

PISA2009 Öğrenci Anketi Tutum Maddelerinin Kültüre Göre Değişen Madde Fonksiyonu Açısından İncelenmesi

Analysis Of Attitude Items In PISA2009 Student Questionnaire In Terms Of Differential Item Functioning Based On Culture

Bilge GÖK** Kübra ATALAY KABASAKAL*** Hülya KELECİOĞLU****

Öz

Bu araştırmada Uluslararası Öğrenci Değerlendirme Programı (PISA) 2009 kapsamında uygulanan öğrenci anketinin kültürlere göre değişen madde fonksiyonu (DMF) gösterip göstermediği Türkiye-Finlandiya, Türkiye-Rusya, Singapur-Yeni Zelanda ve Avustralya-Yeni Zelanda örneklemi kullanılarak incelenmiştir. Araştırmada PISA'da yer alan 11 maddelik okumaya yönelik tutum maddeleri kullanılmıştır. Araştırmaya bu ülkelerin dâhil edilmesinde ülkelerin başarı sıralaması, dil ve kültür özellikleri dikkate alınmıştır. İlk olarak öğrenci anketinin faktör yapısı açıklayıcı faktör analizi ile incelenmiştir. Daha sonra DMF analizleri poly-SIBTEST, ordinal lojistik regresyon ve madde tepki kuramı olabilirlik oranı yöntemleri ile yapılmıştır. Farklı dil ve farklı kültüre sahip ülkelerde maddelerin %91'i DMF gösterirken, bu oran farklı kültür ve aynı dilde %82'ye, aynı kültür ve aynı dilde %37'ye düşmüştür. Bu çalışmada ortaya çıkan sonuç, dil fakından ziyade kültür farklılıklarının maddelerin DMF göstermesinde etkili olduğudur.

Anahtar Kelimeler: Değişen madde fonksiyonu, pisa, ordinal lojistik regresyon, poly-sibtest, olabilirlik oranı.

Abstract

In this study the attitude items in PISA 2009 student questionnaire were analysed in terms of whether or not these items changes based on culture in a comparative way between samples from Turkey-Finland, Turkey-Russia, Singapur-New Zeland and Avustralia-New Zeland. More specifically eleven attitudes items concerning reading were analysed in the study. The choice of countries was based on achievement rank, the characteristics of language and culture. Explatory factor analysis was used to analyse the factor pattern of the questionnaire. Analyses about differential item functioning (DIF) were achieved through the poly-SIBTEST, ordinal logistic regression and item response theory- likelihood ratio. The results showed that the DIF value in the countries with different language and culture is 91%. It was found to be 82% among the countries with the same language and different culture. It was 37% among those countries with the same culture and language. Therefore, not language but the culture seem to cause DIF in the items.

Key Words: Differential item functioning, pisa, ordinal logistic regression, poly-sibtest, likelihood ratio.

GİRİŞ

PISA, öğrencilerin matematik, fen ve okuma becerileri alanlarındaki bilgi ve becerilerinin ölçüldüğü uluslararası en büyük eğitim araştırmalarından biridir. Üç yılda bir

* Bu araştırma III.Ulusal Eğitimde Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

** Arş. Gör. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: bilgeb@hacettepe.edu.tr

*** Arş. Gör. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: katalay@hacettepe.edu.tr

**** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: hulyaebb@hacettepe.edu.tr

yapılan PISA uygulamalarında, her dönem üç konu alanından (okuma becerileri, matematik okuryazarlığı ve fen okuryazarlığı) birine ağırlık verilmektedir. PISA uygulamasına Türkiye 2003 yılında dâhil olmuş ve bu sınavda Türkiye, OECD ülkelerine göre düşük bir performans göstermiştir (EARGED, 2004). Daha sonraki 2006 ve 2009 yılında da dünya sıralamasında ülkenin yerinde önemli bir ilerleme gözlemlenmemiştir. PISA 2009 uygulamalarında 15 yaş grubu öğrencilerin okuma becerileri alanındaki yeterliklerinin, okumaya yönelik ilgi ve tutumlarının ölçülmesi amaçlanmıştır. Okuma, matematik, fen ve problem çözme okuryazarlığını ölçmek amacıyla gerçekleştirilen (Rindermann, 2007) bu sınavlar okul müfredatlarının erişilme düzeylerinden çok, toplum yaşamına etkili katılım için gerekli olan bilgi ve becerileri kapsamaktadır (Berberoğlu ve Kalender, 2005). Aynı zamanda öğrencilerin okulda edindikleri bilgi ve becerileri günlük hayatta karşılaştıkları durumlarda kullanabilme düzeylerini ölçmektedir (EARGED, 2004). Bununla birlikte bu tür sınavların en önemli amaçlarından birisi de öğrenci, öğretmen, veli ve okul anketlerini 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). Bu sınavlarda öğrencilerin bilgi ve beceri düzeylerinin yanı sıra ilgi, tutum ve öz yeterlilik gibi duyuşsal özellikleri de belirlenmektedir. Sınava katılan ülkeler, bu sonuçları değerlendirerek öğretim programlarında düzenlemelere gidebilmektedir.

TIMSS ve PISA gibi uluslararası sınavlar bilişsel alana yönelik testlerden oluşan başarı puanının yanı sıra bu puana etki eden diğer değişkenleri de ölçerek araştırmacılara fen ve matematik başarısına etki eden değişkenlere göre yorum yapma ve program geliştirme uzmanlarına ülke programlarını revize etmede dikkate alınması gereken değişkenlerdeki eksiklikleri görme olanağı sağlamaktadır. Bu açıdan bakıldığında, sadece başarı puanı üzerine odaklanmak yerine başarıyı etkileyebilecek diğer değişkenler üzerinde de önemle durulması gerekmektedir. Bilişsel alandaki başarıyı açıklayan pek çok değişken bulunmaktadır. Bu değişkenler arasında tutum, özyeterlilik gibi duyuşsal özelliklerin yer aldığı çeşitli araştırmalarla ortaya konmuştur (Anıl, 2009). Bu nedenle, öğrencilerin duyuşsal özelliklerinin de en az bilişsel özellikler kadar geçerli ölçülmesi gerekir. PISA, TIMSS gibi uluslararası sınavlarda öğrenci anketi ile ölçülen özellikler öğrencilerin bazı özelliklerine (cinsiyet ve kültür gibi) göre değişiklik gösterebilir. Bu durum ise anketlerin geçerliğini düşürerek, öğrencilerin hem duyuşsal hem de bilişsel özelliklerinin hatalı olarak belirlenmesine yol açar (Atalay Kabasakal ve Kelecioğlu, 2012).

