

## PISA Öğrenci Anketinin Kültürler Arası Eşdeğerliği\*

### Cross-cultural Equivalence of the PISA Student Questionnaire

Mustafa ASİL\*\* Selahattin GELBAL\*\*\*

Ölçme Seçme ve Yerleştirme Merkezi Hacettepe Üniversitesi

Öz

Bu araştırmada Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 kapsamında uygulanan öğrenci anketinin kültürler ve diller arası eşdeğerliği, Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemi üzerinde karşılaştırmalı olarak incelenmiştir. Araştırmanın birinci aşamasında, öğrenci anketinin faktör yapısı Doğrulayıcı Faktör Analizi (DFA) yöntemiyle incelenmiştir. İkinci aşamada ise, Ortalama ve Kovaryans Yapıları (OKY; Sörbom, 1974) metoduna dayanan Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA) kullanılarak anket maddelerinin farklı kültürler ve diller arasında Değişen Madde Fonksiyonu (DMF) gösterip göstermediği araştırılmıştır. Son aşamada, uzman görüşlerine başvurulmuş, maddelerin DMF göstermesinin nedenleri belirlenmeye çalışılmıştır. DFA sonuçları ölçme modelinin tüm örneklerde aynı faktör yapısına sahip olduğunu göstermiştir. ÇG-DFA bulguları ülkeler arasında değişen fonksiyon gösteren maddelerin olduğunu ortaya koymuştur. Ülkeler arasında dilsel ve kültürel farklılıklar arttıkça DMF gösteren maddelerin sayısının da arttığı gözlenmiştir. Maddelerin DMF göstermesinin asıl nedenlerinin çeviri problemleri ve kültürel farklılıklar olduğu sonucuna varılmıştır.

*Anahtar Sözcükler:* PISA, Değişen Madde Fonksiyonu (DMF), Ölçme Eşdeğerliği, Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA), Ortalama ve Kovaryans Yapıları (OKY).

*Abstract*

In this study, cross-lingual and cross-cultural equivalence of the Programme for International Student Assessment (PISA) 2006 Student Questionnaire was assessed comparatively across Turkey, the United States of America, Australia, and New Zealand samples. At the first stage of this study, the factor structure of the questionnaire constructs was examined using Confirmatory Factor Analysis (CFA). At the second stage, Multi-Group Confirmatory Factor Analyses (MG-CFA) based on Mean and Covariance Structures (MACS; Sörbom, 1974) were conducted in order to determine whether the questionnaire items exhibit differential item functioning (DIF) across different cultures or languages. At the last stage, causes of DIF were identified by expert judgments. CFA results showed that the measurement model had the same factor structure across samples. MG-CFA results indicated that some items exhibited DIF between countries. It is observed that as the linguistic and cultural differences increased between countries, the number of DIF items increased. The main reasons for DIF were found to be due to poor translation and differences in cultural relevance.

*Keywords:* PISA, differential item functioning (DIF), measurement equivalence, multi-group confirmatory factor analysis (MG-CFA), mean and covariance structures (MACS).

#### Summary

##### *Purpose*

The OECD Programme for International Student Assessment (PISA) is the most comprehensive global assessment that measures the performance of 15-year-old students in three domains: Reading,

\* Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda 2010 yılında tamamlanan "Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi" adlı doktora tezinden yararlanılarak oluşturulmuştur.

\*\* Dr. Mustafa ASİL, Araştırmacı, Ölçme Seçme ve Yerleştirme Merkezi, asil@hacettepe.edu.tr

\*\*\* Prof. Dr. Selahattin GELBAL, Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, gelbal@hacettepe.edu.tr

Mathematical and Scientific literacy as well as other educational outcomes using test batteries and background questionnaires. More than 50 countries participated in PISA 2006 and measurement instruments had been translated and adapted into about 40 different languages (OECD, 2009).

The accuracy, meaningfulness, and validity of the inferences derived from international large-scale assessments depend on whether the same construct is being measured by different national versions of the assessment instruments (Hambleton, 2005). Previous studies have demonstrated that unless statistical and judgmental evidence is demonstrated, multilingual versions of assessments cannot be assumed to be comparable or equivalent across all studied groups. Cultural or language differences may have a powerful effect on measurement equivalence.

The purposes of this study are to (1) examine the cross cultural measurement equivalence of PISA 2006 student questionnaire constructs across Turkey, the United States of America, Australia, and New Zealand samples both at the scale level and item level, (2) detect the differentially functioning items, and (3) identify the possible sources of DIF.

#### *Method*

Data from PISA 2006 assessment were used for evaluating the equivalence of student questionnaire constructs and items. Data analyses were conducted in three stages. In the first stage, Confirmatory Factor Analyses (CFA) were employed for the pooled sample and for each group separately in order to evaluate whether the same factor structure exists in all countries. In the second stage, Multiple-Group extension of the Confirmatory Factor Analysis with Mean and Covariance Structure (MACS) model was conducted for the detection of differential item functioning (DIF) across different cultures. As proposed by Stark, Chernyshenko, and Drasgow (2006), the loadings (discrimination) and the intercept (location) parameters were simultaneously examined using the likelihood ratio test with a free-baseline model (in which only the parameters of the referent are constrained across groups) and Bonferroni corrected critical p-values.

Even though DIF detection methods are commonly used for assessing construct equivalence of measurement instruments, they do not allow us to identify sources of DIF. Therefore, in the third stage, potential sources of incomparability are examined by judgmental reviews of experts who are proficient in English and Turkish.

#### *Results*

Examination of the factor loadings, goodness of fit indexes, scale reliabilities, and factor correlations indicated that measurement model had the same number of factors and the same number of items and relationship between the items and factors were similar in all countries. DIF analyses with MG-CFA were conducted for three comparisons (Australia vs. New Zealand, Australia vs. the USA, Australia vs. Turkey). Australia-Turkey comparison exhibited the greatest number of DIF (9 items). No DIF items were found for the Australia-New Zealand comparison. Only two items were flagged as showing DIF (lack of measurement equivalence) for the Australia-the USA comparison. Experts found two main reasons for DIF, namely, translation/ adaptation problems and differences in cultural relevance.

#### *Discussion and Conclusion*

This study intended to examine the equivalence of English and Turkish versions PISA 2006 student questionnaire across different cultures. The findings presented here provided strong evidence that as the linguistic and cultural differences increased between countries, the number of DIF items increased as well. Based on the constructs examined, it is concluded that valid comparisons can only be made between Australia and New Zealand.

This study provided a systematic approach to ensure linguistic and cultural equivalence in international comparative assessments by combining quantitative (statistical/psychometric) and qualitative (judgmental reviews) analyses.

