



Uluslararası Öğrenci Değerlendirme Programı (PISA-2012) Okuryazarlık Testlerinin Ölçme Değişmezliğinin İncelenmesi¹

Investigation of Measurement Invariance Of Literacy Tests In The Programme For International Student Assessment (PISA-2012)

Betül Karakoç Alatlı, Gaziosmanpaşa Üniversitesi, betulkarakocalatli@gmail.com
Ömay Çokluk Bökeoğlu, Ankara Üniversitesi, cokluk@education.ankara.edu.tr

Öz. Bu araştırma ile PISA-2012 matematik ve fen okuryazarlığı ile okuma becerileri alt testlerinin dil değişkenine göre ölçme değişmezliğinin incelenmesi amaçlanmıştır. Araştırma tarama modelinde betimsel bir araştırma olması ile birlikte PISA 2012 uygulamasına katılan Avustralya, Fransa, Şangay-Çin ve Türkiye örneklemelerinde 3 no'lu kitapçığı (yayınlanan maddelerin en fazla sayıda ve üç alandan da maddelerin bulunduğu tek kitapçık) alan 2149 öğrenciden elde edilen veriler üzerinden yürütülmüştür. Testlerin farklı dil formlarına ilişkin, faktör yapılarının değişmezliği Çok Gruplu Doğrulayıcı Faktör Analizi (ÇGDFA) ile maddelerin değişen madde fonksiyonu gösterme durumları ise Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM), Simultaneous Item Bias Test (SIBTEST) ve Madde Tepki Kuramı'na dayalı Olabilirlik Oran (MTK-OO) teknikleri ile incelenmiştir. İstatistiksel yöntemlerin yanı sıra olası yanlılık kaynakları uzman görüşleri doğrultusunda belirlenmiştir. Her üç alt test için yapılan ÇGDFA sonuçlarına göre, testlerin dil değişkenine göre yapısal değişmezlik gösterdiği ancak metrik değişmezlik göstermediği belirlenmiştir. Değişen madde fonksiyonu (DMF) analiz sonuçları değerlendirilirken her üç DMF belirleme tekniğine göre ve en az B düzeyinde DMF gösteren maddeler DMF'li kabul edilmiştir. Buna göre matematik ve fen okuryazarlığı ile okuma becerileri testlerinde dil değişkenine göre, DMF'li olarak kabul edilen maddeler sırasıyla testlerin ortalama %35'i, %34'ü ve %22'sini oluştururken, ölçme değişmezliği gösteren maddeler sırasıyla testlerin %24'ü, %7'si ve %36'sını oluşturmaktadır. Her üç alanda da en fazla DMF'li madde, İngilizce ve Çince formu alan gruplar arasında belirlenmiştir. Yanlılık çalışması sonuçlarına göre olası yanlılık kaynakları; maddede yer alan sözcük ya da ifadelerin farklı anlamda kullanımı, madde içeriğine bir dil grubunun aşına olması ve çeviriye bağlı olarak maddelerin farklılaşması olarak belirlenmiştir.

Anahtar Sözcükler: Ölçme değişmezliği, değişen madde fonksiyonu, yanlılık, uluslararası öğrenci değerlendirme programı

Abstract. In this study, it was aimed at examining the measurement invariance of mathematical and science literacy tests and subtests of reading skills of PISA – 2012 in terms of the language variable. The research was designed as a descriptive study. The study was conducted on the data obtained from the test booklet-3 of PISA 2012 (it is the only booklet that had the highest number of items published and also had items from all of the three parts) which was responded by 2149 students from Australia, France, Shanghai – China and Turkey. The invariance of factor structures was examined by Multi-Group Confirmatory Factor Analysis (MGCFA) while the differential item functioning (DIF) was examined by Hierarchical Generalized Linear Modeling (HGLM), Simultaneous Item Bias Test (SIBTEST) and Likelihood Ratio (LR) based on Item Response Theory (IRT) in relation to language forms of the tests. In addition to this, experts' opinions were asked in order to determine the possible bias sources. According to MGCFA results, PISA – 2012 mathematical and science literacy and subtests of reading skills demonstrated structural invariance while there was no metric invariance. The items which demonstrated at least B level DIF according to all three DIF technique results were accepted as showing DIF. According to DIF analyses, the DIF items comprised of 35%, 34% and 22% of mathematical and science literacy tests and subtests of reading skills respectively while the items showing measurement invariance comprised of 24%, 7% and 36% of the tests in terms of language variable. The highest number of DIF items in all three parts was determined between the groups that took English and Chinese forms. As a result of bias analysis, some possible bias sources were as follows; using the words or phrases in different meanings in the items, being familiar to the item content in favor of a language group and the differentiation of the items depending on the translation.

Keywords: Invariance of measurement, differential item functioning, bias, PISA, literacy

¹ Bu çalışma, ilk yazarın doktora tezinden üretilmiştir.

SUMMARY

Introduction

The obligation of explanation by those to whom resources are commended to fulfill a task or duty can be defined as accountability (Baş, 2005). Accountability in education depends mostly on the evaluation of students' learning. Therefore, countries need standard measurement and evaluation systems in order to determine their accountability levels. However, countries such as Turkey make inferences about their educational systems based on international practices due to the lack of a national standard measurement and evaluation system. The fundamental assumption of these kinds of international measurement and evaluation practices is the measurement invariance (Gierl, 2000; Reise, Widaman, and Pugh, 1993). Violation of measurement invariance affect the reliability and validity adversely; thus, it poses as a threat for independent interpretations (Vandenberg and Lance, 2000).

In this research, it was aimed at examining the measurement invariance of reading skills, mathematical and science literacy subtests of PISA-2012 in terms of the language variable. Investigation of measurement invariance of tools used in PISA in order for comparisons and policy changes based on its results to be meaningful. This study is significant since it involves math and science literacy and reading skills test and it reveals a general framework of measurement invariance of PISA practice. In this study, Structural Equation Modelling (SEM), GADM, MTK-00, and SIBTEST were used in harmony in order to investigate the measurement invariance at test, subtest, and item levels. The possible bias sources were determined through taking expert opinions in the bias study conducted on items violating linguistic measurement invariance depending on statistical analyses. Therefore, the validity and relevance of comparisons and interpretations have become questionable.

Method

The study was designed as a descriptive study. The study was conducted on the data obtained from the test booklet-3 of PISA 2012 (it is the only booklet that had the highest number of items published and also had items from all of the three parts) which was responded by 2149 students from Australia, France, Shanghai – China and Turkey. The invariance of factor structures was examined by Multi-Group Confirmatory Factor Analysis (MGCFA) while the differential item functioning was examined by Hierarchical Generalized Linear Modeling (HGLM), Simultaneous Item Bias Test (SIBTEST) and Likelihood Ratio (LR) based on Item Response Theory (IRT) in relation to language forms of the tests. In addition to this, experts' opinions were asked in order to determine the possible bias sources.

Results

According to MGCFA results conducted on math literacy subtest to determine measurement invariance, the three factorial construct of the test demonstrated structural invariance; however, it didn't demonstrate metric invariance. Since this situation calls into bias doubts, measurement invariance analyses at item level was conducted using HGLM, LR and SIBTEST. When the items demonstrating DMF based on three techniques and at least at B level were accepted as the items with DMF, the groups with the highest number of items with DMF were those taking English-Chinese forms (44%) with 11 items. They were followed by groups taking English-Turkish (40%), French-Chinese (40%), Chinese-Turkish(36%), French-Turkish(32%) and English-French(12%) tests. Accordingly, the number of items with DMF was lowest in the groups taking English-French form, which are the main languages of the forms, when compared with groups taking the translated forms. This situation shows that one of the possible reasons of DMF could be the translation. There are six items (24%) in math literacy test, which aren't accepted as the item with DMF in all language groups, showing measurement invariance in terms of language variable (English, French, Chinese, and Turkish). Out of 19 items which didn't

demonstrate measurement invariance in terms of different language groups, 5 were published. According to bias study conducted on these items based on the opinions of 21 experts, only one item was determined not to provide an advantage to a group. Moreover, it was determined that items demonstrated bias in terms of different language groups based on mostly translation (43). Other bias sources were different uses of words or expressions in the item (29), familiarity to the item content (29), item format or stylistic properties (12), cultural differences in terms of the characteristic measured by the item (9), and others (22). Out of 15 science literacy items, 14 (93%) were determined to be items with DMF at least at one comparison in terms of different language groups. Only one item (7%) demonstrated linguistic invariance. Seven items of the reading skills (50%) were determined to be items with DMF at least at one comparison in terms of different language groups. On the other hand, seven items of reading skills (50%) demonstrated measurement invariance in terms of language variable.

Discussion and Conclusion

In the study, MGCFA results, PISA – 2012 reading skills, mathematical and science literacy subtests of demonstrated structural invariance while there was no metric invariance. The items which demonstrated at least B level DIF according to all three DIF technique results were accepted as showing DIF. According to DIF analyses, the DIF items comprised of 35%, 34% and 22% of mathematical and science literacy tests and subtests of reading skills respectively while the items showing measurement invariance comprised of 24%, 7% and 36% of the tests in terms of language variable. The highest number of DIF items in all three parts was determined between the groups that took English and Chinese forms. As a result of bias analysis, some possible bias sources were as follows; using the words or phrases in different meanings in the items, being familiar to the item content in favor of a language group and the differentiation of the items depending on the translation.

GİRİŞ

Eğitim, birçok yapı üzerinde etkisi gözlenen aynı zamanda birçok faktör tarafından etkilenen bir sistemdir. Buna bağlı olarak eğitime ilişkin birçok gösterge (eğitime ayrılan bütçe, okur-yazarlık oranları) ülkelerin gelişmişlik düzeylerinin de bir göstergesi olması sonucunu doğurmaktadır. Eğitim ile gelişmişlik düzeyi arasındaki ilişki, ülkelerin eğitim sistemlerini her boyutta değerlendirmelerini gerektirmektedir. Bu noktada “eğitimde hesapverebilirlik” kavramı ön plana çıkmaktadır. Hesapverebilirlik, “bir işi veya görevi yapmak üzere kendilerine kaynak emanet edilenlerin, kaynakları emanet edenlere açıklamada bulunma zorunluluğu” olarak tanımlanmaktadır (Baş, 2005; akt: Balcı, 2011). Eğitimde hesapverebilirlik ise, norm ve yönetmeliklere uygunluk ile sonuçlara dayalı hesapverebilirlik olmak üç öğeden oluşmaktadır. Öğrenci öğrenmelerinin değerlendirilmesine dayanan sonuçlara dayalı hesapverebilirlik ise diğer iki öğeye göre eğitimde hesapverebilirlik noktasında çok daha önemli olarak görülmektedir (Anderson, 2005). Ulusal boyutta hesapverebilirlik için okul çapında yürütülen ölçme ve değerlendirme çalışmaları yeterli değildir. Özellikle zorunlu eğitim süresince alınan temel derslerdeki öğrenmelerin niteliğini değerlendiren, sık aralıklarla yapılan ve buna göre paydaşlarına geri bildirimler veren standart ölçme ve değerlendirme sistemlerine olan ihtiyaç eğitimde hesapverebilirlik adına oldukça önemli ve gereklidir (Demirtaşlı, 2014). Buna göre Türkiye ulusal düzeyde standart bir ölçme ve değerlendirme sistemine sahip değildir.