PISA'da yer alan ölçme araçları İngilizce olarak hazırlanmakta ve daha sonra uygulanacağı ülkenin diline çevrilmektedir. Bu tür uluslararası uygulamalardan alınan sonuçlara göre ülkelerin eğitim sistemleri hakkında yorumlar yapılabilmesi ve farklı ülkelerin birbirleriyle karşılaştırılabilmesi için ölçülen yapının gruptan bağımsız olması gerekir. Gruptan bağımsızlık, farklı gruplarda yer alan aynı yetenek düzeyindeki kişilerin test puanlarının eşit olması anlamına gelir. Bu koşul sağlanmadan yapılan karşılaştırmalarda ise, görülen farklılığın gerçek durumdan mı yoksa ölçülen yapının gruplarda farklılık

göstermesinden mi kaynaklandığı bilinemeyecektir (Somer, Korkmaz, Dural ve Can, 2009). Yukarıda ifade edilen bu durum PISA, PIRLS, TIMSS gibi bir ölçme aracının farklı bir kültür ya da dile uyarlandığı sınavlarda madde yanlılığı kavramının önemini ifade etmektedir. Bu nedenle özellikle belirli kararların verildiği test uygulamalarında, ölçme araçlarının geçerliği, ölçme sonuçlarına karışan hataların miktarı ve testlerin yansızlığı oldukça önemlidir.

Yanlılık, bir testin geçerliğini etkileyen önemli faktörlerden biridir. Bir sınavda kullanılan ölçme aracının hiçbir gruba avantaj sağlamaması yani yanlı olmaması beklenir. Bir test yanlıysa, bu teste dayalı olarak cevaplayıcılar hakkında verilen kararlar da doğru olmayacaktır. Bazı durumlarda bireylerin ölçmek istediğimiz özelliklerine başka değişkenler karışabilir. Bunlar arasında; cinsiyet, okul türü, sosyo-ekonomik düzey, etnik köken vb. yer alır. Bu değişkenlerin test puanlarına etkisi geçerlik üzerinde bir tehdittir ve test puanlarının yanlı olmasına yol açar. Dolayısıyla bir test kişilerin bağlı bulunduğu herhangi bir grup üyeliği olmaksızın ölçmek istediği özelliği, yeteneği doğru bir şekilde ölçmelidir.

Yanlılık, aynı zamanda ölçme sonuçlarının bir grubun lehine olan sistematik hata içermesidir (Camilli ve Shepard, 1994). Bir test bir grup cevaplayıcının lehineyse, bu testin yanlı olduğu düşünülür. Yanlı değilse, bu durum geçerliğin güçlü bir kanıtıdır. Sonuç olarak, yansız bir ölçme işlemi gerçekleştirmenin ölçek geliştirme sürecinin en önemli basamaklarından biri olduğu söylenebilir.

Madde yanlılığını belirleme sürecinin ilk basamağı olan değişen madde fonksiyonu (DMF) ise, farklı gruplarda (okul türü, eğitim düzeyi, vb.) eşit yeteneğe sahip cevaplayıcıların bir testteki maddeleri doğru cevaplama olasılıklarının farklı olmasıdır (Holland ve Wainer, 1993). DMF belirleme süreci, doğrudan ölçme aracının niteliğine ilişkin önemli bilgiler verebilir. Bir ölçme aracında DMF belirlendiğinde, bu durum aracın ölçmek istediği özellik ile ilgili olmayan boyutları ölçtüğünün bir göstergesi olabilir. Bu açıdan bakıldığında, DMF belirlemek oldukça önemlidir çünkü DMF'nin varlığı gruplar arasındaki farklar hakkında araştırmacıları yanlış yönlendirebilir ve bireyler hakkında hatalı kararlar alınmasına neden olabilir.

DMF çalışmaları, cevaplayıcıların bir testte başarılı olabilmeleri ya da bir ölçek üzerinde kendilerini doğru ifade edebilmeleri için eşit fırsatlara sahip olup olmadıklarını belirlemenin ilk adımıdır. Değişen madde fonksiyonu, madde yanlılığının bir indeksidir. Yanlı bir madde DMF gösterir; ancak bir maddenin DMF göstermesi o maddenin yanlı olduğunu belirtmek için yeterli değildir. DMF istatistiksel bir yöntem iken, madde yanlılığı kavramsaldır ve yoruma açıktır (Camilli ve Shepard, 1994; Clauser ve Mazor, 1998). DMF gösteren maddeler dikkatlice incelenerek, madde güçlüğüne kaynağı yapıyla ilgili değilse, maddenin yanlı olduğuna karar verilir.

DMF, hem iki kategorili hem de çok kategorili puanlanan maddelerde bulunabilir. Hulin, Drasgow ve Parsons (1983) tutum gibi duyuşsal bir özelliği ölçen maddelerde DMF'yi, aynı tutum düzeyine sahip farklı alt gruplarda bulunan bireylerin, bir maddeye olumlu tutum gösterme olasılığının farklı olması olarak tanımlamışlardır (Dodeen ve

Johanson, 2003). Bir başka deyişle, aynı yetenek düzeyindeki kişilerin kişilik ve tutum maddeleri söz konusu olduğunda, maddeyi aynı yönde işaretleme olasılıkları farklılaşıyorsa DMF ortaya çıkar. Çok kategorili MTK modellerinde, iki gruptaki madde parametreleri eşitse, madde cevap fonksiyonu da eşit olacaktır. Madde cevap fonksiyonları referans ve odak grupta eşit değilse, maddenin DMF gösterdiği söylenir (Cohen, Kim ve Baker, 1993).

Çok kategorili DMF belirleme yöntemleri kendi içinde sınıflara ayrılmaktadır. Bunlar arasında ordinal lojistik regresyon, Mantel-Haenszel (MH), standartlaştırma yöntemi, genelleştirilmiş MH, lojistik diskriminant analizi, madde tepki kuramı olabilirlik oranı (MTK-OO), poly-SIBTEST yöntemleri yer almaktadır. Bu araştırmada ordinal lojistik regresyon, MTK-OO ve poly-SIBTEST yöntemleri kullanılmıştır.

Ordinal Lojistik Regresyon (OLR): Ordinal lojistik regresyon, Miller ve Spray (1993) tarafından çok kategorili veriler için önerilmiştir. OLR'nin varsayımı, ölçeğin tek boyutlu olmasıdır (Zumbo ve Gelin, 2003). Zumbo (1999) tarafından önerilen OLR yaklaşımı, grup etkisi ve alt-üst ölçek puanı etkileşimi için DMF testi sağlar. Bu yöntemde değişkenler modele hiyerarşik olarak sırayla eklenir. Modele ilk eklenen değişken toplam puan, ikinci değişken grup değişkeni (cinsiyet vb.), son değişken ise toplam puan grup etkileşimidir. Lojistik regresyon eşitliği aşağıdaki şekildedir:

$$y = b_0 + b_1TOP + b_2GRUP + b_3TOP*GRUP + \varepsilon_i \quad (1)$$

OLR ile DMF belirleme sürecinde standartlaştırılmış regresyon katsayıları (R^2) etki büyüklüğü olarak yorumlanır ve DMF'nin düzeyini verir (Gierl, Jodoin ve Ackerman, 2000). LR sürecinde DMF üç düzeyde belirlenir: $R^2 < 0.13$ ise maddede DMF yoktur ya da ihmal edilebilir düzeydedir (A düzeyi), $0.13 \leq R^2 < 0.26$ ise maddede orta düzeyde (B düzeyi), $R^2 \geq 0.26$ ise maddede önemli düzeyde (C düzeyi) DMF vardır (Zumbo ve Thomas, 1996).