### Giriş

Eğitim, sağlık, ekonomi, sanat ve benzeri alanlardaki artan uluslararası işbirliği, göç, küreselleşme gibi birçok nedenden dolayı psikometri uzmanları, eğitimciler ve diğer alanlardaki araştırmacılar uzun zamandır farklı dillerden/kültürlerden bireyleri değerlendirme durumuyla karşı karşıya kalmaktadırlar. Böyle durumlarda tek bir ölçme ve değerlendirme aracının kullanılması neredeyse imkânsızdır. Bu yüzden, ölçme araçları genelde farklı dillere ve kültürlere uyarlanmaktadır (de Klerk, 2008; Sireci, 2005).

Günümüzde başarı, yetenek, kişilik testleri ve anketlerin diğer dillere uyarlanmasına olan ihtiyaç giderek artmaktadır (Hambleton, 2005; Hambleton ve Patsula, 1998). PISA, TIMSS, PIRLS gibi uluslararası tarama araştırmalarına katılan ülkelerin sayısı her uygulamada artış göstermektedir. Örneğin Türkiye'nin de katıldığı, uluslararası en kapsamlı eğitim araştırması olan PISA uygulamalarına 2003'te 41, 2006'da 57 ve 2009'da 74 ülke katılmıştır.

Testler veya anketler bir dilden başka bir dile uyarlandığında farklı dillerdeki formlarının eşdeğerliğinin sağlandığından emin olunmalıdır. Uyarlanan veya çevirisi yapılan her bir form farklı bir yapıyı ölçebilir, zorluk ya da aşinalık dereceleri bakımından farklılık gösterebilir (Hambleton, 1994, 2002, 2005; Sireci ve Berberoğlu, 2000). Testlerin uyarlanmasında amaç, farklılıkları anlamsal, psikometrik, dilsel ve psikolojik açılardan kabul edilebilir seviyelere indirmektir (Hambleton ve de Jong, 2003). Kullanılan her bir madde bir kültürdeki bireyler için ne anlam ifade ediyorsa, diğer kültürlerdeki bireyler için de aynı şeyi ifade etmelidir (Hui ve Triandis, 1985).

Farklı kültür veya dilden bireyleri karşılaştırmak amacıyla yapılan her bir çalışmada, karşılaştırmanın anlamlı olabilmesi için testlerin ölçtüğü yapıların eşdeğer olması gerekir. Bu tür çalışmalarda ölçme eşdeğerliğinin diğer bir ifadeyle ölçme değişmezliğinin\*\*\* sağlanması temel sayılıdır (Gierl, 2000). Bu nedenle Eğitimde ve Psikolojide Ölçme Standartları (AERA, APA ve NCME, 1999) ve Test Uyarlama Yönergeleri (Hambleton, 1994; ITC, 2001) kültürler arası çalışmalar yapan araştırmacılar farklı dillerdeki ölçme araçlarından elde edilen puanların karşılaştırılabilirliği için kanıtlar sunmalarını istemektedir.

Ölçmelerin eşdeğer olması için gözlenen değişkenlerle örtük değişkenler arasındaki ilişkilerin gruplar üzerinde aynı olması gerekmektedir (Drasgow ve Kanfer, 1985). Diğer bir ifadeyle, ölçülen özellik üzerinde eşit yetenekte olan farklı gruplara ait bireyler maddeyi farklı yanıtlıyorlarsa, o madde eşdeğer değildir ya da farklı işliyordur denilir (Chan, 2000). Ölçme eşdeğerliğinin sağlanamaması Madde Tepki Kuramı literatüründe değişen fonksiyon gösterme olarak adlandırılır. Ölçme değişmezliği madde bazında sağlanamıyorsa 'Değişen Madde Fonksiyonu' (DMF), test bazında sağlanamıyorsa 'Değişen Test Fonksiyonu' (DTF) olarak isimlendirilir. Farklı gruplara ait test puanlarının karşılaştırılabilirliği, yani eşdeğerliği genellikle DMF analizleriyle değerlendirilmektedir (Ercikan ve diğerleri, 2004).

Gruplar arası ölçme eşdeğerliğinin gösterilmesinde ve yapısal eşdeğerliğin incelenmesinde çoklu grup doğrulayıcı faktör analizi (ÇG-DFA) uygun ve önerilen bir metot olup çok boyutlu ölçeklere uygulanabilmesi, çok kategorili verilerin analizinde kullanılabilmesi ve hem tek biçimli (uniform) hem de tek biçimli olmayan (non-uniform) DMF'yi belirleyebilmesi gibi birçok yönden diğer metotlara tercih edilmektedir. Bu yöntemle ortak ve kültüre özgü maddeler tespit edilebilmektedir (Byrne ve Stewart, 2006; Cheung ve Rensvold, 1999; Drasgow ve Kanfer, 1985; Zumbo, Sireci ve Hambleton, 2003). DMF analizlerinde ÇG-DFA'nın MTK ile aynı etkililiğe sahip olduğu hatta verinin çok kategorili olduğu durumlarda daha iyi performans gösterdiği simülasyon çalışmalarıyla kanıtlanmıştır (Stark, Chernyshenko ve Drasgow, 2006).

Test puanlarının karşılaştırılabilirliğini değerlendirmede kullanılan psikometrik teknikler, ölçek seviyesinde analizler ve madde seviyesinde analizler olmak üzere iki sınıfa ayrılabilir. Çalışmalarında Zumbo (2003, 2005) sadece ölçek seviyesinde yapılan incelemelerde yanlış çalışan maddelerin belirlenemeyebileceğini göstermiştir ve testlerin çevirisi ve uyarlanması

\*\*\* Bu makalede ölçme eşdeğerliği ve ölçme değişmezliği aynı anlamda kullanılmıştır.

çalışmalarında ölçme eşdeğerliğini sağlamak için madde bazında (DMF) analizlerin de yapılması gerektiği sonucuna varmıştır.

Uluslararası eğitim araştırmalarının en önemli amaçlarından birisi de öğrenci, öğretmen veli ve okul anketleri kullanarak, öğrenci ve öğretmen özellikleri, özgeçmişleri ve etkinlikleri ve okuldaki öğrenme ortamları hakkında bilgi toplamaktır. Öğrencilerin derslere karşı tutum, ilgi ve değer yargıları önemli öğrenme ürünleri olarak görülmekte, öğrenci performanslarındaki farklılıklar, elde edilen bu bilgiler yardımıyla açıklanmaya çalışılmakta ve bu bilgilere dayalı uluslararası karşılaştırmalar yapılmaktadır (MEB, 2010). Türkiye de dahil olmak üzere diğer birçok ülkede eğitim programları elde edilen bu bilgiler ışığında yeniden yapılandırılmaktadır. TIMSS, PISA veya PIRLS gibi çalışmalarda bu amaçla öğrenci, öğretmen veya veli anketi gibi ölçme araçları uygulanmaktadır. Eğitim politikalarında önemli kararların alınmasına ışık tutacak bu tür uluslararası karşılaştırmaların anlamlı olabilmesi için kullanılan testlerin ve anketlerin ölçtüğü yapıların eşdeğer olması gerekmektedir. İlgili literatür incelendiğinde, uluslararası karşılaştırmalarda daha çok bilişsel testlerin eşdeğerliğinin gösterilmesine emek harcanmakta, fakat anketlerin eşdeğerliğiyle ilgili araştırmalara yeterince önem verilmediği görülmektedir (Schulz, 2003; 2005; 2008).