Türkiye eğitim sistemi incelendiğinde, ortaöğretim ve yükseköğretim kurumlarına öğrenci yerleştirme amacına hizmet eden, öğrencileri başarılarına göre sıralamaktan ibaret, geçerliği ve güvenilirliği sınırlı ölçme ve değerlendirme uygulamalarına rastlanmaktadır. Bu türlü uygulamaların yüksek standartlı, ulusal ölçekli, geçerliği ve güvenilirliği yüksek özellikte olması gerekmektedir. Öğrencileri içinde buldukları gruba göre değerlendiren, mutlak performansları hakkında güvenilir bilgi vermeyen bu tür uygulamaların sonuçlarını temel alarak eğitim sisteminin niteliği hakkında doğru çıkarımlarda bulunmak ise mümkün görünmemektedir. Eğitim sisteminin değerlendirilmesi ve sistemdeki aksaklıkları belirlemeye yönelik sağlıklı analizler yapılabilmesi için, öğrencilerin beceri düzeylerinin farklı ülkelerdeki öğrencilerle karşılaştırılması da, ulusal değerlendirmeler kadar önemli görülmektedir (Demirtaşlı, 2014). Uluslararası boyutta öğrenci başarısının değerlendirilmesi amacı ile Türkiye, Uluslararası Öğrenci Değerlendirme Programı-PISA, Uluslararası Okuma Becerileri Gelişim Projesi (Progress in International Reading Literacy Study-PIRLS) ve Uluslararası Matematik ve Fen Bilimleri Eğilimleri Çalışması (Trends in International Mathematics and Science Study-TIMSS) araştırmalarına katılmıştır, TIMSS ve PISA uygulamalarına katılmaya halihazırda da devam etmektedir.

Uluslararası düzeyde yapılan en büyük eğitim araştırmalarından biri, Ekonomik İşbirliği ve Kalkınma Teşkilâtı (Organisation for Economic Co-Operation and Development-OECD) tarafından düzenlenen PISA’dır. Bu uygulama ile katılımcı ülkelerin kendi eğitim sistemlerini değerlendirmelerine fırsat vermek ve reformlara kaynaklık etmek amaçlanmaktadır. PISA ile katılımcı ülkelerin gündemine “eğitim” kavramı gelmekte ve aynı zamanda eğitim konusunda karar vericilerin eyleme geçmesini desteklemektedir. Bununla birlikte PISA uluslararası uygulama olmasına dayanarak yalnızca ülkelerarası farklılıkları değil, ülke içi farklılıkları da ortaya çıkarmaktadır. Bununla birlikte bu farklılıkların giderilmesi yönünde yapılacak reformlara kaynaklık etmektedir. PISA sonuçlarına dayalı olarak başarılı ülkelerin belirlenmesi üzerine bu ülkelerin eğitim sistemleri incelenebilmekte, daha düşük başarı düzeyine sahip ülkelerin reformlarına örnek oluşturmaktadır (Breakspear, 2012). PISA’nın eğitim politikaları üzerinde etkili olduğunu gösteren araştırmalar hem OECD, hem de bağımsız araştırmacılar tarafından yapılmaktadır (Figazzolo, 2008; Hopkins, Pennock, Ritzen, Ahtaridou ve Zimmer, 2008; Breakspear, 2012).

Eğitimde hesapverebilirlik noktasında bu denli önemli görülen uluslararası ölçme ve değerlendirme uygulamalarına ilişkin yapılan değişikliklerin ve karşılaştırmaların en temel varsayımı ise, ölçme değişmezliğinin sağlanmasıdır (Gierl, 2000; Reise, Widaman ve Pugh, 1993; Vandenberg ve Lance, 2000). Ölçme değişmezliği varsayımı ilgili test veya madde düzeyinde

sağlanmıyorsa, bu durum geçerlik ve güvenilirliği olumsuz yönde etkilemekte, buna bağlı olarak yapılacak bağımsız yorumlar için de tehdit oluşturmaktadır (Vandenberg ve Lance, 2000).

Farklı grupların üyeleri arasında, belirli bir gizil yapıya ilişkin aynı puanlara sahip olan bireylerin, madde ve alt ölçekler düzeyinde de aynı gözlenen puanlara sahip olmaları durumu ölçme değişmezliği olarak tanımlanmaktadır [American Educational Research Association (AERA), American Psychological Association (APA) ve National Council on Measurement in Education (NCME), 1999]. Uluslararası uygulamalar düşünüldüğünde diğer bir ifadeyle uyarlama süreci açısından bu gruplar, farklı dilleri konuşan bireylerdir. Ölçmelerin değişmez olması için, örtük değişkenlerle gözlenen değişkenler arasındaki ilişkilerin farklı gruplarda aynı olması gerekmektedir. Diğer bir ifadeyle, ilgili gruplar, örneğin testin farklı dil formlarını alan gruplar, karşılaştırıldığında, gözlenen puanlar ile gizil yapılar arasında var olan ilişki, gruplar arasında değişmiyor (aynı) ise, gruplar arasında değişmez ölçümler elde edilmiş demektir (Drasgow ve Kanfer, 1985).

Eğitimde ve Psikolojide Ölçme Standartları (AERA, APA ve NCME, 1999) ve Test Uyarlama Yönergeleri (ITC, 2005) incelendiğinde kültürlerarası çalışmalar yapan araştırmacıların farklı dillerdeki ölçme araçlarından elde edilen puanların karşılaştırılabilirliğine (ölçme değişmezliğine) ilişkin kanıtlar sunmaları gerektiği açıkça belirtilmiştir. Ölçme değişmezliği incelemeleri için alanyazında sıklıkla iki yaklaşımın kullanıldığı ve önerildiği görülmektedir (Cheung ve Rensvold, 2000; Little, 1997; Raju, Laffitte ve Byrne, 2002; Lord, 1980; Reise, Widaman ve Pugh, 1993; Stark, Chernyshenko ve Drasgow, 2006; Vandenberg ve Lance, 2000):

a. Yapısal Eşitlik Modellemeleri (YEM)

b. Madde Tepki Kuramı'na (MTK) dayalı Değişen Madde Fonksiyonu (DMF) ve Değişen Test Fonksiyonları (DTF) incelemeleri.

Ölçme değişmezliği analizleri kapsamında test geliştiricilere ve araştırmacılara DMF analizlerinde birden çok tekniği kullanmaları da önerilmektedir (Hambleton 2006). Farklı DMF belirleme tekniklerinden her ne kadar benzer sonuçlar elde edilse de, farklı algoritmalar, farklı eşitleme kriterleri ve maddeyi DMF'li olarak etiketlemek için farklı kesme noktaları kullandıkları için aralarında tam bir uyum gözlenmemektedir (Acar, 2008; Atalay, Gök, Kelecioğlu, Arslan, 2012; Bakan Kalaycıoğlu ve Berberoğlu, 2010; Çepni, 2011; Doğan ve Öğretmen, 2008; Gök, Kelecioğlu ve Doğan, 2010). Bu nedenle bu çalışmada MTK-Olabilirlik Oran (OO) tekniğinin MTK'ya dayalı ve parametrik DMF belirleme tekniği olmasına karşılık KTK'ya dayalı ve parametrik olmayan DMF belirleme tekniği Simultaneous Item Bias Test (SIBTEST) madde düzeyinde ölçme değişmezliğini incelemek amacıyla tercih edilmiştir (Camilli ve Shephard, 1994; Gierl, Khaliq ve Boughton, 1999). SIBTEST tekniği ile DMF'li madde belirlenirken madde dışında kalan diğer maddelere verilen yanıtlar üzerinden bireylerin gerçek puanları kestirilerek bireyler bu puanlara göre eşleştirilmektedir. Bununla birlikte analizlerin gerçekleştirilmesi için gerekli yazılımın ulaşılabilirliği, uygulama ve yorumlanmasının kolaylığı SIBTEST tekniğinin DMF belirleme tekniği olarak seçilmesinde etkili olmuştur.

Kamata (2001), ölçme değişmezliği belirleme çalışmalarında "Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM-Hierarchical Generalized Linear Model)" yaklaşımının kullanılmasını önermektedir. Bu önerisi birçok araştırma ile test edilmiş ve hiyerarşik yapıdaki verilerde ölçme değişmezliğinin saptanması için oldukça avantajlı olduğu görülmüştür (Acar, 2008; Chou, Bentler ve Pentz, 1998; Kamata, Chaimongkol, Genç ve Bilir, 2005; Kim, 2003; Luppescu, 2002; Pan, 2008; Rawls, 2009; Wendorf, 2002). Bu araştırma ile PISA uygulamasının kültürlerarası ölçme değişmezliği test, alt test ve madde düzeyinde belirlenmesi için YEM, GADM ve MTK'nın birlikte kullanıldığı bir çalışmadır. Bu yönüyle de Türkiye'de yapılan ilk çalışmadır. Bununla birlikte çalışmada matematik ve fen okuryazarlığı ile okuma becerileri alt testlerinin birlikte yer alması ile PISA uygulamasının ölçme değişmezliğine ilişkin genel tablonun ortaya çıkartılması da önemli görünmektedir.

