile 0.87 arasında, iki yarı güvenilirlik katsayılarının ise 0.59 ile 0.94 arasında değiştiği görülmüştür. Bu değerler, ölçeğin tüm ülkelerde yeterli güvenilirliğe sahip olduğunu göstermektedir.

Mice ve River okuma parçalarından oluşan altı numaralı alt test için çok gruplu doğrulayıcı faktör analizi sonuçları:

Mice ve River okuma parçalarına ait değerlendirme sorularından oluşan altı numaralı alt testteki maddelerin, ölçtüğü varsayılan kavrama süreçlerine (boyutlar) göre dağılımı Tablo 13'te verilmektedir (Gonzalez ve Kennedy, 2003; Mullis ve diğerleri 2003).

Tablo 13. Mice ve River Okuma Parçalarından Oluşan Altı Numaralı Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımları

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Mice	River
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	2, 5, 10	3, 6
Doğrudan çıkarımlar yapma	1, 3, 7, 9	4, 5, 7
Fikir ve bilgileri birleştirme ve yorumlama	4, 6, 11, 12	8, 9, 10
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	8, 13, 14	1, 2, 11

Tablo 13'te görüldüğü gibi Mice okuma parçasında 14 soru, River okuma parçasında ise 11 soru vardır. Altı numaralı alt teste ait bu yapının test uygulamasına katılan ülkeler arasında eşdeğer olup olmadığını incelemek için 35 ülkeden elde edilen

veriler üzerinden çok gruplu doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 14 hazırlanmıştır.

Tablo 14. Mice ve River Okuma Parçalarından Oluşan Altı Numaralı Alt Test İçin Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2/sd	RMSEA	$\Delta\chi^2_{sd}$	Δ RMSEA
Model A	34279.61 ₁₁₃₁₉	3.02	0.075		
Model B	32424.59 ₁₀₄₆₉	3.09	0.077	1855.02 ₈₅₀ *	0,002
Model C	32050.87 ₁₀₄₄₄	3.06	0.076	2228.74 ₈₇₅ *	0,001
Model D	33927.20 ₁₁₂₉₄	3.01	0.075	352.41 ₂₅ *	0,000

Model A: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.
Model B: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).
Model C: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).
Model D: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

Tablo 14’te görüldüğü gibi modellerin uyum katsayıları birbirine çok yakındır. Özellikle faktör yüklerinin birbirine denk olduğu üzerine kurulmuş olan Model A (RMSEA=0.075) ve hata varyanslarının serbest olduğu durum üzerine kurulmuş olan Model D’nin (RMSEA=0.075; Δ RMSEA=0.000) uyum düzeylerini gösteren RMSEA değerleri birbirine eşittir. Model B (RMSEA=0.077; Δ RMSEA=0.002) ve Model C’nin (RMSEA=0.076; Δ RMSEA=0.001) uyum iyiliğini gösteren RMSEA değerleri, Model A’nın RMSEA değerine, istatistiksel olarak manidar bir farklılık ortaya koymayacak derecede yakındır. Ayrıca modellerin χ^2/sd oranları da aynı oran derecesi (3) içinde ve birbirlerine çok yakındır. Bu durum tüm ülkelerde faktör yapısının birbirine denk olduğu üzerine kurulu bulunan Model A’dan daha iyi bir modelin olmadığını göstermektedir. Modellerin kay-kare değerleri birbirinden farklı olsa da bu farklılıklar kay-kare değerlerinin örneklem büyüklüğünden çok etkilenmesi nedeniyle yorumlanmamıştır. Modellerin uyum iyiliği katsayılarının Model A’nın uyum iyiliği katsayılarından daha iyi olmaması nedeniyle Model A varolan durumu en iyi açıklayan model olarak kabul edilmiştir. Bu nedenle tüm ülkelerde uygulamaların birbirine faktör yapısı olarak denk olduğu kabul edilmiştir. Altı numaralı alt testin yapı geçerliliği evrenselidir.