## Yöntem

### *Araştırmanın Amacı*

Bu çalışmanın amacı, PISA 2006 kapsamında uygulanan öğrenci anketinde yer alan yapıların kültürler arası eşdeğerliğinin hem ölçek hem de madde bazında yapılan analizlerle incelenmesidir. Diğer bir amaç ise kültürler arası DMF gösteren maddelerin tespiti ve DMF'nin olası sebeplerinin uzman görüşlerine başvurularak değerlendirilmesidir.

### *Araştırma Grubu*

Araştırma grubunu PISA 2006 uygulamasına katılan Avustralya (AUS), Yeni Zelanda (YZL), Amerika Birleşik Devletleri (ABD) ve Türkiye (TUR) örneklemelerindeki öğrenciler oluşturmaktadır. Çalışmaya bu ülkelerin alınmasının amacı, aynı kültür-aynı dil (AUS-YZL), farklı kültür-aynı dil (AUS-ABD) ve farklı kültür-farklı dil (AUS-TUR) karşılaştırmalarını yapabilmektir. Çalışma grubuna ait frekans ve yüzdeler Tablo 1'de verilmiştir.

Tablo 1.

### *Çalışma Grubu Frekans ve Yüzdeleri*

Gruplar	Frekans	Yüzde
Avustralya	11.642	48,1
Yeni Zelanda	3.500	14,5
ABD	4.518	18,7
Türkiye	4.535	18,7
Toplam	24195	100,0

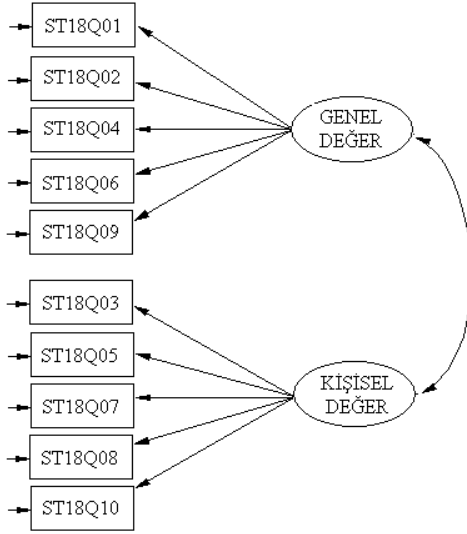
### *Veri Toplama Aracı*

Araştırma, Ekonomik İşbirliği ve Kalkınma Örgütü'nün (OECD) 2006 yılında düzenlemiş olduğu Uluslararası Öğrenci Değerlendirme Programı (PISA) kapsamında uygulanan öğrenci anketinden elde edilen veriler üzerinde yürütülmüştür. Çalışma için gerekli olan veri, OECD PISA internet sitesinden ([www.pisa.oecd.org](http://www.pisa.oecd.org)) elde edilmiştir.

PISA 2006 uygulaması OECD'ye üye 30 ülke ile üye olmayan 27 ülkeyi kapsamaktadır. Bu ülkelerdeki yaklaşık 20 milyon 15 yaş grubu öğrenciyi temsil edecek şekilde yaklaşık 400.000 öğrenci rasgele seçilmiştir. İki-basamaklı tabakalı örnekleme yöntemi kullanılarak önce okullar daha sonra ise öğrenciler rasgele seçilmiştir. Yazılı sınav uygulaması yanında öğrencilere

özgeçmişleri, öğrenme alışkanlıkları, motivasyonları ve fen bilimlerine yönelik tutumlarına ilişkin, doldurması yaklaşık 30 dakika süren bir öğrenci anketi uygulanmıştır. Ankette çoğunluğu iki ve dört kategorili Likert tipinde kodlanmış 37 soru bulunmaktadır (MEB, 2007; OECD, 2005, 2009; Turner ve Adams, 2007).

Bu çalışmada, PISA 2006 öğrenci anketinde ölçülmeye çalışılan öğrencilerin fen bilimlerine yönelik tutumlarından “Bilimsel Sorgulamaya Verilen Destek” alt boyutunun Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemelerinde eşdeğer olarak ölçülüp ölçülmediği belirlenmeye çalışılmıştır. Çalışmada incelenen ölçme modeli Şekil 1’de verilmiştir.



Şekil 1. Bilimsel Sorgulamaya Verilen Destek Alt Boyutu Ölçme Modeli

Bilimsel sorgulamaya verilen destek alt boyutu ölçme modeli her biri beşer maddeli fen bilimlerinin genel değeri (general value of science) ve fen bilimlerinin kişisel (personal value of science) değeri alt boyutlarından oluşmaktadır. Ankette yer alan ST18Q01, ST18Q02, ST18Q04, ST18Q06 ve ST18Q09 kodlu maddeler öğrencilerin fen bilimlerine ilişkin genel değer yargılarını, ST18Q03, ST18Q05, ST18Q07, ST18Q08 ve ST18Q10 kodlu maddeler ise öğrencilerin fen bilimlerine ilişkin kişisel değer yargılarını ölçmektedir. Her iki boyutu da oluşturan tüm maddeler dörtlü Likert tipindedir (OECD, 2007). Maddeler Türkçe formunda tümüyle katılıyorum, katılıyorum, katılmıyorum ve hiç katılmıyorum, İngilizce formunda ise strongly agree, agree, disagree ve strongly disagree şeklinde yanıt kategorilerinden oluşmaktadır. Bilimsel sorgulamaya verilen destek tutumunu ölçen maddelerin Avustralya, ABD ve Yeni Zelanda örneklemelerinde kullanılan İngilizce versiyonu ve Türkiye örnekleminde kullanılan Türkçe versiyonu Ek 1’de verilmiştir.

#### Verilerin Analizi

Bu çalışmada sırasıyla aşağıdaki analizler yapılmıştır:

1. DFA sayıltılarının test edilmesi
2. DFA analizleri
3. ÇG-DFA ile DMF gösteren maddelerin belirlenmesi
4. DMF’nin olası sebeplerinin incelenmesi

Araştırma problemlerine ilişkin analizlere geçilmeden önce her bir ülkeye ait veri seti, kayıp veri, aykırı değerler, doğrusallık (linearity) ve tekli ve çoklu bağlantılılık (singularity, multicollinearity) sayıltıları açısından ayrı ayrı incelenmiştir. Bu incelemeler sonucunda DFA sayıltılarının karşılandığı görülmüştür. Parametre kestirimlerinde ağırlıklandırılmış en küçük

kareler yöntemi (WLS) kullanıldığı (Jöreskog ve Sörbom, 1999, 2001, 2002) ve bu yöntemin de normallik sayılıtsı olmadığı için veri setlerinde çok değişkenli normallik sayılıtsı incelenmemiştir.