Yukarıda belirtilen istatistiksel yöntemlere dayalı ölçme değişmezliği incelemeleri sadece alt gruplarda sistematik olarak farkın olup olmadığını var ise bu farkın manidarlığını ortaya koymaktadır, gruplararası farka ilişkin nedensellik ifade etmezler (Osterlind ve Everson, 2009). Dolayısıyla da, ölçme değişmezliğinin sağlanmadığı istatistiksel olarak belirlense bile bu durum,

ilgili ölçme aracının veya içerisinde yer alan maddelerin yanlı olduğunun, bir başka deyişle mutlaka gruptan birisi için adaletsizlik oluşturduğunun bir göstergesi değildir. Ölçme değişmezliğinin incelendiği çalışmalarda, maddenin veya testin yanlılığına ilişkin yapılan istatistiksel analizler gerekli, ancak yeterli değildir. Bu noktada yanlılık çalışmaları önem kazanmaktadır. Yanlılık, testi alan bireylerin farklı özelliklerinden (cinsiyet, etnik köken vb.) kaynaklanabilir. Bu nedenle istatistiksel yöntemlerle yapılan belirlemelerin uzman görüşleri doğrultusunda değerlendirilmelidir. PISA gibi uluslararası uygulamalarda ise en önemli yanlılık kaynağının, kültürün de en önemli ögesi olan dil değişkeninin olduğu bilinmektedir. Bu doğrultuda PISA 2012 uygulamasında kullanılan okuryazarlık testlerinin ve testlerde yer alan maddelerin dilsel ölçme değişmezliğini sağlamadığının belirlendiği durumlarda, bu durumun nedenleri uzman görüşleri ile sorgulanmıştır. Böylelikle yapılan karşılaştırmaların, yorumların geçerliği ve anlamlılığı sorgulanabilir duruma gelmiştir.

Ülkelerin eğitimde hesapverebilirlikleri adına önemli bir yeri olan PISA-uluslararası eğitim araştırmasına pek çok ülke katılmaktadır. Bu süreçte kullanılan ölçme araçları da katılan ülkeler bağlamında birçok kültüre uyarlanmaktadır. Böylesi önemli bir araştırmanın kültürlerarası ölçme değişmezliğinin test edilmesi ise zorunlu ve gereklidir. Ölçme değişmezliği incelemeleri genel olarak değerlendirildiğinde, Yapısal Eşitlik Modellemesi (YEM)'in alt test düzeyinde, Aşamalı Doğrusal Modelleme ve MTK'nın ise test ve madde düzeyinde ölçme değişmezliği incelemelerinde kullanılması önerilmektedir. Bununla birlikte bu çalışmada madde düzeyinde ölçme değişmezliğinin incelenmesinde birden çok tekniğin kullanılması önerisine dayanarak SIBTEST tekniği DMF belirlemek amacıyla kullanılmıştır. Ayrıca, söz konusu analizler sonucunda ölçme değişmezliğinin sağlanamadığı belirlenirse, bu durumun nedenleri de sorgulanmalıdır. Bu gerekçelerle PISA-2012 uygulamasında kullanılan okuryazarlık testlerinin, dil değişkenine göre ölçme değişmezliğinin ÇGDFA, MTK-00, SIBTEST ve GADM ile incelenmesi bu çalışmanın problemini oluşturmaktadır. Bu durum ülkeler arasında yapılan karşılaştırmaların adil ve anlamlı olabilmesi için bir zorunluluk olarak ortaya çıkmaktadır.

Bu çalışmanın genel amacı PISA-2012 matematik ve fen okuryazarlığı ile okuma becerileri alt testlerinin dil değişkenine göre ölçme değişmezliğinin incelenmesidir. Bu temel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- 1) Matematik ve Fen okuryazarlığı ile okuma becerileri alt testleri dil değişkeni bakımından ölçme değişmezliği göstermekte midir?
 - a) Testlerin faktör yapısı dil değişkeni bakımından ölçme değişmezliği göstermekte midir?
 - b) Testlerde yer alan maddeler dil değişkeni bakımından ölçme değişmezliği göstermekte midir?
- 2) Matematik okuryazarlığı alt testinde dil değişkenine göre ölçme değişmezliği göstermeyen madde(ler) belirlenirse, bu durumun olası nedenleri nelerdir?

YÖNTEM

Araştırma Modeli

PISA-2012 okuryazarlık alt testlerinin dil değişkenine göre ölçme değişmezliğinin incelendiği bu araştırma ile var olan bir durum var olduğu şekliyle ortaya konulması amaçlandığından betimsel bir çalışmadır (Fraenkel ve Wallen, 2006; Karasar, 2011).

Çalışma Grubu

PISA uygulamasının evrenini, genel olarak uygulamaya katılan ülkelerde örgün eğitime devam eden 15 yaş grubu öğrenciler oluşturmaktadır. 2012 yılında yapılan PISA uygulamasında 34'ü OECD üyesi ülke olmak üzere yaklaşık 65 ülke yer almaktadır. Bu çalışmanın genel amacı doğrultusunda, ülkelerin araştırmaya dahil edilmesinde, kültürlerarası en önemli farklılıklardan biri olan dil ögesi dikkate alınmıştır. Bu doğrultuda çalışmanın çalışma grubunu PISA-2012 uygulamasına katılan Avustralya (İngilizce), Fransa (Fransızca), Şangay-Çin (Çince) ve Türkiye

(Türkçe) örneklemelerinden 3 no'lu kitapçığı alan öğrenciler oluşturmaktadır. İngilizce ve Fransızcanın, PISA uygulaması kaynak dilleri olması nedeniyle, kaynak dilde formları alan Avustralya ve Fransa örneklemeleri araştırmaya dâhil edilmiştir. Şangay-Çin ve Türkiye örneklemeleri ise ölçme araçlarının çeviri formlarını almaktadırlar. Çalışma grubunda yer alan öğrencilerin ülkelere göre dağılımı Tablo 1'de yer almaktadır.

Tablo 1. PISA-2012'ye katılan Avustralya, Fransa, Şangay-Çin ve Türkiye örneklemelerinin öğrenci dağılımı

Ülke	Öğrenci Sayısı (3 no'lu kitapçık)		Toplam Öğrenci Sayısı	
	f	%	f	%
Avustralya	1037	48,26	14481	49,73
Fransa	319	14,84	4613	15,84
Şangay-Çin	434	20,20	5177	17,78
Türkiye	359	16,70	4848	16,65
Genel Toplam	2149	100	29119	100

Tablo 2 incelendiğinde Avustralya %48,26 ile çalışma grubunda yer alan ülkeler arasında en büyük örnekleme sahip ülkedir. Bununla birlikte Fransa (%14,84), Şangay-Çin (%20,20) ve Türkiye (%16,70) örneklemelerinin çalışma grubunda benzer büyüklükte yer aldığı ifade edilebilir. Ülkelerin çalışma grubunda yer alma oranları evrende yer alma oranları ile benzerlik göstermektedir.

Veriler ve Elde Edilmesi

Araştırmada kullanılan PISA-2012 verileri OECD'nin resmi sitesinden (www.pisa.oecd.org) elde edilmiştir. OECD tarafından veriler ve puanlama ile ilgili yapılan çalışmalara ek olarak bu çalışmada ölçme değişmezliği analizleri için yanıtlanmayan, ulaşılmayan ve birden fazla seçeneğin yanıtlandığı maddeler yanlış yanıt "0" olarak yeniden kodlanmıştır. İlgili kitapçıkta yer alan maddeler incelendiğinde 52 maddenin ikili, 2 maddenin (M903Q01, S519Q01) ise kısmi puanlandığı belirlenmiştir (OECD, 2014). Bu çalışma kapsamında kısmi puanlanan maddelere ilişkin kısmi puanlar doğru yanıt "1" olarak yeniden kodlanmıştır.

PISA 2012 matematik ve fen okuryazarlığı ile okuma becerileri testlerine ilişkin bilgiler için PISA 2012 teknik raporundan (OECD, 2014) yararlanılmıştır. İlgili alanlardan maddeler içeren toplam 13 kitapçık kullanılmıştır. Öğrencilere bu kitapçıklardan biri rastgele olarak atanmaktadır. Her üç yılda bir yapılan uygulamada ağırlıklı alana ilişkin maddelerden örnek maddeler yayınlanmaktadır. PISA 2012 araştırmasının ağırlıklı alanı matematik okuryazarlığıdır. Çalışma kapsamında DMF gösteren maddelere ilişkin yanlılık çalışmalarının yapılması amaçlanmaktadır. Bu nedenle çalışmada incelenen kitapçıkta olabildiğince fazla sayıda yayınlanan madde yer alması yanlılık çalışmaları için önemli görülmektedir. PISA 2012 uygulaması için 26 matematik test maddesi yayınlanmıştır. Kitapçıklar incelendiğinde bu 26 maddenin bir arada yer aldığı bir kitapçık bulunmamaktadır. Bununla birlikte açıklanan 26 maddeden en fazla 13'ünün bir arada bulunduğu kitapçıkların 1, 3, 4 ve 6 no'lu kitapçıklar olduğu görülmektedir. Bu kitapçıklar içerisinde de araştırmanın amaçları doğrultusunda, her üç alandan da maddelerin yer aldığı tek kitapçık olan 3 no'lu kitapçık çalışma kapsamında incelenmiştir. İlgili kitapçıkta 25 matematik okuryazarlığı, 15 fen okuryazarlığı ve 14 okuma becerileri alanından olmak üzere toplam 54 madde yer almaktadır. Testlerin faktör yapısı incelendiğinde Matematik okuryazarlığı alt testi, "Durumları, problemleri matematiksel olarak formüle etme", "Matematiksel kavramları, gerçekleri, yöntemleri kullanma ve akıl yürütme", "Matematiksel çıktılarını yorumlama, uygulama ve değerlendirme" olmak üzere üç faktörlü yapıya sahiptir. Fen okuryazarlığı alt testi ise, "Bilimsel durumları ayırt etme", "Bilimsel kanıtları kullanma" ve "Olguları bilimsel olarak açıklama" olarak adlandırılan üç faktörden oluşmaktadır. Okuma becerileri alt testi, "Bilgiye ulaşma ve bilgiyi hatırlama", "Bilgileri bir araya getirme ve yorumlama" ve "Kendi düşüncelerini yansıtmaya ve metni değerlendirme" olmak üzere üç faktörlü yapıya sahiptir. Verilerin analizi bu yapılar üzerinden gerçekleştirilmiştir (OECD, 2014).

