ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
2016; 7(1);182-201

Ortak Maddelerin Değişen Madde Fonksiyonu Gösterip Göstermemesi Durumunda Test Eşitlemeye Etkisinin Farklı Yöntemlerle İncelenmesi*

The Study of the Effect of Anchor Items Showing or Not Showing Differantial Item Functioning to Test Equating Using Various Methods

Kadriye Belgin DEMİRUS** Selahattin GELBAL***

Öz
Bu araştırmada ortak maddelerin tamamı cinsiyete göre TB-DMF'li/DMF'siz olduğunda Madde Tepki Kuramı'na dayalı yapılan eşitleme yöntemlerinin performansını karşılaştırmak amaçlanmıştır. Araştırmada DMF'li maddelerin test eşitlemeye etkisi gerçek veri üzerinden, ayrı kalibrasyon yöntemleri ve eşdeğer gruplarda ortak test deseni kullanılarak yatay eşitleme ile ortaya konulmuştur. Araştırmada DMF analizleri "Mantel-Haenszel" yöntemi için EASYDIF programında ve "lojistik regresyon" yöntemi için Zumbo tarafından hazırlanan syntax ile SPSS'de yapılmıştır. Test eşitleme yöntemleri olarak lineer ölçek dönüştürme (moment) yöntemlerinden "ortalama-ortalama" ile "ortalama-sigma" ve karakteristik eğrisi dönüştürme yöntemlerinden "Haebara" ile "Stocking-Lord" kullanılmıştır. Eşitleme yöntemlerinin performansı yetenek kestirimleri arası farka dayalı RMSD eşitleme hataları hesaplanarak değerlendirilmiştir. Madde parametrelerinin ve yeteneğin kestiriminde BILOG-MG, test eşitlemede IRTEQ yazılımı işe koşulmuştur. Çalışmanın verisi oluşturulan fen testi formlarının 1350 8.sınıf öğrencisine uygulamasından elde edilmiştir. Araştırmanın sonucunda ortak maddeler erkekler lehine TB-DMF'li olduğunda en büyük RMSD eşitleme hatasını ortalama-ortalama yöntemi, en küçük hatayı ise ortalama-sigma yöntemi üretmiştir. Ortak maddeler DMF'siz olduğunda ise en büyük hata ortalama-sigma yönteminde, en küçük RMSD eşitleme hatası karakteristik eğrisi yöntemlerinde (Stocking-Lord ve Haebara) birbirine eşit olarak elde edilmiştir.

Anahtar sözcükler: Test eşitleme, değişen madde fonksiyonu, ortak test deseni, eşitleme hatası

Abstract
The purpose of this study to compare the results of equating methods based on Item Response Theory when all the anchor items showing or not showing gender based uniform DIF. In the study the effect of DIF items on test equating presented on real data with horizontal equating using seperate calibration methods and equivalent groups with anchor test design. DIF analysis conducted on EASDIF software for "Mantel-Haenszel" method and on SPSS with syntax presented by Zumbo for "logistic regression" method. For test equating methods "mean- mean", "mean-sigma", "Haebara" and "Stocking-Lord" were used. The performances of the equating methods were evaluated through RMSD equating errors based on difference of ability estimates. BILOG-MG was utilized for the prediction of item parameters and ability, IRTEQ software was utilized for test equating. Data set for the study was obtained from the forms of science test which applicated to 1350 students in 8th grade. According to the results of the study when the anchor items with uniform DIF favored males were used for equating, mean-mean method produced the biggest equating error whereas mean-sigma method produced the smallest. When the anchor items with no-DIF were used for equating the biggest equating error was obtained from mean-sigma method and smallest equating error was obtained from Stocking-Lord ve Haebara methods in equal to each other.

Keywords: Test equating, differential item functioning, anchor test design, equating error

* Bu araştırma "Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi" adlı doktora tezinden üretilmiştir.
** Dr., MEB Öğretmen, Ankara-Türkiye, e-posta: belgindemirus@gmail.com
*** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: sgelbal@gmail.com

Geliş Tarihi: 12.04.2016
DOI: 10.21031/epod.56218 *Kabul Tarihi: 04.06.2016*

GİRİŞ

Ülkemizde ulusal ve uluslararası düzeyde yapılan sınavlarda bireyler ve ülkeler hakkında önemli kararlar alınmaktadır. Bu nedenle test puanlarına dayalı yapılan yorumların gerçeği yansıtması için geçerli ölçmelere ihtiyaç duyulmaktadır. Bir test bireylerde ölçülmesi amaçlanan özelliği başka özellikle karıştırmadan ne kadar doğru ölçebiliyorsa o kadar geçerlidir (Tekin,1991; McDonald, 1999). Cinsiyet, yerleşim yeri, okul türü, kültür, sosyo ekonomik düzey gibi demografik özellikler öğrenci başarısını etkileyen değişkenler olarak araştırmalara konu olmaktadır. Ancak bu değişkenlerde belli bir alt grupta (Ör: kız-erkek) bulunan eşit yetenek düzeyindeki bireylerin puanları arasında, ölçme amacı olmamasına rağmen, gruba bağlı farklılık söz konusuysa yapılan ölçmelerin geçerli olmadığı söylenebilir.

Bir testteki maddelerin aynı yetenek düzeyinde doğru cevaplanma olasılıkları alt gruplara göre farklılık gösteriyorsa, o maddenin değişen madde fonksiyonuna (DMF) sahip olduğu kabul edilmektedir (Camilli ve Shepard, 1994; Embretson ve Reise, 2000). Testlerin psikometrik özellikleri kapsamında ele alınan DMF incelemesi geçerlik için bir kanıttır (Embretson, 2007). Aynı yetenek düzeyinde olmalarına rağmen farklı gruplardaki bireylerin aynı test maddesine verdikleri cevapların farklı olmasının nedenlerini incelemek, hem test maddesinin özelliklerinin hem de farklı gruplarda bulunan bireylerin bilişsel süreçlerinin, test alma stratejilerinin ve bilgi eksikliklerinin anlaşılabilmesini sağlayacaktır. Böylece, test puanlarını etkileyen ancak test ile ölçülmesi amaçlanmayan bozucu değişkenlerin varlığı tespit edilebilecektir. DMF'nin belirlenmesi, düzeltici çalışmaların yapılması, testin yanlılık içermediğinin ortaya konulması testin yapı geçerliğini sağlamada ve test alma koşullarının eşdeğerliğinden emin olmada önemli görülen süreçlerdir (Camilli ve Shepard, 1994).

Literatür incelendiğinde özellikle fen ve matematik gibi sayısal alan maddelerinde cinsiyet değişkenine göre DMF gözlenmektedir. Araştırmalarda erkek ve kızların farklı ilgi alanları ve yeteneklerine sahip olmalarının maddelerin farklı fonksiyonlaşmasına neden olduğu belirtilmektedir. Mantıksal analiz sonucu DMF alt gruplar arası gerçek yetenek farklılığından kaynaklanmadığında madde gruplardan birine avantaj sağlamakta ve yanlı kabul edilmektedir. Bakan Kalaycıoğlu (2008) yaptığı araştırmada ÖSS fen bilimleri testinde erkekler lehine işleyen üç DMF'li fizik maddesinden bir tanesinin yanlılık gösterdiğini saptamıştır. Yanlılık nedeni olarak fizik maddesinde kullanılan otomobillere ve hız konusuna erkek öğrencilerin daha yakın olması gösterilmiştir. Testin amacına uygun olmayan şekilde test maddesinin bazı karakteristik özelliklerinden dolayı erkeklerin kızlara göre maddeyi daha çok doğru cevaplaması madde yanlılığını ortaya çıkarmıştır (Zumbo, 1999). Alt gruplar arası gerçek yetenek farklılığından kaynaklanan değişen madde fonksiyonu yanlılığa neden olmamaktadır. Bir başka ifadeyle maddenin amaca hizmet ettiğini, kız ve erkek öğrenciler için aynı yeteneği ölçtüğünü göstermektedir. Çepni (2011) araştırmasında cebirsel ifadelerle soyut bir biçimde sunulmuş, algoritmik işlemlerle çözülen maddelerin kız öğrenciler lehine; gerçek hayat problemi olarak ifade edilmiş, algoritmik rutin işlemlerle çözülemeyen problemlerin ise erkekler lehine işlediğini tespit etmiştir. DMF gösteren maddelerin yanlı olup olmadıkları uzman kanıları yardımıyla incelendiğinde, DMF sebebi olabilecek gerekçelerin tümü testin ölçmesi amaçlanan matematiksel/sayısal yetenek yapısı içinde kaldığı görülmüş ve yanlı olmadıkları belirtilmiştir. Literatürde erkek öğrencilerin matematiği, kız öğrencilere kıyasla, hayatlarında daha çok kullanılabilir ve daha değerli bir kavram olarak algıladıklarının rapor edildiği, bu nedenle erkeklerin gerçek hayattan alınmış uygulamalı problemlerde denk yetenek düzeylerindeki kız öğrencilere göre daha başarılı olmalarının beklendiği belirtilmiştir

Bireylerin ya da ülkelerin geleceği hakkında önemli kararlar alınan sınavlarda testlerin geçerliğini artırmak için her sınavda DMF analizlerinin yapılarak sonraki sınavlarda yanlı çıkma olasılığı bulunan maddelerin elimine edilmesi gerekmektedir. Yanlılık aynı testin farklı formları arasında güçlük farkı olduğunda da ortaya çıkabilmektedir. Ortaöğretime, lisans ve lisansüstü yükseköğretime giriş, yabancı dil yeterlik belirleme gibi seçme ve/veya yerleştirme gerektiren sınavlar (TEOG, YGS, LYS, YDS, ALES, KPSS) yılda bir kez ya da birden fazla uygulanmaktadır. Bununla birlikte ülkeler arası karşılaştırmaların yapıldığı uluslararası geniş ölçekli testlerde (PISA, TIMMS, PIRLS) farklı

sorular içeren çok sayıda kitapçık bulunmaktadır. Aynı teste ait farklı formların kullanılması durumunda sınava başvuranların kolay/zor formdan hangisini aldığı dikkat edilmesi gereken önemli bir konudur. Çünkü zor formun sorularını çözen bireyler kolay formu çözenlere göre daha düşük puan alabilirler. Böyle bir durumda formları ortak bir ölçeğe yerleştirmek ve karşılaştırılmalarını sağlamak için test eşitleme yapılır. Böylece zor testi alan bireylere karşı yapılan haksızlık önlenerek, test formlarından kaynaklanan yanlılık problemi ortadan kaldırılmaktadır (Angoff, 1971; Hambleton, Swaminathan ve Rogers, 1991; Cook ve Eignor, 1991; Kolen ve Brennan, 2004; Holland ve Dorans, 2006).

