
Eđitimde ve Psikolojide Ölçme ve Deęerlendirme Dergisi

Journal of Measurement
and Evaluation in
Education and Psychology

ISSN:1309-6575

Yaz 2016
Summer 2016

Cilt: 7- Sayı: 1
Volume: 7- Issue: 1



Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology

ISSN: 1309 – 6575

Sahibi

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Derneği (EPODDER)

Editör

Prof. Dr. Selahattin GELBAL

Yardımcı Editör

Dr. Kübra ATALAY KABASAKAL

Genel Sekreter

Doç. Dr. Tülin ACAR

Yayın Kurulu

Doç. Dr. Cem Oktay GÜZELLER
Doç. Dr. Hakan Yavuz ATAR
Doç. Dr. Neşe GÜLER
Doç. Dr. Tahsin Oğuz BAŞOKÇU
Yrd. Doç. Dr. Deniz GÜLLEROĞLU
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Yrd. Doç. Dr. N. Bilge BAŞUSTA
Dr. Nagihan BOZTUNÇ ÖZTÜRK

Dil Editörü

Doç. Dr. Burcu ATAR
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Dr. Ayfer SAYIN

Sekreteryası

Arş. Gör. İbrahim UYSAL
Arş. Gör. Levent ERTUNA
Arş. Gör. Nermin KIBRISLIOĞLU UYSAL

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi (EPOD) yılda iki kez yayımlanan hakemli ulusal bir dergidir. Yayımlanan yazıların tüm sorumluluğu ilgili yazarlara aittir.

İletişim

e-posta: epod@epod-online.org
Web: <http://epod-online.org>

Owner

The Association of Measurement and Evaluation in Education and Psychology (EPODDER)

Editor

Prof. Dr. Selahattin GELBAL

Assistant Editor

Dr. Kübra ATALAY KABASAKAL

Secretary

Assoc. Prof. Dr. Tülin ACAR

Editorial Board

Assoc. Prof. Dr. Cem Oktay GÜZELLER
Assoc. Prof. Dr. Hakan Yavuz ATAR
Assoc. Prof. Dr. Neşe GÜLER
Assoc. Prof. Dr. Tahsin Oğuz BAŞOKÇU
Assist. Prof. Dr. Deniz GÜLLEROĞLU
Assist. Prof. Dr. Derya ÇOBANOĞLU AKTAN
Assist. Prof. Dr. N. Bilge BAŞUSTA
Dr. Nagihan BOZTUNÇ ÖZTÜRK

Language Reviewer

Assoc. Prof. Dr. Burcu ATAR
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Dr. Ayfer SAYIN

Secretariat

Res. Asst. İbrahim UYSAL
Res. Asst. Levent ERTUNA
Res. Asst. Nermin KIBRISLIOĞLU UYSAL

Journal of Measurement and Evaluation in Education and Psychology (EPOD) is a national refereed journal that is published two times a year. The responsibility lies with the authors of papers.

Contact

e-mail: epod@epod-online.org
Web: <http://epod-online.org>

Dizinleme / Abstracting & Indexing

DOAJ (Directory of Open Access Journals), TÜBİTAK Ulakbim Sosyal ve Beşeri Bilimler Veri Tabanı, tei (Türk Eğitim İndeksi)

Hakem Kurulu / Referee Board

| | |
|---|--|
| Adnan ERKUŞ (Emekli Öğretim Üyesi) | Hakan Yavuz ATAR (Gazi Üni.) |
| Adnan KAN (Gazi Üni.) | Halil YURDUGÜL (Hacettepe Üni.) |
| Ahmet TURAN (Pearson) | Hatice KUMANDAŞ (Artvin Çoruh Üni.) |
| Ali BAYKAL (Bahçeşehir Üni.) | Hülya KELECİOĞLU (Hacettepe Üni.) |
| Arif ÖZER (Hacettepe Üni.) | Hüseyin SELVİ (Mersin Üni.) |
| Arkun TATAR (Fatih Sultan Mehmet Vakıf Üni.) | Hüseyin Hüsnü YILDIRIM (Abant İzzet Baysal Üni.) |
| Ayfer SAYIN (Gazi Üni.) | İbrahim Alper KÖSE (Abant İzzet Baysal Üni.) |
| Aylin ALBAYRAK SARI (Hacettepe Üni.) | İbrahim KEKLİK (Hacettepe Üni.) |
| Ayşegül ALTUN (Ondokuz Mayıs Üni.) | İlker KALENDER (Bilkent Üni.) |
| Bayram BIÇAK (Akdeniz Üni.) | İpek ÖZER GÜÇLÜ (Boğaziçi Üni.) |
| Bayram ÇETİN (Gazi Üni.) | İrfan YURDABAKAN (Dokuz Eylül Üni.) |
| Bilge BAŞUSTA UZUN (Mersin Üni.) | İsmail KARAKAYA (Gazi Üni.) |
| Bilge GÖK (Hacettepe Üni.) | Kaan Zülfikar DENİZ (Ankara Üni.) |
| Burcu ATAR (Hacettepe Üni.) | Kübra ATALAY KABASAKAL (Hacettepe Üni.) |
| Burhanettin ÖZDEMİR (Siirt Üni.) | Mediha KORKMAZ (Dokuz Eylül Üni.) |
| Cem Oktay GÜZELLER (Hacettepe Üni.) | Mehmet Hakan GÜNDÖĞDU (Mersin Üni.) |
| Çetin SEMERCİ (Bartın Üni.) | Mehmet KAPLAN (MEB) |
| Deniz GÜLLEROĞLU (Ankara Üni.) | Mehtap ÇAKAN (Gazi Üni.) |
| Deniz Tuğçe Özmen (Ankara Üni.) | Melek Gülşah ŞAHİN (Gazi Üni.) |
| Derya ÇAKICI ESER (Kırıkkale Üni.) | Meliha TUZGÖL DOST (Hacettepe Üni.) |
| Derya ÇOBANOĞLU AKTAN (Hacettepe Üni.) | Meltem ACAR GÜVENDİR (Trakya Üni.) |
| Devrim ALICI (Mersin Üni.) | Meral ALKAN (MEB) |
| Dilara BAKAN KALAYCIOĞLU (ÖSYM) | Metin YAŞAR (Pamukkale Üni.) |
| Dilek GENÇTANRIM (Kırşehir Ahi Evran Üni.) | Mustafa ASİL (University of Auckland) |
| Durmuş ÖZBAŞI (Çanakkale Onsekiz Mart Üni.) | Mustafa İLHAN (Dicle Üni.) |
| Duygu ANIL (Hacettepe Üni.) | Nagihan BOZTUNÇ ÖZTÜRK (Hacettepe Üni.) |
| Duygu GÜNGÖR (İzmir Üni.) | Neşe GÜLER (Sakarya Üni.) |
| Emine ÖNEN (Gazi Üni.) | Neşe ÖZTÜRK GÜBEŞ (Mehmet Akif Ersoy Üni.) |
| Emrah GÜL (Hakkari Üni.) | Nilüfer KAHRAMAN (Gazi Üni.) |
| Emre ÇETİN (Doğu Akdeniz Üni.) | Nizamettin KOÇ (Ankara Üni.) |
| Ergül DEMİR (Ankara Üni.) | Nuri DOĞAN (Hacettepe Üni.) |
| Esin TEZBAŞARAN (İstanbul Üni.) | Nükhet DEMİRTAŞLI (Ankara Üni.) |
| Esin YILMAZ KOĞAR (Hacettepe Üni.) | Okan BULUT (University of Alberta) |
| Esra Eminoğlu ÖZMERCAN (MEB) | Oya SOMER (Ege Üni.) |
| Ezel TAVŞANCIL (Ankara Üni.) | Ömay ÇOKLUK (Ankara Üni.) |
| Fatih KEZER (Kocaeli Üni.) | Ömer KUTLU (Ankara Üni.) |
| Fatih ORCAN (Karadeniz Teknik Üni.) | Ömür Kaya KALKAN (Hacettepe Üni.) |
| Fatma BAYRAK (Hacettepe Üni.) | Önder SÜNBÜL (Mersin Üni.) |
| Fazilet TAŞDEMİR (Recep Tayyip Erdoğan Üni.) | Özlem HASKAN AVCI (Hacettepe Üni.) |
| Funda NALBANTOĞLU YILMAZ (Nevşehir Üni.) | Özlem Yeşim ÖZBEK BAŞTUĞ (Karatekin Üni.) |
| H. Giray BERBEROĞLU (Orta Doğu Teknik Üni.) | Petek AŞKAR (TED Üni.) |
| Güçlü ŞEKERCİOĞLU (Akdeniz Üni.) | Ramazan BAŞTÜRK (Pamukkale Üni.) |
| Göksu GÖZEN (Mimar Sinan Güzel Sanatlar Üni.) | Recep Serkan ARIK (Dumlupınar Üni.) |
| Güliden KAYA UYANIK (Sakarya Üni.) | Reha ALPAR (Hacettepe Üni.) |
| Güler YAVUZ (Adıyaman Üni.) | Salih BİNİCİ (Florida State Üni.) |
| Gülşah BAŞOL (Gaziosman Paşa Üni.) | Seçil Ömür SÜMBÜL (Mersin Üni.) |
| Gülşen TAŞDELEN TEKER (Sakarya Üni.) | Sedat ŞEN (Harran Üni.) |
| H. Canan SÜMER (Orta Doğu Teknik Üni.) | Seher YALÇIN (Ankara Üni.) |
| Hakan ATILGAN (Ege Üni.) | Selahattin GELBAL (Hacettepe Üni.) |
| Hakan KOĞAR (Akdeniz Üni.) | Selda YILDIRIM ((Abant İzzet Baysal Üni.) |
| Hakan SARIÇAM (Dumlupınar Üni.) | |

Sema SULAK (Bartın Üni.)
Seval KIZILDAĞ (Adıyaman Üni.)
Sevda ÇETİN (Hacettepe Üni.)
Sevilay KİLMEN (Abant İzzet Baysal Üni.)
Şener BÜYÜKÖZTÜRK (Hasan Kalyoncu Üni.)
Şeref TAN (Gazi Üni.)
Tahsin Oğuz BAŞOKÇU (Ege Üni.)
Tuncay ÖĞRETMEN (Ege Üni.)

Tülin ACAR (Parantez Eğitim)
Türkan DOĞAN (Hacettepe Üni.)
Ümit ÇELEN (Amasya Üni.)
Yaşar BAYKUL (Emekli Öğretim Üyesi)
Yavuz AKPINAR (Boğaziçi Üni.)
Yeşim ÖZER ÖZKAN (Gaziantep Üni.)
Zekeriya NARTGÜN (Abant İzzet Baysal Üni.)

*Ada göre alfabetik sıralanmıştır. / Names listed in alphabetical order.

Çok Düzeyli Meta-Analiz Yöntemleri Üzerine Bir Çalışma

A Study on Multilevel Meta-Analysis Methods

Sedat ŞEN*

Nihal AKBAŞ**

Öz

Bu çalışmada meta-analiz alanyazınında kullanılan istatistiksel modeller incelenmiştir. Özellikle alanyazında karşılaşılan bağımlılık problemini çözmek için önerilen yöntemlerden birisi olan çok düzeyli meta-analiz modellerinin tanıtılması amaçlanmıştır. Çalışmada ilk olarak geleneksel meta-analiz modellerinin nasıl yapıldığından, daha sonra meta-analizde kullanılan yeni yaklaşımlardan ve çok düzeyli meta-analiz modellerinden bahsedilmiştir. Bu çalışmada daha önce geleneksel meta-analiz modelleriyle analiz edilmiş bir meta-analiz verisi geleneksel, iki düzeyli ve üç düzeyli meta-analiz modelleri ile analiz edilmiştir. Kullanılan veri özel eğitimde sıklıkla çalışılan yaratıcılık yapısı ile psikolojik bir rahatsızlık olan psikotizm arasındaki ilişkiyi içeren çalışmalardan elde edilen etki büyüklüklerini içermektedir. Yapılan üç farklı analizde iki düzeyli meta-analiz sonuçlarının geleneksel (rastgele etkiler modeli) meta-analiz sonuçlarıyla aynı olduğu görülmüş buna ek olarak iki düzeyli modellerin üç düzeye genişletilerek bağımlılık problemini çözmek için nasıl kullanılacağı belirtilmiştir. Çalışmada Türkiye’deki araştırmacılar için meta-analiz alanyazınında geliştirilen bu yeni yöntemler hakkında bilgiler sunulmuştur.