Poly-SIBTEST: SIBTEST parametrik olmayan örtük değişken modelidir. Chang, Mazzeo ve Roussos (1996) SIBTEST yönteminin çok kategorili veriler için farklı formunu geliştirmişlerdir. Poly-SIBTEST örtük değişkenler için tanımlanmasına rağmen, bu yöntem MTK yetenek ve madde parametrelerini kullanmaz. Poly-SIBTEST yönteminde H_0 hipotezi beklenen sonuçların referans ve odak grupta θ yetenek düzeyinde aynı olması durumudur. θ yetenek düzeyinde DMF'nin miktarı 2 no'lu denklemle hesaplanır.

$$B_0 = E_R [Y|\theta] - E_F [Y|\theta] \quad (2)$$

$E_{R,F} [Y|\theta]$ Y'nin θ yetenek düzeyinde X üzerindeki bağlanımını (referans ve odak grup için) temsil eder. Bir maddenin DMF göstermediği durumda $E_R [Y|\theta] = E_F [Y|\theta]$ eşitliği sağlanır. SIBTEST etki büyüklüğü β , her bir madde için DMF miktarı olarak yorumlanır. Pozitif β değeri o maddenin referans grubun lehine, negatif β değeri ise o maddenin odak grubun lehine çalıştığını gösterir. SIBTEST'te β etki büyüklüğü olarak yorumlanır ve bu değer DMF'nin düzeyini ifade eder. $\beta < |0.059|$ ise, maddede ihmal edilebilir düzeyde (A düzeyi), $\beta |0.059-0.088|$ arasında ise orta düzeyde (B düzeyi), $\beta \geq |0.088|$ ise önemli düzeyde (C düzeyi) DMF olduğu kabul edilmektedir (Roussos ve Stout, 1996).

MTK-OO: Bu yöntem MTK'nin değişmezlik ilkesinden dolayı DMF tespiti için ideal bir yoldur. MTK-OO yönteminde sınırlandırılmış ve genişletilmiş model oluşturularak bu modellerin olabilirlik oranları karşılaştırılır. Sınırlandırılmış modelde, madde parametrelerinin odak ve referans grup için eşit olduğu varsayılır. Genişletilmiş model ise, test edilen i. maddenin parametrelerinin odak ve referans grup için farklı olduğu, diğer maddeler için eşit olduğu varsayımına dayalı olarak kurulur. Analiz edilen her bir madde için bir tane sınırlandırılmış model oluşturulur ve bu model genişletilmiş model ile oranlanır.

OlabilirlikOranı(OO)=Olabilirlik*(SınırlandırılmışModel)/Olabilirlik*(GenişletilmişModel).

İki modelin olabilirliklerinin logaritması alınarak G^2 değeri elde edilir (Thissen, 2001). Bu işlem tüm maddeler için tekrarlanır.

$$G^2 = -2LL_s - (-2LL_g) \quad (3)$$

G^2 istatistiği tüm parametrelerin iki grup arasında eşit olup olmadığına dair eş zamanlı bir hipotez testi imkânı sağlar. Serbestlik derecesi MTK modelindeki parametre sayısı olan G^2 değeri, χ^2 dağılımının kritik değeri ile kıyaslanarak hipotez testi gerçekleştirilir. Anlamlı olan χ^2 değeri DMF'nin varlığını gösterir (Thissen, 2001). Araştırma çok kategorili (4 kategorili) veriler üzerinde yapıldığından sonuçlar 4 serbestlik derecesindeki 0.01 anlamlılık düzeyinde χ^2 değeri olan 13.28'e göre yorumlanmıştır.

Bu konuda yapılan çalışmalar incelendiğinde, araştırmada ele alınan yöntemlerin çok kategorili puanlanan maddeler için I.tip hatasının düşük olduğu ve yüksek güce sahip olduğu görülmüştür (Kim & Cohen, 1998; Ankenmann, Witt ve Dunbar, 1999). Bununla birlikte yapılan çalışmalar bu yöntemlerin DMF belirlemede etkili bir yol olduğunu göstermiştir. Tüm DMF yöntemlerinin kendilerine özgü güçlü ve zayıf yönleri vardır. Eğer kültürlerarası karşılaştırma yapılacaksa, yansızlıktan emin olmak için birden fazla yöntemle DMF analizlerinin yapılması gerekir (Hambleton ve Patsula, 1999).

Türkiye'de yapılan büyük ölçekli sınavlarda yer alan maddelerin değişen madde fonksiyonu içerip içermediği incelenmemektedir. Uluslararası sınavlarda ise çeviriden kaynaklanan ve kültüre özgü pek çok sorun yaşanmaktadır. Türkiye'deki yaygın bakış açısı TIMSS, PISA ve PIRLS sınavlarına katılan ülkelerin uluslararası platformda ne düzeyde olduklarını gözleme şansı bulduğu (Berberoğlu ve Kalender, 2005) ve ülkenin eğitimde bulunduğu yerin bu sınavlarla belirlenebileceği yönündedir (Taşdemir ve Taşdemir, 2008). Bu nedenle uluslararası sınavlar üzerinde yapılan çalışmaların artırılması gerekmektedir. Yanlılık, sadece bilişsel testlerde değil aynı zamanda psikolojik testlerde de geçerliği düşüren bir faktördür. Bilişsel yeteneği ölçen testlerde madde yanlığı üzerine çok sayıda çalışma yapılmıştır; ancak psikolojik ölçmelerde bireylerin diğer özelliklerinin ölçme sonuçlarına etkisi bilişsel özellikleri ölçen testler kadar iyi bilinmemektedir. Son yıllarda iki kategorili maddeler üzerinde pek çok DMF çalışmalarının yapıldığı görülmekle birlikte, ülkemizde çok kategorili maddeler üzerinde yapılan çalışmaların sayısı oldukça azdır. Bu

nedenle, yapılan bu çalışmanın hem diğer araştırmacılara hem de test geliştirme uzmanlarına bilgi vereceği düşünülmektedir.

Bu çalışmanın amacı, PISA 2009 öğrenci anketinde yer alan tutum maddelerinin kültüre göre DMF içerip içermediğinin ordinal lojistik regresyon, poly-SIBTEST ve MTK-OO yöntemleri kullanılarak incelenmesidir. Bu çalışmadan elde edilen sonuçların test geliştirme uzmanlarına ve bu alanda çalışan kişilere test yanlılığı ve geçerliği ile ilgili önemli bilgiler vereceği düşünülmektedir. Bu çalışmanın sonunda tutum maddelerinin ele alınan ülkelerdeki öğrencilerin tutumunu yansıtan bir biçimde ölçüp ölçmediğine ilişkin veri sağlanacaktır.