Altı numaralı alt testin faktör yapısının ülkelerarası uyumunun ne oranda olduğunu bulmak amacıyla Model A'nın RMSEA değeri (0.075) incelenmiş ve bu alt testin tüm ülkelerde birbirine orta düzeyde benzer bir faktör yapısıyla ölçme yaptığı belirlenmiştir.

Clay ve Leonardo okuma parçalarından oluşan sekiz numaralı alt teste ilişkin bulgular:

Clay ve Leonardo okuma parçalarından oluşan sekiz numaralı alt testten elde edilen puanların dağılımına ait betimleyici istatistikler ülkelere göre Tablo 15'te verilmektedir.

Tablo 15. Clay ve Leonardo Okuma Parçalarından Oluşan Sekiz Numaralı Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel istatistikler	N	Art. Ort.	Ortanca	SS	Varyans	Çarpıklık	Basıklık	Ranj	Min.	Maks.	Testteki Madde Sayısı	Testin KR20 Güv. Kat.	Testin iki yarı Güv. Kat.
Türkiye	435	8.42	8	4.64	21.61	0.58	-0.34	23	0	23	24	0.80	0.70
Almanya	645	13.38	14	4.43	19.67	-0.31	-0.24	23	1	24	24	0.73	0.78
ABD	321	13.88	15	5.67	32.24	-0.40	-0.75	24	1	25	24	0.85	0.82
0.55 Arjantin	284	6.65	6	4.09	16.79	0.80	0.26	20	0	20	24	0.72	0.76
Belize	293	4.85	4	3.60	13.02	1.58	3.05	19	0	19	24	0.80	0.84
Bulgaristan	293	14.00	15	4.65	21.70	-0.58	-0.31	21	2	23	24	0.66	0.64
Kanada	661	13.00	13	5.00	25.00	-0.22	-0.70	24	0	24	24	0.79	0.75
Çek Cumh.	245	13.19	14	4.40	19.4	-0.45	-0.52	21	2	23	24	0.62	0.55
Fas	258	5.53	5	3.98	15.85	1.75	4.13	22	0	22	24	0.88	0.91
Fransa	293	12.03	12	4.81	23.21	0.08	-0.66	23	1	24	24	0.80	0.83
Hollanda	337	14.21	14	4.12	17.02	-0.36	-0.19	21	3	24	23	0.76	0.73
Hong Kong	427	12.31	13	4.20	17.84	-0.34	-0.38	21	0	21	24	0.75	0.63
İngiltere	267	13.65	14	5.55	30.81	-0.24	-0.79	23	1	24	24	0.82	0.80
İran	630	8.468	7	4.84	23.47	0.64	-0.43	22	0	22	24	0.81	0.81
İskoçya	231	13.69	15	5.17	26.81	-0.52	-0.36	24	0	24	24	0.79	0.73
İsrail	339	12.19	12	5.47	29.96	-0.08	-0.94	23	0	23	24	0.81	0.79
İsveç	476	14.36	15	4.19	17.60	-0.58	0.17	22	1	23	24	0.68	0.64
İtalya	290	14.14	15	4.65	21.65	-0.37	-0.50	21	2	23	24	0.75	0.71
İzlanda	382	12.00	12	4.74	22.55	-0.05	-0.40	22	1	23	24	0.76	0.77
Kıbrıs	256	10.73	10	5.24	27.49	0.10	-1.11	22	1	23	24	0.79	0.74
Kolombiya	428	7.16	6.5	3.88	15.09	0.67	0.05	19	0	19	24	0.72	0.71
Kuveyt	737	6.51	6	3.96	15.71	0.75	0.24	21	0	21	24	0.76	0.75
Letonya	244	13.75	14	4.28	18.39	-0.23	-0.64	20	4	24	24	0.71	0.68
Litvanya	233	14.25	15	4.29	18.45	-0.60	-0.43	20	2	22	24	0.63	0.62
Macaristan	390	14.25	15	4.54	20.637	-0.52	-0.42	21	2	23	24	0.74	0.73
Makedonya	314	9.35	9	5.14	26.43	0.32	-0.47	23	0	23	24	0.74	0.71