Test maddeleri farklı şekilde çalıştığında gruplar arası puanların karşılaştırılması zorlaşmaktadır. Ortak test kullanılarak yapılacak bir eşitlemede ise, söz konusu ortak test DMF'li maddeler barındırıyorsa, bu maddeler ortak testin geçerliğini ve güvenirliğini düşürme eğilimindedir ve grup farklılıklarının doğru biçimde ortaya konmasında ciddi tehlike oluşturmaktadır. Aynı şekilde eşitleme katsayıları da DMF varlığında bozulabilmektedir. Dolayısıyla ortak test kullanıldığında, bu testin farklı test formlarında aynı şekilde işleyip işlemediğini belirlemek için DMF analizi yapılmalı ve test eşitleme katsayılarına yaptıkları etkinin en aza indirilmesi gerekmektedir (Cook ve Paterson, 1987; Kim ve Cohen, 1991; Hidalgo-Montesinos ve Lopez-Pina, 2002).

DMF'li maddelerin testten çıkarılması arzu edilmeyen bir durumdur; çünkü bu durum madde sayısının azalmasına ve kapsam geçerliğinin düşmesine neden olmaktadır. Chu (2002) ve Turhan (2006) araştırmalarında DMF'li maddelerin testten çıkarılmasının eşitleme hatasını arttırdığını; bu artışın DMF'li madde sayısı arttıkça büyüdüğünü belirtmişlerdir. Chu (2002) araştırmasında özellikle kısa bir testten DMF'li maddenin silinmesinin geçerliği daha fazla zedelediği sonucuna ulaşmıştır. Dolayısıyla böyle bir durumda DMF'li maddeleri çıkarmak yerine, DMF'nin farklı yöntemlerle elde edilen eşitleme hatalarına etkisi karşılaştırılarak düşük hata veren yöntem tercih edilebilir.

Yapılan literatür taramasında DMF'li maddeler varlığında test eşitleme konulu araştırmaların oldukça sınırlı sayıda olduğu gözlenmektedir (Chu, 2002; Turhan, 2006; Atalay Kabasakal, 2014; Huggins, 2014). Bununla birlikte madde parametre kayması (MPK, Item Parameter Drift= IPD) değişen madde fonksiyonunun bir çeşidi olarak ele alınmaktadır (Han, 2008). MPK, belirli bir maddenin birden fazla test uygulaması veya zaman boyunca parametrelerinin değişmesidir (Goldstein, 1983). MPK'lı maddelerin madde parametre kestirimlerine, eşitleme katsayılarına, yetenek kestirimlerine etkisini farklı eşitleme yöntemleriyle inceleyen araştırmalar bulunmaktadır (Wells, Subkoviak ve Serlin, 2002; Han, 2008; Babcock ve Albano, 2012). Bununla birlikte ülkemizde yapılan MTK'ya dayalı eşitleme çalışmalarının daha çok eşdeğer olmayan gruplarda ortak madde deseni ile simülasyon veri üzerinden gerçekleştirildiği görülmektedir (Kilmen, 2010; Gök, 2012; Uysal, 2014). Gerçek veri üzerinden yapılan eşitleme çalışmalarında MTK'ya dayalı Rasch modeli ve 2PLM kullanılarak sözde ortak test desenli (Kelecioğlu, 1994), Rasch modeli kullanılarak ortak test desenli (Şahhüseyinoğlu, 2005) ve dikey eşitleme ile (Çetin, 2009) çalışıldığı; KTK'ya dayalı daha çok tek grup desenli (Bozdağ 2007; Öztürk, 2010; Kan, 2010; 2011; Mutluer, 2013) çalışmalar yapıldığı belirlenmiştir.

Bu araştırmada DMF'li maddelerin test eşitlemeye etkisi, diğer araştırmalardan farklı olarak gerçek veri üzerinden, ayrı kalibrasyon yöntemleri ve eşdeğer gruplarda ortak test deseni kullanılarak, ortak maddelerin tamamı DMF'li veya DMF'siz olduğunda yatay eşitleme ile ortaya konulmaktadır. Bu nedenle araştırmanın test eşitleme alanında yapılan çalışmalara önemli bir katkı getireceği düşünülmektedir.

Araştırmanın Amacı

Bu araştırmanın amacı, ortak maddelerle yapılan eşitlemede DMF'nin bireylerin yetenek kestirimlerine etkisini dört eşitleme yönteminin performasını inceleyerek belirlemek ve değerlendirmektir. Bu amaç doğrultusunda ortak maddeler cinsiyete göre erkekler lehine TB-DMF'li veya DMF'siz olduğunda Madde Tepki Kuramı'na dayalı lineer ölçek dönüştürme (moment) yöntemleri "Ortalama-Ortalama (O-O)", "Ortalama-Sigma (O-S)" ve karakteristik eğrisi dönüştürme

yöntemleri "Haebara (HB)", "Stocking-Lord (S-L)" için RMSD eşitleme hataları karşılaştırılmıştır. "Ortak maddeler DMF'li ve DMF'siz olduğu durumlarda O-O, O-S, HB ve S-L eşitleme yöntemleri ile hesaplanan eşitleme hataları nasıldır?" araştırmanın sorusunu oluşturmaktadır.

YÖNTEM

Araştırmanın Türü

Bu araştırma ortak maddelerin DMF içerip içermediğinin araştırılması bakımından betimsel, Madde Tepki Kuramı'na (MTK) dayalı test eşitleme yöntemlerinde eşitleme hatalarının karşılaştırılması bakımından nedensel karşılaştırma türündedir. Nedensel karşılaştırma araştırmalarında, belli bir faktörün gözlenen sonuçlarda farklılık oluşturup oluşturmadığı incelenir; ancak incelenen faktörün sonuçlar üzerinde kesin olarak neden olduğu söylenemez. Bu araştırma olası nedenleri belirleme bakımından değerlidir (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2008).

Çalışma Grubu

Test eşitleme çalışması için evrenden örneklem seçme yoluna gidilmeyip, çalışma grubu üzerinde uygulama yapılmıştır. Araştırmanın çalışma grubunu 2012-2013 eğitim-öğretim yılında sekiz ilde (Ankara, Kütahya, Eskişehir, Denizli, Bursa, Kahramanmaraş, Malatya, İstanbul) 8. sınıfta öğrenim gören 1576 ulaşılabilen öğrenci oluşturmuştur. Ancak bazı okullardan gelen verinin ayrı incelenmesi sonucu, tek boyutluluk ve normallik sayıltılarını karşılamama, atlanan veya ulaşılamayan maddeler barındırma, aykırı standart değerlere sahip olma gibi nedenlerle bu örneklemler araştırmada kullanılmamıştır. Söz konusu durumlara sahip kişiler çıkarıldığında örneklem 1350 öğrenciden oluşmuştur.

Veri Toplama Araçları

Ortak maddelerin DMF'li veya DMF'siz olduğu durumlarda test eşitlemesi yapmak için kullanılacak veri toplama aracının (fen ve teknoloji testi A ve B formu) hazırlanması üç aşamada gerçekleştirilmiştir:

1) DMF'li veya DMF'siz ortak maddelerin ve DMF'siz ortak olmayan maddelerin seçileceği madde havuzunun oluşturulması,

2) Madde havuzundan gerçek uygulama yapılacak DMF'li ve DMF'siz ortak maddeli paralel test formlarının elde edilmesi,

3) Uygulama sonunda maddelerin DMF'li ve DMF'siz olma durumlarının incelenerek eşitleme için test formlarına son halinin verilmesi.

Öncelikle 2000-2012 yılları arasında uygulanan OKS ve SBS fen bilgisi alt testlerinde 2000-2007 yılları OKS'de çıkan 175 madde ve 2008-2012 yılarında SBS'de çıkan 100 madde olmak üzere toplam 275 maddenin cinsiyete dayalı DMF analizi yapılmıştır. Madde Tepki Kuramı'na dayalı yöntemlerden elde dilen DMF istatistikleri incelendiğinde testlerdeki hemen hemen tüm maddelerin DMF gösteren maddeler olarak işaretlendikleri görülmüştür (Çepni, 2011). Bu nedenle bu çalışmada, DMF analizleri için Klasik Test Teorisi'ne dayalı MH ve LR yöntemleri kullanılmıştır.

Bunun için her bir yıla ait evrenden tesadüfi örnekleme yoluyla seçilen N=4000 ile N=20000 arası örneklemlerden; N=750 ve N=1000 büyüklüğünde alt gruplar oluşturulmuştur. Kullanılan iki yönteme göre de B ve C düzeyinde erkekler lehine ve tek biçimli DMF'li çıkan maddeler dikkate alınmıştır. LR yönteminde sadece TB-DMF içeren maddeler için ikinci model ile birinci model arasındaki ikili puanlama için hesaplanan Nagelkerke R^2 değerleri farkı (ΔR^2)$\geq$ 0.010 olan maddeler ile MH yöntemine göre -1 $\leq$IΔMHI< -1.5 aralığında olan maddeler B düzeyinde DMF'li alınmıştır. Aynı şekilde LR yöntemine göre ikinci model ile birinci model arasındaki Nagelkerke R^2 değerleri

farkı (ΔR^2)≥ 0.020 olan maddeler C düzeyinde TB-DMF'li ve MH yöntemine göre -1.5 ≤IΔMHI olan maddeler B düzeyinde referans grup olan erkekler lehine DMF'li kabul edilmiştir. Yapılan analizler sonucunda maddeler DMF'li ve DMF'siz olmak üzere gruplandırılmıştır.

İkinci aşama olarak DMF analizleri yapılan maddelerden gerçek uygulaması yapılacak A ve B formları oluşturulmuştur. Bu formlarda ortak maddeler dışında kalan maddeler DMF'siz olanlardan seçilmiştir. Maddelerin kapsam bakımından paralelliğin sağlanabilmesi için ölçtükleri davranış bakımından kategorilendirilmiştir. Bu kategorilendirme yapılırken bilimsel süreç becerileri konusunda çalışma yapmış, fen bilgisi, ölçme ve değerlendirme, program geliştirme alanlarında yüksek lisans ve doktora düzeyinde uzmanlaşmış beş kişi ve iki fen bilgisi öğretmeninin görüşlerine başvurulmuştur.