Uyarlaması yapılan maddelerin gruplar arasında DMF gösterip göstermediğinin belirlenmesinden önce ölçülen faktör yapısı her bir grup için ayrı ayrı incelenmelidir (Sireci ve Swaminathan, 1996). Çalışmadaki ölçme modelinin her bir ülke verisindeki uyumunun değerlendirilmesinde modele ait faktör yükleri, genel uyum indeks değerleri (Ki-kare, RMSEA, SRMR, CFI ve NNFI), faktör güvenilirlikleri (Hair, Anderson, Tahtam ve Black, 1998) ve faktörler arası korelasyonlar incelenmiştir.

#### ÇG-DFA ile DMF Gösteren Maddelerin Belirlenmesi

Maddelerin ve faktör yapılarının gruplar arasındaki ölçme değişmezliği genellikle ÇG-DFA ile test edilmektedir. Literatürde önerildiği gibi (Byrne, 2008; Little, 1997; Meredith, 1993) ölçme değişmezliği analizleri kovaryans modeli yerine ortalama ve kovaryans yapıları (OKY) metoduyla yapılmıştır.

OKM şu şekilde gösterilebilir:

$$x_{ij} = \tau_j + \lambda_{j1}\xi_{1i} + \lambda_{j2}\xi_{2i} + \dots + \lambda_{jp}\xi_{pi} + \delta_{ij} \quad (1)$$

Bu denklemde  $x_{ij}$ , i'nci bireyin ( $i=1\dots N$ ), j'nci gözlenen değişkendeki puanını ( $j=1\dots N$ ) ifade etmektedir. Her bir yanıtın, regresyon sabiti,  $\tau_j$ , bir veya daha fazla faktör,  $\xi_{pi}$ , ( $p=1\dots P$ ) ve normal dağılmış tesadüfi hata terimlerinin,  $\delta_{ij}$ , lineer bir kombinasyonu olduğu varsayılır. Regresyon katsayıları,  $\lambda_{jp}$ , j maddesinin p faktöründeki faktör yükleri ve regresyon sabiti,  $\tau_j$ , faktör puanının 0 olduğu durumdaki  $x_{ij}$  puanıdır. Çoklu grup uygulamalarında bu modelin birden fazla grupta geçerli olup olmadığı test edilir.

ÇG-DFA ile DMF analizi, temel model ile kademeli olarak model parametreleri gruplar üzerinde eşit olacak şekilde kısıtlanarak veya serbest bırakılarak oluşturulan iç içe modellerin test edilmesini içerir. İç içe modellerin karşılaştırılmasında ise ki-kare değerleri kullanılır. Kısıtlanan veya serbest bırakılan parametrelerin model uyumuna anlamlı bir katkısının olup olmadığına bakılır. Ki-kare değeri anlamsız ise ölçme değişmezliğinin sağlandığı kabul edilir (Wu, Li ve Zumbo, 2007).

Bu çalışmada OKM'ye dayanan ÇG-DFA ile DMF gösteren maddeleri belirleyebilmek için Stark, Chernyshenko ve Drasgow (2006) tarafından önerilen Bonferroni düzeltmeli serbest temel modeli yaklaşım kullanılmıştır. Bu yaklaşımda sadece referans değişkeninin parametrelerinin sınırlandırılıp diğer tüm maddelerin parametrelerinin ( $\lambda, \tau$ ) serbest bırakıldığı bir temel model oluşturulur. Daha sonra ardı ardına her seferde birer maddenin sınırlandırıldığı modellerle temel model arasındaki ki-kare farklılıkları incelenir. Anlamlı olmayan ki-kare farklılıkları maddenin ölçme değişmezliği sağladığını ifade eder. Referans değişkenin belirlenmesinde de yine yazarlar tarafından önerilen yöntem uygulanmıştır.

OKM'de, regresyon sabiti ( $\tau$ ) MTK'daki madde zorluğu parametresini (b), faktör yükü ( $\lambda$ ) ise madde ayırt ediciliği parametresini (a) temsil etmektedir. Faktör yükü ve/veya regresyon sabitlerinde değişmezliğin sağlanamaması DMF'nin varlığını göstermektedir. Regresyon sabitlerindeki farklılıklar tek biçimli (uniform), faktör yüklerindeki farklılıklar ise tek biçimli olmayan (non-uniform) DMF olarak adlandırılmaktadır (Chan, 2000). Faktör yüklerinin ( $\lambda$ ) ve regresyon sabitlerinin ( $\tau$ ) değişmezliği sağlandığında test puanları farklı kültürler arasında problemsiz ve geçerli bir şekilde karşılaştırılabilir (Schmitt ve Kuljanin, 2008; Van de Vijver ve Poortinga, 2005).

Analiz sonucunda maddelerin gruplar arasında değişen fonksiyon göstermesinin kaynağının saptanması amacıyla üniversitelerin yabancı diller eğitimi ya da ölçme-değerlendirme bölümlerinde görev yapan anadili İngilizce veya Türkçe olan ve her iki dilde de yetkin olan 10 akademisyenin

görüşüne başvurulmuştur. Uzman görüşlerini alırken Allalouf, Hambleton ve Sireci'nin (1999) geliştirdiği madde inceleme formu bu çalışma için modifiye edilerek kullanılmıştır. Maddeleri incelemeye önce uzmanlara DMF hakkında özet bilgi verilmiş ve DMF'li maddelere örnekler gösterilmiştir. Tüm uzmanlar birbirlerinden bağımsız olarak maddeleri incelemiş ve madde inceleme formundaki soruları DMF gösteren her bir madde için yanıtlamışlardır.

## Bulgular ve Yorum

### I. Aşama: Doğrulayıcı Faktör Analizi

Bilimsel sorgulamaya verilen destek alt boyutu ölçme modeli faktör yapısının her bir kültür içinde de geçerli olup olmadığını incelemek için model uyumu birleştirilmiş veri ve her bir ülke verisi için ayrı ayrı değerlendirilmiştir. DFA sonucunda elde edilen faktör yükleri 0,05 anlamlılık düzeyinde değerlendirildiğinde hepsinin anlamlı olduğu görülmüştür. Faktör yükleri Avustralya örneklemini için (0,68-0,93), Yeni Zelanda örneklemini için (0,62-0,91), ABD örneklemini için (0,72-0,89) ve Türkiye örneklemini için (0,65-0,83) arasında değerler almıştır. Ölçme modeline ait genel uyum indeks değerleri ise Tablo 2'de özetlenmiştir.

Tablo 2.

*Bilimsel Sorgulamaya Verilen Destek Alt Boyutu Ölçme Modeli Uyum İstatistikleri*

	$\chi^2$	df	RMSEA	CFI	NNFI	SRMR
Birleşik veri	2665,099*	34	0,057	0,950	0,934	0,079
Avustralya	1512,769*	34	0,061	0,960	0,947	0,090
Yeni Zelanda	475,228*	34	0,061	0,949	0,932	0,092
ABD	475,351*	34	0,054	0,946	0,929	0,078
Türkiye	469,880*	34	0,053	0,943	0,925	0,073

Not. sd = serbestlik derecesi, RMSEA= root mean square error of approximation , CFI= comparative fit index , NNFI=non-normed fit index , SRMR = standardized root mean square residual.