Tablo 15. Devam

Moldova	295	10.94	11	4.57	20.93	0.12	-0.62	23	0	23	24	0.75	0.77
Norveç	293	10.57	11	4.77	22.8	-0.17	-0.46	22	0	22	24	0.68	0.70
Romanya	304	11.64	12	5.22	27.3	0.01	-0.87	22	1	23	24	0.81	0.80
Rusya Fed.	336	13.10	14	4.35	18.94	-0.33	-0.53	20	2	22	24	0.72	0.67
Singapur	592	12.89	14	5.49	30.21	-0.25	-0.97	23	1	24	24	0.85	0.80
Slovakya	319	12.57	13	4.93	24.39	-0.25	-0.64	22	1	23	24	0.78	0.73
Slovenya	253	11.22	11	4.67	21.86	0.04	-0.53	21	0	21	24	0.78	0.80
Yeni Zelanda	215	12.93	14	5.70	32.58	-0.39	-1.00	22	1	23	24	0.84	0.83
Yunanistan	204	12.61	13	4.81	23.23	-0.36	-0.16	23	0	23	24	0.78	0.78

Sekiz numaralı alt testin uygulandığı ülkelerin çarpıklık katsayıları incelendiğinde bu değerlerin -0.60 ile 1.75 arasında değiştiği görülmüştür. Çarpıklık değeri 1.00'in üzerinde olan iki ülke bulunmaktadır. Bu iki ülke Belize ve Fas'tır. Bu iki ülkenin diğer ülkelere oranla ortalamalarının çok düşük olması nedeniyle dağılımlarının sağa çarpık oldukları düşünülmüştür. Basıklık katsayıları incelendiğinde değerlerin -1.11 ile 4.13 arasında değiştiği görülmüştür. Yine Belize ve Fas hariç tüm ülkeler için dağılımların normal ya da normale çok yakın olduğu görülmektedir. Belize ve Fas uygulamalarından elde edilen puanların dağılımlarının basıklık katsayısı yüksek çıkmıştır. Bu durum incelendiğinde basıklık katsayılarının yüksek çıkmasına yol açan durumun bu iki ülkenin bu alt teste ait ortalamalarının çok düşük olması ve puanların ranjlarının düşük olmasıdır. Belize ve Fas'tan elde edilen veriler analize bu haliyle alınmıştır. Ölçeğin tüm ülkeler için KR20 güvenilirlik katsayılarının 0.62 ile 0.88 arasında değiştiği, iki yarı güvenilirlik katsayılarının ise 0.55 ile 0.91 arasında değiştiği görülmüştür. Genel olarak uygulamanın tüm ülkelerde güvenilir ölçme yaptığı görülmüştür.

Clay ve Leonardo okuma parçalarından oluşan sekiz numaralı alt test için çok gruplu doğrulayıcı faktör analizi sonuçları:

Clay ve Leonardo okuma parçalarına ait değerlendirme sorularından oluşan sekiz numaralı alt testteki maddelerin, ölçtüğü varsayılan kavrama süreçlerine (boyutlar) göre dağılımı Tablo 16'da verilmektedir (Gonzalez ve Kennedy, 2003; Mullis ve diğerleri 2003).

Tablo 16. Clay ve Leonardo Okuma Parçalarından Oluşan Sekiz Numaralı Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımları

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Clay	Leonardo
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	1, 8, 9	1, 3, 9
Doğrudan çıkarımlar yapma	2, 3, 4, 5, 7	2, 5, 6, 8
Fikir ve bilgileri birleştirme ve yorumlama	6, 10, 11, 13	4, 7, 10
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	12	11, 12