PISA-2012 matematik okuryazarlığı maddelerinin İngilizce ve Türkçe formlarını alan öğrenci gruplarının karşılaştırılması sonucu DMF'li olduğu saptanan ve açıklanan maddeler üzerinde olası yanlışlık kaynakları hakkında uzman görüşlerine başvurulmuştur. Bu amaç doğrultusunda DMF gösteren maddelerde dil değişkeni bakımından yanlışlık kaynaklarının neler olabileceğine ilişkin literatür taraması yapılmıştır. Yanlışlık kaynakları göz önünde bulundurularak altı maddeden oluşan madde inceleme formu araştırmacı tarafından geliştirilmiştir. Uzmanlara görüşleri alınmadan önce DMF ve yanlışlık hakkında bilgi verilmiş, bu bilgiler uzman görüş formu yönergesinde ayrıca ele alınmıştır. Çalışma kapsamında 14 ölçme ve değerlendirme uzmanı (altı uzmanın lisans eğitimi matematik alanında, dört uzmanın lisans eğitimi yabancı dil eğitimi, dört uzmanın ise lisans eğitimi diğer alanlardandır), üç yabancı dil eğitimcisi ve dört matematik eğitimcisi olmak üzere toplam 21 uzmanın görüşüne başvurulmuştur. Uzmanlar DMF gösteren beş madde için ayrı ayrı görüş bildirmiştir.

Verilerin Analizi

Bu çalışmada sırasıyla şu analizler yapılmıştır;

1. Veri setinin kayıp veri ve uçdeğer bakımından incelenmesi
2. ÇGDFA analizleri
3. GADM analizleri
4. MTK-OO analizleri
5. SIBTEST analizleri
6. DMF'nin olası nedenlerinin incelenmesi

Verilerin analizine geçilmeden önce veri seti kayıp veri ve uç değerler bakımından incelenmiş ve gerekli düzeltmeler yapılmıştır. ÇGDFA öncesinde veriler normallik ve çoklu bağlantı varsayımları açısından incelenmiş ve varsayımların karşılandığı belirlenmiştir (Tabachnick ve Fidell, 2007). Farklı dillere uyarlanan ölçme araçlarında yer alan maddelerin gruplar arasında ölçme değişmezliği gösterip göstermediğinin incelenmesinden önce, ölçülen özelliğe ilişkin faktör yapısının her bir grup için ayrı ayrı incelenmesi gerekmektedir (Sireci ve Swaminathan, 1996). Her bir alt test için öğrenci performansına ilişkin faktör yapısı düşünme süreçleri dikkate alınarak oluşturulmuştur. Belirlenen faktör yapıları doğrultusunda da ölçme modelleri oluşturulmuştur. Bir sonraki aşamada ise ölçme modellerinin, önce her bir ülke için ayrı ayrı ve ardından tüm ülke verilerinin birleştirilmesi sonucu elde edilen veri seti için bir kez daha olmak üzere uyumu incelenmiştir. Ölçme modelinin uyumunun değerlendirilebilmesi için kurulan modelin bir bütün olarak kabul edilebilir bir model olup olmadığını belirlemek amacıyla uyum iyiliği istatistikleri (Goodness of fit statistics) incelenmiştir. Çok Gruplu Doğrulamalı Faktör Analizi ile ölçme değişmezliği iç içe geçmiş (nested) dört hiyerarşik modelin veya hipotezin test edilmesi ile ortaya konulmaktadır. İlk adımda, ölçme modeli için yapısal değişmezlik test edilirken faktör yükleri, regresyon sabitleri ve hata varyanslarının serbest tahminlenmesine izin verilerek, yalnızca gruplar için faktör sayısı ve yüklenme örüntüsünün (loading pattern) sınırlandırıldığı "Model A" uyumu incelenmektedir. Metrik değişmezlik parametre sınırlamasının ilk yapıldığı adımdır. Faktör yüklerinin gruplar arasında sınırlandırıldığı, regresyon sabitleri ve hata varyanslarının ise serbest tahminlenmesine izin verildiği bu model "Model B" olarak adlandırılmıştır. Metrik değişmezliğin test edilmesi amacıyla, Model A ile Model B arasında model uyumundaki değişim değerlendirilmektedir. Alan yazında model uyumundaki değişimin değerlendirilmesi için, Ki kare (X^2) ve CFI değerlerindeki değişimin incelenmesi önerilmektedir. Ki kare değerindeki değişimin, iki modelin serbestlik dereceleri arasındaki farka göre belirlenen serbestlik derecesine göre manidarlığı incelenir ve manidar bir değişim olduğu belirleniyor ise metrik değişmezliğin sağlanmadığı yorumu yapılır. Ki-kare değerinin örneklem büyüklüğüne duyarlı olması nedeniyle ölçme değişmezliği analizlerinde uyum iyiliği indekslerinden CFI indeksindeki değişimin de değerlendirilmesi önerilmektedir. CFI indeksindeki değişim $-0,01 \leq \Delta CFI \leq 0,01$ aralığında yer alıyor ise bu durum değişmezlik için gerekli olan koşulun sağlandığı anlamına gelmektedir. Diğer bir ifadeyle CFI indeksindeki değişim belirtilen aralıkta ise, faktör yüklerinin gruplar arasında değişmez olduğu, metrik değişmezliğin sağlandığı kabul edilir. (Cheung ve Rensvold, 2002; Hu ve Bentler, 1998; Rensvold ve Cheung, 1998; Vandenberg ve

Lance, 2000; Wu ve diğerleri, 2007). (Brown, 2006; Şencan, 2005, Şimşek, 2007; Tabacknick ve Fidell, 2007). ÇGDFA için LISREL 8.8 programından yararlanılmıştır.

Hiyerarşik yapıların bulunduğu veri gruplarında ölçme değişmezliğinin incelenmesi için Kamata (2001) tarafından önerilen Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM) analizi yapılmıştır. Analiz öncesinde çoklu bağlantı, varyansların homojenliği, hataların bağımsızlığı, düzey 1 hatalarının normalliği, düzey 2 hatalarının normalliği varsayımları bakımından veri seti incelenmiş ve varsayımların karşılandığı görülmüştür (Hofmann, Griffin ve Gavin, 2000; Raudenbush ve Bryk, 2002). Analiz için HLM7 programı kullanılmıştır.

PISA-2012 matematik ve fen okuryazarlığı ile okuma becerileri alt testlerinde yer alan maddelerinin dil değişkeni bakımından ölçme değişmezliği gösterip göstermediği MTK'ya dayalı DMF belirleme tekniklerinden Olabilirlik Oran Testi Tekniği (MTK-OO) ile incelenmiştir. Analiz öncesi veri seti tekboyutluluk ve yerel bağımsızlık varsayımları bakımından incelenmiş ve varsayımların karşılandığı sonucuna ulaşılmıştır. (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1989). Ayrıca MTK analizleri kapsamında her bir alt test için model veri uyumu incelenmiştir. Buna göre matematik ve fen okuryazarlığı testleri 3 parametrelilikle uyum gösterirken okuma becerileri testi 2 parametrelilikle uyum göstermiştir. MTK-OO tekniği ile DMF analizlerini yapabilmek için IRTLRDIF programından yararlanılmıştır.

SIBTEST tekniği ise KTK'ya dayalı parametrik olmayan bir DMF belirleme yöntemidir. Diğer tekniklerden farklı olarak herhangi bir varsayımı yoktur. Analiz için SIBTEST 1.7 programı kullanılmıştır (Abbott, 2007).

Tüm alt testlerde yer alan maddelerin dil değişkenine göre DMF gösterip göstermediğini incelemek üzere yapılan DMF analizlerinde, dil grupları analize dahil edilirken PISA 2012 sonuçlarına göre ilgili alanda daha başarılı olan grup odak grup olarak kodlanmıştır. Bununla birlikte anlamlılık düzeyi 0,05 olarak belirlenmiştir.

BULGULAR

Matematik Okuryazarlığı Alt Testinin Dil Değişkenine Göre Ölçme Değişmezliğine İlişkin Bulgular

Bu başlık altında “Matematik okuryazarlığı alt testi dil değişkeni bakımından ölçme değişmezliği göstermekte midir?” sorusu doğrultusunda sırasıyla test düzeyinde ölçme değişmezliğine, madde düzeyinde ölçme değişmezliğine ve madde yanlılığı çalışmasına ilişkin bulgular ve yorumlar yer almaktadır.

PISA-2012 matematik okuryazarlığı alt testinin farklı dil formlarının aynı faktör yapısına sahip olup olmadığı ÇGDFA ile belirlenmiştir. ÇGDFA öncesinde matematik okuryazarlığı alt testinin üç faktörlü yapısının her bir dil grubunda doğrulanıp doğrulanmadığı DFA ile test edilmiştir ve analiz sonuçlarına Tablo2’de yer verilmiştir.

Tablo2. Matematik Okuryazarlığı Testine İlişkin Kurulan Modelin Uyum İndeksleri

Dil Grubu	İstatistikler								
	X^2	Sd	X^2/sd	RMSEA	CFI	GFI	SRMR	AGFI	NNFI
İngilizce	637,02	269	2,37	0,036	0,98	0,95	0,036	0,94	0,97
Fransızca	477,23	269	1,77	0,049	0,96	0,89	0,053	0,87	0,95
Çince	312,41	269	1,16	0,019	0,99	0,95	0,038	0,93	0,99
Türkçe	358,43	269	1,33	0,030	0,98	0,93	0,045	0,91	0,98

Tablo2’de yer alan, kurulan modelin her bir dil grubu için uyum iyiliği indeksleri incelendiğinde; model uyumunun değerlendirilmesinde kullanılan kabul edilebilir aralıklarda ($X^2/sd \leq 2$, $GFI \geq 0.90$, $CFI \geq 0.95$, $NNFI \geq 0.95$, $SRMR \leq 0.05$ ve $RMSEA \leq 0.05$) yer almakta ve matematik okuryazarlığı alt testi, her bir dil grubunda iyi bir uyum göstermektedir.

İlk adımda, ölçme modeli için yapısal değişmezlik test edilirken faktör yükleri, regresyon sabitleri ve hata varyanslarının serbest tahminlenmesine izin verilerek, yalnızca gruplar için faktör sayısı ve yüklenme örüntüsünün sınırlandırıldığı “Model A”nın uyumu incelenmiştir

(Vandenberg ve Lance, 2000; Wu ve diğeri, 2007). Tablo 3'te Model A'ya ilişkin uyum iyiliği istatistikleri yer almaktadır.