Gerçek uygulamaya hazır formların son hali Tablo 1'de gösterilmektedir.

Tablo 1. Gerçek Uygulamada Kullanılan A ve B Formları Madde Sayıları

Formlar	DMF'siz 15 Madde	Ortak Maddeler		DMF'siz 15 Madde	Toplam Madde Sayısı
		8 DMF'li Madde	6 DMF'siz Madde		
A	√	√	√		29
B		√	√	√	29

Uygulama sonrası yapılan DMF analizi sonuçlarına göre ortak maddelerde TB-DMF'li çıkması öngörülen iki madde TBO-DMF'li, iki madde de DMF'siz çıktığı için bu dört madde A ve B formundan çıkarılmışlardır. Bununla birlikte ortak maddelerde DMF'siz çıkması öngörülen iki madde DMF verdikleri için formlardan çıkarılmışlardır. Ortak olmayan maddelerde ise DMF'siz çıkması öngörülen maddelerden bu durumu ihlal edenlere rastlanmıştır. Bu ihlal iki maddede sadece tek bir formda (A ve B düzeyinde) olduğu için kapsam geçerliğini düşürmemek adına maddeler formlara kabul edilmiştir. İki madde ise her iki formda da DMF'li bulunduğu için çıkarılmıştır. Yapılan değişiklikler sonrası A ve B formlarının son durumu Tablo 2'de verilmiştir:

Tablo 2. Eşitleme Yapılacak A ve B Formlarının Son Durumdaki Madde Sayıları

Formlar	DMF'siz 13 Madde	Ortak Maddeler		DMF'siz 13 Madde	Toplam Madde Sayısı
		4DMF'li Madde	4 DMF'siz Madde		
A	√	√	√		21
B		√	√	√	21

Tablo 2'ye göre eşitleme yapılacak formlardan DMF'li ortak maddeli A ve B formları ile DMF'siz ortak maddeli A ve B formları 17'şer maddeden oluşmaktadır.

Araştırma Deseni

Bu araştırmada test eşitleme için eşdeğer gruplarda uygulanan ortak test/madde deseni (equivalent groups (EQ) design with anchor test) kullanılacaktır. Bu desen iki örneklem grubu ortak bir evrenden alındığı zaman ortaya çıkmaktadır (Holland ve Dorans, 2006). Eşdeğer gruplar ortak madde deseninde eşdeğer yetenek düzeyinde rastgele seçilmiş iki grup bir testin farklı iki formunu

almaktadır. Öğrencilerin öğretim programına uygun olarak hazırlanmış bu formlar ortak maddeler barındırmaktadır.

Kullanılan Programlar

1) DMF gösteren maddelerin belirlenmesi amacıyla Mantel-Haenzel yöntemi EASYDIF (González Padilla, Hidalgo, Gómez-Benito ve Benítez, 2011) programı ile, Lojistik regresyon yöntemi ise Zumbo (1999) tarafından hazırlanan syntax kullanılarak SPSS programı ile gerçekleştirilmiştir.

2) MTK varsayımlarından tek boyutluluk varsayımı için LISREL (Jöreskog ve Sörborn, 1986) programı kulanılarak doğrulayıcı faktör analizi yapılmış, yerel bağımsızlık varsayımı için STATISCA programı ile elde edilen tetrakorik korelasyonlardan model veri uyumu test edilmiştir.

3) BILOG-MG (Zimowski, Muraki, Mislevy ve Bock, 1996) programı ile 2 parametreli lojistik modelin veri ile uyumlu olduğu tespit edilerek, beklenen a posteriori (Expected A Posteriori, EAP) parametre kestirim yöntemi kullanılarak madde parametreleri kalibre edilmiş ve yetenek kestirimleri gerçekleştirilmiştir.

4) MTK'ya dayalı ortalama-ortalama, ortalama-sigma ve karakteristik eğrisi yöntemleri için IRTEQ (Han, 2009) programı kullanılarak iki form arasında eşitleme yapmak için gerekli eşitleme katsayıları (A ve B) elde edilmiştir.

5) Her bir eşitleme yöntemi için eşitleme katsayıları kullanılarak EXCEL programı ile RMSD eşitleme hataları değerlendirme kriteri olarak hesaplanmıştır.

Verilerin Analizi

Bu başlıkta test eşitleme yapılacak A ve B formlarına (ADMF'li, BDMF'li ve ADMF'siz, BDMF'siz) ait gerçekleştirilen analizlere yer verilmiştir. Tablo 3'te eşitleme yapılacak A ve B formlarına ait KTK'ya dayalı betimsel istatistikler görülmektedir.

Tablo 3. Eşitlenecek A ve B Formlarına ait KTK'ya Dayalı Betimsel İstatistikler

	A FORMU Ortak Maddeler		B FORMU Ortak Maddeler	
	DMF'li	DMF'siz	DMF'li	DMF'siz
Madde Sayısı	17	17	17	17
Ortak Madde Sayısı	4	4	4	4
N	691	691	659	659
A. Ortalama	8.441	9.036	8.470	9.089
A. Ortalama SH	0.139	0.150	0.143	0.143
Mod	7	8	8	7
Medyan	8	9	7	9
Standart Sapma	3.674	3.953	3.681	3.693
Varyans	13.505	15.626	13.550	13.644
Çarpıklık (SH)	0.276 (0.093)	0.087 (0.093)	0.192 (0.095)	0.055 (0.095)
Çarpıklık / SH	0.026	0.935	2.021	0.579
Basıklık (SH)	-0.831(0.186)	-0.878(0.186)	-0.849(0.19)	-0.813(0.190)
Minimum	1	1	1	1
Maksimum	17	17	17	17
Ortalama rjx (biserial)	0.416	0.483	0.427	0.435
Ortalama güçlük p	0.496	0.531	0.498	0.534

Tablo 3 incelendiğinde, eşitleme yapılacak formlardan A DMF'li ortak maddeli ile B DMF'li ortak maddeli ve A DMF'siz ortak maddeli ile B DMF'siz ortak maddeli formların toplam 17'şer

maddeden oluştuğu görülmektedir. Her bir formda 4 ortak madde bulunmaktadır. Eşitlenecek formların ortalama güçlük (p) ve nokta çift serili ayırt edicilik indeksleri, r_{jx} (biserial), bakımından birbirine benzer olduğu ve 0.50 civarında elde edildikleri görülmektedir. Formların ortalama-mod-medyan değerlerinin birbirine yakın olması, çarpıklık basıklık katsayılarının sıfır civarında olması, çarpıklık katsayısının hatasına bölünmesiyle (çarpıklık/SH) elde edilen z istatistiklerinin α= 0.01 için 2.58'den küçük çıkması sebepleriyle tüm formların normal dağılıma uygun oldukları söylenebilir (Büyüköztürk, 2005).

Eşitleme yapılacak formların MTK'ya dayalı madde parametre ve yetenek kestirimleri yapılmadan önce MTK varsayımları test edilmiş ve veriye uygun MTK parametre kestirim modeline karar verilmiştir. MTK varsayımlarının testi aşağıda başlıklar halinde açıklanmıştır.

Tek Boyutluluk Varsayımın Test Edilmesi

Test eşitlemede kullanılan A ve B formlarının OKS ve SBS'de çıkmış fen ve teknoloji maddelerinden oluştuğu bilinmektedir. Söz konusu formların tek boyutlu olma hipotezinin incelenmesi için LISREL programı ile doğrulayıcı faktör analizi yapılmıştır. Burada amaç, kurulan modele ilişkin beklenen kovaryans matrisi ile örneklemden gözlenen kovaryans matrisi arasındaki uyumun çıkarımını yapmaktır. Model veri uyumu ölçütlerine ilişkin araştırmada elde edilen değerlere Tablo 4'te yer verilmiştir.

Tablo 4. A ve B Formlarına Ait Doğrulayıcı Faktör Analizi Sonuçları

Model Uyum Ölçütleri	A FORMU		B FORMU		Beklenen Uyum Değeri
	A DMF'li	A DMF'siz	B DMF'li	B DMF'siz	
χ^2 / Sd	1.771	1.881	2.275	1.603	< 3
RMSEA	0.033	0.036	0.044	0.030	0'a yakın
NFI	0.92	0.94	0.88	0.93	1'e yakın
NNFI	0.96	0.97	0.92	0.97	1'e yakın
CFI	0.97	0.97	0.93	0.97	1'e yakın
IFI	0.97	0.97	0.93	0.97	1'e yakın
RFI	0.91	0.93	0.86	0.91	1'e yakın
GFI	0.97	0.96	0.95	0.97	1'e yakın
AGFI	0.96	0.95	0.94	0.96	1'e yakın

Tablo 4'te görüldüğü gibi ki-kare'nin serbestlik derecesine bölünmesiyle elde edilen χ^2/Sd oranları tüm testler için 3'ten küçüktür. Model veri uyumu değerlendirilmesinde ki-kare/sd oranının 3'ten küçük olması modelin veriyle uyumlu olduğuna işaret etmektedir (Carmines ve Mclver, 1981). RMSEA değerleri 0'a yakın ve 0.05'ten küçük; diğer indeksler ise beklenildiği gibi 1'e yakın elde edilmiştir. Tüm formlara ait istenilen uyum indeksi değerlerinin elde edilmesi gerekçesiyle, fen sorularından oluşan formların tek bir özelliği ölçtüğü sonucuna ulaşılmıştır.

Yerel Bağımsızlık Varsayımın Test Edilmesi

Bu araştırmada A ve B formlarında tek boyutluluk varsayımı karşılanmış olmakla birlikte, bununla yakından ilişkili yerel bağımsızlık varsayımının testi yoluna da gidilmiştir. Crocker ve Algina (1986:342-343) yerel bağımsızlığın testinde, belli yetenek düzeylerinde elde edilen madde puanları arasındaki ikili korelasyonların karşılaştırılmasını önermiştir. Bu nedenle araştırmada, tüm gruptaki maddeler arası tetrakorik korelasyonlar, daha dar (homojen) yetenek dağılımında olan alt ve üst yetenek gruplarından elde edilen maddeler arası tetrakorik korelasyonlarla karşılaştırılmıştır. Elde edilen korelasyonlara ilişkin betimsel istatistikler Tablo 5'te verilmiştir.