DMF içeren maddeleri belirlemek, olası yanlılık nedenlerini belirlemenin ilk adımıdır. Bir kültürden diğer bir kültüre uyarlanan ölçme araçlarının kültüre göre DMF içerip içermediği incelenmediğinde, o ölçme aracının ölçmek istediği özelliği ölçtüğüne ilişkin kesin bir yargıya varılamaz. Bu nedenle hem bilişsel hem de duyuşsal özelliklerin ölçülmesini amaçlayan ölçme araçlarının geliştirilmesinde DMF çalışmalarına yer verilmesi oldukça önemlidir. Böylece ortaya çıkan DMF'nin sebepleri tespit edilerek daha sonra yapılacak çalışmalarda gerekli önlemlerin alınmasını sağlanabilir. Bireyler arasındaki farklılıklar üzerine odaklanan psikolojik özelliklerin ortaya çıkarılmasını amaçlayan çalışmalarda genellikle çeviri ve adaptasyon sorunlarına rastlanmaktadır. Bu tür çalışmalarda ve uluslararası sınavlarda ele alınan psikolojik yapıların ölçülmesinde kültüre özgü değişen madde fonksiyonuna rastlanmaktadır (Asil, 2010; Atalay Kabasakal, 2012).

Türkiye'de uluslararası sınavlar üzerinde yapılan çalışmalar öğrenci anketinde ülkeler arasında cinsiyete ve kültüre göre değişen fonksiyon gösteren maddelerin olduğunu ve farklı ülkelerde dilsel ve kültürel farklılıklar arttıkça DMF gösteren maddelerin sayısının arttığını göstermiştir. Bu doğrultuda Türkiye'de çok kategorili puanlanan maddeler üzerinde yapılan çalışmalar aşağıda özetlenmiştir. Asil ve Gelbal (2012) yaptıkları çalışmada, PISA 2006 uygulamasında yer alan öğrenci anket maddelerinin farklı kültürler ve diller arasında DMF gösterip göstermediğini Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemi üzerinde karşılaştırmalı olarak Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA) ile incelemişlerdir. ÇG-DFA bulguları ülkeler arasında değişen fonksiyon gösteren maddelerin olduğunu ve ülkeler arasında dilsel ve kültürel farklılıklar arttıkça DMF gösteren maddelerin sayısının da arttığı göstermiştir. Öğretmen ve Başokcu (2014) yaptıkları çalışmada TIMSS 2011 8.sınıf öğrenci anketinde "Matematiğe Karşı Tutum" maddelerinin ülke ve cinsiyete göre DMF gösterip göstermediğini Türkiye ve Güney Kore örneklemi üzerinde parametrik ve nonparametrik MTK yöntemlerini kullanarak incelemişlerdir. Araştırmanın sonucunda yöntemlerin benzer sonuçlar verdiği ve her bir ülke için cinsiyete dayalı karşılaştırmaların yöntemlere göre farklılık gösterdiği görülmüştür. Başokcu ve Öğretmen (2013) Öğretmenlik mesleğine ilişkin yeterlik algısını ölçmeye yönelik geliştirilen ölçme aracının alt popülasyonlar için madde bazında farklı çalışıp çalışmadığını değişen madde fonksiyonlarını inceleyerek belirlenmişlerdir. Analizler hem ölçek toplam puanları hem de ölçeğin alt boyutlarının toplam puanları göz önüne alınarak ayrı ayrı yapılmış ve sonuç olarak cinsiyete göre 7 madde de değişen fonksiyonu belirlenmiştir. Atalay Kabasakal ve Kelecioğlu (2012) yaptıkları çalışmada PISA 2006

öğrenci anketinde yer alan fen bilimlerine ilişkin tutum maddeleri, cinsiyete ve kültüre göre değişen madde fonksiyonu gösterip göstermediğini Amerika Birleşik Devletleri ve Türkiye örneklemelerinde incelemiştir. Cinsiyete göre yapılan analizlerde her iki yöntemde ortak ve önemli düzeyde DMF gösteren madde sayısı Türkiye’de 7, ABD’de 13’tür. İki ülke arasında kültüre göre her iki yöntemde yapılan analizlerde 30 maddede önemli düzeyde DMF görülmüştür. Analizler sonucu iki yöntem karşılaştırıldığında, yöntemlerin DMF belirlemede bir uyum gösterdiği; ancak düzeylerinde farklılıklar olabileceği sonucuna varılmıştır. Yukarıda ifade edilen çalışmaların sonuçları genel olarak incelendiğinde uluslararası sınavların kültür, dil ve cinsiyet gibi değişkenlere göre değişen madde fonksiyonu içerdiğini göstermektedir.

Maddelerin farklı gruplarda farklı fonksiyon göstermesi, grupların dil ve kültür farklılıklarından ya da başarı düzeylerinin farklı olmasından kaynaklanabilir. Bu nedenle, PISA 2009 öğrenci anketinde yer alan okumaya yönelik tutum maddeleri karşılaştırılırken, başarı sırası, dil ve kültür yönünden karşılaştırma olanağı sağlayacak ülkeler arasından bir seçim yapılmıştır. Bu ülkeler, araştırmada ele alınan değişkenler açısından çeşitli düzeylerde benzerlik ve farklılık gösteren Türkiye, Finlandiya, Rusya, Avustralya, Yeni Zelanda ve Singapur’dur. Türkiye PISA 2009’da 39. sırada yer alırken, Finlandiya 3.sırada yer almıştır. Türkiye ve Finlandiya farklı başarı düzeylerine ve farklı dil ve kültüre sahip iki ülkedir. Bunun yanında Türkiye başarı düzeyi benzer fakat farklı dil ve kültüre sahip olan 41.sıradaki Rusya ile karşılaştırılmıştır. Bunlara ek olarak araştırma kapsamında başarı düzeyleri yüksek ve benzer olan aynı dil ve kültüre sahip Avustralya (7.sırada) ve Yeni Zelanda (9.sırada) ile yine başarı düzeyi benzer ve yüksek olan ve aynı dil ve farklı kültüre sahip Singapur (5.sırada) ve Yeni Zelanda (7.sırada) ele alınmıştır.

YÖNTEM

Çalışma Grubu

PISA 2009 uygulamasında, ulusal merkezler tarafından tabakalı örnekleme yöntemine göre örneklemin çatısı oluşturulur. Bu uygulamaya 65 katılımcı ülkeden, 15 yaş grubu yaklaşık 26 milyon öğrenciyi temsilen 475.460 öğrenci katılmıştır. Bu araştırmada Türkiye, Rusya, Finlandiya, Singapur, Yeni Zelanda ve Avustralya örneklemeleri ele alınmıştır. Öğrencilerin buldukları ülkelerin kültürel ve dilsel özelliklerini tam olarak yansıttıklarından emin olabilmek için testin uygulandığı ülkede doğan ve evlerinde testin uygulandığı dilin konuşulduğunu belirten öğrenciler değerlendirmeye alınmıştır. Bununla birlikte analizde kullanılacak olan yöntemlerin kayıp verilerle çalışmamasından dolayı kayıp veriler çıkartılmıştır. Kayıp ve aykırı verilerin çıkarılması sonucu çalışma grubuna ait frekans ve yüzdeler Tablo 1’de verilmiştir.