Tablo 16'da görüldüğü gibi Clay ve Leonardo okuma parçalarında 12'şer madde bulunmaktadır. Tablo 16'da gösterilen sekiz numaralı alt teste ait bu yapının test uygulamasına katılan ülkeler arasında eşdeğer olup olmadığını incelemek için 35 ülkeden elde edilen veriler üzerinden çok gruplu doğrulayıcı faktör analizi yapılmıştır. Analize kovaryans matrisinin negatif çıkmasına yol açtığı için Hollanda ve Yunanistan'dan toplanan veriler dahil edilmemiştir. İki ülkenin, kovaryans matrislerinin negatif çıkmasına yol açan sebebi bulabilmek amacıyla veriler incelenmiştir. Wothke'ye (1993) göre kovaryans matrislerinin negatif çıkmasının sebepleri sınırlıdır ve bunlar sırasıyla şunlardır. Birinci sebep değişkenler arasındaki yüksek doğrusal bağlantıdır. Bu durumda analize giren değişkenler birbirleriyle çok yüksek ilişki göstermektedirler. Böyle bir analizde yüksek ilişki gösteren değişkenlerden birisi analizlerden çıkarılır. İkinci sebep ise analiz için kullanılan başlangıç değerlerinin negatif olmasıdır. Bu durumda faktör analizine başlarken maddelerin tahmin edilen faktör yüklerinin analize uygun olmaması söz konusudur. Böyle bir durumda madde faktör yüklerinin yeniden tanımlanması gerekir. Üçüncü durum ise örneklemin varyansıdır. Örneklemin çok küçük olduğu durumlarda örneklem varyansının kararsızlaştığı tespit edilmiştir. Bu tür durumlarda örneklem varyansının sabitlenmesi ya da örneklemin büyütülmesi yoluna gidilmektedir. Dördüncü durum ise kayıp verilerin miktarının çok olmasıdır. Bu durumda kayıp verilerin analizin şartlarına uygun değerlerle değiştirilmesi önerilmektedir. Fakat kayıp verilerin oranının verilerden daha çok olduğu durumlarda bu yol kullanılabilirliğini yitirmektedir. Matrislerin negatif çıkmasına yol açan bir diğer durum ise maddelerin varyanslarının sıfır olmasıdır. Bu durumda ya madde tamamen boş bırakılmış ya da

herkes tarafından aynı doğrultuda işaretlenmiştir. Yukarıdaki olasılıklar dahilinde bu analizde Hollanda ve Yunanistan verilerine ait matrislerin negatif çıkmasına sebep olan durum araştırılmıştır. Olası tüm durumlar gözden geçirilmiş ve sorunun kayıp verilerin çokluğundan kaynaklandığı tespit edilmiştir. Bu iki ülkeden elde edilen veriler incelendiğinde iki ülkede de kimi maddelerde kayıp veri oranı %30'un üzerindedir. Bu maddelerin Hollanda ve Yunanistan verilerinden çıkarılarak analize devam edilmesi yapı geçerliliğinin birçok katılımcı ülke arasındaki denkleğinin belirlenmeye çalışıldığı bir durumda mümkün olmadığından bu ülkelerin analizlerden çıkarılması kararı alınmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 17 hazırlanmıştır.

Tablo 17. Clay ve Leonardo Okuma Parçalarından Oluşan Sekiz Numaralı Alt Test İçin Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2/sd	RMSEA	$\Delta\chi^2_{sd}$	Δ RMSEA
Model A	25293.44 ₁₀₆₆₉	2.37	0.061		
Model B	24355.03 ₉₈₆₉	2.47	0.065	938.41 ₈₃₀ *	0,004
Model C	24264.89 ₉₈₄₄	2.46	0.065	1028.55 ₈₂₅ *	0,004
Model D	25163.46 ₁₀₆₄₄	2.36	0.061	129.98 ₂₅ *	0,000