Tablo 5. Farklı Yetenek Gruplarından Elde Edilen Maddeler Arası Tetrakorik Korelasyonlara Ait Betimsel İstatistikler

	Yetenek Grupları	Puan Aralıkları	N	Ortalama Korelasyon	Minimum	Maksimum
A Formu (DMF'li Ortak Maddeli)	Tüm	1-17	691	0.233	-0.107	0.566
	Üst	12-17	187	-0.059	-1.000	0.573
	Alt	1-6	187	-0.101	-0.731	0.533
B Formu (DMF'li Ortak Maddeli)	Tüm	1-17	659	0.250	-0.007	0.472
	Üst	11-17	178	-0.103	-1.000	0.555
	Alt	1-6	178	-0.084	-0.622	0.440
A Formu (DMF'siz Ortak Maddeli)	Tüm	1-17	691	0.297	-0.027	0.574
	Üst	12-17	187	-0.115	-1.000	0.677
	Alt	1-6	187	-0.073	-0.539	0.424
B Formu (DMF'siz Ortak Maddeli)	Tüm	1-17	659	0.258	-0.007	0.472
	Üst	12-17	178	-0.111	-1.000	0.595
	Alt	1-7	178	-0.064	-0.634	0.380

Tablo 5 incelendiğinde formlarda homojen olan üst ve alt gruptan elde edilen korelasyonların ortalamasının heterojen olan tüm gruba ilişkin ortalamadan düşük ve sıfıra oldukça yakın olduğu gözlenmektedir. Ranj küçüldükçe korelasyonların düşmesi yerel bağımsızlığın sağlandığını desteklemektedir.

Kalibrasyon Yöntemi ve Madde Parametreleri ile Yetenek Kestirimleri

Araştırmada farklı gruplardan elde edilen madde parametrelerini aynı ölçeğe dönüştürmek için ayrı kalibrasyon yöntemi kullanılmıştır. Ayrı kalibrasyon yönteminde uygun parametre sayısı belirlenerek (1, 2, 3 PLM) farklı gruplara ait madde parametreleri arasında doğrusal ilişki ortaya konulmaktadır. Elde edilen A ve B katsayıları ile iki grubun madde parametre ve yetenek kestirimleri ortak ölçeğe dönüştürülmektedir. Araştırmada yetenek kestirimi içinse maddelerin tamamı doğru veya yanlış yapıldığı durumlarda tüm modellere göre kestirim yapabilen, önceden belirlenmiş bir θ önsel dağılımının ortalamasını kullanan ve kolay hesaplanabilen Bayes modellerden EAP yöntemi tercih edilmiştir (Embretson ve Reise, 2000).

Model Veri Uyumunun Kontrolü

Madde Tepki Kuramı modellerinin uyum iyiliği değerlendirmelerinde tek boyutluluk, yerel bağımsızlık, hız testi olmama gibi varsayımların test edilmesi dışında hangi modelin veri ile uyumlu olduğunun raporlaştırılması da gerekmektedir. Düşük model-veri uyumu varlığında yetenek ve madde parametrelerinde değişmezlikten söz edilmesi mümkün değildir. Uyum iyiliği analizinde artıkların (residuals) bilgisine başvurmak çok değerli bir yoldur. Artıkların analizi yolunda tüm yetenek düzeylerinde her bir maddenin beklenen ve gözlenen doğru cevaplama oranları arasındaki fark incelenmekte ve bu fark madde karakteristik eğrileri ile açık bir şekilde gözlenmektedir. Madde karakteristik eğrilerinde artıklar küçük ve rastgele dağılım gösterdiğinde model veri uyumunun yakalandığı sonucuna ulaşılmaktadır (Hambleton, Swaminathan ve Rogers, 1991).

Bu araştırmada A DMF'li ortak maddeli, A DMF'siz ortak maddeli, B DMF'li ortak maddeli ve B DMF'siz ortak maddeli formların tüm maddelerinde 1, 2 ve 3 parametreli lojistik modellerden elde edilen madde karakteristik eğrileri incelenerek artıkların analizi değerlendirilmiştir. Madde karakteristik eğrilerinde artıkların küçük ve rastgele dağılım gösterdiği modelin 2PLM olduğu gözlenmektedir.

Model varsayımları testinde bir diğer yol olarak BILOG-MG programı ile 1, 2 ve 3 PLM için elde edilen χ^2 istatistiğinin anlamlılık düzeyi incelenerek testin ve maddelerin kullanılan modellerle

uyumu incelenmiştir. Büyük χ^2 değerleri beklenen (teorik) ve gözlenen (ampirik) MKE'ler arasında farkın büyüklüğüne dolayısıyla model veri uyumsuzluğuna işaret etmektedir. H_0 hipotezinin manidarlık testinde hesaplanan p değerinin 0.01'den büyük çıkması modelde beklenen ve gözlenen dağılımlar arasında uyum olduğunun göstergesi olmuştur (Hambleton ve Swaminathan, 1985).

Tablo 6. A ve B Formlarının 1, 2, 3PLM'e Dayalı Uyum İyiliği Sonuçları

Formlar	Toplam χ^2 (SH) / Model İle Uyumlu Madde Sayısı		
n_A= 691, n_B = 659, N=17	1PLM	2PLM	3PLM
A Formu (DMF'li Ortak Maddeli)	447.0 (120.0)/ 6	163.8 (122.0)*/ 17	249.2 (111.0)/ 12
B Formu (DMF'li Ortak Maddeli)	263.9 (124.0)/ 10	144.7 (126.0)*/ 17	252.8 (121.0)/ 11
A Formu (DMF'siz Ortak Maddeli)	472.3 (129.0)/ 6	249.8 (128.0)/ 13	329.2 (125.0)/ 11
B Formu (DMF'siz Ortak Maddeli)	290.6 (127.0)/ 10	180.1 (127.0)/ 15	317.5 (130.0)/ 10

*p> 0.01

Tablo 6'da verilen testlerin bütününe ait Ki-kare (SH) ve anlamlılık (p) değerlerine bakıldığında DMF'li ortak maddeli A ve B formunun ki-kare (p>0.01) ile 2PLM'ye göre model veri uyumu sağladığı; diğer formların sözkonusu modellerle (1PLM, 2PLM, 3PLM) uyumsuz olduğu görülmektedir. MTK model uyumu için yapılan pek çok çalışmada madde düzeyindeki uyuma bakılmaktadır. Bir testte bir tip MTK modeli ile tüm maddelerin uyumlu çıkmaması oldukça olası bir durumdur (Kang ve Cohen, 2007). Araştırmada maddelerin tek tek model uyumu incelendiğinde, iki parametreli lojistik modelle uyumlu (p>0.01) madde sayısının tüm testlerde en fazla olduğu gözlenmektedir. Yapılan incelemeler sonucunda madde parametreleri ve yeteneklerin kestirilmesinde 2PLM'nin kullanılmasının uygun olacağına karar verilmiştir.

Madde parametre ve yeteneklerinin 2PLM'ye dayalı kestirimlerinden sonra A ve B formlarından elde edilen MTK'ya dayalı betimsel istatistikler aşağıda Tablo 7'de verilmiştir.

Tablo 7. A ve B Formlarının MTK'ya Dayalı Betimsel İstatistikleri

	A FORMU Ortak Maddeler		B FORMU Ortak Maddeler	
	DMF'li	DMF'siz	DMF'li	DMF'siz
Madde Sayısı	17	17	17	17
Ortak Madde Sayısı	4	4	4	4
θ Ortalama	-0.001	-0.001	-0.001	-0.001
θ St. Sapma	0.034	0.034	0.034	0.034
θ Çarpıklık (SH)	0.323 (0.093)	0.224 (0.093)	0.270(0.095)	0.163(0.095)
θ Çarpıklık / SH	3.473*	2.408	2.842*	1.715
θ Basıklık (SH)	-0.749(0.186)	-0.721(0.186)	-0.664(0.190)	-0.641(0.190)
μ(b) Toplam	0.152	-0.053	0.084	-0.118
μ(b) Ortak Madde	0.164	-0.705	0.108	-0.756
μ(a) Toplam	1.011	1.148	0.975	0.998
μ(a) Ortak Madde	0.529	1.132	0.839	0.905

* α= 0.01 için >2.58 (normal dağılım yok)

Tablo 7 incelendiğinde tüm formların yetenek (θ) ortalama ve standart sapmalarının eşit olduğu görülmektedir. DMF'li ortak maddeli formların normal dağılımdan saptığı, sağa çarpık dağılım gösterdikleri, DMF'siz ortak maddeli formların ise normal dağılım koşulunu sağladıkları göze çarpmaktadır.

A ve B Formlarının KTK ve MTK'ya Dayalı Güvenirlik Katsayıları

İki testi alan bireylerin puan dağılımları eşitse ve bu iki test eşit derecede güvenilirse test puanlarının karşılaştırılabilir olduğu söylenebilir (Lord, 1950). Araştırmada her bir 1-0 puanlamalı eşitlenecek form için klasik test kuramına dayalı KR-20 ve madde tepki kuramına dayalı Lord'un güvenirlik katsayıları hesaplanmıştır. Güvenirlik analizi sonuçları Tablo 8'de görülmektedir.

Tablo 8. A ve B Formlarının KTK ve MTK'ya Dayalı Güvenirlik Katsayıları

Formlar	KTK'ya Dayalı KR-20 Güvenirlik Katsayısı	MTK'ya Dayalı Lord'un Güvenirlik Katsayısı		
		1PLM	2PLM	3PLM
A Formu (DMF'li Ortak Maddeli)	0.744	0.736	0.778	0.729
B Formu (DMF'li Ortak Maddeli)	0.752	0.742	0.762	0.719
A Formu (DMF'siz Ortak Maddeli)	0.792	0.782	0.800	0.772
B Formu (DMF'siz Ortak Maddeli)	0.756	0.748	0.766	0.737

Tablo 8 incelendiğinde MTK'ya dayalı Lord'un güvenirlik katsayıları en yüksek 2PLM'de elde edilirken, 1PLM ve 3 PLM'de Lord'un güvenirlik katsayısıları KR-20 güvenirlik katsayılarından düşük bulunmuştur. Özdemir'in (2004) araştırmasında benzer şekilde 2PLM'de Lord'un güvenirlik katsayısı yardımıyla elde edilen güvenirliğin, KTK'ya dayalı KR-20 güvenirliğinden yüksek sonuçlar verdiği bulunmuştur.