\* $p < 0,01$

Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri alt boyutlarından oluşan iki faktörlü ölçme modelinin uyumu Tablo 2'de verilen ki-kare ve indeks değerleri kullanılarak incelenmiştir. RMSEA ve SRMR değerleri sırasıyla 0,08 ve 0,10'un altında ve CFI ile NNFI değerleri de 0,90'ın üstünde olduğundan iki faktörlü ölçme modeline ait indeks değerlerinin hepsinin kabul edilebilir uyum kriterlerini sağladığı görülmüştür (Brown, 2006).

Ölçme modelinin her bir ülkedeki uyumunun değerlendirilmesinde, modeldeki faktörlerin güvenilirlikleri ve faktörler arası ilişkiler de incelenmiştir. Tüm ülkeler için fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerinin yapı güvenilirliklerinin (0,87-0,92) aralığında ve açıklanan varyans oranlarının (0,60-0,72) aralığında değerler aldığı ve yeterli düzeyde olduğu (Hair, Anderson, Tahtam ve Black, 1998) tespit edilmiştir. Faktörler arası korelasyonlar incelendiğinde, fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerinin ülkeler arasında (0,82-0,87) aralığında korelasyon gösterdiği ve tüm ülkelerde oldukça yüksek ilişkili olduğu gözlenmiştir.

Model uyum indeksleri, faktör yükleri, faktör ve madde güvenilirlikleri incelendiğinde, ölçme modelinin her bir kültürde iki faktörlü olduğu ve maddelerin faktörlerle aynı örüntüye sahip olduğu sonucuna varılmıştır. Bununla birlikte, iki faktörlü ölçme modelinin tüm gruplarda oldukça iyi uyum vermesi, model parametrelerinin gruplar arasında eşdeğer olacağı anlamına gelmemelidir.

### II. Aşama: ÇG-DFA ile DMF Analizi

Bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin DMF analizi sonuçları Tablo 3'te verilmiştir. Referans değişkeni olarak kullanılan maddeler tabloda yer almamaktadır.

Tablo 3.

*Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Alt Boyutu DMF Analizi Sonuçları*

		$\chi^2$	$\Delta\chi^2$	$\Delta sd$	p	DMF
AUS - YZL	Serbest temel model	1987,998	-	-	-	-
	GD					
	ST18Q02	1988,229	0,231	2	0,891	Yok
	ST18Q04	1991,048	3,050	2	0,218	Yok
	ST18Q06	1989,148	1,150	2	0,563	Yok
	ST18Q09	1991,181	3,183	2	0,204	Yok
	KD					
	ST18Q03	1993,574	5,576	2	0,062	Yok
	ST18Q05	1989,523	1,525	2	0,466	Yok
	ST18Q07	1989,085	1,087	2	0,581	Yok
ST18Q10	1988,490	0,492	2	0,782	Yok	
AUS - ABD	Serbest temel model	1988,120	-	-	-	-
	GD					
	ST18Q01	1990,034	1,914	2	0,384	Yok
	ST18Q04	1990,287	2,167	2	0,339	Yok
	ST18Q06	1988,643	0,523	2	0,770	Yok
	ST18Q09	1994,381	6,261	2	0,044	Yok
	KD					
	ST18Q03	2016,864	28,744	2	0,000	Var
	ST18Q07	1988,147	0,027	2	0,987	Yok
	ST18Q08	1999,802	11,682	2	0,003	Var
ST18Q10	1989,499	1,379	2	0,502	Yok	
AUS - TUR	Serbest temel model	1982,649	-	-	-	-
	GD					
	ST18Q01	2022,892	40,243	2	0,000	Var
	ST18Q02	2035,419	52,770	2	0,000	Var
	ST18Q04	2014,059	31,410	2	0,000	Var
	ST18Q09	2060,827	78,178	2	0,000	Var
	KD					
	ST18Q03	2074,006	91,357	2	0,000	Var
	ST18Q05	1999,324	16,675	2	0,000	Var
	ST18Q08	2007,537	24,888	2	0,000	Var
ST18Q10	2014,681	32,032	2	0,000	Var	

Not. sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD = fen bilimlerinin kişisel değeri.

Avustralya ve Yeni Zelanda örneklemeleri üzerinde yapılan karşılaştırmada, fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri maddelerine ait ki-kare farklılıkları Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde değerlendirildiğinde hiçbir maddenin DMF içermediği sonucuna varılmıştır.

Avustralya ve ABD karşılaştırmasında ise ST18Q03 ve ST18Q08 kodlu maddelerin DMF gösterdiği gözlenmiştir. Bu ülkelerde aynı dil kullanılmasına karşın kültürel açıdan farklılıklar olduğundan, bu iki maddenin Avustralya ve ABD'deki öğrenciler tarafından aynı şekilde



algılanmadığı ve dolayısıyla DMF gösterdiği düşünülmüştür.

Avustralya ve Türkiye örneklemeleri arasında ST18Q07 kodlu madde hariç diğer tüm maddeler DMF göstermiştir. Bu ülkelerde hem kullanılan dil hem de kültürel açıdan büyük farklılıklar olduğundan, bu maddelerin Avustralya ve Türkiye'deki öğrenciler için aynı anlamı vermediği ve dolayısıyla DMF gösterdiği düşünülmüştür. Avustralya-Türkiye karşılaştırmasına ait ki-kare farklılıklarının diğer karşılaştırmalara oranla oldukça yüksek olduğu görülmüştür.

### III. Aşama: DMF'nin Olası Sebepleri

Maddelerin Avustralya-Türkiye ve Avustralya-ABD arasında DMF göstermesinin olası nedenlerine yönelik uzman görüşleri iki kategoride toplanıp Tablo 4'te özetlenmiştir.

Tablo 4.