Model A: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

Model B: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

Model C: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

Model D: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

Tablo 17'de görüldüğü gibi faktör yüklerinin tüm ülkelerde eşit olduğu üzerine kurulu olan Model A'nın RMSEA değeri 0.061 olarak hesaplanmıştır. Faktör yapısının tüm ülkelerde farklı olduğu üzerine kurulu bulunan modelin uyum iyiliği katsayılarının daha iyi bir uyuma işaret edip etmediğini belirlemek amacıyla bu modellerin RMSEA değerleri incelenmiştir. Model B'nin RMSEA değeri 0.065 (Δ RMSEA=0.004) olarak bulunmuş, Model C'nin RMSEA değeri 0.065 (Δ RMSEA=0.004) olarak belirlenmiş ve Model D'nin RMSEA değeri 0.061 (Δ RMSEA=0.000) olarak bulunmuştur. Ayrıca tüm modellerin χ^2/sd oranları da 5'ten düşük ve birbirine yakın bulunmuştur. Model B, Model C ve Model D faktör yük değerlerinin veya hata varyanslarının ülkelerarasında eşit olduğu üzerine kurulan Model A'dan istatistiksel olarak anlamlı ve daha iyi bir

uyum göstermemiştir. Bu sonuçlar doğrultusunda sekiz numaralı ölçek için kabul edilen model, Model A olmuştur. Buna göre sekiz numaralı alt ölçeğin, uygulamaya katılan 35 ülkede dört ayrı kavrama süreciyle ifade edilen boyutlarda ölçme yaptığı söylenebilir. Sekiz numaralı alt testin yapı geçerliliği evrenseldir.

Sekiz numaralı alt testin faktör yapısının tüm ülkelerde birbirine ne derecede uyumlu olduğunu belirlemek amacıyla Model A'nın RMSEA değeri (0.061) incelenmiş, bu alt testin faktör yapısının tüm ülkelerde birbirine orta derecede benzerlikte olduğu belirlenmiştir.

TARTIŞMA

Sıfır, üç, altı ve sekiz numaralı alt testin faktör yapısının tüm ülkelerde birbirine denk olması ile ilgili bulgular, Brown ve arkadaşlarının (2005) yaptıkları çalışmadan elde ettikleri bulguları desteklememektedir. Brown ve arkadaşları, PIRLS maddelerinin farklı kültürlerde farklı IRT modellerine uygun ölçme yaptıklarını bulmuşlar, bu bulguya dayalı olarak ölçeğin faktör yapısının da bu ülkelerde farklı olduğu yorumunda bulunmuşlardır. Bu farkı bulmuş olmalarının olası sebeplerinden biri, çalışmalarını, PIRLS 2001 uygulamasına katılmış 35 ülkeden gelişmiş, gelişmemiş ve diğer ülkelerden ikişer örnek seçerek gerçekleştirmiş olmalarıdır. Oysa bu çalışmada, tüm ülkeler dahil edilerek çok daha geniş bir veri havuzu yaratılmıştır. Bu veri havuzu içindeki farklılıkların ortaya çıkabilmesi olasılığı, çok daha dar kapsamlı ve özellikle iki uçtan seçilmiş gruplar arasında fark elde etme olasılığından daha küçüktür. Bu nedenle PIRLS 2001 testinin alt testlerinden elde edilen bulgular ile Brown ve arkadaşlarının bulguları birbirini desteklememektedir. Ayrıca bu çalışmada kullanılan alt ölçeklerin bazı maddeleri kısmen doğru, oldukça doğru ve tamamen doğru gibi puanlanırken faktör analizi çalışmasında homojenliği sağlayabilmek için tüm bu maddeler 1 ve 0 şeklinde puanlanır hale dönüştürülmüştür. Bu durum maddelerden elde edilen varyansların bir kısmının kaybolmasına yol açmıştır. Bu durum Brown ve arkadaşlarının (2005) yapmış oldukları çalışmanın sonuçları ile bu çalışmanın sonuçlarının birbirlerini desteklememesinin bir sebebi olarak görülebilir.