Tablo3. Matematik Okuryazarlığı Testine İlişkin ÇG DFA Sonuçları

Model	X^2	Sd	X^2/sd	GFI	CFI	NNFI	SRMR	RMSEA	$\Delta X^2 (\Delta sd)$	ΔCFI
A	1785,09	1076	1,66	0,93	0,98	0,97	0,04	0,04	-	-
B	2292,14	1151	1,99	0,90	0,96	0,96	0,09	0,04	507,06 (75)	0,02

Tablo 3'te yer alan Model A'nın uyum iyiliği istatistikleri incelendiğinde, modelin iyi bir uyum gösterdiği görülmektedir. Bu durum testin dil değişkeni bakımından farklı gruplarda yapısal değişmezlik gösterdiği anlamına gelmektedir. Yapısal değişmezliğin sağlanması ile hiyerarşik adımlar doğrultusunda metrik değişmezliğin test edilmesi adımına geçilmiştir. Faktör yüklerinin gruplar arasında sınırlandırıldığı "Model B" ile Model A arasında model uyumundaki değişim değerlendirilmiştir. Tablo3 incelendiğinde $\Delta X^2=507,06$ 'dır. Bu değer, Δsd dikkate alınarak belirlenen ve kritik değer olan $X^2_{(75, 0.05)} = 96,21$ değerinden büyüktür. Buna göre faktör yükleri sabitlendiğinde, model uyumunun manidar bir şekilde farklılaştığı belirlenmiştir. Ki-kare değerinin örneklem büyüklüğüne duyarlı olması nedeniyle ölçme değişmezliği analizlerinde uyum iyiliği indekslerinden CFI indeksindeki değişimin de değerlendirilmesi önerilmektedir. Tablo 3 incelendiğinde matematik okuryazarlığı alt testi için iki model arasındaki CFI farkı 0,02 olarak hesaplanmıştır bu değişim kritik aralık olan $-0,01 \leq \Delta CFI \leq 0,01$ aralığında yer almamaktadır. CFI değerindeki değişim de X^2 değerindeki değişime benzer olarak, matematik okuryazarlığı alt testinin farklı dil grupları arasında metrik değişmezlik göstermediğine işaret etmektedir. Değişmezlik analizleri hiyerarşik bir yapı gösterdiğinden, metrik (zayıf) değişmezliğin sağlanmadığı adımda analize son verilmiş, güçlü ve katı değişmezlik kontrollerine geçilmemiştir.

Metrik değişmezliğin sağlanmadığında bu gibi durumlar madde yanlılığının bir göstergesi olarak kabul edilmektedir (Johnson, 1998; Prelow ve diğeri, 2000). Buna göre matematik okuryazarlığı testi için madde yanlılığı şüphesi ortaya çıkmıştır. Bu doğrultuda öncelikle madde düzeyinde ölçme değişmezliği analizleri yapılarak analiz sonuçlarına göre açıklanan maddeler üzerinden madde yanlılığı çalışması yapılmıştır.

PISA 2012 Matematik okuryazarlığı alt testinde yer alan maddelerin dil değişkeni bakımından ölçme değişmezliği bir diğer ifadeyle DMF gösterme durumları GADM, SIBTEST ve MTK-00 istatistiksel teknikleri ile incelenmiştir. DMF belirleme çalışmalarında analizler, araştırma kapsamındaki dil gruplarının ikili kombinasyonları (İngilizce-Fransızca, İngilizce-Çince, İngilizce-Türkçe, Fransızca-Çince, Fransızca-Türkçe, Çince-Türkçe) üzerinden yürütülmüştür. Analizler sonucunda her üç tekniğe göre ve en az B düzeyinde DMF gösteren maddeler DMF'li olarak kabul edilmiştir. Tablo 4'te matematik okuryazarlığı alt testinin farklı dil formlarını alan öğrencilerden elde edilen verilere göre testte yer alan ve dil değişkeni bakımından DMF'li kabul edilen maddeler ve bu maddelerin hangi dil grubu lehine DMF gösterdiği sunulmuştur.

Tablo 4. Matematik okuryazarlığı test maddelerine ilişkin DMF analiz sonuçları

Madde No	DMF Gösteren Maddeler ve Avantajlı Dil Grubu					
	İngilizce-Fransızca	İngilizce-Çince	İngilizce-Türkçe	Fransızca-Çince	Fransızca-Türkçe	Çince-Türkçe
1	-	İngilizce	İngilizce		Fransızca	-
2	-	-	-	Çince	-	-
3	-	İngilizce	-	Fransızca	-	Türkçe
4	İngilizce	İngilizce	İngilizce	Fransızca	-	-
5	-	-	-	-	Türkçe	-
6*	-	-	-	-	-	-
7	-	-	-	Fransızca	-	Türkçe
8	-	-	Türkçe	Çince	Türkçe	-
9*	-	-	-	-	-	-
10	-	Çince	Türkçe	-	Türkçe	-
11	Fransızca	-	-	-	-	-
12	-	-	Türkçe	-	Türkçe	Türkçe
13	Fransızca	Çince	Türkçe	-	-	-
14	-	Çince	-	Çince	-	Çince
15*	-	-	-	-	-	-
16	-	-	Türkçe	-	Türkçe	Türkçe
17	-	İngilizce	İngilizce	-	-	-
18*	-	-	-	-	-	-
19	-	İngilizce	İngilizce	Fransızca	Fransızca	-
20	-	Çince	İngilizce	Çince	Fransızca	Çince
21	-	-	-	-	-	Çince
22*	-	-	-	-	-	-
23	-	Çince	-	Çince	-	Çince
24*	-	-	-	-	-	-
25	-	İngilizce	-	Fransızca	-	Çince
Toplam	3(%12)	11(%44)	10(%40)	10(%40)	8(%32)	9(%36)

*Tüm dil gruplarına göre ölçme değişmezliği gösteren madde

Tablo 4'e göre DMF analizleri incelendiğinde DMF'li kabul edilen madde sayısı en yüksek 11 madde (%44) ile İngilizce-Çince formu alan gruplar arasında belirlenmiştir. Bunu sırasıyla İngilizce-Türkçe (%40), Fransızca-Çince (%40), Çince-Türkçe (%36), Fransızca-Türkçe (%32) ve İngilizce-Fransızca (%12) formu alan gruplar izlemiştir. Buna göre testin çeviri formunu alan gruplara göre test formunu kaynak dilde olan İngilizce-Fransızca formu alan gruplarda DMF'li kabul edilen madde sayısı en az sayıdadır. Bu durum DMF'nin olası nedenlerinden birinin çeviriye bağlı olabileceğini göstermektedir.

PISA 2012 matematik okuryazarlığı maddelerinden 19 tanesinin (%76) farklı dil gruplarında ölçme değişmezliği göstermediği yapılan ikili karşılaştırmalar sonucunda belirlenmiştir (Tablo 4). Bütün dil gruplarının ikili karşılaştırmalarında ortak olarak DMF'li olarak kabul edilmeyen bir diğer ifadeyle dil (İngilizce, Fransızca, Çince ve Türkçe) değişkenine göre ölçme değişmezliği gösteren, altı (%24) matematik okuryazarlığı maddesi (6, 9, 15, 18, 22 ve 24. maddeler) bulunmaktadır. Bu maddelerden 15 (M903Q01), 18 (M918Q05) 22 (M924Q02) ve 24. (M995Q02) maddeler yayınlanmıştır.

PISA-2012 Matematik okuryazarlığı maddelerinden dil değişkenine göre ölçme değişmezliği göstermeyen maddelerde olası yanlışlık kaynakları

Yanlılık çalışması, matematik okuryazarlığı maddelerinden İngilizce ve Türkçe formlarını alan gruplara göre DMF'li kabul edilen ve yayınlanan maddeler (M00FQ01, M918Q01, M918Q02, M923Q01, M923Q03) ile gerçekleştirilmiştir. Çalışma kapsamında 21 uzman, görüşlerini araştırmacı tarafından geliştirilen "Uzman Görüş Formu" ile iki aşamada bildirmişlerdir. Birinci aşamada uzman görüşleri doğrultusunda maddelerin yanlışlık durumları ve yanlışlık kaynakları belirlenmiştir. İkinci adımda ise daha doğru kararlar verebilmek amacıyla uzmanlardan ikinci kez aynı form üzerinden görüş alınmıştır. Buna göre DMF'li maddelerin bir dil grubuna avantaj sağlama durumlarına ilişkin ikinci uygulama sonucu elde edilen uzman görüşlerinin dağılımına Tablo 5'te yer verilmiştir.

Tablo 5. DMF'li maddelerin bir gruba avantaj sağlama durumlarına göre dağılımı

	Maddeler					
	M00FQ01	M918Q01	M918Q02	M923Q01	M923Q03	
Bir dil grubuna avantaj sağlamaz	15	7	7	3	3	
Bir dil grubuna avantaj sağlar	İngilizce	6	14	13	18	18
	Türkçe	-	-	1	-	-

Tablo 5'te yer alan uzman görüşlerine göre, 15 uzman M00FQ01 no'lu maddenin bir dil grubuna avantaj sağlamadığını belirtmiştir. M918Q01, M918Q02, M923Q01 ve M923Q03 no'lu maddeler için sırasıyla 14, 14, 18 ve 18 uzman maddenin bir dil grubuna avantaj sağladığını belirtmiştir. Maddelerin hangi dil grubuna avantaj sağladığı incelendiğinde ise bir görüşün dışındaki diğer tüm görüşlerin maddenin İngilizce formu alan grup lehine avantaj sağladığı belirlenmiştir. Tablo 6'da ise uzmanlar tarafından bir dil grubuna avantaj sağladığı belirtilen maddelerde olası yanlışlık kaynaklarına ilişkin uzman görüşlerinin dağılımı yer almaktadır.