#### Maddelerin DMF Göstermesinin Olası Nedenlerine İlişkin Uzman Görüşleri Dağılımı

		ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
AUS-TUR	Çeviri/Uyarlama Problemleri	3	9	10	3	8	7	-	8	4	8
	Kültürel Farklılıklar	0	0	6	0	3	0	-	7	6	0
AUS-ABD	Kültürel Farklılıklar	-	-	7	-	-	-	-	7	-	-

Tablodan da görülebileceği gibi, uzmanlara göre maddelerin DMF göstermesinin asıl nedenleri çeviri/uyarlama problemleriyle kültürel farklılıklar olarak belirlenmiştir. Avustralya-Türkiye karşılaştırması için uzmanlar DMF gösteren maddelerin tamamında çeviri/uyarlama problemleri olduğu yönünde görüş bildirmişlerdir. Avustralya-ABD karşılaştırmasında her iki soru için de yedi uzman kültürel farklılıkların DMF'ye neden olabileceğini belirtmiştir. Elde edilen bu sonuçlar literatürle de paralellik göstermektedir. Ölçme araçlarının kültürler arası eşdeğerliğiyle ilgili çalışmalar incelendiğinde, maddelerin kültürler ve diller arası DMF göstermesinin olası nedenleri arasında çeviri ve uyarlama problemleri (Allalouf, Hambleton & Sireci, 1999; Ercikan, 1998, 2002; Ercikan & diğerleri, 2004; Grisay & diğerleri, 2007; Grisay & Monseur, 2007; Hambleton, 1994; Sireci & Allalouf, 2003; Yıldırım, 2006; Yıldırım & Berberoğlu, 2009; Van de Vijver & Tanzer, 2004) ve kültürel farklılıklar (Allalouf, Hambleton & Sireci, 1999; Gierl & Khaliq, 2001; Grisay 2003; Grisay & Monseur, 2007; Vijver & Tanzer, 2004; Walker, 2007) en önemli sebepler arasında yer almaktadır.

Uzmanlar maddelerin değişen fonksiyon göstermesinin olası nedenlerini; sözcük veya cümle düzeyinde çeviri hataları, orijinal cümledeki anlam belirsizliği, belirsiz anlam içeren sözcük veya ifadelerin kullanılması, kelimelerin farklı dillerdeki değişken zorluğu (aşinalığı), dile veya kültüre özgü ifadelerin kullanılması, kelimelerin kültürlerde farklı anlamlar içermesi şeklinde tanımlamışlardır.

Sözcük veya cümle düzeyinde çeviri hatasına örnek olarak İngilizce formunda "natural world" olarak verilen bir ifadenin Türkçeye "doğal dünya" olarak çevrilmesi verilebilir. Uzmanlar bu ifadenin doğa ya da tabiat olarak çevrilmesi gerektiğini belirtmişlerdir. Ankette kullanılan "some concepts - bazı kavramlar" ya da "things around me - çevremdeki şeyler" gibi ifadelerin farklı kişiler ve kültürler için farklı anlamlar içerebileceği, bu yüzden de bu maddelerin DMF göstermiş olabileceği yönünde uzmanlar ortak görüşte bulunmuşlardır.

Görüşlerine başvurulmuş akademisyenlerden Amerikalı uzmana göre örneğin "things around me" ifadesi Amerika'da somut şeyler için kullanılmaktadır. Türk kültüründe ise "şey"

kelimesi hem somut hem de soyut anlamlar içermektedir. Benzer şekilde bu ifade Avustralya kültüründe de soyut anlamlar içerecek şekilde kullanılıyor olabilir. Başka bir maddede kullanılan "adult" kelimesi Türkçe formunda "yetişkin" olarak tercüme edilerek kullanılmıştır. "Adult" kelimesi İngilizcede günlük dilde daha sık kullanıldığından, uzmanlar 15 yaş grubu öğrencileri için bu iki kelimenin aynı zorlukta olmayabileceği ve Avustralyalı ve Türk öğrenciler tarafından farklı şekilde algılanabileceği, bunun da DMF kaynağı olabileceği şeklinde görüş bildirmişlerdir. İngilizce formunda "social benefits" olarak verilen ve Türkçeye "toplumsal yararlar" olarak çevrilen ifadenin de farklı kültürdeki bireyler için farklı anlamlar içerebileceği uzmanlar tarafından belirtilmiştir.

### Sonuç ve Öneriler

Öğrencilerin fen bilimlerine yönelik tutumlarından birisi olan bilimsel sorgulamaya verilen destek alt boyutu PISA uygulamalarında, fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörleriyle ölçülmüştür. İki faktörlü ölçme modelinin madde ortalamaları ve standart sapmaları incelendiğinde, tüm öğrencilerin fen bilimlerine genel ve kişisel olarak değer verdikleri belirlenmiştir. Tüm ülke öğrencilerinin yanıtlarında benzer bir örüntü olduğu gözlenmiştir. Analizler sonucunda ölçme modelinin her bir kültürde iki faktörlü olduğu ve maddelerin faktörlerle benzer örüntüye sahip olduğu sonucuna varılmıştır.

Bilimsel sorgulamaya verilen destek alt boyutu ölçme modeline ait maddeler için Avustralya ve Yeni Zelanda örneklemi arasında yapılan analizler sonucunda hiçbir maddenin DMF içermediği sonucuna varılmıştır. Avustralya ve ABD örneklemi karşılaştırmasında, fen bilimlerinin kişisel değeri faktörüne ait ST18Q03 ve ST18Q08 kodlu iki maddenin bu kültürler arasında eşdeğer olmadığı belirlenmiştir. Fen bilimlerinin genel değerini ölçen maddelerde DMF tespit edilmemiştir. Avustralya ve Türkiye karşılaştırmasında ise ST18Q07 kodlu madde hariç diğer tüm maddelerin bu kültürler arasında değişen fonksiyon gösterdiği görülmüştür. Maddelere ait ki-kare farklılık değerlerinin, diğer bir ifadeyle DMF miktarlarının, diğer karşılaştırmalara oranla oldukça yüksek olduğu tespit edilmiştir.

Bu çalışmadaki DMF analizlerinin ortaya koyduğu en önemli sonuç, dil ve kültür farklılıklarının maddelerin DMF göstermesinde oldukça etkili olabileceğidir. DMF analizleri bu farklılıklara göre incelendiğinde Tablo 5'te verilen sonuçlara ulaşılmıştır.

Tablo 5.

#### *Kültür ve Dil Farklılıklarına Göre DMF Gösteren Maddelerin Dağılımı*

	Madde sayısı	DMF'li madde sayısı	% DMF
Aynı dil – Benzer kültür (AUS-YZL)	10	0	0
Aynı dil – Farklı kültür (AUS-ABD)	10	2	20
Farklı dil – Farklı kültür (AUS-TUR)	10	9	90

Tablodan da görüldüğü gibi kültürel ve dilsel farklılıklar arttıkça değişen fonksiyon gösteren maddelerin sayısı da artmaktadır. Elde edilen bulgular benzer çalışmalar tarafından da desteklenmektedir (Ercikan ve Koh, 2005; Grisay ve diğerleri, 2007; Grisay ve Monseur 2007).

Özetle, DFA sonuçları ölçme modelinin faktör yapısının tüm kültürlerde benzer yapıda olduğunu desteklemiştir. Diğer yandan ortalama ve kovaryans yapıları metoduyla yapılan ÇG-DFA sonuçları ise Avustralya-ABD ve Avustralya-Türkiye karşılaştırmalarında bazı maddelerin eşdeğer olmadığını, yani DMF gösterdiğini ortaya koymuştur. Bilimsel sorgulamaya verilen destek tutumuna ilişkin, ülkeler veya kültürler arası karşılaştırmaların sadece Avustralya ve Yeni

Zelanda arasında yapılabileceği sonucuna varılmıştır. Avustralya-Türkiye ve Avustralya-ABD karşılaştırmalarının tüm ölçek bazında yapılması ise anlamlı olmayacaktır. ITC Test Uyarlama Yönergelerinde (Yönerge 21; ITC, 2001) de farklı gruplar üzerindeki karşılaştırmaların ancak ölçme değişmezliğinin sağlandığı ölçekler için yapılabileceği belirtilmektedir.