Anahtar Kelimeler: Meta-analiz, çok düzeyli meta-analiz, bağımlılık, özel eğitim

Abstract

Main purpose of this study is to examine different models that are used to analyze meta-analytic data sets. Particularly, it was aimed to introduce the multilevel models that have been proposed as one of the best ways to handle dependency problem. Firstly, traditional meta-analysis techniques were described then new approaches provided in meta-analysis were explained with the focus on multilevel meta-analysis models. Three different methods (traditional meta-analysis, two-level and three-level meta-analysis models) were used to analyze a meta-analysis data set that has been used in a previous study. This data set included effect sizes collected from a common phenomenon in special education which examines the relationship between creativity and one of the psychological disorders, psychoticism. Analyses with three different methods showed that results of two-level meta-analysis yielded the same results as traditional meta-analysis did. In addition, it was shown that how a three-level meta-analysis model can be used to solve dependency problem as an extension of two-level meta-analysis model. Necessary information about the new approaches developed in meta-analysis literature was provided for researchers in Turkey.

Keywords: Meta-analysis, multilevel meta-analysis, dependency, special education

GİRİŞ

Meta-analiz terimi ilk olarak Glass (1976) tarafından ortaya atılmıştır. Meta-analiz, belirli bir alanda yapılmış çalışmaların bulgularında yer alan özet istatistikleri kullanarak araştırma konusu ile ilgili genel durumu ortaya koyan ve son yıllarda uygulanmasında yaygın bir artış olan istatistiksel bir yöntemdir. Kısaca belirtmek gerekirse bir konu hakkındaki büyük resmi ortaya koymak için sistematik şekilde yapılan nicel bir incelemedir. Örneğin öğrencilerin cinsiyeti ile matematik dersindeki başarıları arasında bir ilişki olup olmadığını anlamak isteyen bir araştırmacı, bu iki değişkeni (cinsiyet ve matematik başarıları) içeren bütün çalışmaları araştırabilir. Böyle bir durumda belirtilen araştırma konusu içerisine giren çalışmalardan nicel araştırma yöntemleri kullanılmış olanlar seçilerek bu çalışmaların bir meta-analizi yapılır. Yapılan bu meta-analiz yardımıyla

* Yrd. Doç. Dr., Harran Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri, Şanlıurfa-Türkiye, e-posta: sedatsen@harran.edu.tr

** Arş. Gör., Harran Üniversitesi, Eğitim Fakültesi, Özel Eğitim, Şanlıurfa-Türkiye, e-posta: nihalakbass@hotmail.com

cinsiyetin matematik başarısı üzerinde ortalama ne kadar etkisi olduğu hakkında fikir edinilebilir. Meta-analiz bir nevi farklı çalışmaların tek bir çalışmada özetlenmesidir. İlk olarak farklı çalışmalardan elde edilen bulguları birleştirme işlemi bu çalışmalarda rapor edilen korelasyon katsayılarının özetlenmesine (Pearson, 1904), *p* değerlerinin birleştirilmesine (Fisher, 1932; Tippet, 1931) ya da bağımsız çalışmalardan elde edilen aritmetik ortalamaların ortalamasını almaya (Cochran, 1954) dayalı olarak yapılmaktaydı. Glass bu değerler yerine etki büyüklüğü (effect size) değerlerini kullanmayı önermiştir (Glass, 1976; 1977). Önceleri etki büyüklüğü terimi, standartlaştırılmış aritmetik ortalama farkı olarak tanımlansa da özellikle sosyal bilimlerde ilişkiisel araştırma yapan araştırmacılar korelasyon gibi istatistikleri etki büyüklüğü değeri olarak kullanmaya başlamıştır.

Glass'ın (1976) ortaya attığı ilk meta-analiz uygulamalarından bu yana özellikle son 20 yılda meta-analiz yöntemlerinde birçok gelişme yaşanmıştır. İlk olarak regresyon ve ANOVA gibi istatistiksel yöntemler kullanılarak yapılan meta-analizlerin arasına ilerleyen yıllarda çok değişkenli ve çok düzeyli istatistiksel modeller de katılmıştır. Meta-analizde karşılaşılan en önemli sorunlar arasında yayın yanlılığı (Begg, 1994) ve etki büyüklüklerinin bağımlı (dependency) olması durumları sayılabilir. Bu iki konuda araştırmacılar halen çözüm yolları aramaktadır.

Bir araştırmacının, araştırma konusu ile ilgili yapılmış ve ölçütlerine uygun olan bütün çalışmaları meta-analiz çalışmasında kullanması gerekir. Uygulamada karşılaşılan durum ise meta-analize yalnızca yayınlanmış ve manidar (significant) sonuçlar bulunmuş çalışmaların katılmasıdır. Yayınlanmamış ve manidar sonuçlar bulunmamış (non-significant) çalışmaların meta-analiz yapılırken göz ardı edilmesinden dolayı sonuçların yanlı olduğu düşünülmektedir. Bu problem meta-analiz alanyazınında yayın yanlılığı (publication bias, file-drawer problem) olarak bilinmektedir (Rosenthal, 1979). Yayın yanlılığını azaltmak için araştırma konusu ile ilgili yapılmış tezler ve raporlar dahil tüm çalışmalara ulaşmak gerekmektedir. Alanyazınında yayın yanlılığı sorununu tespit etmek ve çözüm sunabilmek için birçok yöntem geliştirilmiştir (Hedges, 1984, 1992; Dear ve Begg, 1992; Iyengar ve Greenhouse, 1988; Silliman, 1997). Bu sorunu tamamıyla ortadan kaldırmak olası gözükme de meta-analize başlamadan önce yayın yanlılığının çalışma sonuçlarını etkilemeyecek düzeyde olduğu ortaya konulmalıdır.

Meta-analiz alanyazınında üzerinde çözüm aranan başka bir problem de etki büyüklükleri arasında bağımlılık olmasıdır. Geleneksel meta-analiz yöntemleri etki büyüklükleri arasında bir bağımsızlık olduğunu varsaydığı için olası bir bağımlılık durumu, analizlerin sonuçları açısından problem oluşturabilir. Eğitim araştırmalarında sıklıkla gözlenen bağımlılık problemi birçok durumda ortaya çıkabilir (Scammacca, Roberts ve Stuebing, 2014). Örneğin aynı araştırma grubu tarafından yapılan farklı çalışmaların birbirleriyle bağlantılı olması beklenir. Benzer olarak aynı ülkede yapılan çalışmaların da birbirleriyle bağlantılı olabileceği söylenebilir (Van den Noortgate, López-López, Marín-Martínez ve Sánchez-Meca, 2013). Alanyazınında gösterildiği gibi (Van den Noortgate vd., 2013) farklı çalışmalar arasında bağımlılık olabileceği gibi aynı çalışma içerisindeki etki büyüklükleri arasında da bağımlılık olması söz konusudur. Örneğin meta-analiz yapılan konu hakkında bazı çalışmalar sadece bir etki büyüklüğü içerirken bazı çalışmalar birden fazla etki büyüklüğü içerebilmektedir. Eğer aynı çalışmadan gelen birden fazla etki büyüklüğü aynı örnekleme kullanıyorsa bu etki büyüklükleri birbirlerinden bağımsız olarak düşünülemez (Becker, 2000; Gleser ve Olkin, 1994). Bunun bir adım ötesi olarak bazı araştırmacılar (Littell, Corcoran ve Pillai, 2008) aynı çalışmadaki etki büyüklüklerinin farklı örneklemler kullandıklarında birbirinden bağımsız olarak kabul edilebileceğini öne sürse de aynı çalışma içerisinde farklı örneklem kullandığı halde rapor edilen etki büyüklüklerinin de bağımsızlık durumunu bozduğu belirtilmektedir. Birbirleriyle bağımlı olan etki büyüklükleri meta-analiz yaparken araştırmacıya çok bilgi sunmaz (Van den Noortgate vd., 2013). Örneğin birbirleriyle korelasyonu çok yüksek olan iki etki büyüklüğü ele alınırsa bu iki etki büyüklüğünü birlikte meta-analize eklemenin çok doğru olmadığı açıktır. Meta-analiz alanyazınında etki büyüklükleri arası bağımlılık sorununu aşmak için birçok yöntem önerilmiştir (Scammacca vd., 2014).

Bu çalışmada, yukarıda bahsedilen problemlerden bağımlılık konusu üzerinde durulmuş ve bu problemin etkilerini azaltmak için geliştirilmiş olan çok düzeyli meta-analiz yöntemlerinden

bahsedilmiştir. Çok düzeyli meta-analiz yöntemlerinden önce geleneksel anlamda uygulanan bir meta-analiz çalışmasının nasıl yapıldığından bahsetmenin yararlı olacağı düşünülmektedir.

Derleme çalışmalarının aksine meta-analiz araştırmasına konu olan çalışmaların nicel çalışmalar olması gerekmektedir. Meta-analize katılan çalışmaların ampirik çalışma olmaları, nicel veri üretmeleri, karşılaştırılabilir istatistikler içermeleri, aynı değişkenleri içermeleri ve bu değişkenler arası ilişkileri incelemeleri gibi özelliklere sahip olması gerekir (Lipsey ve Wilson, 2000). Geleneksel ve yeni meta-analiz yöntemleri bazı farklılıklar gösterse de genel olarak şu aşamalar takip edilerek uygulanmaktadır:

- Araştırmanın konusu ve problem durumu belirlenmelidir.
- Araştırma ölçütleri belirlenmelidir.
- Çalışmaların nasıl seçileceğine karar verilmelidir.
- Kullanılacak etki büyüklüğüne karar verilmelidir.
- Uygun istatistiksel analizler seçilmelidir.
- Varsa çalışma konusuna giren değişkenler belirlenmelidir.
- Yukarıda belirtilen işlemlerin detayları meta-analiz bulgularıyla beraber rapor edilmelidir.

Bu listede bahsedilen aşamalar hakkında daha fazla bilgi için okuyucuların Dinçer (2013) ile Lipsey ve Wilson (2000) kaynaklarına göz atmaları tavsiye edilir.

Meta-Analizde Yeni Yaklaşımlar

Eğer bir araştırma konusunda toplanan etki büyüklükleri arasında bağımsızlık varsayımına aykırı bir durum yoksa heterojenlik durumuna göre sabit etkiler modeli ya da rastgele etkiler modeli seçilerek yukarıda bahsedilen adımlar üzerinden bir meta-analiz yapılabilir. Eğer bir meta-analiz çalışmasında etki büyüklükleri arasında bir bağımlılık söz konusu ise (örneğin aynı çalışmadan elde edilen birden fazla etki büyüklükleri olduğunda) analizlere başlamadan önce bu problemle nasıl başa çıkılacağına karar verilmesi gerekir. Alanyazına bakıldığında etki büyüklükleri arasında yukarıda bahsedilen bağımlılık problemlerinden biri gözlemlendiğinde genel olarak üç yoldan birisi izlenmektedir (Van den Noortgate vd., 2013). Bu yollardan birincisi, bağımlılık problemini yok sayıp analizlere etki büyüklüklerinin bağımsız olduğunu varsayarak devam etmektir. Bu durum bulguların geçerliği açısından sorunludur ve tavsiye edilmez (Scammacca vd., 2014). Eğer çok az sayıda çalışma birden fazla etki büyüklüğüne sahipse bu durumda bağımsızlık varsayımı yapmak sorun olmayabilir (Hedges, 2007). Bu yolu tercih eden araştırmacının ek olarak hassaslık analizleri yapması gerekmektedir (Greenhouse ve Iyengar, 2009). İkinci yol olarak bağımlılık problemini ortadan kaldırma yolu tavsiye edilmektedir. Bu yolda aynı çalışmadaki birden fazla etki büyüklüğü sayısı her çalışma için bir etki büyüklüğü olacak şekilde ayarlama yapılır. Bunu yapabilmek için bazı araştırmacılar etki büyüklüklerinden birisini rastgele seçerek ya da tüm etki büyüklüklerinin ortalama değerini alarak analizlere devam etmektedir. Veri kaybına yol açtığı için bu ikinci yol da pek önerilmemektedir (Cheung, 2014; Marín-Martínez ve Sánchez-Meca, 1999; Scammacca vd., 2014). Son yol olarak bağımlılığı istatistiksel olarak modelleme yöntemi tavsiye edilmektedir (Hox, 2002). Sonuncu yöntem, genel olarak çok değişkenli (multivariate) modellerin ve çok düzeyli (multilevel) modellerin kullanımını içerir. Çok değişkenli model yaklaşımı ilk olarak Hedges ve Olkin (1985) tarafından ortaya atılmıştır. Daha sonra Raudenbush, Becker ve Kalaian (1988) tarafından geliştirilen çok değişkenli meta-analiz modellerinden biri olan çok değişkenli sabit etkiler regresyon (multivariate fixed effects regression) modeli uygulamada çok karmaşık oldukları için araştırmacılar tarafından çok fazla kullanılmamıştır (Van den Noortgate vd., 2013). Hedges ve Olkin (1985) ile Raudenbush vd. (1988) yaklaşımları örneklem varyans-kovaryans matrisini kullanarak bağımlılık problemini çözmeye çalışmışlardır. Bağımlı etki büyüklükleri ile başa çıkmak için son yıllarda çok düzeyli modellerin kullanıldığı yaklaşımlar geliştirilmiş (Hox, 2002; Hox ve de Leeuw, 2003) ve bu yaklaşımların araştırmacılar tarafından sıklıkla tercih edildiği gözlenmiştir. Bu yaklaşımlara göre geleneksel meta-analiz modelleri çok düzeyli modellerin özel bir durumu olarak

gösterilmiştir. Yani birinci düzeyde örneklem varyansı, ikinci düzeyde de çalışma düzeyi varyansı açıklanacak şekilde geleneksel meta-analiz verisi çok düzeyli modellerle analiz edilebilir duruma getirilmiştir. Takip eden kısımda bağımlılık problemini aşmak için bağımlı etki büyüklüklerinin istatistiksel olarak modellenmesine dayanan çok düzeyli meta-analiz yaklaşımlarından bahsedilmektedir.