Tablo 6. DMF'li maddelerde olası yanlışlık kaynaklarına ilişkin uzman görüşlerinin dağılımı

	Maddeler					Toplam
	M00FQ01	M918Q01	M918Q02	M923Q01	M923Q03	
Maddede yer alan sözcük ya da ifadelerin farklı anlamda kullanımı	3	7	7	6	6	29
Madde içeriğine bir dil grubunun aşına olması	1	3	3	11	11	29
Madde formatının veya biçimsel özelliğinin bir dil grubuna avantaj sağlaması	2	2	2	3	3	12
Madde ile ölçülen beceriler açısından kültürel farklılıkların olması	1	2	2	2	2	9
Çeviriye bağlı olarak maddelerin farklılaşması	3	10	10	10	10	43
Diğer	-	4	4	7	7	22

Tablo 6 incelendiğinde uzman görüşleri doğrultusunda, maddelerin farklı dil gruplarına göre en çok çeviriye (43) bağlı olarak yanlışlık gösterdiği belirlenmiştir. Bununla birlikte maddede yer alan sözcük ya da ifadelerin farklı anlamda kullanımı (29), madde içeriğine bir dil grubunun

aşına olması (29), madde formatının veya biçimsel özelliğinin bir dil grubuna avantaj sağlaması (12), madde ile ölçülen beceriler açısından kültürel farklılıkların olması (9) ve diğerleri (22) yanlışlık kaynağı olarak ifade edilmiştir. PISA matematik okuryazarlığı maddeleri üzerinde yapılan yanlışlık çalışmaları incelendiğinde de (Çet, 2006; Yıldırım ve Berberoğlu, 2009; Yıldırım, 2006) benzer sonuçlar elde edildiği görülmektedir.

Fen Okuryazarlığı Alt Testinin Dil Değişkenine Göre Ölçme Değişmezliğine İlişkin Bulgular

PISA-2012 fen okuryazarlığı alt testinin farklı dil formlarını alan gruplarda aynı faktör yapısına sahip olup olmadığı ÇGDFA ile test edilmiştir. Analiz öncesinde üç faktörlü yapının her bir dil grubu verisi ile uyumluluğu DFA ile incelenmiştir. Analiz sonuçları Tablo 7’de yer almaktadır.

Tablo 7. Fen okuryazarlığı testine ilişkin kurulan modelin uyum indeksleri

Dil Grubu	İstatistikler								
	X^2	sd	X^2/sd	RMSEA	CFI	GFI	SRMR	AGFI	NNFI
İngilizce	79,17	87	0,91	0,000	1,00	0,99	0,021	0,99	1,00
Fransızca	92,00	87	1,06	0,013	1,00	0,96	0,042	0,95	0,99
Çince	119,64	87	1,38	0,029	0,96	0,96	0,041	0,95	0,95
Türkçe	80,12	87	0,92	0,000	1,00	0,97	0,039	0,96	1,00

Tablo 7’de yer alan uyum iyiliği indeksleri incelendiğinde indekslerin kabul edilebilir aralıklarda yer aldığı görülmektedir. Fen okuryazarlığı alt testinin dil değişkeni bakımından ölçme değişmezliği gösterip göstermediğini belirlemek için yapılan ÇGDFA sonuçları ise Tablo 8’de yer almaktadır.

Tablo 8. Fen Okuryazarlığı Alt Testine İlişkin ÇGDFA Sonuçları

Model	X^2	Sd	X^2/sd	GFI	CFI	NNFI	SRMR	RMSEA	ΔX^2	Δsd	ΔCFI
A	480,75	357	1,35	0,96	0,97	0,96	0,047	0,02	-	-	
B	679,20	402	1,69	0,94	0,93	0,92	0,085	0,04	198,45	45	0,04

ÇGDFA ile ölçme değişmezliği analizlerinin ilk adımı olan yapısal değişmezliğin incelenmesi kapsamında “Model A”ya ilişkin uyum iyiliği indekslerinin ($X^2/sd < 2$, $GFI > 0.90$, $CFI > 0.95$, $NNFI > 0.95$, $SRMR \leq 0.05$ ve $RMSEA < 0.05$) kabul edilebilir aralıklarda yer aldığı görülmektedir. Bu durum, fen okuryazarlığına ilişkin kurulan üç faktörlü yapının testin farklı dil formlarını alan gruplarda yapısal değişmezlik gösterdiği şeklinde yorumlanabilir. Yapısal değişmezlik sağlandığı için metrik değişmezlik test edilmiştir. Bu amaçla gruplar arasında faktör yüklerinin sınırlandırıldığı “Model B”ye ilişkin uyum iyiliği indeksleri ile Model A’ya ilişkin indeksler karşılaştırılmıştır. Model uyumundaki değişim, X^2 ve CFI indekslerindeki fark incelenerek değerlendirilmiştir. Tablo 8’e göre $\Delta X^2=198,45$ olarak hesaplanmış, bu değer Δsd ’ye göre belirlenen kritik değer $X^2_{(45, 0.05)} = 66,63$ değerinden büyüktür. Buna göre gruplar arasında faktör yükleri sabitlendiğinde, model uyumu manidar şekilde farklılaşmaktadır. CFI indeksindeki değişim incelendiğinde ise (Tablo 8) ΔCFI değeri 0,04 olarak belirlenmiştir ve bu değer metrik değişmezlik için gerekli koşul olan $\pm 0,01$ aralığında yer almamaktadır. Hem X^2 hem de CFI değerlerine göre metrik değişmezlik değerlendirildiğinde PISA 2012 fen okuryazarlığı alt testi dil değişkeni bakımından metrik değişmezlik göstermemektedir.

Fen okuryazarlığı alt testinde yer alan maddelerin de DMF gösterme durumları GADM, SIBTEST ve MTK-OO istatistiksel teknikleri ile incelenmiştir. Analizler sonucunda her üç tekniğe göre ve en az B düzeyinde DMF gösteren maddeler DMF’li olarak kabul edilmiştir. Tablo 9’da her bir ikili grup için DMF’li kabul edilen fen okuryazarlığı maddelerine yer verilmiştir.

Tablo 9. Fen okuryazarlığı test maddelerine ilişkin DMF analiz sonuçları

Madde No	DMF Gösteren Maddeler ve Avantajlı Dil Grubu					
	İngilizce-Fransızca	İngilizce-Çince	İngilizce-Türkçe	Fransızca-Çince	Fransızca-Türkçe	Çince-Türkçe
1			İngilizce			
2		Çince				Çince
3		Çince				
4			İngilizce			
5		İngilizce				Türkçe
6	İngilizce	İngilizce	Türkçe		Fransızca	
7	Fransızca	Çince			Fransızca	Çince
8			İngilizce			
9	İngilizce	İngilizce	Türkçe			
10	Fransızca	Çince	İngilizce		Türkçe	
11						
12	Fransızca		İngilizce			
13		İngilizce				
14	İngilizce	İngilizce				
15		İngilizce				Türkçe
Toplam	6 (%40)	10 (%67)	7(%47)	0(%0)	3 (% 20)	4(%40)

Tablo 9 incelendiğinde DMF'li kabul edilen madde, testin %67'si ile, en çok İngilizce-Çince formu alan gruplar arasında belirlenmiştir. Fransızca-Çince formu alan gruplara göre hiçbir maddenin DMF'li olarak kabul edilmemesi ise dikkat çekmektedir. PISA 2012 fen okuryazarlığı alt testinde yer alan maddelerin tümünün farklı dil formlarını alan gruplara göre ölçme değişmezliği göstermediği yapılan DMF analizleri sonucunda belirlenmiştir.

Okuma Becerileri Alt Testinin Dil Değişkenine Göre Ölçme Değişmezliğine İlişkin Bulgular

PISA-2012 okuma becerileri alt testinin farklı dil formlarını alan gruplarda aynı faktör yapısına sahip olup olmadığı ÇGDFA ile test edilmiştir. analiz öncesinde üç faktörlü yapının her bir dil grubu verisi ile uyumluluğu DFA ile incelenmiştir. Analiz sonuçları Tablo 10'da yer almaktadır.

Tablo 10. Okuma Becerileri Alt Testine İlişkin ÇGDFA Sonuçları

Model	X ²	Sd	X ² /sd	GFI	CFI	NNFI	SRMR	RMSEA	ΔX ²	Δsd
A	524,88	296	1,77	0,96	0,94	0,92	0,045	0,038	-	-
B	764,65	338	2,26	0,93	0,88	0,88	0,077	0,049	239,77	42

Tablo10'a göre testin farklı dil formlarını alan gruplar arasında faktör yapısının sınırlandırıldığı Model A'ya ilişkin uyum iyiliği indeksleri incelendiğinde, modelin iyi uyum gösterdiği görülmektedir. Bu durum okuma becerileri alt testinin farklı dil gruplarında yapısal değişmezliği sağladığı şeklinde yorumlanmaktadır. Ölçme değişmezliği analizleri hiyerarşik bir yapı gösterdiğinden yapısal değişmezlik sağlandığı için metrik değişmezlik test edilmiştir. Faktör yüklerinin gruplar arasında sınırlandırıldığı "Model B" okuma becerileri alt testi için de kurulmuştur. Model A ile Model B arasında model uyumundaki değişim incelenmiştir. Tablo 10'a göre ΔX²=198,45 olarak hesaplanmış, bu değer Δsd dikkate alınarak belirlenen ve kritik değer olan

$X^2_{(45, 0.05)} = 66,63$ değerinden büyüktür. Buna göre faktör yükleri sabitlenmesi ile model uyumu manidar bir şekilde farklılaşmaktadır. Tablo 10 incelendiğinde Model B ve Model A arasındaki CFI farkının 0,06 olarak hesaplandığı görülmekte ve $\pm 0,01$ aralığında bulunma kriterini karşılamamaktadır. Her iki kritere göre de okuma becerileri alt testi farklı dil gruplarında metrik değişmezlik göstermemektedir.

Daha önce de belirtildiği gibi ÇGDFA ile değişmezlik incelemesi hiyerarşik bir yapı gösterdiğinden, metrik değişmezliğin sağlanmaması ile birlikte analize son verilmiş, güçlü ve katı değişmezliğe ilişkin analizler yapılmamıştır.