Johansone ve Foy (2004) yaptıkları çalışmada Avrupa Birliği üyesi ülkeler ve aday olarak bekleyen ülkelerin PIRLS 2001 testi başarılarının birbirine denk olduğunu

bulmuşlardır. Bu karşılaştırmayı faktör yapılarının birbirine denk olup olmadığını test etmeden yapmışlardır. Bu çalışmada ise faktör yapılarının birbirine denk olduğu ve ortalamaların birbirleriyle doğrudan karşılaştırılabileceği gösterilmiştir. Böylece Johansone ve Foy'un (2004) çalışmalarında eksik bıraktıkları önemli bir psikometrik teknik boşluk giderilmiştir.

Görüldüğü gibi sıfır, üç, altı ve sekiz numaralı alt testler tüm ülkelerde birbirine denk faktör yapısıyla ölçme yapmaktadır. Fakat bu denkliğin derecesi çok yüksek değildir. Dört alt testin kabul edilmiş olan Model A için RMSEA değerleri 0.061 ile 0.075 arasında değişmektedir. Bu değerler faktör yapılarının uygulamaya katılan ülkelerde birbirilerine orta düzeyde benzerlik gösterdiğine işaret etmektedir.

Blum ve Guerin-Pace (2000) ile Blum ve arkadaşları (2001) kültürden kaynaklanan yanlılıklar nedeniyle uygulamaya katılan ülkelerde yapı geçerliliklerinin birbirine denk olmayabileceği endişesini bildirmişlerdir. Yapılan bu çalışma ile PIRLS 2001 uygulamasının kültürlerarası geçerliliğinin orta düzeyde sağlandığı gösterilmiştir. Böylece Blum ve Guerin-Pace (2000) ile Blum ve arkadaşlarının (2001) bildirdikleri endişeler giderilmiştir. PIRLS 2001 uygulamasına katılan 35 ülkede faktör yapılarının eşitliğinin dolayısı ile de ölçeğin faktör yapısının katılımcı ülkelerde birbirine denkliğinin orta düzeyde sağlanabildiği, kültürel farklılıkların bu 35 ülkenin uygulamalarında geçerliliği bozan bir etken olmadığı görülmüştür.

Uyumun orta derecede çıkmış olmasının nedenlerini saptamak için uyumsuzluğa yol açmış maddelerin incelenmesi şarttır. Fakat PIRLS 2001 maddeleri açıklanmamıştır. Maddelerin içeriklerinin açıklanması gerekliliği, Hilton'un (2006) önerileriyle ve eleştirileriyle de uyumludur.

SONUÇLAR ve ÖNERİLER

Sıfır, üç alt ve sekiz numaralı alt testlerden oluşan PIRLS 2001 uygulamasından elde edilen verilerin 35 ülke bağlamında karşılaştırılması sonucunda, uygulamanın tüm ülkelerde birbirine yapısal olarak denk olduğu, dolayısı ile PIRLS 2001 testinin ölçme değişmezliğinin tüm ülkelerde sağlanabildiği görülmüştür. Bu değişmezliğin derecesinin orta düzeyde olduğu görülmüştür.

Buna göre PIRLS 2001 testine dayalı olarak ülkelerin, hem alt test hem de toplam puanı bakımından başka bir puan dönüştürmesine gerek kalmadan karşılaştırılması

mümkündür. Ölçme değişmezliği, başka bir ifadeyle ölçeğin yapı geçerliliğinin tüm ülkelerde birbirine denk olduğu kanıtlandığından 35 ülkenin ortalamalarının büyüklük bakımından sıraya dizilebileceği söylenebilir. Dolayısıyla bu ortalamaların birbirleriyle karşılaştırılabilmesi ve daha büyük ortalamaya sahip olan ülkelerin başarılarının daha yüksek olduğunun söylenmesi mümkündür.

Araştırma sürecinde karşılaşılan durumlar bağlamında yapılabilecek öneriler aşağıda listelenmiştir.

* PIRLS 2001 başarı testinin maddeleri iki farklı yapıdadır. Daha sonraki araştırmalarda bu farklı türdeki maddelerin her birinin faktör yapıları ayrıca incelenerek gözlenen farklılıkların hangi tür maddelerden kaynaklanabileceği araştırılmalıdır.