Tablo 1. *Çalışma Grubuna ait Frekans ve Yüzdeler*

Ülkeler	Frekans	Yüzde	Ülkeler	Frekans	Yüzde
Türkiye	4531	13.9	Singapur	2090	6.4
Rusya	4606	14.1	Yeni Zelanda	3820	11.7
Finlandiya	5213	15.9	Avustralya	12446	38.1

Bununla birlikte, araştırmada karşılaştırma yapılan ülkelere ilişkin karşılaştırma ölçütleri Tablo 2’de verilmiştir.

Tablo 2. *Karşılaştırma Yapılan Ülkeler ve Karşılaştırma Ölçütleri*

Ülkeler	Başarı Düzeyi ve Sırası	Kültür	Testin Uygulandığı Dil*
Türkiye-Finlandiya	Farklı (39–3)	Farklı	Farklı (Türkçe-Fince)
Türkiye-Rusya	Benzer (düşük) (39-41)	Farklı	Farklı (Türkçe-Rusya)
Avustralya-Yeni Zelanda	Benzer (yüksek) (7–9)	Aynı	Aynı (İngilizce)
Singapur-Yeni Zelanda	Benzer (yüksek) (5–7)	Farklı	Aynı (İngilizce)

*Testin uygulandığı dil ile aynı dili konuşan öğrenciler çalışmaya alınmıştır.

Veri Toplama Araçları

Araştırmada, PISA 2009 öğrenci anketinden elde edilen veriler kullanılmıştır. Araştırma kapsamında incelemeler öğrenci anketinde yer alan 11 tutum maddesi üzerinde yapılmıştır. PISA 2009 öğrenci anketinde yer alan st24 kodlu maddeler okumaya yönelik tutumu ölçmektedir. Araştırmada kullanılan veriler OECD PISA internet sitesinden elde edilmiştir (www.pisa.oecd.org).

Verilerin Analizi

Verilerin analizinde izlenen aşamalar şu şekildedir:

PISA 2009’da yer alan “okumaya yönelik tutum” 11 madde ile ölçülmüştür. Bu 11 maddenin tek boyutlu olup olmadığını incelemek amacıyla açımlayıcı faktör analizi yapılmıştır. Faktör analizi sonucunda ölçeği oluşturan 11 maddenin tek boyutta toplandığı özdeğerler ve çizgi grafikleri incelenerek karar verilmiştir. Ölçeğin tek boyutta açıkladığı varyans oranı Avusturya’da %60.751, Finlandiya’da %57.633, Yeni Zelanda’da %56.553, Rusya’da %42.238, Singapur’da %52.239, Türkiye’de %39.879’dur.

Ülkelere göre değişen madde fonksiyonu OLR, poly-SIBTEST ve MTK-OO yöntemleri ile incelenmiştir. Analizler sonucu maddelerin hangi düzeyde DMF gösterdiği ve hangi grubun lehine çalıştığı belirtilerek tablolaştırılmıştır.

Verilerin analizinde, OLR için Zumbo (1999) tarafından yazılan betik, poly-SIBTEST için SIBTEST yazılımı, MTK-OO için ise IRTLRFIF yazılımı kullanılmıştır.

BULGULAR

Bu çalışmada PISA 2009 öğrenci anketinde yer alan okumaya yönelik 11 tutum maddesinin Türkiye-Finlandiya, Türkiye-Rusya, Avustralya-Yeni Zelanda ve Singapur-Yeni Zelanda örneklemelerinde kültüre göre değişen madde fonksiyonu içerip içermediği ordinal lojistik regresyon, poly-SIBTEST ve MTK-OO yöntemleri ile incelenmiştir.

Singapur-Yeni Zelanda örneğine ait OLR, Poly-SIBTEST ve MTK-OO yöntemlerinden elde edilen analiz sonuçları Tablo 3’te verilmiştir.

Tablo 3. *Singapur ve Yeni Zelanda Örnekleminden Elde Edilen Sonuçlar*

Tutum maddeleri	OLR		Poly-SIBTEST		MTK-OO		Avantajlı grup
	ΔR^2	Düzyey	β_u	Düzyey	g^2	DMF	
ST24Q01	1.73	C	0.374	C	35.4	Var	Singapur
ST24Q02	3.21	C	0.59	C	211.6	Var	Singapur
ST24Q03	0.33	C	1.017	C	25.7	Var	Singapur
ST24Q04	0.78	C	0.552	C	80	Var	Singapur
ST24Q05	0.08	-	-1.378	C	56	Var	Yeni Zelanda
ST24Q06	1.83	C	0.377	C	21.2	Var	Singapur
ST24Q07	2.41	C	1.985	C	157	Var	Singapur
ST24Q08	0.87	C	0.453	C	151	Var	Singapur
ST24Q09	1.59	C	0.582	C	59.6	Var	Singapur
ST24Q10	0.03	-	0.984	C	8.3	Yok	Singapur
ST24Q11	0.28	C	0.973	C	38.6	Var	Singapur

Not: Referans grup: Singapur Odak grup: Yeni Zelanda

Tablo 3 incelendiğinde, OLR, Poly-SIBTEST ve MTK-OO yöntemlerine göre 9 maddenin (1, 2, 3, 4, 6, 7, 8, 9, 11. maddeler) C düzeyinde DMF gösterdiği görülmüştür. Bununla birlikte 5. maddede OLR yöntemine göre, 10. maddede ise hem OLR hem de MTK-OO yöntemine göre DMF olmadığı bulunmuştur. DMF gösteren bu maddelerin hangi gruba avantaj sağladıkları incelediğinde ise, 5. madde dışında diğer tüm maddelerin Singapur lehine işlediği sadece 5. maddenin Yeni Zelanda lehine işlediği belirlenmiştir. Singapur ve Yeni Zelanda farklı kültüre sahip iki ülke olmakla birlikte aynı dili kullanmaktadır. Bu nedenle, PISA 2009 öğrenci anketinde yer alan 11 tutum maddesinden 9'unun bu iki ülkedeki öğrenciler tarafından aynı şekilde algılanmadığı ve DMF gösterdiği düşünülmektedir. Bununla birlikte ankette yer alan maddelerin DMF göstermesinin nedenleri arasında, araştırmada ele alınan bu iki ülkenin aynı dili konuşuyor olsa bile, kültürel farklılıklardan ve anketlerde kültüre özgü bazı kelimelerin yer alması gösterilebilir.

Türkiye-Finlandiya örneğine ait OLR, Poly-SIBTEST ve MTK-OO yöntemlerinden elde edilen analiz sonuçları Tablo 4'te verilmiştir.