Eşitleme Hatasının Hesaplanması

Bu araştırmada A ve B formlarının ortak maddeleri DMF'li ve DMF'siz olduğu durumlarda yapılan eşitlemelere karışan hata eşitlik (1) ile gösterilen RMSD kriteri kullanılarak hesaplanmıştır. Bu kriter bireylerin gerçek yetenek düzeyi ile almadıkları testten kestirilen yetenek düzeyleri arası farkın kare ortalamalarının karekökü dayanmaktadır (Paek ve Young, 2005; Harris ve Crouse, 1993):

$$RMSD = \sqrt{\frac{\sum_i f_i(\theta^* - \theta)^2}{\sum_i f_i}} \qquad (1)$$

θ = Gerçek yetenek düzeyi

θ^*= Kestirilen yetenek düzeyi

f = Yetenek düzeylerine ait frekans

Eşitlik (3.2)'de her bir bireyin kestirilen yetenek düzeyi (θ^*) ile gerçek yetenek düzeyi (θ) farkının karesi hesaplanmaktadır. Daha sonra elde edilen değerler toplanarak yetenek düzeylerine ait frekansa bölünüp, sonucun karekökleri alınmaktadır. Bulunan RMSD değerlerinin küçük olması, gerçek ve kestirilen yetenek düzeyleri arasındaki farkın da küçük olduğunu; yani eşitleme hatasının az olduğunu belirtmektedir.

IRTEQ programıyla madde parametreleri eşitlenirken iki farklı grup ele alınmıştır. Bunlar temel grup ve karşılaştırılacak gruptur. Yine aynı şekilde yeniden ölçekleme yapılacak temel test ve karşılaştırılacak test söz konusudur. Burada süreç şöyle işlemektedir: Araştırmada temel test B formudur. Temel test formu aynı zamanda yeniden ölçeklendirilen test formudur. A formu ise hedef form ya da karşılaştırılacak test formudur. Temel test formunu alan; ancak bu formla ortak maddeler içeren A formunu almayan bireylerin gerçek yeteneklerinden (θ) yola çıkarak, almadıkları A formundaki kestirilen yetenekleri (θ^*) aşağıdaki eşitlik (2) ile hesaplanmıştır:

$$\theta^* = A\theta + B \qquad (2)$$

θ = Bireyin gerçek yetenek düzeyi (B formu)

θ^*= Bireyin almadığı testteki kestirilen yetenek düzeyi (A formu)

A = Eşitleme denkleminin eğimi

B = Eşitleme denkleminin sabiti

O-O, O-S, Haebara ve S-L eşitleme yöntemlerine ait A ve B eşitleme katsayıları ile temel formu (B) alan bireylerin karşılaştırma yapılacak formdaki (A) yeni yetenekleri kestirilmiş; aralarındaki farklar ile eşitleme hataları belirlenmeye çalışılmıştır.

BULGULAR

Ortak maddelerin TB-DMF'li ve DMF'siz olduğu durumlarda ortalama-ortalama, ortalama-sigma, Haebara ve Stocking Lord yöntemlerine göre yetenek kestirimleri arası farka dayanan RMSD eşitleme hataları Tablo 9'da verilmiştir.

Tablo 9. Ortak Maddelerin DMF'li ve DMF'siz Olduğu Durumlarda Dört Yönteme göre Yapılan Eşitlemelerin RMSD Eşitleme Hataları

Eşitleme Yöntemleri	RMSD Eşitleme Hataları	
	Ortak Maddeler DMF'li	Ortak Maddeler DMF'siz
Ortalama-Ortalama	0.518	0.202
Ortalama-Sigma	0.221	0.248
Haebara	0.485	0.181
Stocking-Lord	0.468	0.181

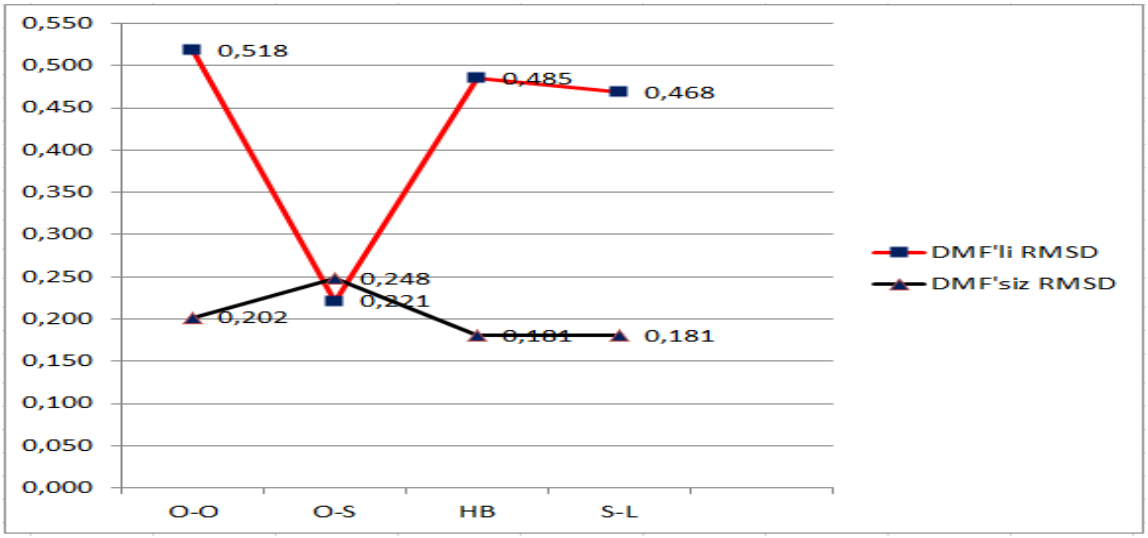
Tablo 9'da ortak maddelerin DMF'siz olduğu durumlarda HB ve S-L yöntemlerine ait eşitleme hataları birbirine eşit elde edilirken diğer hata değerlerinin farklılaştığı göze çarpmaktadır. Eşitleme hatalarını hesaplamada kullanılan A ve B eşitleme katsayıları Tablo 10'da verilmiştir.

Tablo 10. Ortak Maddelerin DMF'li ve DMF'siz Olduğu Durumlarda Dört Eşitleme Yöntemine göre Elde Edilen A ve B Eşitleme Katsayıları

Eşitleme Yöntemleri	Eşitleme Katsayıları			
	Ortak Maddeler DMF'li		Ortak Maddeler DMF'siz	
	A	B	A	B
Ortalama-Ortalama	1.59	-0.01	0.80	-0.10
Ortalama-Sigma	1.25	0.03	0.76	-0.13
Haebara	1.55	0.05	0.81	-0.07
Stocking-Lord	1.53	-0.05	0.81	-0.07

Tablo 10'a bakıldığında ortak maddeler DMF'siz olduğunda HB ve S-L yöntemlerine ait A ve B eşitleme hataları birbirine eşit elde edilirken diğer diğer durumlarda katsayıların farklılaştığı göze çarpmaktadır. Araştırmanın ana problemini oluşturan dört eşitleme yöntemine göre RMSD eşitleme hatalarının karşılaştırılması yapılırken Tablo 9'da elde edilen sonuçlar grafiğe aktarılarak somutlaştırılmıştır.

Araştırmanın sorusuna ilişkin bulgular Şekil 1'de yer almaktadır.



Şekil 1. Ortak Maddeler DMF'li veya DMF'siz Olduğunda O-O, O-S, HB ve S-L Yöntemlerine Göre Eşitleme Hataları (RMSD)

Şekil 1'de görüldüğü gibi ortak maddeler TB-DMF'li olduğunda yetenek kestirimlerine ait en büyük RMSD eşitleme hatası ortalama-ortalama yönteminde elde edilirken, en küçük hata ortalama-sigma yönteminde bulunmuştur. Kolen ve Brennan (2004) b parametreleri kestirimlerinin a parametreleri kestirimlerinden daha kararlı olmaları nedeniyle O-S yöntemini O-O yöntemine göre bazen daha tercih edilebilir bulmaktadır. O–O yöntemi ölçekleri eşitlemek için direkt olarak a ve b parametrelerinin ortalamalarını kullanmakta, A ve B eşitleme katsayılarını daha ileri rafine edememektedir (Baker ve Al-Karni, 1991). Araştırmada O-O yönteminin en büyük eşitleme hatasını verme nedeninin, ortak maddelerde a ve b parametreleri ortalamalarının A ve B formlarında gösterdiği farklılık olduğu düşünülmektedir. Diğer yöntemlerde eşitlenen iki formun ortak maddeleri arasında (O-S yönteminde b parametre standart sapmalarında, karakteristik eğri yöntemlerinde karakteristik eğrileri arasında) daha küçük farklar olmasının eşitleme hatalarının daha küçük çıkmasına neden olduğu söylenebilir.

Bununla birlikte ortak maddelerin TB-DMF'li olduğu durumda Haebara ve Stocking-Lord karakteristik eğrisi eşitleme yöntemlerinin birbirine çok yakın eşitleme hatalarına sahip olduğu gözlenmektedir. HB yöntemi S-L yönteminden 0.017 farkla daha büyük RMSD eşitleme hatasına sahiptir. S-L yöntemi karakteristik eğrisi yöntemleri içerisinde en az hataya sahip eşitleme yöntemi olmuştur. HB ve S-L yöntemlerinin birbirine yakın eşitleme hataları vermesinin hesaplanma şekillerinin benzerliğinden kaynaklandığı söylenebilir. İki yöntem de madde karakteristik eğrileri arasıdaki farka dayanmaktadır. Haebara yönteminde belli bir yetenek düzeyindeki bireyler için her bir maddenin madde karakteristik eğrileri arasındaki farkın karelerinin toplamı kullanılırken S-L yönteminde her bir maddenin madde karakteristik eğrileri toplamları farkının karesi işe koşulmaktadır.

Ortak maddeler DMF'siz olduğunda en büyük RMSD eşitleme hatası ortalama-sigma yönteminde, en küçük hata karakteristik eğrisi yöntemlerinde (HB ve S-L) ve eşit değerlerde elde edilmiştir.

Bununla beraber karakteristik eğrisi yöntemlerinin moment yöntemlerine göre daha doğru ve kararlı sonuçlar ürettiği görülmektedir. S-L ve HB yöntemlerinin birbirine açık bir üstünlüğü gözlenmemektedir. Diğer taraftan O-O yönteminin a ve b'nin sadece ortalamalarını kullandığı için güçlük parametresinin standart sapmasını kullanan O-S yöntemine göre daha az hassas olduğu ve böylece daha küçük hata ürettiği söylenebilir.