Avustralya-Türkiye karşılaştırmasında, maddelerin DMF göstermesinin asıl nedeninin çeviri ve uyarlama hataları olduğu belirlenmiştir. Ayrıca kültürel farklılıklardan dolayı, bazı maddelerin bu ülke öğrencileri tarafından farklı şekilde algılanmış olabileceği sonucuna varılmıştır. Belirsiz anlam içeren ifadelerin kullanılmasından dolayı iki maddenin Avustralyalı ve ABD'li öğrenciler için aynı anlamı ifade etmediği ve DMF gösterdiği tespit edilmiştir. Kültürel ve dilsel farklılıklar arttıkça ölçme değişmezliği sağlamayan maddelerin sayısının ve sergiledikleri DMF miktarının arttığı gözlenmiştir.

Maddelerin kültürel farklılıklardan çok, çeviri ve uyarlamadan kaynaklı problemlerden dolayı DMF gösterdiği sonucu dikkate alındığında, özellikle ülkemizde PISA uygulamalarından sorumlu MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı'nın uygulamalarda görev alan çevirmenlerin seçimi, eğitimi ve PISA çalışmalarında çevirmenlerle alan uzmanları ve ölçme-değerlendirme uzmanlarının ortak çalışmasının sağlanması konusuna daha fazla önem vermesi gerektiği görülmektedir.

Bu çalışmada, DMF gösteren maddelerin tespit edilmesinde sadece bir yöntem kullanılmıştır. Hambleton ve Patsula (1999), kültürler arası karşılaştırmalarda, testin farklı dil versiyonlarının eşdeğerliğinden emin olabilmek için bir veya birden fazla yöntemle DMF analizlerinin yapılmasını önermektedirler. Araştırmacılar, Madde Tepki Kuramına dayalı DMF belirleme yöntemlerini, Mantel-Haenszel ya da lojistik regresyon gibi diğer yöntemleri de kullanarak değişen fonksiyon gösteren maddeleri birden fazla yöntemle belirleyip sonuçları karşılaştırabilirler.

#### Kaynakça

- Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the Causes of DIF in Translated Verbal Items. *Journal of Educational Measurement*, 36, 185-198.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Byrne, B. M. (2008). Testing for Multigroup Equivalence of a Measuring Instrument: A Walk Through the Process. *Psicothema*, 20 (4), 872-882.
- Byrne, B. M., & Stewart, S. M. (2006). The MACS Approach to Testing for Multigroup Invariance of a Second-Order Structure: A Walk Through the Process. *Structural Equation Modeling*, 13(2), 287-321.
- Chan, D. (2000). Detection of Differential Item Functioning on the Kirton Adaptation-Innovation Inventory Using Multiple-Group Mean and Covariance Structure Analyses. *Multivariate Behavioral Research*, 35, 169-199.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method. *Journal of Management*, 25, 1-7.
- De Klerk, G. (2008). Cross Cultural Testing. In M. Bron, C.D. Foxcroft & R. Butter (Eds.), *Online Readings in Testing and Assessment*, International Test Commission, <http://www.intestcom.org/Publications/ORTA.php>
- Drasgow, F., & Kanfer, R. (1985). Equivalence of Psychological Measurement in Heterogeneous Populations. *Journal of Applied Psychology*, 70(4), 662-680.
- Ercikan, K. (1998). Translation Effects in International Assessments. *International Journal of Educational Research*, 29, 543-553.

- Ercikan, K. (2002). Disentangling the Sources of Differential Item Functioning in Multilingual Assessments. *International Journal of Testing*, 2 (3&4), 199-215.
- Ercikan, K., Gierl, M. J., McCreith, T., Puhan, G., & Koh, K. (2004). Comparability of Bilingual Versions of Assessments: Sources of Incomparability of English and French Versions of Canada's National Achievement Tests. *Applied Measurement in Education*, 17(3), 301-321.
- Ercikan, K., & Koh, K. (2005). Examining the Construct Comparability of the English and French Versions of TIMSS. *International Journal of Testing*, 5(1), 23-35.
- Gierl, M. J. (2000). Construct Equivalence on Translated Achievement Tests. *Canadian Journal of Education*, 25(4), 280-296.
- Gierl, M., J., & Khaliq, S., N. (2001). Identifying Sources of Differential Item and Bundle Functioning on Translated Achievement Tests: A Confirmatory Analysis. *Journal of Educational Measurement*, 38(2), 164-187.
- Grisay, A. (2003). Translation Procedures in OECD/PISA 2000 International Assessment. *Language Testing*, 20(2), 225-240
- Grisay, A., de Jong, J. H. A. L., Gebhardt, E., Berenzer, A., & Halleux-Monseur, B. (2007). Translation Equivalence Across PISA Countries. *Journal of Applied Measurement*, 8 (3), 249-266.
- Grisay, A. & Monseur, C. (2007). Measuring the Equivalence of Item Difficulty in the Various Versions of an International Test. *Studies in Educational Evaluation*, 33, 69-86.
- Hair, J. F., Tatham, R. L., Anderson, R. E. & Black, W. (1998). *Multivariate Data Analysis* (5th ed.). New York: Prentice Hall.
- Hambleton, R.K. (1994). Guidelines for Adapting Educational and Psychological Tests: A Progress Report. *European Journal of Psychological Assessment*, 10, 229-240.
- Hambleton, R.K. (2002). Adapting Achievement Tests into Multiple Languages for International Assessments. In A. Porter y A. Gamoran (Ed.), *Methodological Advances in Large-scale Cross-national Education Surveys* (pp. 58-79) Washington: National Academy of Sciences.
- Hambleton, R.K. (2005). Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Hillsdale, NJ: Lawrence Erlbaum.
- Hambleton, R.K., & de Jong, J. H. A. L. (2003). Advances in Translating and Adapting Educational and Psychological Tests. *Language Testing*, 20, 127-134.
- Hambleton, R. K., & Patsula, L. (1998). Adapting Tests for Use in Multiple Languages and Cultures. *Social Indicators Research*, 45, 153-171.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the Validity of Adapted Tests: Myths to be Avoided and Guidelines for Improving Test Adaptation Practices. *Journal of Applied Testing Technology*, 1(1), 1-30.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-Cultural Psychology: A Review and Comparison of strategies. *Journal of Cross-Cultural Psychology*, 16(2), 131-152.
- International Test Commission (2001). *International Test Commission Guidelines for Test Adaptation*. London: Author.
- Jöreskog, K. G., & Sörbom, D., (1999). *Lisrel 8.30: Structural Equation Modeling with the Simplis Command Language*, Chicago, IL: Scientific Software International, Inc.
- Jöreskog, K. G., & Sörbom, D. (2001). *Lisrel 8: User's Reference Guide*. Chicago: Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (2002). *PRELIS 2: User's Reference Guide*. Chicago: Scientific Software International.
- Little, T. D. (1997). Mean and Covariance Structures (MACS) Analyses of Cross-cultural Data: Practical