Tablo 4. *Türkiye-Finlandiya Örnekleminden Elde Edilen Analiz Sonuçları*

Tutum maddeleri	OLR		Poly-SIBTEST		MTK-OO		Avantajlı grup
	ΔR^2	Düzyey	β_u	Düzyey	g^2	DMF	
ST24Q01	6.02	C	0.569	C	587.9	Var	Türkiye
ST24Q02	0.51	C	-1.035	C	285.4	Var	Finlandiya
ST24Q03	0.22	B	-0.603	C	192.5	Var	Finlandiya
ST24Q04	0.2	B	0.704	C	1276.9	Var	Türkiye
ST24Q05	0.4	C	-1.562	C	218.9	Var	Finlandiya
ST24Q06	8.57	C	0.417	C	156.3	Var	Türkiye
ST24Q07	0.32	C	-0.6	C	50	Var	Finlandiya
ST24Q08	1.46	C	0.827	C	670.7	Var	Türkiye
ST24Q09	0.35	C	0.795	C	2134.1	Var	Türkiye
ST24Q10	0.34	C	-0.12	C	316.2	Var	Finlandiya
ST24Q11	4.06	C	-0.028	-	1838.7	Var	-

Not: Referans grup: Türkiye Odak grup: Finlandiya

Tablo 4’te, 10 maddenin OLR, poly-SIBTEST ve MTK-OO yöntemlerine göre DMF gösterdiği bulunurken, 3. ve 4. maddelerin OLR analizine göre B düzeyinde poly-SIBTEST yöntemine göre ise C düzeyinde DMF gösterdiği tespit edilmiştir. DMF gösteren maddelerin hangi grubun lehine çalıştıkları incelendiğinde ise, 5 maddenin (1, 4, 6, 8 ve 9. maddeler) Türkiye lehine işlediği görülürken, diğer 5 maddenin de (2, 3, 5, 7 ve 10) Finlandiya lehine işlediği belirlenmiştir.

Türkiye-Finlandiya arasında yapılan karşılaştırmada, 11. madde dışında diğer tüm maddelerin tüm yöntemlere göre DMF göstermesinin nedeni, bu iki ülkede hem kullanılan dil hem de kültürel özelliklerin birbirinden oldukça farklı olmasından dolayı bu maddelerin iki ülkede yaşayan öğrenciler tarafından farklı anlamlar ifade etmesinden kaynaklanıyor olabilir. Uluslararası testlerde karşılaşılan en önemli sorun, çeviriden kaynaklı sorunlardır. Bununla birlikte, maddelerin çevirisi doğru olsa bile bazı kelimelerin o ülkede yaşayan öğrencilerin aşına olduğu kelimeler olmaması, madde içeriğinin/formatının ya da uzunluğunun değişmesi ve ülkeler arasında kültürel farklılıkların olması bu maddelerin kültüre göre DMF göstermesinin nedenleri arasında sayılabilir. Bununla birlikte Finlandiya PISA 2009 okuma becerileri alanında en başarılı 3. ülke iken, Türkiye ise 39. sırada yer almıştır. Çok farklı başarı düzeylerine sahip olan bu iki ülkenin eğitim programları da birbirinden farklıdır. Schulz’a (2005) göre eğitim sistemlerindeki farklılıklar da öğrencilerin soruları nasıl algıladığına ve yorumladığına etki edebilir.

Türkiye-Rusya örnekleme ait OLR, Poly-SIBTEST ve MTK-OO yöntemlerinden elde edilen analiz sonuçları Tablo 5’te verilmiştir.

Tablo 5. *Türkiye-Rusya Örnekleminden Elde Edilen Analiz Sonuçları*

Tutum maddeleri	ORL		Poly-SIBTEST		MTK-OO		Avantajlı grup
	ΔR^2	Düzyey	β_u	Düzyey	g^2	DMF	
ST24Q01	10.18	C	-0.096	C	234.7	Var	Rusya
ST24Q02	1.68	C	0.003	-	384	Var	-
ST24Q03	0.71	C	-0.801	C	173.1	Var	Rusya
ST24Q04	0.4	C	0.219	C	747.9	Var	Türkiye
ST24Q05	5.08	C	0.086	B	301.3	Var	Türkiye
ST24Q06	12.54	C	-0.103	C	127.3	Var	Rusya
ST24Q07	2.23	C	-0.122	C	179	Var	Rusya
ST24Q08	10.8	C	-0.065	B	102.5	Var	Rusya
ST24Q09	0.7	C	0.349	C	1431.3	Var	Türkiye
ST24Q10	1.08	C	-0.304	C	61.6	Var	Rusya
ST24Q11	10.18	C	-0.059	A	323.2	Var	Rusya

Not: Referans grup: Türkiye

Odak grup: Rusya

Tablo 5’den elde edilen bilgiler incelendiğinde, 10 maddede tüm yöntemlere göre DMF bulunduğu görülürken, poly-SIBTEST yöntemine göre 5. ve 8. maddelerde B düzeyinde, 11. madde de ise A düzeyinde DMF bulunduğu OLR yöntemine göre ise tüm maddelerde C düzeyinde DMF bulunduğu görülmüştür. Bu maddelerde görülen DMF’nin

hangi grubun lehine işlediğine bakıldığında, 7 maddenin (1, 3, 6, 7, 8, 10 ve 11.maddeler) Rusya lehine işlediği, 3 maddenin (4, 5 ve 9. maddeler) ise Türkiye lehine işlediği belirlenmiştir.

Türkiye-Finlandiya örneklemelerinde olduğu gibi Türkiye-Rusya örneklemelerinde de üç yönteme göre 10 maddenin DMF göstermesinin nedeni, iki ülkede hem kullanılan dilin hem de kültürel özelliklerin birbirinden oldukça farklı olmasıdır. Araştırmadan elde edilen bu bulgu, Allalouf, Hambleton ve Sireci (1996), Ercikan (2002) ve Elosua ve Jauregú (2007) tarafından yapılan çalışmada elde edilen bulgu ile de desteklenmektedir. Farklı dil ve farklı kültüre sahip iki ülke olan Türkiye-Finlandiya ve Türkiye-Rusya örneklemelerinde tüm yöntemlere göre 11 tutum maddesinin 10'unda DMF görülmüştür. Buna göre Van de Vijer ve Tanzer (2004) DMF'nin olası nedenlerinin yetersiz çeviriler, orijinal maddedeki anlam belirsizliği, madde içeriğinin belirli kültürlerdeki aşinalığı veya madde yazılış tarzı ile ilgili kültüre özgü imalar olabileceğini ifade etmişlerdir. PISA, TIMSS ve PIRLS gibi uygulamalarda yapılan DMF çalışmalarında da, DMF'nin nedenlerine çeviri problemlerinin, kültürel farklılıkların ve eğitim programlarındaki farklılıkların neden olabileceği sonucuna varılmıştır (Yıldırım ve Berberoğlu, 2009; Le, 2009).

Yeni Zelanda-Avustralya örneğine ait OLR, Poly-SIBTEST ve MTK-OO yöntemlerinden elde edilen analiz sonuçları Tablo 6'da verilmiştir.