* Ayrıca PIRLS 2001 uygulamasından elde edilen veriler çok çeşitli yetenek gruplarından toplanmıştır. Çeşitli yetenek grupları için analizlerin tekrar edilmesi ve bunların birbirleriyle karşılaştırılması, gözlenen faktör farklılaşmalarının daha çok hangi gruplardan kaynaklandığını görebilmeye yardımcı olacaktır.

* Düşük yetenek grubunda bulunan öğrencilerin verilerin güvenilirliğine olan etkisi düşünüldüğünde bu yetenek grubundan öğrenciler analiz dışı bırakılarak analizlerin tekrar edilmesi faktör yapılarındaki değişimleri yetenek değişkeninden bağımsız olarak inceleyebilmeye imkan tanıyabilir.

KAYNAKÇA

- Baykul, Y. (2000). **Eğitimde ve Psikolojide Ölçme. Klasik Test Teorisi ve Uygulaması.** Ankara: ÖSYM Yayınları.
- Brown, G., Micklewright, J., Schnepf, S. V., Waldmann, R. (2005), Cross-National Surveys of Learning Achievement: How Robust are the Findings?, Southampton Statistical Sciences Research Institute Applications & Policy Working Paper, A05/05. United Kingdom
- Brown, T. A. (2006). **Confirmatory Factor Analysis for Applied Research.** New York: The Guilford Press.
- Cheung, G., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 233-255.

- Cheung, G., & Rensvold, R. B. (2000). Testing measurement invariance using critical values of fit indices: A Monte Carlo study. Web: http://division.aonline.org/rm/cheung_files/cheung.htm adresinden 10 Nisan 2009 tarihinde alınmıştır
- Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 281-302.
- Foy, P. (2004), Intraclass Correlation And Variance Components As Population Attributes And Measures Of Sampling Efficiency In PIRLS 2001. The 1st IEA International Research Conference. Yayınlanmamış kongre bildirisi. Web: <http://www.iea-dpc.org/download/ieahq/IRC2004/Foy.pdf> adresinden 24 Mart 2008 tarihinde alınmıştır.
- George, D., & Mallery, P. (2001). **SPSS for Windows: A Simple Guide and Reference**. Boston: Allyn and Bacon.
- Gonzalez, E. J. and Kennedy, A. M. (2003). PIRLS 2001 User Guide for the International Database. Chestnut hill, Boston College.
- Hilton, M. (2006). Measuring standards in primary English: Issues of validity and accountability with respect to PIRLS and National Curriculum test scores. *British Educational research Journal*, 32 (6), 817-837.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 1-55.
- Johansone ve Foy, 2004 PIRLS 2001 Results In The Context Of The European Union Expansion, The 1st IEA International Research Conference. Yayınlanmamış kongre bildirisi.
Web: http://www.iea-dpc.org/download/ieahq/IRC2004/Johansone_Foy.pdf 24 Mart 2008 Tarihinde Alınmıştır.
- Kline, R. B. (1998). **Principles and Practice of Structural Equation Modeling**. New York: The Guilford Press.
- Martin, M. O. Mullis, I. V. S. Kennedy, A. M. (2003). PIRLS 2001 Technical Report, International Study Center, Lynch School of Education, Boston College, USA
- Mullis, Ina V.S., Martin, Michael O., Gonzalez, Eugenio J., Ann M. (2003), Prils 2001 International Report, International Study Center, Lynch School of Education,

Boston College, USA

Raykov, T., & Marcoulides, G. A. (2006). **A First Course in Structural Equation Modeling**. Routledge Inc.

Rodriguez, M. C. (2006). Construct Equivalence of Multiple-Choice and Constructed-Response Items: A Random Effects Synthesis of Correlations. *Journal of Educational Measurement* , 163-184.

Thompson, B. (2004). **Exploratory and Confirmatory Factor Analysis**. Washington: American Psychological Association.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. K. A. Bollen, & J. S. Long içinde, **Testing structural equation models** (s. 256-293). Newbury Park, CA: Sage Publications.