- and Theoretical Issues. *Multivariate Behavioral Research*, 32, 53-76.
- MEB (2007). *PISA 2006 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Raporu*. MEB, Ankara.
- MEB (2010). *PISA 2009 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Raporu*. MEB, Ankara.
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrica*, 58, 525-543.
- OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1 and Volume 2*. Paris: OECD Publications.
- OECD (2009). *PISA 2006 Technical Report*. Paris: OECD Publications.
- PISA Web Site: [www.pisa.oecd.org](http://www.pisa.oecd.org)
- Schmitt, N., & Kuljanin, G. (2008). Measurement Invariance: Review of Practice and Implications. *Human Resource Management Review*, 18, 210-222.
- Schulz, W. (2003). Validating Questionnaire Constructs in International Studies. Two Examples from PISA 2000. *Paper Presented at the Annual Meetings of the American Educational Research Association (AERA)*, Chicago, 21-25 April.
- Schulz, W. (2005). Testing Parameter Invariance for Questionnaire Indices Using Confirmatory Factor Analysis and Item Response Theory. *Paper Presented at the Annual Meetings of the American Educational Research Association (AERA)*, San Francisco, 7-11 April.
- Schulz, W. (2008). Questionnaire Construct Validation in the International Civic and Citizenship Education Study. *Paper presented to the 3rd IEA International Research Conference*, Taipei, September.
- Sireci, S. G. (2005). Using Bilinguals to Evaluate the Comparability of Different Language Versions of a Test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 93-115). Hillsdale, NJ: Lawrence Erlbaum.
- Sireci, S. G., & Allalouf, A. (2003). Appraising Item Equivalence Across Multiple Languages and Cultures. *Language Testing* 20(2), 148-166.
- Sireci, S. G., & Berberoğlu, G. (2000). Using Bilingual Respondents to Evaluate Translated – Adapted Items. *Applied Measurement in Education*, 13(3), 229-248.
- Sireci, S. G., & Swaminathan, H. (1996). Evaluating Translation Equivalence: So What's the Big Dif? *Paper Presented at the Annual Meeting of the Northeastern Educational Research Association*, Ellenville, NY, October.
- Sörbom, D. (1974). A General Method for Studying Differences in Factor Means and Factor Structures between Groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting Differential Item Functioning with Confirmatory Factor Analysis and Item Response Theory: Toward a Unified Strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston MA: Allyn & Bacon.
- Turner, R., & Adams, R. J. (2007). The Programme for International Student Assessment: an Overview. *Journal of Applied Measurement*, 8(3), 237-248.
- Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and Methodological Issues in Adapting Tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39-63). Hillsdale, NJ: Lawrence Erlbaum.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and Equivalence in Cross-Cultural Assessment. *European Review of Applied Psychology*, 54, 119-135.
- Walker, M (2007). Ameliorating Culturally Based Extreme Response Tendencies to Attitude Items. *Journal of Applied Measurement*, 8(3), 267-278.

- Wu, Amery D., Li, Zhen & Zumbo, Bruno D. (2007). Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration with TIMSS Data. *Practical Assessment Research & Evaluation*, 12(3). Available online: <http://pareonline.net/pdf/v12n3.pdf>
- Yıldırım, H. H (2006). "The Differential Item Functioning (DIF) Analysis of Mathematics Items in the International Assessments Programs." Yayınlanmamış Doktora Tezi. Orta Doğu Teknik Üniversitesi, Ankara.
- Yıldırım, H. H. & Berberoğlu, G. (2009). Judgmental and Statistical Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing*, 9(2), 108-121.
- Zumbo, B. D., (2003). Does Item-Level DIF Manifest Itself in Scale-Level Analyses? Implications for Translating Language Tests. *Language Testing*, 20 (2), 136-147.
- Zumbo, B. D. (2005). Manifestation Of Differences In Item-Level Characteristics In Scale-Level Measurement Invariance Tests Of Multi-Group Confirmatory Factor Analyses. *Journal of Modern Applied Statistical Methods*, 4 (1), 275-282.
- Zumbo, B. D., Sireci, S. G., & Hambleton, R. K. (2003). Re-Visiting Exploratory Methods for Construct Comparability: Is there something to be Gained From the Ways of Old? *Paper Presented in the Symposium Construct Comparability Research: Methodological Issues and Results, National Council on Measurement in Education, April, Chicago, IL.*

Ek 1.

*Australya, Yeni Zelanda, ABD ve Türkiye Öğrenci Anketlerinde Bilimsel Sorgulamaya Verilen Destek Tutumunu Ölçmek İçin Kullanılan Maddeler*

	How much do you agree with the statements below? (Please tick one box in each row)	Aşağıdaki ifadelere ne ölçüde katılıyorsunuz? (Her sırada sadece bir kutuyu işaretleyiniz)
ST18Q01	a) Advances in science and technology usually improve people's living conditions	Fen bilimleri ve teknolojiadaki ilerlemeler genellikle insanların yaşam koşullarını iyileştirir
ST18Q02	b) Science is important for helping us to understand the natural world	Fen bilimleri, doğal dünyayı anlamamıza yardımcı olması açısından önemlidir
ST18Q03	c) Some concepts in science help me see how I relate to other people	Fen bilimlerindeki bazı kavramlar, diğer insanlarla nasıl bir bağlantı içinde olduğumu görmeme yardımcı olur
ST18Q04	d) Advances in science and technology usually help improve the economy	Fen bilimleri ve teknolojiadaki ilerlemeler genellikle ekonominin iyileştirilmesine yardımcı olur
ST18Q05	e) I will use science in many ways when I am an adult	Bir yetişkin olduğumda fen bilimlerinden birçok yönde yararlanacağım
ST18Q06	f) Science is valuable to society	Fen bilimleri toplum için yararlıdır
ST18Q07	g) Science is very relevant to me	Fen bilimleri beni çok ilgilendiriyor
ST18Q08	h) I find that science helps me to understand the things around me	Fen bilimlerinin, çevremdeki şeyleri anlamada bana yardımcı olduğu görüşündeyim
ST18Q09	i) Advances in science and technology usually bring social benefits	Fen bilimleri ve teknolojiadaki ilerlemeler genellikle toplumsal yararlar sağlar
ST18Q10	j) When I leave school there will be many opportunities for me to use science	Okulu bitirdikten sonra, fen bilimlerinden yararlanmam için pek çok fırsat olacaktır