Tablo 6. Yeni Zelanda-Avustralya Örneğinden Elde Edilen Analiz Sonuçları

Tutum maddeleri	OLR		Poly-SIBTEST		MTK-OO		Avantajlı grup
	ΔR^2	Düzyey	β_u	Düzyey	g^2	DMF	
STQ01	0.21	B	0.06	-	22.6	Var	-
STQ02	0.26	B	-0.167	C	18.4	Var	Avustralya
STQ03	0.36	C	-0.01	-	7.6	Yok	-
STQ04	0.26	B	0.079	B	7.6	Yok	Yeni Zelanda
STQ05	0.49	C	0.095	C	4.8	Yok	Yeni Zelanda
STQ06	0.86	C	0.133	C	20.1	Var	Yeni Zelanda
STQ07	0.79	C	0.513	C	27.5	Var	Yeni Zelanda
STQ08	0.24	B	0.054	-	15.9	Var	-
STQ09	0.9	C	0.143	C	20.4	Var	Yeni Zelanda
STQ10	0.13	A	-0.009	-	15.1	Var	-
STQ11	0.52	C	-0.003	-	21.8	Var	-

Not: Referans grup: Yeni Zelanda

Odak grup: Avustralya

Tablo 6 incelendiğinde, OLR, poly-SIBTEST ve MTK-OO yöntemlerine göre 4 maddenin (2, 6, 7 ve 9. maddeler) OLR ve POLY-SIBTEST yöntemlerine göre 6 maddenin ve MTK-OO yöntemine göre ise 8 maddenin DMF gösterdiği görülmüştür. DMF gösteren maddelerin 5'i Yeni Zelanda lehine çalışırken, 1'i ise Avustralya lehine çalışmaktadır. Avustralya ve Yeni Zelanda'da aynı dil kullanıldığı ve benzer kültürel özelliklere sahip olduğu için maddelerin bu ülkelerdeki öğrenciler tarafından aynı şekilde anlaşıldığı ve diğer ülkelere göre en az DMF gösteren maddeye rastlandığı düşünülmektedir. Bununla birlikte

bu iki ülkenin başarı sıralamaları incelendiğinde, iki ülkenin benzer ve yüksek başarı düzeyine sahip iki ülke olduğu görülmektedir.

Araştırma kapsamında ele alınan ülkelerde dil ve kültürel farklılıklara göre her üç yöntemde de DMF gösteren (OLR ve Poly-SIBTEST'te C düzeyinde olan) maddelerin dağılımı incelenmiş ve sonuçlar Tablo 7'de verilmiştir.

SONUÇLAR ve TARTIŞMA

Tablo 7. *Kültür ve Dil Farklılıklarına Göre DMF Gösteren Maddelerin Oranı*

Çalışma Grubu	Madde sayısı	DMF'li madde sayısı	%DMF
Singapur-Yeni Zelanda (Farklı kültür-aynı dil)	11	9	81.81
Türkiye-Finlandiya (Farklı kültür-farklı dil)	11	10	90.91
Türkiye-Rusya (Farklı kültür-farklı dil)	11	10	90.91
Yeni Zelanda-Avustralya (Aynı kültür-aynı dil)	11	4	36.36

Tablo 7 incelendiğinde, kültürel ve dilsel farklılıklar arttıkça değişen madde fonksiyonu gösteren maddelerin sayısının da arttığı görülmektedir. Farklı dil ve farklı kültüre sahip ülkelerde maddelerin %91'i DMF gösterirken, bu oran farklı kültür ve aynı dilde %82'ye, aynı kültür ve aynı dilde %37'ye düşmüştür. Sonuçların ülkelerin başarı sırasından etkilenmediği görülmüştür. Bununla birlikte en az DMF'li madde sayısının Yeni Zelanda ve Avustralya arasında yani aynı kültür ve aynı dile sahip ülkeler arasındagörüldüğü söylenebilir. DMF'li madde sayısının %90'ı geçen ülkelerdeülkelerin ortak özelliğine bakıldığında, bu ülkelerin farklı kültüre sahip olduğu görülmektedir. Bununla birlikte ülkelerin aynı ya da farklı dile sahip olmasının DMF'li madde sayısı üzerindeki etkisinin göz ardı edilecek düzeyde olduğu görülmüştür. Bu çalışmada ortaya çıkan sonuç, dil fakından ziyade kültür farklılıklarının maddelerin DMF göstermesinde etkili olduğudur. Bu bulgu Asil (2010)'un yaptığı çalışmada anket maddelerin kültürel farklılıklardan çok çeviri ve uyarlamadan kaynaklı problemlerden dolayı DMF göstermesine dair elde ettiği sonuç ile tutarlılık göstermektedir. Bu da çeviri ve uyarlama sürecinin ne kadar önemli olduğunun, uyarlama sürecinde kültürel eşdeğerliğin sağlanması gerektiğinin göstergesidir.

PISA, PIRLS ve TIMSS gibi uluslararası uygulamalarda yapılacak karşılaştırmalarda anket maddelerinin DMF gösterip göstermediğinin incelenmesi ve maddelerin DMF göstermesinin nedenleri araştırılmalıdır. Bununla birlikte, çeviri ya da kültürden kaynaklanan sorunları gidermek için uluslararası sınavlarda görev alan kişilerin bu konuda uzman olması ve soruların çevrilmesinde titiz ve dikkatli olmaları gerekmektedir. Özellikle bu tür sınavlarda görev alan çevirmenlerin seçimine ve çevirmenler ile ölçme değerlendirme uzmanlarının işbirliği içinde çalışmasına büyük önem verilmelidir. Bu doğrultuda özellikle ülkeler arası yapılacak karşılaştırmalarda, kültüre özgü DMF içeren maddelerin üzerinde durularak karşılaştırmaların yapılması önerilebilir.

Bu çalışmada OLR, MTK-OO ve poly-SIBTEST yöntemleri kullanılmıştır. Bundan sonra yapılacak çalışmalarda farklı yöntemlerin (Mantel-Haenzsel, genelleştirilmiş Mantel Haenzsel, çoklu grup doğrulayıcı faktör analizi ve MTK'ye dayalı yöntemler vb.) kullanılması ve bu çalışmada olduğu gibi birden fazla yöntemle elde edilen sonuçların karşılaştırılması önerilebilir.

KAYNAKLAR

- Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement, 36*, 185–198.
- Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programı (pisa)'nda türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim, 34*, 87-100.
- Asil, M. (2010). *Uluslararası öğrenci değerlendirme programı (PISA) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Asil, M., ve Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim, 37*(166), 236-249.
- Atalay, K. ve Kelecioğlu, H. (2012). PISA 2006 Öğrenci anketinde yer alan maddelerin değişen madde fonksiyonu açısından incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 45*(2), 77-96
- Ankenmann, R. D., Witt, E. A., & Dunbar, S. B. (1999). An investigation of the power of the likelihood ratio goodness-of-fit statistic in detecting differential item functioning. *Journal of Educational Measurement, 36*, 277-300.
- Berberoğlu, G. ve Kalender, İ. (2005). Öğrenci başarısının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi. *Eğitim Bilimleri ve Uygulama, 4*(7), 21-35.
- Basokçu, O. & Öğretmen, T. (2014). Comparison of parametric and nonparametric item response techniques in determining differential item functioning in polytomousscale. *American Journal of Theoretical and Applied Statistics, 3*(2), 31-38.
- Camili, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage Publications.
- Chang, H., Mazzeo, J., & Roussos, L. A. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement, 33*, 333–353.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31–44.
- Cohen, A. S., Kim, S. -H., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17*(4), 335-350.
- Dodeen, H., & Johanson, G. A. (2003). An analysis of sex-related differential item functioning in attitude assessment. *Assessment & Evaluation in Higher Education, 28*(2), 129–134.
- EARGED. (2004). *Öğrenci başarısını belirleme programı (PISA-2003), ulusal ön rapor*. Ankara: MEB-Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Ercikan, K. (2002). Disentangling the sources of differential item functioning in multilingual assessments. *International Journal of Testing, 2*(3&4), 199–215.
- Elousa, P., & López-Jauregui, A. (2007). Potential sources of differential item functioning. *International Journal of Testing, 7* (1), 39–52.
- Gierl, M. J., Jodoin, M., & Ackerman T. (2000). *Performance of Mantel-Haenzsel, simultaneous item bias test and logistic regression when the proportion of DIF items is large*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, USA.
- 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.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.
- Kim, S. -H., & Cohen, A. S. (1998). Detection of differential item functioning under the graded response model with the likelihood ratio test. *Applied Psychological Measurement, 22*(4), 345-355.
- Le, L. T. (2009). Investigation gender differential item functioning across countries ABD test languages for PISA science items. *International Journal of Testing, 9*(2), 122–133.