Metrik değişmezliğin sağlanmadığı durumda madde yanlılığı şüphesi ile birlikte DMF belirleme analizlerinin de gerekliliği ortaya çıkmıştır. Okuma becerileri alt test maddelerinin dil değişkeni bakımından DMF gösterme durumları GADM, SIBTEST ve MTK-00 teknikleri ile incelenmiştir. İlgili analizlerin sonuçlarına göre Tablo 11'de her bir ikili grup için DMF'li kabul edilen okuma becerileri maddelerine yer verilmiştir

Tablo 11. Okuma Becerileri Test Maddelerine İlişkin DMF Analiz Sonuçları

Madde No	DMF Gösteren Maddeler ve Avantajlı Dil Grubu					
	İngilizce-Fransızca	İngilizce-Çince	İngilizce-Türkçe	Fransızca-Çince	Fransızca-Türkçe	Çince-Türkçe
1	Fransızca	-	-	Fransızca	-	-
2	-	-	-	-	-	-
3	Fransızca	Çince	Türkçe	-	Türkçe	Türkçe
4	İngilizce	-	-	-	-	-
5	-	İngilizce	İngilizce	Fransızca	-	-
6	-	-	-	-	-	-
7	-	-	-	-	-	-
8	-	-	-	-	-	-
9	İngilizce	-	-	-	-	Çince
10	İngilizce	İngilizce	İngilizce	-	-	-
11	-	Çince	-	Çince	-	-
12	-	Çince	-	-	-	-
13	-	-	-	-	-	-
14	-	-	-	-	-	-
Toplam	5 (%36)	5 (%36)	3 (%21)	3 (%21)	1(%7)	2 (%14)

Tablo 11 incelendiğinde okuma becerileri testinde yer alan 14 maddeden dil değişkenine göre DMF'li kabul edilen madde en çok İngilizce-Çince (5 madde) ve İngilizce-Fransızca(5 madde) formu alan gruplar arasında belirlenmiştir. Fransızca-Çince formu alan gruplarda ise diğer gruplara göre en az sayıda (1 madde) DMF'li kabul edilen madde belirlenmiştir. Dil değişkenine göre ölçme değişmezliği gösteren maddeler ise 2, 6, 7, 8, 13, 14 no'lu maddelerdir. Testin %43'ü ölçme değişmezliği göstermektedir. Okuma becerileri alt testinde diğer alt testlere (matematik ve fen okuryazarlığı) göre çok daha fazla ölçme değişmezliği gösteren madde bulunmaktadır. Ancak testin tamamı için ölçme değişmezliği gösterdiği söylenemez. Her üç alan için elde edilen bulgular genel olarak Tablo 12'de değerlendirilmiştir.

Tablo 12.PISA-2012 Okuryazarlık testlerinin ölçme değişmezliğine ilişkin genel değerlendirme

	Matematik Okuryazarlığı	Fen Okuryazarlığı	Okuma Becerileri
Faktör yapılarının değişmezliği	Yapısal	Yapısal	Yapısal
Ortalama DMF'li madde	%35	%34	%22
Sayıcı en fazla DMF'li madde	İngilizce-Çince	İngilizce-Çince	İngilizce-Çince ve İngilizce-Fransızca
Sayıcı en az DMF'li madde	İngilizce-Fransızca	Fransızca-Çince	Fransızca-Türkçe
Tüm karşılaştırmalarda ortak olan DMF'li madde	-	-	-
Ölçme değişmezliği gösteren madde yüzdesi	%24	%7	%50
DMF'li maddelerin en fazla avantaj sağladığı grup	Çince	İngilizce	Çince
DMF'li maddelerin en az avantaj sağladığı grup	Fransızca	Fransızca	Fransızca

Tablo 12'de de görüldüğü gibi PISA 2012 okuryazarlık testleri genel olarak değerlendirildiğinde; tüm alt testlerin faktör yapısının dil değişkenine göre sadece yapısal değişmezlik gösterdiği belirlenmiştir. Tüm alt testlerde dil değişkenine göre farklı sayılarda değişmezlik gösteren ve göstermeyen maddeler olduğu belirlenmiştir. Okuma becerileri testinde yer alan maddeler diğer test maddelerine göre dil değişkeni açısından çok daha fazla ölçme değişmezliği göstermektedir. Ölçme değişmezliği göstermeyen madde sayısı ise yine okuma becerileri alt testinde en az sayıdadır. Fen okuryazarlığı testinde ise ölçme değişmezliği gösteren madde sayısı tek madde ile en azdır.

TARTIŞMA ve SONUÇ

Bu bölümde, araştırma bulgularından çıkarılan sonuçlar araştırma amaçları doğrultusunda ele alınmıştır. Bununla birlikte sonuçlara dayalı olarak önerilere yer verilmiştir.

PISA 2012 matematik okuryazarlığı alt testine ilişkin kurulan üç faktörlü model farklı dil grupları arasında yapısal değişmezlik gösterirken, metrik değişmezlik göstermemektedir. Farklı dil gruplarının ikili karşılaştırmalarında maddelerin ortalama %35'inin DMF'li olduğu kabul edilmiştir. Sayıcı en fazla DMF'li madde İngilizce ve Çince formu alan gruplar arasında bulunmaktadır. Bu durum çeviriye bağlı ve kültürel farklılıklar ile ilişkilendirilebilir. Sayıcı en az DMF'li madde ise İngilizce ve Fransızca formu alan gruplar arasında belirlenmiştir. İngilizce ve Fransızca dillerinin PISA uygulaması kaynak dilleri olması nedeniyle bu grupların çeviri formu almamaları bu gruplar arasında DMF'li maddenin en az sayıda olmasının bir nedeni olarak düşünülebilir. DMF'li olarak belirlenen maddeler gruplar arasında benzerlik gösterse de bütün karşılaştırmalarda ortak olarak DMF gösteren madde yoktur. İkili karşılaştırmaların hiçbirinde DMF'li maddelerin belli bir gruba çok daha fazla avantaj sağladığı görülmemiştir. Bununla birlikte DMF'li maddeler, sayıcı en fazla olandan en aza doğru sırasıyla Çince, Türkçe, İngilizce ve Fransızca formu alan gruplar lehine DMF göstermektedir. Matematik okuryazarlığı testi maddelerinin %24'ü dil değişkenine göre ölçme değişmezliği göstermektedir. Maddelerin İngilizce (Avustralya) ve Türkçe (Türkiye) formunu alan gruplara göre DMF gösteren maddeler üzerinde olası yanlışlık kaynaklarını belirlemek üzere yapılan yanlışlık çalışmasında uzmanlar, maddelerde yer alan sözcük ya da ifadelerin farklı anlamda kullanımını, madde içeriğine bir dil grubunun aşına olmasını ve çeviriye bağlı olarak maddelerin farklılaşmasını olası yanlışlık kaynağı olarak belirtmişlerdir. Elde edilen bu sonuçlar PISA uygulaması matematik okuryazarlığı maddelerinin farklı gruplarda ölçme değişmezliğine ilişkin yapılan çalışmalardan (Çet, 2006; Kankaras ve Moors, 2014; Yıldırım ve Berberoğlu, 2009; Yıldırım, 2006) elde bulgularla benzerlik göstermektedir.

Fen okuryazarlığına ilişkin kurulan üç faktörlü modelin farklı dil grupları arasında yapısal değişmezlik gösterirken, metrik değişmezliği göstermediği belirlenmiştir. On beş fen okuryazarlığı maddesinin ortalama %34'ü DMF'li olarak belirlenmiştir. DMF'li madde sayısı en

fazla İngilizce ve Çince formları alan gruplar arasında, en az Fransızca ve Çince formu alan gruplar arasında belirlenmiştir. Dil değişkenine göre DMF'li kabul edilen fen okuryazarlığı maddeleri, ikili karşılaştırmalarda belli bir gruba belirgin bir şekilde çok daha fazla avantaj sağlamamaktadır. İngilizce formu alan öğrenciler lehine en fazla, Türkçe ve Çince formu alan öğrenciler lehine eşit, Fransızca formu alan öğrenciler lehine sayıca en az madde DMF'li olarak kabul edilmiştir. Fen okuryazarlığı maddelerinin %7'si dil değişkeni açısından ölçme değişmezliği göstermektedir.

Bu sonuçlar, farklı yıllarda uygulanan PISA fen okuryazarlığı maddelerinin farklı dil gruplarına göre ölçme değişmezliğinin incelendiği araştırmalardan (Grisay ve diğerleri, 2007; Kankaras ve Moors, 2014; Ulutaş, 2012) elde edilen bulgularla benzerlik göstermektedir. Le (2006) yaptığı çalışmada PISA 2006 fen okuryazarlığı maddelerinin ölçme değişmezliğini farklı değişkenlere göre incelemiş en fazla DMF'li madde dil değişkenine göre belirlenmiştir. Bu durum, bu araştırmada elde edilen bulguları da desteklemektedir. Bu çalışma kapsamında fen okuryazarlığı maddelerinin farklı dil formlarını alan öğrencilere göre olası DMF gösterme nedenleri, ilgili maddeler yayınlanmadığından, tespit edilememiştir. Ancak ilgili araştırmalar incelendiğinde Le (2009)'ye göre çeviri problemleri ile birlikte kültürel ve eğitim programlarındaki farklılıklar, Ulutaş (2012)'a göre maddelerin içerik, beceri ve düşünme süreçleri ile birlikte madde formatlarına aşına olma durumları olası DMF nedenleri olarak belirlenmiştir.

Okuma becerileri testi farklı dil grupları arasında yapısal değişmezlik gösterirken, metrik değişmezlik göstermemektedir. On dört maddelik okuma becerileri alt testine ilişkin yapılan DMF analizleri sonucunda testin ortalama %22'si DMF'li olarak belirlenmiştir. Sayıca en fazla DMF gösteren madde İngilizce-Çince ve İngilizce-Fransızca formlarını alan gruplar arasında iken en az Fransızca ve Türkçe formları alan gruplar arasında belirlenmiştir. DMF'li olarak kabul edilen maddelerin en fazla Çince formu alan grup lehine DMF gösterdiği belirlenmiştir. Bununla birlikte DMF'li maddeler, birbirine yakın değerler olmakla birlikte sırasıyla sayıca en fazladan en aza doğru İngilizce, Türkçe ve Fransızca formları alan öğrenciler lehine DMF göstermektedir. Bütün gruplarda ortak olarak DMF'li kabul edilen madde bulunmamaktadır. Okuma becerileri maddelerinin %50'si dil değişkenine göre ölçme değişmezliği göstermektedir.

Elde edilen bu bulgular PISA okuma becerileri maddelerinin farklı dil formları üzerinde yürütülen ölçme değişmezliği çalışmalarından (Kankaras ve Moors; 2014; Özmen, 2014) elde edilen bulgular ile benzerlik göstermektedir. Grisay ve Monseur (2007) kültürel farklılıklar minimize edilse, çeviri ne kadar dikkatle yürütülse de uyarlanan bir testin ölçme değişmezliği göstermemesi problemi ile her zaman karşılaşılacağını belirtmiştir. Bu çalışma kapsamında okuma becerileri maddelerinin farklı dil formlarında olası DMF nedenlerine ilişkin bir çalışma, ilgili maddeler yayınlanmadığından, gerçekleştirilememiştir. Ancak PISA okuma becerilerine ilişkin Grisay ve Monseur (2007) tarafından yapılan yanlılık çalışmasına göre madde formatı ve çeviri problemleri olası yanlılık kaynağı olarak belirtilmektedir.