Ortak maddeler TB-DMF'li olduğunda O-O, HB ve S-L yöntemlerinde RMSD eşitleme hataları ortak maddelerin DMF'siz olduğu duruma göre büyük çıkma eğilimindedir. O-S yönteminde ise tam tersi bir durum elde edilmiştir. Bu bulguların değerlendirmeleri yöntemler tek tek ele alınarak aşağıda ayrıntılı olarak açıklanmıştır.

O-O Yöntemine Ait Bulgular

Şekil 1'e göre O-O yönteminde A ve B formlarının ortak maddeleri TB-DMF'li olduğunda eşitleme hatası RMSD, ortak maddelerin DMF'siz olduğu duruma göre 0.316 farkla daha yüksek bulunmuştur. Ortalama-ortalama yönteminde denklem eğimi (A) ayırıcılık parametrelerinin ortalaması ile hesaplanmaktadır. Daha sonra, ortak maddelerden kestirilen b parametrelerinin ortalaması ile B katsayısı elde edilmektedir. Aşağıda hesaplamaları verilmiş A ve B eşitleme katsayılarından hareketle O-O yönteminden elde edilen eşitleme hataları tartışılmıştır.

A ve B formlarında **DMF'li** ortak maddelerin a parametreleri ortalamaları oranlanarak elde edilen A eşitleme katsayısı ve b parametreleri ortalamalarıyla hesaplanan B eşitleme katsayısı formül değerleri yerine konulduğunda aşağıdaki gibidir:

$A = \mu(a_B) / \mu(a_A)$ ve,

$\mu(a_B) = 0.839$

$\mu(a_A) = 0.529$

A= 1.586

$B = \mu(b_A) - A.\ \mu(b_B)$ ve,

$\mu(b_A) = 0.164$

$\mu(b_B) = 0.108$

$B = 0.164 - 1.586(0.108)$

$B = -0.007 \approx -0.01$

A ve B formlarında **DMF'siz** ortak maddelerden hesaplanan A ve B eşitleme katsayıları ise aşağıdaki gibidir:

$A = \mu(a_B) / \mu(a_A)$ ve,

$\mu(a_B) = 0.905$

$\mu(a_A) = 1.132$

A= 0.80

$B = \mu(b_A) - A.\ \mu(b_B)$ ve,

$\mu(b_A) = -0.705$

$\mu(b_B) = -0.756$

$B = -0.705 - 0.80.(-0.756)$

B= -0.10

Ortak maddeler DMF'li veya DMF'siz olduğunda yukarıda hesaplanan A ve B katsayıları incelendiğinde, B katsayılarının (-0.01 ve -0.10) birbirine oldukça yakın oldukları göze çarpmaktadır. A katsayıları karşılaştırıldığında ise DMF'li ortak maddelerle eşitlemede A katsayısı daha büyük bulunmuştur. Bunun nedeninin, a parametreleri ortalamasının kestirilen A formunda keskin şekilde azalması olduğu söylenebilir. Bir başka ifadeyle DMF'li ortak maddelerle eşitlemede A eşitleme katsayısının büyümesi ($A.\theta+B$) ile hesaplanan yeteneğin büyük çıkmasına neden olmaktadır. Bu durumda DMF'li formda RMSD eşitleme hatası da artmaktadır. Ortak maddeler DMF'siz olduğunda a ve b parametreleri ortalamalarının iki formda daha benzer değerlere sahip olmalarının eşitleme hatasını küçülttüğü düşünülmektedir.

O-S Yöntemine Ait Bulgular

Şekil 1'e göre O-S yönteminde A ve B formlarının ortak maddeleri DMF'siz olduğunda eşitleme hatası RMSD, ortak maddelerin TB-DMF'li olduğu duruma göre 0.027 farkla daha yüksek bulunmuştur. O-S yöntemi ortak maddelerden kestirilen b (güçlük) parametrelerinin ortalama ve standart sapmalarına dayalı bir yöntemdir. A ve B formlarında **DMF'li** ortak maddelerin b parametreleri standart sapmaları oranlanarak elde edilen A eşitleme katsayısı ve b parametreleri ortalamalarıyla hesaplanan B eşitleme katsayısı formül değerleri yerine konulduğunda aşağıdaki gibi elde edilmiştir:

$A = \sigma(b_A) / \sigma(b_B)$ ve $B = \mu(b_A) - A.\ \mu(b_B)$ ve,

$\sigma(b_A) = 1.758$, $\sigma(b_B) = 1.406$ → **A= 1.250**

$\mu(b_A) = 0.164$, $\mu(b_B) = 0.108$ → B= 0.164- 1.250. (0.108) → **B= 0.029 = 0.03**

A ve B formlarında **DMF'siz** ortak maddelerin b parametreleri standart sapmaları oranlanarak elde edilen A eşitleme katsayısı ve b parametreleri ortalamalarıyla hesaplanan B eşitleme katsayısı formül değerleri yerine konulduğunda aşağıdaki gibi elde edilmiştir:

$A = \sigma(b_A) / \sigma(b_B)$ ve $B = \mu(b_A) - A.\ \mu(b_B)$ ve,

$\sigma(b_A) = 0.580$, $\sigma(b_B) = 0.761$ → **A= 0.760**

$\mu(b_A) = -0.705$, $\mu(b_B) = -0.756$ → B= (-0.705) – 0.760(-0.756) → **B= -0.13**

Sonuç olarak O-S yönteminde A eşitleme katsayısı ortak maddeler DMF'siz olduğunda ortak maddelerin DMF''li olduğu duruma göre azalmış; çünkü Form B'deki b parametreleri varyansı büyümüştür. Ortak maddeler DMF'li olduğunda ise Form B'deki b parametreleri varyansı küçülmüş ve A katsayısı artmıştır. B eşitleme katsayısı ise ortak maddeler DMF'siz olduğunda ortak maddelerin DMF''li olduğu duruma göre azalmıştır. DMF'siz ortak maddeli formda A ve B katsayılarının ayrı ayrı veya birlikte düşüş göstermeleri kestirilen yeteneğin (θ^*) azalmasına böylece gerçek ve kestirilen yetenekler arası farkın büyümesine neden olmaktadır. Bu durumda gerçek ve kestirilen yetenekler arası farka dayalı hesaplanan RMSD eşitleme hatası da DMF'siz ortak maddeli formda büyümektedir. O-S yönteminin DMF'siz ortak maddelerle eşitlemede daha büyük hata vermesi temel form B'deki DMF'siz ortak madde b parametrelerinin varyans artışına ya da form A'da kestirilen b parametrelerinin varyans düşmesine bağlanmıştır.

HB Yöntemine Ait Bulgular

Şekil 1 incelendiğinde ortak maddeler TB-DMF'li olduğu durumda HB yönteminden elde edilen RMSD eşitleme hatası ortak maddelerin DMF'siz olduğu duruma göre 0.304 farkla daha büyük bulunmuştur. Heabara yöntemi eşitleme yapılan A ve B formlarında her bir maddenin madde karakteristik eğrileri arası farka dayanmaktadır Belli bir yetenek düzeyindeki bireyler için her bir maddenin madde karakteristik eğrileri arasındaki farkın karelerinin toplamı alınmaktadır (Kolen ve Brennan, 2004). Dolayısıyla Haebara yöntemiyle eşitlemeden elde edilen bu bulgu eşitleme yapılan A ve B formlarının karakteristik eğrileri arasındaki farklılık bulunması ile ilişkilendirilmektedir. Ortak maddelerin DMF'li olduğu durumda A ve B formlarının karakteristik eğrilerinin arasında ortak maddelerin DMF'siz olduğu duruma göre daha büyük fark olduğu düşünülmektedir.

S-L Yöntemine Ait Bulgular

Şekil 1'e göre ortak maddeler TB-DMF'li olduğu durumda S-L yönteminden elde edilen RMSD eşitleme hatası, ortak maddelerin DMF'siz olduğu duruma kıyasla 0.287 farkla daha büyük bulunmuştur. S-L yönteminde elde edilen bu bulgu diğer bir karakteristik eğrisi dönüştürme yöntemi HB ile de tutarlılık göstermektedir. Bu bulgu HB yönteminde olduğu gibi eşitleme yapılan A ve B formlarının karakteristik eğrileri arasındaki farklılık bulunması ile ilişkilendirilmektedir (Kaskowitz ve De Ayala (2001).

SONUÇLAR ve TARTIŞMA

Literatürde ortak maddelerde TB-DMF'nin yetenek kestirimlerine etkisini Madde Tepki Kuramı'na dayalı "iki moment (Ortalama-Ortalama ve Ortalama-Sigma)" ve "iki karakteristik eğrisi (Haebara ve Stocking-Lord)" eşitleme yöntemlerinin tamamını birbiriyle karşılaştırarak ortaya koyan araştırma bulunmamaktadır. Yakın zamanda Huggins (2014) tarafından bu araştırmaya bağımsız değişkenleri ile benzerlik gösteren bir çalışma ortaya konulmuştur. Huggins'in (2014) çalışmasında ortak maddelerde DMF'nin eşitlemede grup değişmezliği üzerine etkileri incelenmiştir. Eşitleme değişmezliğinin eksiliği durumunda aynı puana sahip; ancak farklı gruptaki bireylerin eşitlenmiş puanlarının farklı çıkacağı ve bu durumun alt gruplardan birine avantaj sağlayacağı belirtilmiştir. Araştırmada ortak maddelerde TB-DMF söz konusuyken eşitleme değişmezliğine ilişkin RMSD eşitleme hataları dört yöntemde çok benzer elde edilmiştir. S-L ve HB yöntemleri O-O ve O-S ile karşılaştırıldığında daha küçük RMSD eşitleme hataları vermiştir. O-O yöntemi ise O-S yöntemine göre daha küçük RMSD değerleri üretmiştir. Arşaştırmanın moment yöntemleri bakımından bu çalışma ile farklılık göstermesine bağımlı değişkenlerin farklılığının neden olduğu söylenebilir.