Çok düzeyli meta-analiz

Birçok bilim alanında birbiri içerisinde yuvalanmış ya da kümelenmiş yapılar bulmak olasıdır. Örneğin eğitimde birçok öğrenci bir sınıf içerisinde, birçok sınıf bir okul içerisinde, okullar bölgeler içerisinde ve bölgeler de şehirler içerisinde kümelenmiştir için burada bir hiyerarşik yapıdan bahsedilebilir. Bu hiyerarşik yapı yüzünden aynı sınıfta olan öğrencilerin şartları benzer olduğu için aralarında bir bağımlılık söz konusudur. Örneğin öğrencilerin başarılarının karşılaştırıldığı bir araştırma durumunda sosyo-ekonomik düzeyi yüksek olan bölgelerdeki okulların öğrencileri ile sosyo-ekonomik düzeyi düşük olan bölge okullarındaki öğrencileri buldukları bölgeden bağımsızmış gibi düşünerek başarı açısından karşılaştırma yapılması ile bu düzeyler göz ardı edilmiş olur ve çalışmanın bulguları yanıltıcı bilgiler sunabilir. Sosyo-ekonomik düzeyi yüksek olan okul öğrencileri daha fazla olanaklara sahip olacağı için onların diğer bölge okullarına göre daha avantajlı olması beklenebilir. Bir bölgeyi kendi içerisinde, diğer bölgeleri de kendi içerisinde değerlendirip bir üst düzeyde de tüm bölgeler arası farklılığa bakma olanağı veren hiyerarşik doğrusal modeller (çok düzeyli modeller), bu tarz hiyerarşik yapıların bulunduğu durumların doğru bir şekilde analiz edilebilmesi için geliştirilmiştir (Bryk ve Raudenbush, 1992; Goldstein, 1987; Hox, 2002; Kreft ve de Leeuw, 1998; Raudenbush ve Bryk, 2002; Snijders ve Bosker, 1999). Bu modeller hem belirli bir zaman noktasındaki hiyerarşik (çok düzeyli) yapıları hesaba katmak için hem de birden fazla zaman noktasında elde edilen boylamsal (longitudinal) hiyerarşiye sahip verileri analiz etmek için kullanılabilir (Hedeker ve Gibbons, 2006; Singer ve Willett, 2003; Verbeke ve Mollenberghs, 2000).

Çok düzeyli modeller, verilerdeki bağımlılığı hesaba katmayan regresyon modellerine alternatif olarak daha güçlü ve daha doğru analizler sunabilmek için ortaya atılmış olup son yıllarda birçok alanda kullanılmaya başlanmıştır. Geleneksel meta-analiz yöntemleri daha çok regresyon (metaregresyon) ve ANOVA yaklaşımlarını kullanmakta (Cooper ve Hedges, 1994; Lipsey ve Wilson, 2000) ve etki büyüklüklerini bağımsız veriler olarak ele almaktadır. Yukarıda bahsedilen hiyerarşik yapıların olduğu durumlarda bu geleneksel yöntemleri kullanmak uygun olmayabilir. Bu nedenle araştırmacılar (Bryk ve Raudenbush, 1992; Hox ve de Leeuw, 2003; Raudenbush ve Bryk, 1985), hiyerarşik yapının sebep olduğu bağımlılık durumlarında çok düzeyli modelleri tercih etmişlerdir. Meta-analiz verileri hiyerarşik yapıda gösterilebildiği için bu alanda da çok düzeyli modeller kullanılabilir. Örneğin bir çalışma içerisinde birden çok etki büyüklüğü değeri elde etmek çok sık karşılaşılan bir durumdur. Bu durumlarda aynı çalışmadan elde edilen birden fazla etki büyüklüğünün birbirlerine bağımlı olduğu söylenebilir. Aynı çalışmalar içerisindeki etki büyüklükleri, o çalışmalara kümelenmiş (yuvalanmış) olarak düşünüldüğünde çok düzeyli model yaklaşımlarını meta-analiz için kullanmak çok doğru bir uygulama olmaktadır. Cheung'un (2014) çalışmasında bahsedildiği üzere meta-analiz yapılacak çalışmalar arasında bilinen bir bağımlılık varsa çok değişkenli yaklaşımlar kullanılırken bilinmeyen bir bağımlılık söz konusu olduğunda çok düzeyli modelleri kullanmak gereklidir (s. 211). Genelde çok düzeyli modellere ihtiyaç olup olmadığına sıfır ile bir arasında değerler alan sınıf içi korelasyon (intra-class correlation) değeri hesaplanarak karar verilir. Eğer elde edilen sınıf içi korelasyon değeri sıfır ise hiyerarşik düzeylerden kaynaklı sınıflar içi değişimden söz edilemez. Diğer bir deyişle bu değer sıfır çıktığı zaman bağımlılık söz konusu değildir. Eğer sınıf içi korelasyon değeri yüksek (1'e yakın) çıkarsa o zaman çok düzeyli model yaklaşımını kullanmamız uygun olur.

Yukarıda bahsedildiği gibi son yıllarda meta-analiz çalışmalarında çok düzeyli modeller sıklıkla kullanılmaya başlanmıştır ve geleneksel meta-analiz istatistikleri yerine de kullanılabilir hale gelmiştir (Kim, 2008). Bunun nedeni geleneksel meta-analiz ile iki düzeyli meta-analizin birbirine eşdeğer ortalama etki büyüklüğü tahmin değerini vermesidir. Bu durum, çalışmalar içerisinde birden

fazla etki büyüklüğü olması durumunda da aynıdır. Her ne kadar iki düzeyli meta-analiz geleneksel meta-analiz ile eşdeğer sonuçlar verse de çok düzeyli modellerin sunduğu birçok avantaj vardır. Çok düzeyli meta-analiz modelinin avantajları arasında bir çalışmadan elde edilen birden fazla etki büyüklüğünü analiz edebilmek, çalışmalarda ortak olan bazı değişkenleri modele ekleyebilmek, kayıp veri problemlerini daha etkili bir şekilde aşabilmek ve gerektiğinde iki düzeyli modele ek düzeyler ekleyebilmek vardır. Bu açılardan çok düzeyli meta-analiz modellerinin geleneksel meta-analiz modellerine göre daha esnek olduğu söylenebilir (Hox ve de Leeuw, 2003).

İki düzeyli rastgele etki modelinin ayrıntıları Hox (2002) tarafından gösterilmiştir. İki düzeyli meta-analiz modelinde birinci düzey katılımcıları, ikinci düzey ise çalışmaları temsil eder (Hox, 2002). İki düzeyli meta-analiz modelindeki ikinci düzey varyansı, etki büyüklüğü tahminlerinde çalışmalar arası farklılıkları yansıtmaktadır. Burada birinci düzeydeki varyans değerleri doğrudan etki büyüklüklerinden hesaplanmaktadır. Standart iki düzeyli hiyerarşik doğrusal modellerde ham veri kullanarak birinci ve ikinci düzeydeki varyans ve değişken katsayıları hesaplanabilir. Meta-analizlerde ham veriler yerine çalışmalardan elde edilen özet veriler (ortalama ve standart sapma) kullanıldığı için standart çok düzeyli modellerle çok düzeyli meta-analiz modelleri birbirinden ayrılmaktadır. Bu nedenle çok düzeyli meta-analiz modellerinde birinci düzeye ait varyans değerleri biliniyor olarak varsayılır ve bu değerler veride bulunan etki büyüklükleri ve örneklem büyüklükleri kullanılarak hesaplanır. Çok düzeyli modellerde ham veri kullanılırken çok düzeyli meta-analiz modellerinde ham veri yerine farklı çalışmalardan gelen etki büyüklükleri kullanılmaktadır. Bu nedenle farklı çalışmalardan gelen özet veriler, farklı ölçeklerde oldukları için etki büyüklük değerleri hesaplanarak aynı ölçek üzerine konulur.

Birinci düzey çalışma içi (within study) ve ikinci düzey ise çalışmalar arası (between study) olarak ele alınırsa iki düzeyli meta-analiz modeline ait eşitlikler aşağıdaki gibi sunulabilir.

Birinci düzey:

$$d_i = \pi_i + e_i, \quad (1)$$

ikinci düzey:

$$\pi_i = \beta_0 + r_i, \quad (2)$$

birleştirilmiş eşitlik:

$$d_i = \beta_0 + r_i + e_i \quad (3)$$

şeklinde gösterilebilir (Hox ve de Leeuw, 2003; Konstantopoulos, 2011; Van den Noortgate vd., 2013). Bu eşitliklerde $e_i \sim N(0, v_i)$, $r_i \sim N(0, \tau^2)$ olduğu ve v_i örneklem varyansı değerlerinin bilindiği varsayımları kabul edilmektedir. Burada d_i sembolü normal dağılım gösteren etki büyüklüklerini simgelemektedir. Bu etki büyüklüklerinin ortalaması π_i ve varyansı v_i olarak kabul edilir. İkinci düzeyde popülasyon parametresi genel bir ortalama (β_0) etrafında dağılım gösterir. Burada r_i çalışmalara özgü rastgele dağılımı gösterir (Konstantopoulos, 2011, s.8). Birleştirilmiş eşitlikle tek bir eşitliğe indirgenen bu model çok düzeyli model alanyazınında boş model (empty model) ya da koşulsuz model (unconditional model) olarak adlandırılır. Çalışmanın amacına göre bu modele bağımsız değişkenler (moderators) eklenebilir. Eğer meta-analiz çalışması üçüncü bir düzey gerektirirse ya da araştırmacı üç düzeyli bir meta-analiz modeli uygulamak isterse iki düzeyli modele üçüncü bir düzey eklenerek üç düzeyli meta-analiz yapılabilir. Konstantopoulos (2011) ve Van den Noortgate vd., (2013) iki düzeyli rastgele etki modelinin genişletilmiş hali olan üç düzeyli meta-analiz modelini önermişlerdir. Üç düzeyli meta-analiz modelinde üçüncü düzey sonuç düzeyi (outcome level) olarak adlandırılmıştır. Birinci düzey çalışmalar içi düzeyi, ikinci düzey çalışmalar arası düzeyi açıklamak için kullanılırken üçüncü düzey de etki büyüklüklerinin bir küme (cluster) içerisinde korelasyonuna izin vermektedir. Üç düzeyli meta-analiz modeli aynı çalışma içerisinde yer alan bağımlı etki büyüklüklerinin o çalışma içerisinde kümeleşmesine (toplaşmasına) izin vererek bir düzey oluşturur. Burada birinci düzey meta-analiz için toplanan etki büyüklükleri arası varyasyonu (yukarıda bahsedilen hesaplanan varyansı), ikinci düzey etki büyüklüklerinin kümeleştikleri

çalışmalar içerisindeki varyasyonu ve üçüncü düzeyde bağımlı etki büyüklüklerinin kümeleştiği farklı çalışmalar arasındaki varyasyonu açıklamak için kullanılır. Üç düzeyli modelde birinci düzey varyansı etki büyüklüklerinden hesaplandığı için ‘*düzye 1 bilinen varyans*’ olarak adlandırılır. Asıl tahmin edilmeye çalışılan çeşitlilik ikinci düzeyde ve üçüncü düzeyde olduğu için bilinen örneklem varyansına ek olarak Düzey-2 ve Düzey-3 varyansları tahmin edilir. Kısaca belirtmek gerekirse ikinci düzey varyansı çalışma içi varyansı (within-study variance) ve üçüncü düzey varyansı da çalışmalar arası varyansı (between-study variance) olarak adlandırılır. Üç düzeyli meta-analiz eşitlikleri aşağıdaki gibi verilebilir. Burada üç düzey olduğu için yalnızca *i* alt indisini kullanmak yerine *i* ve *g* alt indislerinin kullanılması gereklidir. *i* harfi ikinci düzey birimlerini temsil ederken *g* harfi de üçüncü düzey birimlerini temsil eder. Üç düzeyli meta-analiz eşitlikleri aşağıdaki gibi yazılabilir.