- Milli Eğitim Bakanlığı [MEB] (2010). *PISA 2009 Uluslararası öğrenci değerlendirme programı ulusal ön rapor*. Ankara: MEB Yayınları.
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107–122.
- Öğretmen, T. & Başokçu, T. (2013). Öğretmen öz yeterlilik ölçeğinde değişen madde fonksiyonlarının ağırlıklandırılmış cevap modeli ile belirlenmesi. *Ege Eğitim Dergisi*, 14(2), 63-78.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21(5), 667–706.
- Roussos, L.L., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33(2), 215–230.
- Schulz, W. (7–11 April 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, USA.
- Somer, O., Korkmaz, M., Dural, S., & Can, S. (2009). Detection of measurement equivalence by structural equation modeling and item response theory. *Turkish Journal of Psychology*, 24(64).
- 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.
- PISA Web Site: www.pisa.oecd.org adresinden 08.06.2012 tarihinde indirilmiştir.
- Taşdemir, M., ve Taşdemir, A. (2008). A comparison of Turkish primary school students' achievement in science and maths subjects. *Journal of Qafqaz University*, 22(2). 190-198.
- Thissen, D. (2001). *IRTLDIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. University of North Carolina at Chapel Hill: L. L. Thurstone Psychometric Laboratory.
- Yıldırım, H. H. & Berberoğlu, G. (2006). Judgmental and statistical analyses of the PISA 2003 mathematics literacy items. *International Journal of Testing*, 9(2), 108–121.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D., & Gelin, M. N. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement*, 63, 65–74.
- Zumbo, B. D., & Thomas, D. R. (1996). *A measure of effect size using logistic regression procedures*. Paper presented at the National Board of Medical Examiners, Philadelphia, PA.

EXTENDED ABSTRACT

Introduction

The PISA is one of the most comprehensive international assessment of students in the fields of math, science and reading skills. The study aims at analysing the reading skills, interest and attitude of fifteen year-old students in the PISA 2009. International tests such as TIMSS ve PISA have both cognitive-based achievement scores and results about other variables which have effects on the achievement score. Therefore, they provide the program developers with the opportunity to recognize insufficient points to be taken into consideration in the revision of educational programs. In other words, instead of focusing only on achievement score the other related variables should also be considered. There are many variables that accounts for achievement in the cognitive area. These variables include attitude, self-efficacy and other affective variables. Therefore, students' affective qualities should also be measured in addition to cognitive ones. The PISA items are originally written in English and then, are translated into the language of the related country. Because of this, the concept of item bias is very important in PISA and other similar evaluation settings where the items are adapted into a different culture or language. Because items may not be translated properly or adaptation is not sufficient. It is common that a measurement tool is not expected to be biased or to be advantageous for any group. The variables measured by

the international student assesstment tools such as PISA or TIMSS may vary based on student characteristics (i.e. gender or culture). It leads to lower validity of the measurement and students' affective and cognitive qualities may be erroneously measured. One of the first steps to determine the level of bias in the measurement tool is to identify those items with differential item functioning (DIF). Especially in the development of the measurement tools which are used to measure both the cognitive and affective characteristics of subjects the analysis to identify such items are important. Based on the results of the analyses the reasons for the DIF can be revealed and these are avoided in future studies. Research indicates that international test items concerning the measurement of psychological patterns involve item functioning based on culture. This study aims at idenfying whether or not the attitude items in regard to reading in PISA 2009 student questionnaire involve culture-based DIF using the techniques of the ordinal logistic regression, poly-SIBTEST and item response theory-likelihood ratio. It is expected that the results of the study will inform test development specialists about the test bias and validity. On the other hand, the study will reveal whether or not attitude items in the test mentioned above measure the student attitude in an unbiased way.

Method

The sample of the study includes the attitude items in PISA used for the students from Turkey, Finland, Russia, Singapour, New Zealand and Australia. The varied functioning of items in different groups may stem from the linguistic or cultural differences or achievement levels of student groups. Therefore, the country groups used in the study were determined so as to make a comparison among them. For instance, Turkey is in the rank of 39 in PISA 2009, whereas Finland is in the third rank. Furthermore, these two countries have different achievement levels as well as diffierent linguistic and cultural characteristics. In addition, another comparison was made between Turkey and Russion which is in the rank of 41 in PISA 2009. Therefore, although these two countries have similar achievement level in PISA 2009, there are linguistic and cultural differences between them. Another comparison was made between Australia and New Zealand both of which have higher levels of achievement in PISA 2009 and which have similar linguistic and cultural background and between Singapour and New Zealand which have similar levels of achievement in PISA 2009 and which have similar linguistic but different cultural background.

Results and Discussion

The results of three DIF techniques indicate that more cultural and linguistic differences more the number of items with DIF. More specifically, it is found that 91% of the items involved DIF among the countries with different linguistic and cultural background, whereas it was 82% of items that involved DIF among the countries with different cultural but similar linguistic background. Only 37% of the items were found to involve DIF if countries have similar linguistic and cultural background. It was further found that the achievement rank of the countries in PISA 2009 has no effect on the level of DIF in the items. Therefore, it is determined that less items involved DIF in the students questionnaires used for the Austrialian and New Zealanders who have similar linguistic and cultural background. The rate of items with DIF used for the students groups with different cultural background is more than 90% of the total items. However, linguistic differentiation does not have significant effect on the number of items with DIF. Therefore, not linguistic background but cultural background lead to DIF in the items. This finding is consistent with that of the study by Asil (2012). Therefore, translation and adaptation of items are very significant and cultural coordiantion should be achieved in these processes. The level of DIF in the international tests such as PISA, PIRLS and TIMSS should be revealed and the resons

should be identified. On the other hand, translation of items should be carried out by expert translators who can take into account linguistic and cultural factors. In addition, translators should work in coordination with measurement experts. Therefore, the items used for different country samples should be analysed in terms of DIF.