Bu araştırmada ortak maddeler DMF'siz olduğunda karakteristik eğrisi yöntemlerinin moment yöntemlerine göre daha doğru ve kararlı sonuçlar üretmesi bulgusunu destekleyen pek çok çalışma bulunmaktadır (Baker ve Al-Karni, 1991; Hanson ve Béguin, 2002; Kim ve Cohen, 1992; Ogasawara, 2001a, 2001b; Kim ve Lee, 2004; Kim ve Lee, 2006; Karkee ve Wright, 2004; Kolen ve Brennan, 2004; Kim ve Kolen, 2006, Kilmen, 2010). Bu araştırmada S-L ve HB yöntemleri ortak maddelerin DMF'siz ele alındığı düşünülen diğer araştırmaların bulgularına zıtlık oluşturmayacak şekilde birbirine eşit eşitleme hataları üretmiştir. Karakteristik eğrisi eşitleme yöntemlerinden birinin diğerine açık ve önemli bir üstünlüğü bulunmamaktadır (Karkee ve Wright, 2004:12) ve ikisinin de kullanımı önerilmektedir. Yapılan araştırmaların bazılarında Stocking-Lord yöntemi (Gök, 2012), bazılarında ise Haebara yöntemi (Kim ve Colen, 2006) daha doğru sonuçlar üretmiştir. Bununla birlikte literatürde, iki yöntemin birbirine çok yakın ve hatta bazı koşullarda birbiriyle eşit hatalar ürettiği bulgusunu destekleyen araştırmalar bulunmaktadır (Way ve Tang, 1991; Hanson ve Béguin, 2002; Lee ve Ban 2010; Uysal, 2014). Holland ve Dorans'a (2006:201) göre eşdeğer olmayan gruplarda ortak madde deseninde iki örneklemdeki bireyler arasında çok ufak farklılıklar olduğunda, tüm lineer ölçekleme ve eşitleme yöntemleri benzer sonuçlar verme eğilimindedir (aynı durum lineer olmayan ölçekleme ve eşitleme yöntemlerinde de geçerlidir). Bu araştırmada O-O ve O-S lineer eşitleme yöntemleri de birbirine oldukça yakın eşitleme hataları üretmiştir. O-O yönteminde elde edilen eşitleme hatası ise daha az çıkmıştır. Araştırma bulgusunu destekler nitelikte Baker ve Al-Karni (1991) ortalamaların genellikle standart sapmalardan daha az hassas olması ve daha kararlı değerler üretmeleri nedeniyle ortalama-ortalama yöntemini ortalama-sigma yöntemine göre daha tercih edilebilir bulmaktadır. Bununla birlikte O-O ve O-S yöntemini karşılaştıran ampirik araştırmalar yeterli ve ikna edici değildir; bu nedenle burada önerilen yaklaşım iki prosedürü de dikkate almak ve eşitleme gerçekleştiğinde iki yöntemin uygulanmasından kaynaklanan ham puandan ölçek puanı elde etme değişimini karşılaştırmaktır (Kolen ve Brennan, 2004:167). Bu araştırmada ortak maddeler DMF'siz olduğunda incelenen dört eşitleme yöntemi içinde en büyük hatayı O-S yöntemi üretmiştir. Literatürde bu bulguyu destekleyen araştırmalar bulunmaktadır (Speron, 2009; Kilmen, 2010, Gök, 2012).

Bu araştırmada O-S yöntemi için elde edilen bulgu Han'ın (2008) araştırma bulgularıyla A katsayısı için farklılık B katsayısı için benzerlik göstermektedir. Han (2008) araştırmasında iki test

uygulaması arasında madde parameter kayması anlamına gelen MPK'yi DMF'nin bir çeşidi olarak ele almıştır. O-S yönteminde TB-MPK'nın büyüklüğü ve TB-MPK'lı madde oranı arttığında (40 maddeli testte 10 ortak maddenin %50'si MPK'lı iken) A eşitleme katsayısının azaldığı belirtilmiş; buna neden olarak MPK nedeniyle Form 2'de ortak madde b parametreleri varyansının artması gösterilmiştir. Benzer şekilde B katsayısı TB-MPK'lı madde sayısı ve TB-MPK büyüklüğü artınca, Form 2'de b parametre ortalamasının azalmasına bağlı olarak büyümüştür. Bu araştırmada da Form B'de b parametreleri ortalaması düştüğü için B katsayısı büyümüştür. Araştırmalarda A katsayılarında yaşanan farklılık; iki araştırmanın birebir aynı koşullara sahip olmamasına bağlanabilir. Örneğin Han'ın araştırmasında bu araştırmadan farklı olarak uzun test (40 maddeli) ve büyük örneklem (5000) kullanılmıştır. Yine iki araştırmada DMF'li ortak madde oranlarında farklılıklar bulunmaktadır. Bu araştırmada tüm ortak maddeler DMF'lidir. Han'ın araştırmasında ise ortak maddelerin %50'si MPK'lı iken bulunan A katsayısı, %10'u MPK'lı iken bulunan A katsayısından daha küçük elde edilmiştir. Yani araştırmada MPK'lı madde oranı arttıkça ortak maddelerin güçlük parametre varyansları artmış ve A katsayısı azalmıştır. Böylece bu araştırmadan farklı olarak O-S yönteminde eşitleme hatasının büyümüştür. Bu durumun nedeni DMF-MPK farklılığı başta olmak üzere iki araştırmadaki bağımsız değişkenlerin farklılığı olarak düşünülebilir. Wyse ve Reckase (2012) iki karakteristik eğrisi (S-L ve HB), iki moment (O-O ve O-S) ve eş zaman eşitleme yöntemlerini karşılaştırdıkları araştırmalarında O-S yönteminin zayıf performans göstermesine neden olarak, bir ortak maddenin güçlüğünde yaşanan ciddi değişimin temel formun güçlük parametreleri varyansını büyük çapta artırmasına bağlamışlardır. Araştırmada problemli maddenin testten çıkarılmasının güçlük farkını azaltarak eşitlemeyi iyileştirdiği belirtilmiştir.

Yapılan eşitleme çalışmalarında sıklıkla S-L yönteminin tek başına kullanıldığı görülmektedir. Örneğin Chen'in (2013) araştırma bulguları bu araştırmada S-L yöntemi için elde edilen bulguyu destekler niteliktedir. Chen (2013) simülatif veri ile yürüttüğü araştırmasında S-L yöntemiyle eşitlemede ortak maddelerin a ve b parametrelerinde ayrı ayrı ele alınan MPK'nın yetenek kestirimlerine etkisini incelemiştir. Araştırmada ortak maddelerin MPK'lı olması durumunda MPK'lı madde sayısı ve MPK büyüklüğüne bakmaksızın eşitleme yanlılığı (BIAS) ve RMSE eşitleme hataları yüksek bulunmuştur. MPK'lı maddelerin silinmesi durumunda yanlılık ve hata düşmüştür. Wells, Subkoviak ve Serlin'nin (2002) araştırma bulguları da elde edilen sonuçla paralellik göstermektedir. Araştırma sonuçlarında S-L yöntemiyle eşitlemede a ve b parametrelerinde ayrı ayrı ve birlikte kayma olduğunda, madde parametrelerinde kayma olmadığı duruma göre yetenek kestirimlerindeki hata artmıştır. Atalay Kabasakal (2014) S-L yöntemine göre DMF'li ortak maddelerin testten çıkarılması durumunda yetenek kestirimlerinde elde edilen hatanın 24 farklı simülasyon koşulunda genel olarak arttığını bulmuştur. Bu artış DMF bulunan test ve DMF etki büyüklüğüne göre değişmektedir. Ancak bu araştırmanın koşullarıyla benzer şekilde; DMF'li maddeler ortak testte olduğunda, küçük örneklem (500) ve kısa testlere (20) ait bazı simülasyonlarda % 10 oranındaki DMF'li ortak maddelerin silinmesi yetenek kestirimlerinde hatayı (RMSE) düşürmüştür. İki araştırmanın koşulları kısmen benzer seçildiğinde sonuçlarda tutarlılık gözlendiği söylenebilir.

Ortak maddelere dayalı eşdeğer gruplar deseninde eşitleme yaparken ortak maddelerin tamamı erkekler lehine TB-DMF'li olduğunda tüm yöntemler içinde O-S yöntemi, karakteristik eğrisi yöntemleri uygulanacaksa S-L yönteminin kullanılması önerilebilir. Ortak maddelerin tamamı DMF'siz olduğunda ise tüm yöntemler içinde S-L ve HB yöntemlerinden herhangi birisinin, moment yöntemleri içinden ise daha kararlı olan O-O yönteminin kullanılması önerilebilir. Böylece uygulamalarda büyük eşitleme hatasının önüne geçilebilir ve sınavların sonunda alınan kararların doğruluğu artırılmış olur. Diğer araştırmacılar tarafından aynı anda hem DMF'li hem de DMF'siz ortak maddeler varlığında benzer bir çalışma yürütülebilir. DMF'li ortak maddeleri çıkarmanın eşitleme hataları üzerine etkisi bakılabilir. Farklı demografik özellikler değişken olarak ele alınıp DMF analizlerinde TBO DMF'li maddeler de ele alınabilir. Ortak maddelerde DMF'ye ek olarak "madde yanlılığı" varlığında test eşitleme hataları uzun ve kısa testler birlikte kullanılarak karşılaştırılabilir. Farklı kapsamda testlerle çalışılarak eşitleme hatalarının ölçülen özellik bakımından değişip değişmediği çalışılabilir. Test eşitleme deseni ve eşitleme yöntemleri

değiştirilerek araştırma tekrarlanabilir. Simülasyon ve gerçek veri birlikte uygulanarak farklı bir araştırma yapılabilir.

KAYNAKÇA

Angoff, W.H. (1971). Scales, Norms, and Equivalent Scores. In Thorndike, R.L. (Ed.) *Educational Measurement, 508-600.* American Council on Education, US: Washington D.C.

Atalay Kabasakal, K. (2014). D*eğişen madde fonksiyonunun test eşitlemeye etkisi*. Yayımlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü.

Babcock, B., & Albano, A. D. (2012). Rasch scale stability in the presence of item parameter and trait drift. *Applied Psychological Measurement, 36*(7), 565-580.

Bakan Kalaycıoğlu, D. (2008). *Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi.* Yayımlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Baker, F.B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. J*ournal of Educational Measurement, 28*(2), 147-162.

Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi*. Yayımlanmamış Yüksek Lisans Tezi, Mersin: Mersin Üniversitesi, Sosyal Bilimler Enstitüsü.

Büyüköztürk, Ş. (2005). Sosyal bilimler için veri analizi el kitabı. 5. Baskı Ankara: Pegema Yayıncılık.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2008). Bilimsel araştırma yöntemleri. 2. Baskı Ankara: Pegem Akademi.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Volume 4, Thousand Oaks, CA: Sage Publications.

Carmines, E.G., & McIver, S.P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt and E.F. Borgatta (Eds.), *Social measurement: current issues, 65-115.* Beverly Hills, California: Sage Publications, Inc.

Chu, K. (2002). *Equivalent group test equating with the presence of differential item functioning.* Unpublished Doctoral Dissertation. US: Florida State University.