Birinci düzey:

$$d_{ig} = \pi_{ig} + e_{ig}, \quad (4)$$

ikinci düzey:

$$\pi_{ig} = \beta_{0g} + r_{0g}, \quad (5)$$

üçüncü düzey:

$$\beta_{0g} = \gamma_{00} + u_{0g}, \quad (6)$$

birleştirilmiş eşitlik

$$d_{ig} = \gamma_{00} + u_{0g} + r_{0g} + e_{ig} \quad (7)$$

şeklinde verilebilir (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate vd., 2013). Bu eşitliklerde $e_{ig} \sim N(0, v_i)$, $r_{0g} \sim N(0, \tau^2)$ olduğu ve v_i 'nin bilinen örneklem varyansı olduğu varsayımları kabul edilmektedir. İki düzeyli modelde olduğu gibi d_{ig} sembolü normal şekilde dağılım gösteren etki büyüklüklerini simgelemektedir. Örneğin d_{12} ikinci çalışma içerisindeki birinci etki büyüklüğü değerini simgeler. Bu etki büyüklüklerinin ortalaması π_{ig} ve varyansı v_i olarak kabul edilir. İkinci düzeyde popülasyon parametresi genel bir ortalama (β_{0g}) etrafında dağılım gösterir. Burada r_{0g} çalışmalara özgü rastgele dağılımı gösteren Düzey-2 varyansdır. Üçüncü düzeyde Düzey-3 birimleri genel bir ortalama (γ_{00}) etrafında dağılım gösterir. Burada u_{0g} sıfır aritmetik ortalamaya ve çalışmalar arası varyansa sahip olan Düzey-3 varyansdır. Eğer araştırmada ikinci düzey ve üçüncü düzeye ait değişkenler varsa bunlar da o düzeylerdeki heterojenliği açıklamak için eklenebilir (Cheung, 2014). Geleneksel meta-analizde olduğu gibi çok düzeyli meta-analizlerde de etki büyüklükleri arası heterojenliğin varlığını kontrol etmek yararlı olacaktır. Bunun için daha önce belirtilen Q istatistiği ve I^2 değeri kullanılabilir (Cheung, 2014). Üç düzeyli meta-analiz modelinde hem Düzey-2 için ($I^2_{(2)}$) hem de Düzey-3 için ($I^2_{(3)}$) I^2 değerleri hesaplayabiliriz (Cheung, 2014, s.215). Bu iki I^2 değeri Düzey-2 ve Düzey-3'teki sınıf içi korelasyon değerlerini hesaplamada kullanılabilir. Üç düzeyli meta-analiz hakkındaki ayrıntılı açıklamalar Konstantopoulos (2011), Van den Noortgate vd., (2013) ve Cheung (2014) çalışmalarında yer almaktadır. Bu üç çalışmada önerilen üç düzeyli modeller arasında bazı farklılıklar bulunmaktadır. Birincisinde (Konstantopoulos, 2011) önerilen üç düzeyli model Hox (2002) tarafından önerilen modelin genişletilmiş hali iken ikincisi (Van den Noortgate vd., 2013) çok değişkenli modellerin aksine tek değişkenli (univariate) bir üç düzeyli model önerisidir. Bunlardan sonuncusu (Cheung, 2014) çok düzeyli meta-analizin yapısal eşitlik yaklaşımı ile nasıl yapılacağı hakkında bilgiler sunmaktadır. Son on yıldır bu çok düzeyli meta-analiz modellerinin uygulanmasında fark edilebilir bir artış gözlenmektedir (Acar ve Sen, 2013; Bornmann, Mutz, Hug ve Daniel, 2011; de la Torre, Camilli, Vargas ve Vernon, 2007; Geeraert, van den Noortgate, Grietens ve Onghena, 2004; Kim, 2008; Lebeda, Zabelina ve Karkowski, (2016); Marsh, Bornmann, Mutz, Daniel, ve O'Mara, 2009; Van den Bussche, Wim Van den Noortgate ve Reynvoet, 2009; Yeager, Fong, Lee ve Espelage, 2015).

Meta-analiz alanyazınında yapılan çalışmalar incelendiğinde, son yıllarda çok düzeyli modellerin kullanımında artış olmasına karşın eğitim ve psikoloji alanlarında uygulanan çalışmaların çoğunda geleneksel meta-analiz yöntemlerinin kullanıldığı ve bağımsızlık varsayımının ihlal edildiği görülmüştür. Geleneksel meta-analiz ve çok düzeyli meta-analiz yöntemlerinin karşılaştırılmasının alana katkı sağlayacağı düşünülmektedir. Bu çalışmada geleneksel meta-analiz yöntemleri ile çok düzeyli meta-analiz yöntemlerinden elde edilen sonuçların karşılaştırılması amaçlanmaktadır. Bu doğrultuda çalışmada aşağıdaki sorulara yanıt aranmıştır.

1. Bağımsızlık varsayımını ihlal eden geleneksel meta-analiz yöntemiyle elde edilen sonuçlar nasıldır?
2. İki düzeyli meta-analiz ile geleneksel meta-analiz yöntemlerinden elde edilen meta-analiz sonuçları arasında nasıl bir ilişki vardır?
3. Üç düzeyli meta-analiz modeli ile elde edilen sonuçlar ile geleneksel ve iki düzeyli meta-analiz yöntemlerinden elde edilen sonuçlar arasında nasıl bir ilişki vardır?

Araştırmanın Önemi

Çok düzeyli modellerin kullanımı avantajları nedeniyle meta-analiz alanyazınında giderek artmaktadır. Alanyazına getirdiği katkılara karşın araştırmacıların çoğu bu modeller yerine geleneksel meta-analiz yöntemlerini uygulamaya devam etmektedir. Bunun başlıca sebeplerinden birisi bu modellerin uygulanmasının karmaşık olması ve bu modelleri anlaşılır bir şekilde anlatan çalışma sayısının az olmasıdır. Bu araştırmanın amacı meta-analiz uygulamalarında son yıllarda öne çıkan çok düzeyli meta-analiz yöntemlerini tanıtmak ve gerçek bir veri seti ile uygulama yapmaktır. Bu şekilde meta-analizde kullanılan en yeni yaklaşımların Türkiye'deki araştırmacılara tanıtılması sağlanmaktadır. Meta-analiz konusunda çalışma yapmak isteyen araştırmacılara yardımcı olacak bilgilerin sunulması sebebiyle çalışmanın önem taşıdığı düşünülmektedir.

YÖNTEM

Araştırmanın Modeli

Bu çalışma, meta-analiz uygulamalarında karşılaşılan bağımlılık sorununun çözümünde kullanılan çok düzeyli meta-analiz yöntemlerinin etkililiğinin incelendiği bir temel araştırmadır. Temel araştırmalar yeni bilgi elde etmek amacıyla yapılan teorik veya deneysel çalışmaları içerirken evrene genelleme amacı gütmemektedir (Karasar, 2007). Aynı zamanda temel araştırmalar bir konu ya da yöntem hakkındaki var olan bilgileri derleyerek, seçilen konu ya da yöntemi açıklama ve yorumlama amacı güden çalışmaları da içermektedir. Bu çalışmada da meta-analiz uygulamalarındaki geleneksel ve yeni yaklaşımlara ilişkin alanyazındaki bilgiler derlenmiş ve bir uygulama neticesinde açıklamalarda bulunulmuştur.

Veriler ve Verilerin Düzenlenmesi

Bu çalışmada daha önce Acar ve Runco (2012) tarafından toplanan ve geleneksel meta-analiz yöntemleriyle analiz edilen bir veri kullanılmıştır. Metodolojik çalışmalarda yeni bir yöntem geliştirildiğinde ya da bu yeni yöntem açıklandığında daha önce geliştirilmiş yöntemlerle analiz edilen verileri kullanarak analiz yapmak ve eski yöntem ile yeni yöntem arasında karşılaştırma yapmak çok yaygın bir uygulamadır (Cheung, 2014; Hox, 2002). Bu sebeple bu çalışmada da daha önce geleneksel bir yöntemle analiz edilmiş ve okuyucunun ulaşabileceği bir çalışmanın verisinin kullanılması tercih edilmiştir.

Bu çalışmada kullanılan veri özel eğitim alanında sıklıkla çalışılan bir yapı olan yaratıcılık üzerine yapılmış bir çalışmadan (Acar ve Runco, 2012) alınmıştır. Acar ve Runco (2012) çalışmalarında yaratıcılık (creativity) ve psikotizm (psychoticism) değişkenleri arasındaki ilişkiyi geleneksel meta-analiz yöntemiyle incelemek üzere ilgili alandaki yayınlanmış araştırmalara ulaşmaya çalışmışlardır.

Yaratıcılık ve çeşitli psikolojik rahatsızlıklar (örn: şizofreni) arasındaki ilişki birçok çalışmada gösterilmiştir (Bachtold, 1980; Chavez-Eakle, Lara ve Cruz-Fuentes, 2006; Cox ve Leon, 1999; Ludwig, 1992; Post, 1994). Bu ilişki için sıklıkla psikiyatristler tarafından Isaac Newton'a otizm teşhisi konulduğu üzerine söylentiler olması örnek gösterilen bir durumdur. Yapılan çalışmaların sayısının çokluğu bu konuda bir meta-analiz yapma ihtiyacı doğurmuştur. Bu gereksinimden yola çıkarak Acar ve Runco (2012) yaratıcılık ve psikolojik bir rahatsızlık olan psikotizm arasındaki ilişkiyi araştırmak üzere bir meta-analiz çalışması yapmışlardır. Bu çalışmada da aynı veri seti kullanıldığı için Acar ve Runco (2012) tarafından yapılan meta-analiz çalışmasındaki aynı etki büyüklük değerleri analiz edilmiştir. Acar ve Runco (2012) yaptıkları araştırma sonucunda yaratıcılık ve psikotizm konularını içeren toplamda 32 nicel çalışmanın meta-analize dahil edilmesine karar vermişlerdir. Bu çalışmalarda toplam katılımcı sayısı 6771 olarak rapor edilmiştir. Bu 32 çalışmanın bazıları bir etki büyüklüğü içerirken bazıları ise birden fazla etki büyüklüğü içermekte olup toplamda 119 etki büyüklüğü analizlere dahil edilmiştir. İki yapı arasındaki ilişki incelendiği için etki büyüklüğü olarak Pearson korelasyon katsayısı kullanılmıştır. Toplanan çalışmalarda ortak olarak sunulan sekiz farklı değişken belirlenmiş ve analizlere eklenmiştir. Bunlar yaratıcılık ölçeği, yaratıcılık test türü, yaratıcılık indisi, psikotizm ölçüsü ölçeği, alan (domain), örneklem özellikleri, cinsiyet ve yaş değişkenlerini içermektedir.