Öneriler

Uygulayıcılar noktasında araştırma sonuçlarına dayalı olarak PISA gibi kapsamlı bir eğitim araştırmasında asıl uygulama öncesi yapılan pilot uygulama verileri üzerinde DMF analizleri ile birlikte analiz sonuçlarına göre yanlılık kaynakları belirlenmeli, gerekli düzeltme ve düzenlemeler yapılmalıdır. Birçok ülkenin katıldığı bu türlü uygulamalarda yer alan maddelerin hazırlanma aşamasında içeriğe belli bir grubun aşına olma veya olmama sorununa mutlaka dikkat edilmelidir. Bununla birlikte birçok dile uyarlanan PISA uygulamasında yer alan maddelerin uyarlanma aşamasında bir diğer yanlılık kaynağı olan çeviri problemleri de oldukça önemli görünmektedir. Bu nedenle testlerin ulusal versiyonlarını hazırlayan kişiler titizlikle seçilmelidir. İlgili kişiler yanlılık kaynakları konusunda gerekli eğitimleri almalıdır.

Araştırma sonuçları doğrultusunda araştırmacılara yönelik öneriler şu şekilde özetlenebilir: Bu araştırma PISA 2012 okuryazarlık maddelerinin yer aldığı 3 no'lu kitapçıktan elde edilen veriler üzerinde gerçekleştirilmiştir. Yayınlanan maddelerin yer aldığı diğer kitapçıklar üzerinde de benzer ölçme değişmezliği analizleri yapılmalıdır. Bu araştırmada DMF belirleme analizlerinden GADM, SIBTEST ve MTK-OO teknikleri kullanılmıştır. Diğer DMF

belirleme teknikleri ile benzer çalışmalar yapılarak analiz sonuçlarının tutarlılığı karşılaştırılabilir. Bu çalışmada olası yanlışlık kaynakları matematik okuryazarlığı maddelerinin İngilizce ve Türkçe formlarını alan gruplarda dil değişkenine göre DMF'li olarak belirlenen ve yayınlanan maddeler üzerinde uzman görüşleri doğrultusunda yapılmıştır. Benzer çalışmalar diğer dil gruplarına göre belirlenen DMF'li maddeler üzerinde de yürütülmelidir.

KAYNAKÇA

- Abbott, M.L. (2007). A confirmatory approach to differential item functioning on an esl reading assessment. *Language Testing*, 24(1), 7-36 <http://ltj.sagepub.com/content/24/1/7> adresinden alınmıştır.
- Acar, T. (2008). *Maddenin farklı fonksiyonlaşmasını belirlemede kullanılan genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oranı tekniklerinin karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anderson, J. A. (2005). Accountability in Education. *Education Policy series Vol. 1*. Unesco International Academy of Education, International Institute for Educational Planing. <http://unesdoc.unesco.org/images/0014/001409/140986e.pdf> adresinden alınmıştır
- Atalay, K., Gök, B., Kelecioğlu, H. ve Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43: 270-281
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Breakspear, S. (2012), The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers*, 71, OECD Publishing. <http://dx.doi.org/10.1787/5k9fdqffr28-en> adresinden
- Bakan Kalaycıoğlu, D. ve Berberoğlu, G. (2010). Differential item functioning analysis of the science and mathematics items in the university entrance examinations in Turkey. *Journal of Psychoeducational Assessment*, 20, 1-12.
- Baş, H. (2005). *Hesap verme sorumluluğu ve kamu yönetimi ve kontrol kanunu, 20. Türkiye Maliye Sempozyumu Türkiye'de Yeniden Mali Yapılanma (23-27 Mayıs)*. Pamukkale Üniversitesi, İktisadi Ticari Bilimler Fakültesi'nde sunulan bildiri.
- Camilli, G. and Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Cheung, G. W. and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-cultural Psychology*, 31(2), 188-213.
- Çet, S. (2006). *A Multivariate analysis in detecting differentially functioning items through the use of programme for international student assessment (PISA) 2003 mathematics literacy items* (Yayımlanmamış doktora tezi). Ortadoğu Teknik Üniversitesi, Ankara
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, Mantel Haenszel, Lojistik Regresyon ve Madde Tepki Kuramı yöntemleriyle incelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Demirtaşlı, R. N. (2014). Öğrenme, öğretim ve değerlendirme arasındaki ilişkiler. N. Demirtaşlı (Ed), *Eğitimde Ölçme ve Değerlendirme İçinde (2. Baskı) (3-29)*. Edge Akademi: Ankara
- Drasgow, F. and Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662-680.
- Embretson, S.E. and Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Figazzolo, L. (2008). Impact of PISA 2006 on the Education Policy Debate Education International.???<http://download.eiie.org/docs/IRISDocuments/Research%20Website%20Documents/2009-00036-01-E.pdf> adresinden alınmıştır.
- Gierl, M.J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education*, 25(4), 280-296.
- Gierl, M. H., Khaliq, S. N. and Boughton, K. (1999). *Gender differential item functioning in mathematics and science: prevalence and policy implications*. Paper presented at the Annual Meeting of the Canadian Society for the Study of Education, Canada
- Gök, B., Kelecioğlu, H. ve Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35, 3-16.

- Grisay, A. and Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33, 69-86. <http://www.sciencedirect.com/science/article/pii/S0191491X07000077?#> adresinden alınmıştır.
- Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44, 182-188.
- Hambleton, R.K., and Swaminathan, H. (1989). *Item response teory: principles and applications*. USA: Kluwer Nijhoff Publishing.
- Hofmann, D. A., Griffin, M. A., & Gavin, M. A. (2000). The application of hierarchical linear modeling to organizational research. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods In Organizations: Foundations, Extensions And New Directions*. San Francisco, CA: Jossey-Bass Inc.
- Hopkins, D., Pennock, D., Ritzen, J., Ahtaridou, E. and Zimmer, K. (2008). External evaluation of the policy impact of PISA. OECD doc. EDU/PISA/GB(2008)35/REV1. Paris:OECD.
- Hu, L. ve Bentler, P. (1995). Evaluating Model Fit. R. Hoyle (Ed). *Structural equation modeling* (pp. 76-99). *Concept, Issues and Application*. Thousand Oaks: SagePublications.
- International Test Commission (2005). *International test commission guidelines for test adaptation*. London: Author.
- Johnson, T. P. (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. ZUMA-Nachrichten Spezial, 1-40. <http://www.ssoar.info/ssoar/handle/document/49730> adresinden alınmıştır.
- Kamata, A. (2001). Item Analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38(1), 79-93.
- Kamata, A., Chaimongkol, S., Genç, E. and Bilir, K. (2005). *Random-effect differential item functioning across group unites by the hierarchical generalized linear model*. This paper was presented at the annual meeting of the American Educational Research Association, Montreal, Canada <http://garnet.acns.fsu.edu/~akamata/papers/AERA%202005.pdf> adresinden alınmıştır.
- Kankaras, M. and Moors, G. (2014). Analysis of cross-cultural comparability of PISA 2009 scores. *Journal of Cross-Cultural Psychology*, 45(3) 381-399. <http://jcc.sagepub.com/content/45/3/381.full.pdf+html> adresinden alınmıştır.
- Karasar, N (2011). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım
- Kim, W. (2003). *Development of a differential item functioning (dif) procedure using the hierarchical generalized linear model: A comparison study with logistic regression procedure* (doctoral dissertation). The Pennsylvania State University, Pennsylvania.
- Le, L.T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing*, 9(2), 122-133.
- Luppescu, S. (2002). *DIF detection in HLM*. Paper Presented at The Annual Meeting of The American Educational Research Association, New Orleans.
- OECD (2014). *PISA 2012 technical report*. Paris: OECD Publications. <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf> adresinden alınmıştır.
- Osterlind, S., J. and Everson, H., T. (2009). *Differential item functioning*. (2nd). Thousand Oaks. CA: SAGE Publications, Inc. doi: <http://dx.doi.org/10.4135/9781412993913>
- Pan, T. (2008). *Using the multivariate multilevel logistic regression model to detect dif: A comparison with Hglm and Logistic Regression dif detection methods* (Doctoral dissertation) Michigan State University, Michigan.
- Prelow, H. M., Tein, J.Y., Roosa, M. W. & Wood, J. (2000). Do Coping Styles Differ Across Sociocultural Groups? The Role of Measurement Equivalence in Making This Judgment. *American Journal of Community Psychology*, 28(2), 225-244. Web: <http://www.springerlink.com/media> adresinden alınmıştır.
- Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage.
- Raju, N. S., Laffitte, L. J. and Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 527-529.
- Reise, S.P., Widaman, K.F., and Pugh, R.H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement equivalence. *Psychological Bulletin*, 114, 552-566.
- Rensvold, R. B., & Cheung, G. W. (1998). Testing Measurement Models for Factorial Invariance: A Systematic Approach. *Educational and Psychological Measurement*, 58, 1017-1034.

- Sireci, S. G. and 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.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayınları
- Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları*. Ankara: Ekinoks
- Stark, S., Chernyshenko, O. S. and 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. and Fidell, L. S. (2007). *Using multivariate statistics*. Fifth Edition. Pearson: AB
- Ulutaş, S. (2015). A study on detecting of differential item functioning of PISA 2006 science literacy items in Turkish and American samples. *Eurasian Journal of Educational Research*, 58, 41-60
- Vandenberg, R. J. and Lance, C. E. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Yıldırım, H. H. (2006). *The differential İtem functioning (dif) analysis of mathematics İtems in the İnternational assessment programs*. (Doctoral dissertation). Middle East Technical University, Ankara
- Yıldırım, H., H. and Berberoğlu, G. (2009). Judgmental and statistical dif analyses of the PISA-2003 mathematics literacy items. *International Journal of Testing*, 9, 108-121
- Wendorf, C. A. (2002). Comparisons of structural equation modeling and hierarchical linear modeling approaches to couples' data. *Structural Equation Modeling*, 9, 126 - 140.
- Wu, A. D., Li, Z. and Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: a demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, 12, 1-26.