Cook, L. L., Eignor, D. R., & Taft, H. L. (1985). *A comparative study of curriculum effects on the stability of IRT and conventional item parameter estimates*. Princeton NJ: Educational Testing Service.

Cook, L. L., & Paterson, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11, 225-244.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* FL, Orlando: Harcourt Brace Jovanovich, Inc.

Çepni, Z. (2011). D*eğişen madde fonksiyonlarının sibtest, Mantel Haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. Yayımlanmamış Doktora Tezi. Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması.* Yayımlanmamış Doktora Tezi. Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* USA: Lawrence Erlbaum Associates.

Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation prosedure? *Educational Researcher, 36*(8), 449-455.

Goldstein, H. (1983). Measuring changes in educational attainment over time: problems and possibilities. *Journal of Educational Measurement, 20*(4), 369-377.

González, A.; Padilla, J.L.; Hidalgo, M.D. Gómez-Benito , J. & Benítez, I. (2011). Easydif: Software for analysing differential item functioning using the Mantel-Haenszel and standardization procedures. *Applied Psychological Measurement*, 35, 483-484.

Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*. Yayımlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* US: Kluwer-Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications.

Han, K.T. (2008). *Impact of item parameter drift on test equating and proficiency estimates.* Unpublished Doctoral Dissertation. US: University of Massachusetts.

Han, K. T. (2009). IRTEQ: Windows application that implements IRT scaling and equating (computer programme). *Applied Psychological Measurement, 33*(6), 491-493.

Hanson, B.A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement, 26*(1), 3-24.

Harris, D. J., & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education, 6*(3), 195-240.

Hidalgo Montesinos, M. D., & Lopez Pina, J. A. (2002). Two-stage equating in differential item functioning detection under the graded response model with the Raju area measures and Lord statistic. *Educational and Psychological Measurement. 62*(1), 32.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational Measurement (4th Edition). 187-220.* Westport, CT: American Council on Education and Praeger.

Huggins, A. C. (2014). The effect of differential item functioning in anchor items on population invariance of equating. *Educational and Psychological Measurement. 74(*4), 627-658.

Jöreskog, K.G., & Sorböm, D. (1986). *Prells a program for multivariate data screening and data summarization: A preprocessor for Lisrel*. Mooresville, Ind.: Scientific Software Inc.

Kan, A. (2010). Test eşitleme: Aynı davranışları ölçen, farklı madde formlarına sahip testlerin istatistiksel eşitliğinin sınanması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(1), 16-21.

Kan, A. (2011). Test eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim, 36*(160), 38-51.

Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement, 31*(4), 331-358.

Karkee, T. B., & Wright, K. R. (2004). *Evaluation of linking methods for placing three-parameter logistic item parameter estimates onto a one parameter scale.* Paper presented at the Annual Meeting of the American Educational Research, US: San Diego, California, April 16.

Kaskowitz, G. S., & De Ayala, R. J. (2001). The effect of error in item parameter estimates on the test response function method of linking. *Applied Psychological Measurement*, 25, 39-52.

Kelecioğlu, H. (1994). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma.* Yayımlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Kilmen, S. (2010). *Madde tepki kuramına dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması.* Yayımlanmamış Doktora Tezi, Ankara: Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü.

Kim, S., & Cohen, A. S. (1991). Effects of linking methods on detection of DIF. Paper presented at the Annual Meeting of the National Council on Measurement in Education, US: Chicago, IL.

Kim, S., & Cohen, A. S. (1998). A comprasion of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22, 131-143.

Kim, S., & Lee, W. C. (2004). *IRT scale linking methods for mixed-format tests.* ACT Research Report 2004–5. US: IA, ACT Inc.

Kim, S., & Lee, W. C. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement, 43*(1), 53–76.

Kim, S., & Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education, 19*(4), 357-381.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling and linking: Methods and practices. (*2nd Edition). New York: Springer- Verlag.

Mcdonald, Roderick P.(1999). Test theory: A unified treatment. New Jersey: Lawrence Erlbaum Associates, Publishers.

Mutluer, C. (2013). *Yıl içinde farklı dönemlerde yapılan akademik personel ve lisansüstü eğitimi giriş sınavı (ales) puanlarına ilişkin bir test eşitleme çalışması.* Bolu: Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü.

Ogasawara, H. (2001a). Item response theory true score equating and their standard errors. *Journal Of Educational Behavioral Statistics, 26*(1), 31-50.

Ogasawara, H. (2001b). Least square estimations of item response theory linking coefficients. *Applied Psychological Measurement, 25*(4), 3-21.

Özdemir, D. (2004). Çoktan seçmeli testlerin klasik test teorisi ve örtük özellikler teorisine göre hesaplanan psikometrik özelliklerinin iki kategorili ve ağırlıklandırılmış puanlanması yönünden karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Sayı:* 26,117-123.

Öztürk, N. (2010). A*kademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma.* Yayımlanmamış Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Paek, I., & Young, M. J. (2005). Investigation of student growth recovery in a fixed-item linking procedure with a fixed-person prior distribution for mixed-format test data. *Applied Measurement in Education, 18*(2), 199-215.

Şahhüseyinoğlu, D. (2005). *İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması*. Yayımlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.

Tekin, H. (1991). Eğitimde ölçme ve değerlendirme. Ankara: Yargı Yayınları.

Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning.* Unpublished Doctoral Dissertation. US: Florida State University.

Uysal, İ. (2014). *Madde tepki kuramı'na dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması*. Yayımlanmamış Yüksek Lisans Tezi, Bolu: Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü.

Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement. 26*(1), 77–87.

Zumbo, B.D.A. (1999). *Handbook on the theory and methods of differantial item functioning: Logistic regression modelling as a unitary framework for binary and likert-type item scores.* Ottowa: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *Bilog–mg: Multiple-group IRT analysis and test maintenance for binary items.* Chicago: Scientific Software International.

EXTENDED ABSTRACT

Introduction

It is expected that examinees who are at the same ability level but different subgroups should have the same probability of answering the item correctly. If they do not, that item will have differential item functioning (DIF). Statistical findings of DIF must always be followed up by logical analyses to interpret the cause of differential difficulty (Camilli ve Shepard, 1994). DIF statistics may occur from two reasons: the real ability difference between subgroups or item characteristic features. After the logical investigations, if item characteristic features that are not supposed to be measured cause differential difficulty, the item would be biased against examinees of the affected subgroup. According to Embretson (2007) DIF as a psychometric property is evidence in the validity system. A number of studies (Yurdugül, 2003; Bekçi, 2007; Bakan Kalaycıoğlu, 2008; Asil, 2010; Çepni, 2011) discuss detection of the item bias is considered very important in improving the validity and reliability of test scores.

Bias could also come out when between two forms of the test have differences in difficulty. It is important which form (easy or difficult) examinees did get. Because examinees who take the easy form can get high scores than examinees who take the difficult one. In this case equating is used to adjust for differences on difficulty among test forms so that scores on the forms can be used interchangeably. Thus bias that caused by test forms would be removed (Angoff, 1971; Hambleton, Swaminathan ve Rogers, 1991; Cook ve Eignor, 1991; Kolen ve Brennan, 2004; Holland ve Dorans, 2006).

While equating with anchor test, which has DIF items, these items are inclined to reduce validity and reliability of the anchor test. This phenomenon endangers the correct expression of group differences. Similarly equating coefficients may be seriously affected by the presence of the DIF items. Accordingly, when anchor test was used it has to be investigated if anchor items processed in same way between different test forms. This investigation can reduce the potential effects of DIF items on test equating coefficients (Kim ve Cohen, 1991; Cook ve Paterson, 1987; Hidalgo-Montesinos ve Lopez-Pina, 2002).

Removing DIF items from test is undesirable; because it reduces the number of items and content validity. Chu (2002) and Turhan (2006) confirmed this conclusion and they also reported that equating error increases proportionally when number of DIF items increase. Chu (2002) pointed out that removing DIF items from a short test impairs validity of the test harder. Thus, instead of

removing DIF items from the test, a comparison of DIF effect to equating errors obtained with various methods can be done and method with lowest error can be selected. The purpose of this study to compare the results of equating methods based on Item Response Theory when all the anchor items showing or not showing gender based uniform DIF.

Method

The effect of DIF items on test equating presented on real data with horizontal equating using seperate calibration methods and equivalent groups with anchor test design. Data set for the study was obtained from the forms of science test which applied to 1350 students in 8th grade. A and B forms of science test consisted of 29 items each. The items extracted from OKS and SBS science tests applied between 2000-2012. In total 275 items were analyzed for detecting DIF. During DIF research, gender was chosen as group variable. DIF analysis conducted on EASYDIF software for "Mantel-Haenszel" (MH) method and on SPSS with syntax presented by Zumbo for "logistic regression" (LR) method. DIF items, which were at the B and C level on both MH and LR methods, uniform and favored males were chosen in the research. After the application of A and B forms, items that were not consistent with terms mentioned above were removed. At last each form, consisted of 4 anchor items and 17 items in total, were equated. The equating methods "mean-mean", "mean-sigma", "Haebara" and "Stocking-Lord" were used which depend on Item Response Theory (IRT). Before the estimation of item parameter and ability, IRT assumptions were tested. Two parameter logistic model was chosen in the form of IRT parameter estimation model. BILOG-MG was utilized for the estimation of item parameters and ability, IRTEQ software was utilized for test equating. The performances of the equating methods were evaluated through Root Mean Square Differences (RMSD) equating errors which is based on difference of ability estimates

Results and Discussion

According to the results of the study when the anchor items with uniform DIF favored males were used for equating, mean-mean method produced the biggest RMSD equating error whereas mean-sigma method produced the smallest. When the anchor items with no-DIF were used for equating the biggest RMSD was obtained from mean-sigma method and smallest RMSD was obtained from Stocking-Lord ve Haebara methods in equal to each other.

When the anchor items showed DIF; mean-mean, Haebara and Stocking-Lord methods produced larger values of RMSD than condition of anchor items showed no DIF. In the study conflicting result was obtained for the mean-sigma method. When anchor items were showing no DIF, RMSD increased amount of 0.027. This is because mean-sigma method was affected from the change of standart deviation in the b-parameter. In this stage, decrease in standart deviation of b parameter values of the anchor items influenced by DIF. Thus, large scaling coefficients A and B which are functions of the mean and the standard deviation of a set of b parameter values of the anchor items was obtained in basic form B.