Verilerin Analizi

Acar ve Runco (2012) topladıkları meta-analiz verisinde bazı çalışmalarda birden fazla etki büyüklüğü olmasına karşın bu etki büyüklüklerini bağımsız kabul ederek geleneksel meta-analiz yöntemlerinden birisini (metaregresyon) kullanmışlardır. Acar ve Runco'nun (2012) kullandıkları veride 32 çalışmadan 119 etki büyüklüğü değeri elde edilmiş olması bu etki büyüklüklerinin çoğunun bağımlı olduğunu göstermektedir. Bu nedenle aynı veri bu çalışmada çok düzeyli meta-analiz yöntemleriyle analiz edilmiştir. Çok düzeyli analizler için etki büyüklüklerinin çalışmalar içerisinde yuvalanmış olduğu hiyerarşik bir yapı kullanılmıştır. Acar ve Runco'nun (2012) çalışmasında olduğu gibi bu çalışmadaki analizlerde de etki büyüklüğü olarak Pearson korelasyonunun dönüştürülmüş hali olan Fisher korelasyonu kullanılmıştır. Pearson korelasyonuna karar verilerek toplanan etki büyüklükleri normal dağılım göstermedikleri için analizler yapılırken Fisher korelasyonuna dönüştürülür (Hedges ve Olkin, 1985; Rosenthal, 1994) ve analizler dönüştürülmüş şekilde yapılır. Bu çalışmada Pearson korelasyon (r) değerlerinden Fisher korelasyon (Zr) değerlerini elde etmek için Excel programındaki FISHER formülü kullanılmıştır. Etki büyüklüğü olarak Fisher korelasyonu kullanıldığı için varyansın tersi olan $1/(N-3)$ değeri ağırlık olarak alınmıştır. Önce Acar ve Runco'nun (2012) çalışmasında yaptığı gibi meta-analiz verisi geleneksel meta-analiz modelleri ile analiz edilmiştir. Daha sonra geleneksel meta-analizle eşdeğer sonuçlar sunması beklenen iki düzeyli meta-analiz yöntemi aynı veri setine uygulanmıştır. Son olarak da verideki bağımlılık probleminin üstesinden gelebilmek için daha karmaşık olan üç düzeyli meta-analiz modeli uygulanmıştır. Üç düzeyli analizde Düzey-1 ham verisine sahip olmadığımız katılımcılara, Düzey-2 çalışmalar içerisindeki etki büyüklüklerine ($N=119$), Düzey-3 de çalışmalara ($n=32$) karşılık gelmektedir. Çok düzeyli analiz alanyazınındaki yaygın olan yaklaşımı takip ederek önce boş modeller (empty models), daha sonra da Acar ve Runco'nun (2012) çalışmasında kullandığı verideki değişkenlerden cinsiyet ve yaş değişkenlerinin bağımsız değişken olarak kullanıldığı ek modeller analiz edilmiştir. Cinsiyet değişkeni kadın, erkek ve karma olacak şekilde üç kategori halinde kodlanmıştır. Yaş değişkeni de veriyi toplayan araştırmacılar tarafından 30 yaş altı, 30 yaş üstü ve rapor edilmeyenler şeklinde üç kategori halinde kodlanmıştır. Bu değişkenler yapay (dummy) kodlama şeklinde analizlere dahil edilmiştir. Cinsiyet değişkeni için 'karma (kadın ve erkek)' kategorisi, yaş değişkeni için de 'belirtilmeyen yaş' kategorisi referans kategorisi olarak ele alınmıştır. Bulgular kısmında geleneksel yöntemle iki düzeyli meta-analiz bulguları arasında ve iki düzeyli meta-analiz ile üç düzeyli meta-analiz bulguları arasında karşılaştırmalar yapılmıştır. Geleneksel meta-analiz Lipsey ve Wilson'un (2000) ürettiği MEANES ve METAREG adlı SPSS makroları kullanılarak yapılmıştır. Çok düzeyli analizlerde Konstantopoulos'un (2011) önerdiği çok düzeyli analiz yaklaşımı ve aynı çalışmada sunulan SAS kodu kullanılmıştır. Analizler SAS programındaki *proc mixed* fonksiyonu aracılığıyla maksimum olabilirlik (maximum likelihood)

yöntemi kullanılarak yapılmıştır. Bu çalışmadaki çok düzeyli meta-analizler için kullanılan SAS kodları Ekler kısmında sunulmaktadır.

```
3.00 Extremes  (<=-.45)
 1.00  -3 . 2
 1.00  -2 . 5
 .00   -2 .
 4.00  -1 . 5668
 2.00  -1 . 33
 2.00  -0 . 67
 8.00  -0 . 11223334
11.00   0 . 00012222344
12.00   0 . 555666677889
16.00   1 . 001111222333444
10.00   1 . 5555666778
11.00   2 . 00000133344
11.00   2 . 5677777799
 8.00   3 . 02223334
 6.00   3 . 567778
 1.00   4 . 0
 3.00   4 . 578
 3.00   5 . 334
 1.00   5 . 6
5.00 Extremes  (>=.71)
```

Şekil 1. Fisher Korelasyonu (Z_r) Etki Büyüklük Değerlerini Gösteren Kök ve Yaprak Grafiği

BULGULAR

İlk olarak Acar ve Runco (2012) çalışmasında rapor edilen geleneksel meta-analizin tekrarı yapılarak aynı bulgular elde edilmeye çalışılmıştır. Şekil 1’de etki büyüklüklerinin dağılımını gösteren kök ve yaprak (stem and leaf plot) grafiği sunulmuştur. Verideki en küçük etki büyüklüğü değeri -.485 ve en büyük etki büyüklüğü değeri de .829 olarak bulunmuştur. Bu veriyi kullanarak MEANES (Lipsey ve Wilson, 2000) SPSS makrosu ile elde edilen sabit etkiler ortalama etki büyüklüğü değeri .139 ve rastgele etkiler modeline göre ortalama etki büyüklüğü değeri .161 çıkmıştır. Bu SPSS makrosu aynı zamanda heterojenlik testi olan Q istatistiğinin de elde edilmesini sağlamaktadır. Kullanılan meta-analiz verisindeki etki büyüklükleri ile hesaplanan $Q_{(118)}$ değeri 625.672 olarak hesaplanmış ve manidar bulunmuştur ($p < .001$). Bu değer sıfır hipotezi olan tüm etki büyüklüklerinin homojen olduğu hipotezini istatistiksel olarak reddetmektedir. Bu sonuç eldeki etki büyüklüklerinin heterojenlik gösterdiğini ve verilerin rastgele etkiler modeli seçilerek analiz edilebileceğini göstermektedir. Buraya kadar olan sonuçların hepsi Acar ve Runco’nun (2012, s.346) çalışmasındaki sonuçlar ile örtüşmektedir.

Q istatistiğine göre rastgele etkiler modelinin kullanılabilmesi belirlenmiş olup çalışmada yaş ve cinsiyet değişkenlerini kullanarak SPSS METAREG makrosu ile metaregresyon analizi yapılmıştır. Rastgele etkiler metaregresyon modeline göre bulunan sonuçlar Tablo 1’de sunulmuştur. MEANES sonuçlarında olduğu gibi ortalama etki büyüklük tahmin değeri .161 bulunmuş ve R^2 değeri .028 olarak hesaplanmıştır. Tablo 1’de sunulan değişkenlerin hiçbirisi 0.05 düzeyinde manidar bulunmamıştır.

Tablo 1. Rastgele Etkiler Modeli Sonuçları

| Değişkenler | Tahmin Değerleri |
|-------------|------------------|
| Sabit | 0.071 |
| Kadın | 0.005 |
| Erkek | -0.025 |
| 30 Yaş Üstü | 0.053 |
| 30 Yaş Altı | 0.111 |

Tablo 2. Çalışma Başına Düşen Etki Büyüklüğü Sayısı (n=32, N=119)

| Çalışma Numarası | Etki Büyüklüğü Sayısı | Çalışma Numarası | Etki Büyüklüğü Sayısı | Çalışma Numarası | Etki Büyüklüğü Sayısı | Çalışma Numarası | Etki Büyüklüğü Sayısı |
|------------------|-----------------------|------------------|-----------------------|------------------|-----------------------|------------------|-----------------------|
| 1 | 4 | 9 | 4 | 17 | 1 | 25 | 13 |
| 2 | 2 | 10 | 10 | 18 | 3 | 26 | 2 |
| 3 | 3 | 11 | 2 | 19 | 2 | 27 | 10 |
| 4 | 3 | 12 | 5 | 20 | 3 | 28 | 1 |
| 5 | 10 | 13 | 1 | 21 | 1 | 29 | 2 |
| 6 | 1 | 14 | 1 | 22 | 5 | 30 | 1 |
| 7 | 2 | 15 | 10 | 23 | 5 | 31 | 1 |
| 8 | 3 | 16 | 1 | 24 | 4 | 32 | 3 |

Tablo 2’de çalışma başına düşen etki büyüklüğü sayısı sunulmuştur. Çalışma başına düşen etki büyüklüğü sayısı en az 1 iken en fazla 13 olarak gözükmektedir. Tablo 2’de özetlenen bu meta-analiz verisi hiyerarşik yapısı sebebiyle çok düzeyli meta-analiz yapmaya uygun gözükmektedir. Bu sebeple geleneksel meta-analize ek olarak bu veri önce iki düzeyli daha sonra da üç düzeyli meta-analiz modeli ile analiz edilmiştir.

Tablo 3. Çok Düzeyli Koşulsuz Meta-Analiz Model Sonuçları

| Model Türü | İki Düzeyli Koşulsuz Model | Üç Düzeyli Koşulsuz Model |
|---------------------|----------------------------|---------------------------|
| Sabit etkiler | | |
| β_0 | 0.161 | 0.139 |
| Varyans Bileşenleri | | |
| İkinci Düzey | 0.042 | 0.016 |
| Üçüncü Düzey | - | 0.023 |

Tablo 4. Cinsiyet ve Yaş Değişkenlerinin Eklendiği Çok Düzeyli Meta-Analiz Model Sonuçları

| Model Türü | İki Düzeyli Model | Üç Düzeyli Model |
|---------------------|-------------------|------------------|
| Sabit etkiler | | |
| β_0 | 0.071 | 0.082 |
| Kadın | 0.005 | 0.034 |
| Erkek | -0.025 | 0.026 |
| 30 Yaş Üstü | 0.053 | 0.015 |
| 30 Yaş Altı | 0.111 | 0.067 |
| Varyans Bileşenleri | | |
| İkinci Düzey | 0.043 | 0.016 |
| Üçüncü Düzey | - | 0.024 |

Çok düzeyli analiz sonuçları Tablo 3’te ve Tablo 4’te sunulmuştur. Tablo 3’te koşulsuz yani hiçbir bağımsız değişkenin eklenmediği modellerin sonuçları sunulurken Tablo 4’te yaş ve cinsiyet değişkenlerinin de eklenmesiyle gerçekleştirilen çok düzeyli analiz sonuçları verilmiştir. Tablo 3’teki koşulsuz iki düzeyli meta-analiz sonuçlarından görüldüğü üzere iki düzeyli rastgele etkiler meta-analiz modeli ile geleneksel meta-analiz (rastgele etkiler modeli) modeli aynı ortalama etki büyüklüğü değerini ($\beta_0 = .161$) vermiştir. İki düzeyli model analizine göre ikinci düzey varyansı 0.042 olarak bulunmuştur (Bkz. Tablo 3). Bu sonuca göre, çok düzeyli meta-analiz modeli etki büyüklükleri arası değişimin çalışma düzeyindeki farklılıklardan kaynaklandığını göstermektedir. Q istatistiği kullanarak hesaplanan I^2 değeri 81.14 çıkmıştır ve çeşitliliğin yaklaşık yüzde 81’inin çalışma düzeyinde açıklandığını göstermektedir. Cinsiyet ve yaş değişkenleri eklenerek uygulanan iki düzeyli model analizlerine göre ortalama etki büyüklüğü değeri .071 olarak tahmin edilmiştir

(Tablo 4). Yine Tablo 4'te görüldüğü üzere ikinci düzey varyansı 0.043 çıkmıştır. Buna ek olarak cinsiyet ve yaş kategorilerine ait sabit etki (fixed effect) değerleri de Tablo 4'te sunulmuştur. Bu değerlerin hiçbirisi 0.05 düzeyinde manidar bulunmamıştır.

Tablo 3'te görüldüğü üzere, koşulsuz üç düzeyli meta-analiz modeli sonuçlarına göre ortalama etki büyüklüğü değeri .139 çıkmıştır. Tahmin edilen heterojenlik düzeyleri ikinci düzeyde 0.016 ve üçüncü düzeyde 0.023 olarak bulunmuştur. Bu varyans değerleri toplam çeşitliliğin Düzey-2'de %1.6 ve Düzey-3'te de %2.3 oranında açıklandığını, geri kalan %96 varyansın örneklem hatasından kaynaklandığını göstermektedir. Burada Düzey-3 varyansının Düzey-2 varyansından yüksek olduğu görülmektedir. Tablo 4'te cinsiyet ve yaş değişkenleri eklenerek yapılan üç düzeyli meta-analiz sonuçları verilmiştir. Bu sonuçların hiçbirisi 0.05 düzeyinde manidar bulunmamıştır.

SONUÇLAR ve TARTIŞMA

Bu çalışmada yaratıcılık ve psikotizm üzerine toplanmış 32 çalışma içeren bir meta-analiz verisinin üç farklı model ile analizleri yapılmıştır. Bu analizlerden birincisi daha önce de başka bir çalışmada (Acar ve Runco, 2012) yapılan geleneksel meta-analiz modelini, ikincisi iki düzeyli meta-analiz modelini ve sonuncusu da üç düzeyli meta-analiz modelini içermektedir. Bu analizlerin yapılmasındaki amaçlardan biri, geleneksel meta-analiz sonuçlarıyla hiyerarşik bir yapıda sunulabilen iki düzeyli meta-analiz sonuçlarının aynı çıktığını göstermektir. Bir başka amaç ise son yıllarda geliştirilen ve araştırmacılar tarafından sıklıkla kullanılmaya başlanan çok düzeyli meta-analiz modelleri hakkında bilgi sunulmasıdır.

Yapılan meta-analizlerle yaratıcılık ve psikotizm yapıları arasında nasıl bir ilişki olduğu sorusu cevaplanmaya çalışılmıştır. Geleneksel meta-analiz sonuçlarına göre yaratıcılık ve psikotizm arasında küçük düzeyde ($r = .161$) bir ilişki olduğu belirlenmiştir (Cohen, 1988). Alanyazında bahsedildiği gibi bu çalışmada da iki düzeyli meta-analiz sonuçları ile geleneksel meta-analiz sonuçları tutarlılık göstermektedir (Bateman ve Jones, 2003; Hox, 2002). Bu çalışmada üç düzeyli meta-analiz modeli bağımsızlık varsayımını sağlamak için kullanılmıştır. Üç düzeyli koşulsuz meta-analiz modeli de yaratıcılık ve psikotizm arasında düşük bir ilişki ($r = .139$) olduğunu göstermiştir. Bu sonucun geleneksel ve iki düzeyli meta-analiz sonuçlarından daha düşük olduğu görülmüştür. Koşulsuz modellere yaş ve cinsiyet değişkenleri eklenerek yapılan analizlere göre bu iki değişken de manidar bulunmamıştır. Bu sonuca dayanarak katılımcıların yaş ve cinsiyetlerinin yaratıcılık ve psikotizm arasındaki ilişkiyi etkilemedikleri söylenebilir. Ayrıca varyans tahmin değerlerine göre yaratıcılık ve psikotizm arasındaki varyasyonun (çeşitliliğin) %2'sinin ikinci düzeyde ve %3'ünün üçüncü düzeyde açıklandığı söylenebilir. Bu sonuçlar alanyazında sunulan yaratıcılık düzeyinin erkekler ve kadınların kişilik özelliklerine göre farklılık gösterdiği sonuçları ile tutarlılık göstermemektedir (H. J. Eysenck, 1993; H. J. Eysenck & S. B. G. Eysenck, 1976).

Meta-analiz uygulamalarında sıklıkla karşılaşılan problemlerden birisi de etki büyüklüklerinin birbirlerine bağımlı olmasıdır. Bu problemi aşmak için meta-analiz alanyazınında pek çok yöntem önerilmektedir. Bu yöntemler arasında istatistikçiler tarafından en çok tavsiye edileni meta-analiz verisinin çok düzeyli modellerle analiz edilmesi olmuştur (Bateman ve Jones, 2003; Hox, 2002; Hox ve de Leeuw, 2003). Geleneksel meta-analiz modellerinde kayıp veriye sahip çalışmalar veri setinden çıkarılarak analiz edilmektedir. Çok düzeyli modellerde kayıp olan kısımlar rastgele (missing at random) olduğu varsayılarak analizler eldeki mevcut verilerle yapılmaktadır. Bu gibi avantajlar çok düzeyli yöntemlerin tavsiye edilmesinde rol oynamaktadır. Yurtdışında giderek artan kullanımına karşın genelde çok düzeyli modellerin özelde de çok düzeyli meta-analiz modellerinin Türkiye'deki araştırmacılar tarafından pek fazla tercih edildiği söylenemez. Bunun başlıca sebepleri bu modellerin uygulanmasının üst düzey istatistik bilgisi gerektirmesi ve bu modelleri açıklayan Türkçe bir yayının mevcut olmaması görülmektedir. Bu çalışma ile hem meta-analizlerdeki bağımlılık problemi hakkında hem de bu problemle baş etme yolları hakkında bilgiler sunulmuştur. Aynı zamanda son yıllarda öne çıkan çok düzeyli modelleri bir veri analizi ile tanıtmak amaçlanmıştır. Alanyazında farklı meta-analiz yöntemlerinden elde edilen bulguları karşılaştıran çalışmalar (Cheung, 2014; Hox, 2002; Bateman ve Jones, 2003) oldukça azdır. Bu çalışmada yapılan

analizler neticesinde iki düzeyli meta-analiz sonuçları ile geleneksel meta-analiz sonuçlarının aynı çıktığı gösterilmiştir. Bu bulgular alanyazındaki bulgularla tutarlılık göstermektedir (Hox, 2002; Bateman ve Jones, 2003). Araştırmacılar, iki düzeyli model ile geleneksel modeller arasında tercih yaparken bu ikisi arasında uygulama ve yorumlama açısından daha kolay olan geleneksel modelleri tercih ediyor olsalar da sağladığı esneklikler düşünüldüğünde çok düzeyli meta-analizlerin uygulanması daha yerinde bir seçim olacaktır (Hox, 2002).

Bu çalışmada gösterildiği gibi iki düzeyli meta-analiz modeli kolaylıkla üç düzeyli hale getirilebilir. İki düzeyli modelin avantajlarını korumanın yanında üç düzeyli modellerin sağladığı diğer bir avantaj da üçüncü düzeyde çalışmalar arası korelasyona izin vererek çalışmalar arası varyansı hesaba katabilmesidir. Gerçek dünyada genelde açıklanamayan çalışmalar arası varyansa sahip olmak çok olasıdır (Hox ve de Leeuw, 2003). Bu sebeple çok düzeyli meta-analiz modellerinin ve özellikle de üç düzeyli meta-analiz modellerinin kullanılması daha doğru bir uygulama olacaktır. Üç düzeyli modeller çok yeni geliştirildiği için meta-analiz çalışmalarında hak ettiği yeri edinememiş olsa da alanyazında kullanımı giderek artmaktadır.

Belirli bir konuda meta-analiz yapmayı planlayan araştırmacılar için birçok öneride bulunulabilir. Meta-analize başlarken farklı modellerin farkında olan araştırmacının bir model üzerinde karar kılması gerekmektedir. Bu çalışmada bahsedilen üç farklı yöntemden hangisinin seçilmesi gerektiğinin cevabı araştırmacılara bırakılsa da toplanan etki büyüklükleri arasında bağımlılık problemi olması durumunda tercih edilmesi gereken yöntemin üç düzeyli meta-analiz modeli olduğu görüşü hakimdir (Cheung, 2014; Konstantopoulos, 2011). Eğer bir araştırmacı üç düzeyli meta-analiz modelini uygulamaya karar verirse yapısal eşitlik modeline dayanan yaklaşımı (Cheung, 2014) ya da hiyerarşik doğrusal model yaklaşımlarından (Konstantopoulos, 2011; Van den Noortgate vd., 2013) birini tercih edebilir. Kendi alanlarında olduğu gibi istatistiksel yöntemler konusunda da araştırmacıların bilgilerini güncel tutmaları çok önemlidir. Araştırmalarda sıklıkla tercih edilen yöntemlerden birisi olan meta-analiz yönteminde de bu durum geçerlidir. Diğer metodolojik gelişmelere paralel olarak meta-analiz yöntemlerinde de son yıllarda yeni yaklaşımlar geliştirilmiştir. Bunlar arasında çok değişkenli modellerin, çok düzeyli modellerin ve yapısal eşitlik modellerinin meta-analiz modellerine entegre edilmesi sayılabilir. Bu yaklaşımların araştırmacılara sağladığı yararlar sayesinde alanyazında karşılaşılan etki büyüklükleri arasındaki bağımlılık gibi problemleri çözmek mümkündür. Bu nedenle araştırmacıların çok düzeyli meta-analiz modelleri gibi yeni yaklaşımlar hakkında kendilerini güncellemeleri önerilir. Bu çalışmada yapıldığı gibi yeni metodolojik gelişmelerin anlaşılabilir bir dille Türkiye'deki araştırmacılara sunulması çok önemlidir. Bu çalışmada sunulan yaklaşımların Türkiye'deki araştırmacılar tarafından anlaşılıp kullanılması durumunda yapılan meta-analiz çalışmalarını güçlendirip geçerliliğini artırması olasıdır. Alanyazında geliştirilmiş olan birçok çok düzeyli meta-analiz modeli bulunmaktadır (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate vd., 2013). Bu karmaşık çok düzeyli modeller, analizlerin gerçekleştirilebilmesi için geleneksel meta-analiz modellerinden daha büyük örneklemelere ihtiyaç duymaktadır. Araştırmacılara bu farklı çok düzey yaklaşımlarının farklı veri koşullarında (farklı örneklem büyüklüğü ve farklı bağımlılık durumlarında) nasıl sonuçlar verebileceğini karşılaştırmaları önerilmektedir.

KAYNAKÇA

- Acar, S., & Runco, M. A. (2012). Psychoticism and creativity: A meta-analytic review. *Psychology of Aesthetics, Creativity and the Arts*, 6, 341–350.
- Acar, S., & Sen, S. (2013). A multilevel meta-analysis of the relationship between creativity and schizotypy. *Psychology of Aesthetics, Creativity and the Arts*, 7, 214–228.
- Bachtold, L. M. (1980). Psychoticism and creativity. *Journal of Creative Behavior*, 14, 242–248.
- Bateman, I. J. & Jones, A. P. (2003). Contrasting conventional with multi-level modelling approaches to meta-analysis: Exception consistency in UK woodland recreation values, CSERGE Working Paper EDM, No. 03-01.
- Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley and S. G. Brown (Eds.) *Handbook of Applied Statistics and Mathematical Modeling* (pp. 499-525). San Diego, CA: Academic Press.

- Begg, C. B. (1994). Publication bias. In H. Cooper & L. Hedges (Eds.), *Handbook of research synthesis* (pp. 399–409). New York, NY: Sage Publication.
- Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, 5(3), 346–359.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Chavez-Eakle, R. A., Lara, M. C., & Cruz-Fuentes, C. (2006). Personality: A possible bridge between creativity and psychopathology? *Creativity Research Journal*, 18, 27–38.
- Cheung, M. W.L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19, 211–229.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10, 101–129.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Cooper, H., & Hedges, L.V. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.
- Cox, A. J., & Leon, J. L. (1999). Negative schizotypal traits in the relation of creativity to psychopathology. *Creativity Research Journal*, 12, 25–36.
- de la Torre, J., Camilli, G., Vargas, S., & Vernon, R. F. (2007). Illustration of a multilevel model for meta-analysis. *Measurement and Evaluation in Counseling and Development*, 40, 169–180.
- Dear, K. B., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science*, 7, 237–245.
- Dinçer, S. (2013). *Meta-analize giriş*. Ankara: Anı Yayıncılık.
- Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry*, 4, 147–178.
- Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. London, England: Hodder & Stoughton.
- Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.). London: Oliver and Boyd.
- Geeraert, L., Van den Noortgate, W., Grietens, H., & Onghena, P. (2004). The effects of early prevention programs for families with young children at risk for physical child abuse and neglect. A meta-analysis. *Child Maltreatment*, 9, 277–291.
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3–8.
- Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351–379.
- Gleser, L. J. & Olkin, I. (1994). Stochastically dependent effect sizes. In *The handbook of research synthesis*, H. Cooper and L. V. Hedges (Eds.). New York: Russell Sage Foundation.
- Goldstein, H. (1987). *Multilevel models in education and social research*. London, UK: Griffen.
- Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In *The handbook of research synthesis*, H. Cooper and L. V. Hedges (Eds.). New York: Russell Sage Foundation.
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.
- Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational and Behavioral Statistics*, 9(1), 61–85.
- Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics*, 17, 279–296.
- Hedges, L. V. (2007). Meta-analysis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 919–953), Amsterdam, the Netherlands: Elsevier.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hox, J. J., & de Leeuw, E. D. (2003). Multilevel models for meta-analysis. In S. P. Reise & N. Duan (Eds.), *Multilevel modelling: Methodological advances, issues, and applications* (pp. 90–111). Mahwah, NJ: Lawrence Erlbaum.
- Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, 3, 109–117.
- Karasar, N. (2007). *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler*. (17. Baskı). Ankara: Nobel Yayın Dağıtım.
- Kim, J. S. (2008). Examining the effectiveness of solution-focused brief therapy: A meta-analysis. *Research on Social Work Practice*, 18, 107–116.
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis? *Research Synthesis Methods*, 2, 61–76.
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Lebuda, I., Zabelina, D. L., & Karwowski, M. (2016). Mind full of ideas: A meta-analysis of the mindfulness–creativity link. *Personality and Individual Differences*, 93, 22–26.

- Lipsey, M. W., & Wilson, D. (2000). *Practical meta-analysis*. Thousand Oaks: CA, SAGE Publications, Inc.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.
- Ludwig, A. M. (1992). Creative achievement and psychopathology: Comparison among professions. *American Journal of Psychotherapy*, 46, 330–356.
- Marín-Martínez, F., & Sánchez-Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *The Spanish journal of psychology*, 2, 32–38.
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79, 1290–1326.
- Pearson, K. (1904). Report on certain enteric fever inoculations. *British Medical Journal*, 2, 1243–1246.
- Post, F. (1994). Creativity and psychopathology. A study of 291 worldfamous men. *British Journal of Psychiatry*, 165, 22–34.
- Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, 103, 111–120.
- Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics*, 10, 75–98.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). London, UK: Sage.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Review of educational research*, 84(3), 328–364.
- Silliman, N. P. (1997). Nonparametric classes of weight functions to model publication bias. *Biometrika*, 84(4), 909–918.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.
- Snijders, T.A.B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Ltd.
- Tippett, L. H. C. (1931). *The methods of statistics*. London: Williams and Norgate.
- Van den Bussche, E., Van den Noortgate, W., Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, 135, 452–477.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45, 576–594.
- Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.
- Yeager, D. S., Fong, C. J., Lee, H. Y., & Espelage, D. L. (2015). Declines in efficacy of antibullying programs among older adolescents: A developmental theory and a threelevel meta-analysis. *Journal of Applied Developmental Psychology*, 37, 36–51.

EXTENDED ABSTRACT

Introduction

The meta-analysis term was first introduced by Glass in 1976. Meta-analysis can be defined as a collection of quantitative studies on a certain topic in order to make a summary of these research studies. Previously, a number of researchers have proposed to combine the results from quantitative studies including correlation, *p*-values and mean values. However, Glass (1976) has proposed to combine effect sizes from different quantitative studies. After his first meta-analysis model, extensions of this model have been developed using moderators that are particular to collected studies.

Dependency and publication bias have been two most common problems observed in meta-analysis applications. Publication bias is said to exist when a meta-analyst includes only published studies rather than reaching all of the studies on the research topic. Dependency problem is said to exist when the effect sizes are dependent on each other due to common studies. Namely, the dependency problem occurs when studies have more than one effect size. These multiple effect sizes nested

within a study seem to violate the independence assumption. In the presence of dependency problem, it would not be appropriate to use traditional meta-analysis as if effect sizes are independent. A number of methods have been proposed to deal with dependency problem in meta-analysis including ignoring the dependency and doing analyses as if they are independent, combining multiple effect sizes from a study into a single effect size (median or mean of all effect sizes) and modeling dependency using multivariate and multilevel models.

Multivariate and multilevel models have been integrated into meta-analysis models in order to deal with dependency problem that occurs due to multiple effect sizes within a study. Multivariate models have not been preferred much by researchers due to complexity. Popularity of multilevel models has increased since first introduction by Hox (2002). Hox proposed a two-level meta-analysis model as an alternative to traditional random effects meta-analysis model. It has been showed that meta-analysis data set can be shown with a two-level hierarchical structure. It has also been shown that two-level meta-analysis model yields the same results as traditional meta-analysis models. Although two-level modeling approach is more complex than traditional meta-analysis model, there are many advantages of multilevel approach over traditional one. Two-level model can be extended to three-level model in order to deal with dependency problem in meta-analysis (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate et al., 2013). Flexibility of multilevel models allows us to add moderators into the model and explain variability at within-study and between-study levels.

The use of multilevel models in meta-analysis plays an important role in solving dependency problem. Given that importance, their popularity has increased recently in meta-analysis literature. Main purpose of this study is to examine the approaches developed for solving the dependency problem that occurs in meta-analysis literature. Particularly, it was aimed to introduce the multilevel modeling approach that has been proposed as one the best ways to handle this problem.

Method

Three different methods (traditional meta-analysis, two-level and three-level meta-analysis models) were used to analyze a meta-analysis data set that has already been analyzed with a traditional meta-analysis model in a previous study. This data set included effect sizes collected from a common phenomenon in special education which examines the relationship between creativity and one of the psychological disorders, psychoticism. Totally 119 effect sizes were collected from 32 studies. It was obvious that there was a dependency problem in this data set because most of these 32 studies have multiple effect sizes. For this reason this meta-analysis data set was deemed appropriate for multilevel analyses. Before proceeding to multilevel analyses, a traditional method was used to analyze this data set as it was done in original study (Acar & Runco, 2012). In addition, the same data set was analyzed with a two-level meta-analysis model and a three-level meta-analysis model.

Results and Discussion

Analyses with three different methods showed that results of two-level meta-analysis yielded the same results as traditional meta-analysis did. This finding showed us that two-level model can be used instead of traditional random effects model in meta-analysis applications. In addition, a three-level meta-analysis model was used by adding an extra level to two-level model. It was shown that how a three-level meta-analysis model can be used to solve dependency problem as an extension of two-level meta-analysis model. As in other research areas, new approaches have been developing in the meta-analyses literature in order to provide solutions to some methodological problems or take the place of old approaches. It is researchers' responsibility to follow new research trends and developments in the methodology. To our knowledge, multilevel models have not attracted the attention of researchers in Turkey although their popularity has increased much in other countries. This study attempted to describe multilevel models within meta-analysis framework in a simple language to attract the attention of researchers in Turkey.

Ekler

Ek 1. Çok düzeyli Meta-Analizlerde Kullanılan SAS Program Kodları

Ek 1A. İki Düzeyli Koşulsuz Meta-Analiz Modeli SAS Kodu

```
proc mixed data=acarrunco covtest;
class esid;
model zr= / solution ddfm=bw notest;
random int / sub = esid;
repeated / group=esid;
parms (0.01)

(0.012987013) (0.012987013) (0.012987013) (0.012987013)
(0.011494253) (0.011494253) (0.008849558) (0.008849558)
(0.008849558) (0.010309278) (0.010309278) (0.010309278)
(0.013513514) (0.013513514) (0.013513514) (0.013513514)
(0.013513514) (0.010752688) (0.010752688) (0.010752688)
(0.010752688) (0.010752688) (0.027027027) (0.011111111)
(0.011111111) (0.020408163) (0.015151515) (0.004807692)
(0.00286533) (0.002012072) (0.00286533) (0.000904977)
(0.029411765) (0.029411765) (0.029411765) (0.029411765)
(0.029411765) (0.029411765) (0.029411765) (0.029411765)
(0.029411765) (0.029411765) (0.019607843) (0.019607843)
(0.010309278) (0.010309278) (0.010309278) (0.010309278)
(0.010309278) (0.016949153) (0.010309278) (0.027027027)
(0.027027027) (0.027027027) (0.027027027) (0.027027027)
(0.027027027) (0.027027027) (0.027027027) (0.027027027)
(0.027027027) (0.027027027) (0.015384615) (0.029411765)
(0.029411765) (0.029411765) (0.021276596) (0.021276596)
(0.022222222) (0.022222222) (0.022222222) (0.010309278)
(0.012987013) (0.012987013) (0.012987013) (0.012987013)
(0.012987013) (0.00456621) (0.004545455) (0.004504505)
(0.002985075) (0.008849558) (0.016949153) (0.035714286)
(0.016949153) (0.035714286) (0.008928571) (0.008928571)
(0.008928571) (0.008928571) (0.011111111) (0.011111111)
(0.011111111) (0.011111111) (0.010309278) (0.010309278)
(0.010309278) (0.010309278) (0.010309278) (0.002252252)
(0.002409639) (0.010309278) (0.010309278) (0.010309278)
(0.010309278) (0.010309278) (0.010309278) (0.010309278)
(0.010309278) (0.010309278) (0.010309278) (0.00390625)
(0.002583979) (0.001972387) (0.003773585) (0.00729927)
(0.002770083) (0.002770083) (0.002770083)
/eqcons=2 to 120; run;
```

Ek 1B. İki Düzeyli Model Kodu

```
proc mixed data=acarrunco covtest;
class esid;
model zr= yas cinsiyet / solution ddfm=bw notest;
random int / sub = esid;
repeated / group=esid;
parms (0.01)
```

BURAYA EK1.'DEKİ KODDA PARANTEZLER İÇERİSİNDE VERİLEN 119 ÖRNEKLEM VARYANSI EKLENMELİDİR.

```
/eqcons=2 to 120; run;
```

Ek 1C. Üç Düzeyli Koşulsuz Model SAS Kodu

```
proc mixed data=acarrunco;  
class studyID esid;  
model zr= / solution ddfm=bw notest;  
random int / sub = studyID;  
random int / sub = esid(studyID);  
repeated / group=esid(studyID);  
parms (0.01) (0.01)
```

BURAYA EK1.'DEKİ KODDA PARANTEZLER İÇERİSİNDE VERİLEN 119 ÖRNEKLEM VARYANSI EKLENMELİDİR.

```
/eqcons=3 to 121; run;
```

Ek 1D. Üç Düzeyli Model SAS Kodu

```
proc mixed data=acarrunco;  
class studyID esid;  
model zr= yas cinsiyet/ solution ddfm=bw notest;  
random int / sub = studyID;  
random int / sub = esid(studyID);  
repeated / group=esid(studyID);  
parms (0.01) (0.01)
```

BURAYA EK1.'DEKİ KODDA PARANTEZLER İÇERİSİNDE VERİLEN 119 ÖRNEKLEM VARYANSI EKLENMELİDİR.

```
/eqcons=3 to 121; run;
```

Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Bazı Değişkenler Açısından İncelenmesi*

Examination of Teachers' Self-Efficacy Perceptions on Educational Measurement and Evaluation in Terms of Some Variables

Gökhan BAŞ **

Ömer BEYHAN ***

Öz

Bu çalışmanın temel amacı, ilköğretim okulu ve liselerde görev yapan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarını cinsiyet, mesleki kıdem, eğitim durumu ve görev yapılan öğretim kademesi değişkenlerine göre incelemektir. Çalışma, Niğde il merkezinde bulunan toplam altı ilköğretim okulunda ve dört lisede gerçekleştirilmiştir. Araştırmaya, amaçsal örnekleme yöntemine göre seçilen öğretmenler (n = 211) katılmış olup, araştırmada “tarama modeli” kullanılmıştır. Araştırmanın amacına dayalı olarak, yüzde, standart sapma, bağımsız gruplar *t* testi, tek yönlü varyans (ANOVA) analizi istatistik teknikleri kullanılmıştır. Çalışmada veri toplamak için “Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algı Ölçeği” kullanılmıştır. Çalışmada elde edilen sonuçlara göre, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının gerek bilgi gerekse de beceri boyutunda düşük düzeyde olduğu saptanmıştır. Bununla birlikte, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının cinsiyete, mesleki kıdeme, eğitim durumuna ve görev yapılan öğretim kademesi değişkenlerine göre anlamlı bir şekilde farklılaştığı sonucuna ulaşılmıştır.

Anahtar Kelimeler: Eğitimde ölçme ve değerlendirme, özyeterlik algısı, lise, ilköğretim okulu, öğretmenler.

Abstract

The main purpose of this study was to analyse elementary and high school teachers' self-efficacy perceptions on educational measurement and evaluation with respect to gender, occupational experience, educational level and instruction stage. The study was carried out in the city centre of Niğde province from six elementary schools and four high schools. Teachers (n = 211), selected according to the purposive sampling method, participated in the study. The “survey method” was adopted in the research. In accordance with the purpose of the study, percentage, mean, standard deviation, independent samples *t* test, one-way ANOVA (variance) analyses were employed in the study in order to examine the data obtained in terms of some variables. “The Scale for Measurement and Evaluation Self-Efficacy in Education” was used in order to collect data for the study. According to findings of the study, it was found that teachers participated in the study had low level of self-efficacy in educational measurement and evaluation both in knowledge and skill sub-dimensions. It was also found that teachers' self-efficacy perceptions on educational measurement and evaluation statistically differed in relation to gender, occupational experience, educational level, and instruction stage which teachers work in variables.

Key Words: Educational measurement and evaluation, self-efficacy perception, high school, elementary school, teachers.

GİRİŞ

Eğitimde temel amaç, bireyde istendik yönde davranış değişikliği meydana getirmek ve nihai olarak bireyi topluma kazandırmaktır (Ertürk, 1972). Bu amaç doğrultusunda; bilgiyi tüketen değil, üretip

* Bu çalışma, Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi'nde düzenlenen “III. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi”nde (19-20 Eylül 2012) sözlü bildiri olarak sunulmuştur.

** Yrd. Doç. Dr., Niğde Üniversitesi, Eğitim Fakültesi, Niğde-Türkiye, e-posta: gokhanbas51@gmail.com

*** Doç. Dr., Necmettin Erbakan Üniversitesi, Eğitim Fakültesi, Konya-Türkiye, e-posta: obeyhan@konya.edu.tr

yorumlayarak sürece etkin bir şekilde katılan bireyler yetiştirilmesi istenmektedir. Eğitimin bu belirlenen amaçları gerçekleştirip gerçekleştirmediğinin anlaşılması, öğrencilerde meydana gelen bu davranış değişikliğinin niteliği veya niceliğinin tespit edilmesini gerektirmektedir (Karaca, 2008). Öğrencilerin istedik davranışları kazanıp kazanmadığı, kazandıysa kazanma düzeyi ve yönü ölçme ve değerlendirme işlemiyle anlaşılmaktadır (Bayram, 2011).

İlgili alanyazında ölçme, kısaca, varlık veya olayların belli bir özelliğe sahip oluş derecelerini belirleme işlemi veya bir niteliğin gözlenip gözlem sonucu sayı ya da sembollerle gösterilmesi (Özçelik 1998; Turgut, 1988) olarak tanımlanırken, değerlendirme ise ölçme sonuçlarına göre zihinsel yargılarla öznel sonuçlar çıkarmak; ölçümlerden anlam çıkarmak ve ölçülen nesnelere hakkında bir değer yargısına ulaşmak (Turgut, 1988; Linn ve Gronlund 1995) olarak nitelendirilmektedir. Ölçme ve değerlendirme, genel anlamıyla, öğrencilerin nasıl öğrendiklerini bildirmek için kullanılan ve eğitimin etkililiği hakkında ilgili kişilere bilgi veren bir dönüt sistemidir (Yaman ve Karamustafaoğlu, 2011). Öğretme ve öğrenme sürecinde ölçme ve değerlendirme sonucunda elde edilen bilgiler, öğrenciler ve eğitim öğretim süreci ile ilgili birçok kararda veri olarak kullanılmaktadır (Airasian, 1994; Kubiszyn ve Borich, 2003). Ölçme ve değerlendirme süreci ile aynı zamanda eğitimde programların istenilen başarıyı gösterip göstermediği, öğrencilerden beklenen bilgi, beceri ve tutumların gelişip gelişmediği tespit edilmektedir. Bununla birlikte, ölçme ve değerlendirme ile eğitim öğretim sürecinin sürekli izlenmesi ile her aşamada ortaya çıkan sorunları tespit ve düzenlenme imkânı da sağlanmaktadır (MEB, 2009). Bu bağlamda, uygulanan eğitim programlarının etkililiğini ortaya çıkarmak için, öğrencileri sürecin basında, süreç sırasında ve sürecin sonunda değerlendirilerek belirlenen hedef ve davranışların ne kadarına ulaşabildiklerini saptamak ve eksiklikleri ortaya çıkarmak için öğretmenlerin sahip olması gereken yeterlik alanı eğitimde ölçme ve değerlendirme özyeterlikleridir (Atılgan, Kan ve Doğan, 2007).

Eğitimde ölçme ve değerlendirme uygulamalarının niteliği büyük ölçüde, öğretmenlerin bu alandaki özyeterliliğine bağlıdır (Çakan, 2004; Pektaş, 2010; Bayram, 2011; Kılınç, 2011). Ölçme ve değerlendirme konusunda sahip olunan yeterlik; ölçme ve değerlendirme faaliyetlerinin her aşamasında yöntem, teknik ve ilkelerin yerinde ve zamanında kullanılmasına bağlıdır. Bu niteliğin artırılması için öğretmenlerin ölçme ve değerlendirme konusunda yeterli eğitimi almış olmaları ve bu bilgileri etkin bir şekilde kullanabilmeleri gerekmektedir (Erdemir, 2007). Öğretmenlerin ölçme ve değerlendirme konusunda amaca uygun test türünü seçme, ölçülecek hedefi tespit etme, testlerin güvenilirlik ve geçerliği, temel test istatistikleri, test puanlarının nasıl kullanılması gerektiği, notların yararlı olacak şekilde kullanılması ve veliye etkili bir iletişim tekniği kullanarak iletilmesi bilgi ve becerilerine sahip olması beklenmektedir (Daniel ve King 1998; Kubiszyn ve Borich, 2003). Öğretmenlerin ölçme ve değerlendirmeyi gerektiği şekilde uygulayabilmesi, bu alanda becerilerini geliştirmiş ve olumlu tutumlar kazanmış, öğrenci başarısını izleme ve değerlendirme yeterliliğine sahip, yeterli düzeyde ölçme ve değerlendirme bilgi ve becerisiyle donanımlı olmaları; öğrenci başarısını değerlendirme faaliyetlerinin daha nitelikli olması, eğitimin kalitesinin yüksek olması genel olarak meslek yaşantıları için bir zorunluluk ortaya çıkmaktadır (Pilten, 2001; Karaca 2003; Pektaş, 2010). Öğretmenlerin sahip oldukları eğitimde ölçme ve değerlendirmeye ilişkin yüksek özyeterlik algısının sınıftaki öğretimi önemli ölçüde etkilediğini göstermektedir (Kılınç, 2011; Yavuz, 2011). Diğer yandan, ölçme ve değerlendirme yeterliklerini kazanmış bir öğretmenle, bu bilgileri kazanmamış bir öğretmen arasında önemli farklılıklar bulunmaktadır. Ölçme ve değerlendirme tekniklerine uyarak not veren bir öğretmen hem daha az hatalı yargılara ulaşmış olmakta, hem de kendi öğretim yöntemlerini değerlendirip geliştirebilmektedir. Çünkü iyi bir ölçme, ders programının ve öğrenme-öğrenme sürecinin, ayrıntılarına kadar analiziyle başlamaktadır. Zira, böylesi bir analiz öğretmenlerin kendi öğretme ve öğrenme etkinliklerini geliştirmesine yardım edebilmektedir (Turgut, 1988). Özellikle, öğretmenlerin bu anlamda eğitimde ölçme ve değerlendirmeye ilişkin yüksek özyeterlik algısına sahip olması oldukça önemli görülmektedir. Nitekim eğitimde ölçme ve değerlendirmenin etkili bir şekilde uygulanabilmesi için öğretmenlerin öğrencilerin gerek hazırbulunuşluk düzeylerini gerek öğrenme eksikliklerini belirlemek, gerekse program sonunda öğrencilerin ulaştığı öğrenme düzeyini saptamak için ölçme aracı geliştirmeleri gerekmektedir. Ölçme araçları, eğitim sistemine ve uygulamalarına ilişkin verilen çeşitli kararların

yerinde ve isabetli olması için gereken bilgileri ortaya koyması bakımından oldukça önemlidir. Ölçme araçlarının etkin bir biçimde kullanılabilmesi için öğretmenlerin ölçme değerlendirme yöntem ve tekniklerini bilmeleri gerekmektedir (MEB, 2009; Yavuz, 2011). Nitekim öğretmenlerin eksik ya da yetersiz bilgi ve beceriye sahip oldukları ölçme ve değerlendirme tekniklerini uygulamaları sakıncalı sonuçlar doğurabilir. Faydalı bir ölçme ve değerlendirme tekniği yanlış şartlarda kullanıldığı ya da doğru zamanda yanlış kullanıldığı için fayda yerine zarara neden olabilmektedir. Bu, en başta öğrenci olmak üzere, öğretim programına, öğretim ortamına yönelik yanlış ya da eksik kararların alınması şeklinde gerçekleşebilmektedir. Bu alandaki eksiklerin ve ihtiyaçların saptanması, gerekli önlemlerin alınması ve dolayısıyla eğitimin kalitesinin artırılması, öğretmenlerin kendilerini geliştirmelerine olanak sağlanması için öğretmenlerin eğitimde ölçme ve değerlendirme uygulamaları düzenli olarak izlenmeli ve eksiklikleri gidermeye yönelik tedbirler alınmalıdır (Çakan, 2004). İlgili alanyazın gözden geçirildiğinde, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının incelendiği çalışmaların sayısının oldukça az olduğu görülmektedir (örn., Aydın, 2001; Çakan, 2004; Karaca, 2004; Birgin, 2007; Gelbal ve Kelecioğlu, 2007; Birgin ve Gürbüz, 2008; Buldur, 2009; Pektaş, 2010; Yavuz, 2011). Ayrıca, yapılan bu çalışmaların da genel olarak öğretmen adayları üzerinde gerçekleştirildiği anlaşılmaktadır (örn., Karaca, 2004; Birgin, 2007; Birgin ve Gürbüz, 2008; Buldur, 2009; Pektaş, 2010; Yavuz, 2011). Hali hazırda, görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik yeterlik algılarının incelendiği çalışmaların oldukça az sayıda olması (örn., Aydın, 2001; Çakan, 2004; Gelbal ve Kelecioğlu, 2007), böylesi bir araştırmaya olan ihtiyacı daha da artırmaktadır. Bu anlamda, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının saptanmasına yönelik daha fazla sayıda araştırmanın yapılması oldukça önemli görülmektedir. Zira görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının saptanması, okullarda daha sağlıklı ölçme ve değerlendirme sürecinin tesis edilmesi için oldukça hayattır. Öğretmen adaylarının eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının belirlenerek, burada görülen eksikliklerin düzeltilmesine dönük eğitimlerin zamanında verilmesi oldukça önemli olmakla birlikte bu konuda yapılan çalışmalar ise manidardır. Ancak, özellikle eğitim sistemi içerisinde görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının belirlenmesi de çok büyük bir önem taşımaktadır. Nitekim, sistem içerisinde 850 bin 960 kişilik bir öğretmen kadrosu görev yapmaktadır (bkz. MEB, 2015). Bu kadar geniş bir kitleye sahip öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının belirlenerek, gerekli önlemlerin alınması ve bu noktada eğitimden geçirilmeleri oldukça önemlidir. Görev yapmakta olan öğretmenlerin böylesi bir analizin yapılması, dolayısıyla sistemin ölçme ve değerlendirme ögesinin sağlıklı bir biçimde işlememesine yol açabileceği göz önünde bulundurulmalıdır. Sonuçta, alanyazında görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının belirlenmesine dönük çalışmaların oldukça az sayıda olması böylesi bir çalışmaya olan ihtiyacı bir kat daha artırmıştır. Buradan hareketle, Milli Eğitim Bakanlığı (MEB) içerisinde görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının belirlenmesine karar verilmiştir.

Araştırmanın Amacı

Yapılan bu araştırmanın amacı, ilköğretim okulları ve liselerde görev yapmakta olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarını incelemektir. Bu sebeple araştırmanın problem cümlesini ise, “öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları hangi düzeydedir?” sorusu teşkil etmektedir. Araştırmanın problemine cevap arayabilmek amacıyla aşağıdaki alt-problemlere çalışmada yanıt aranacaktır:

1. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları ne düzeydedir?
2. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları cinsiyet, mesleki kıdem, eğitim durumu ve görev yapılan öğretim kademesi değişkenlerine göre anlamlı bir farklılık göstermekte midir?

YÖNTEM

Bu araştırmada, Milli Eğitim Bakanlığına (MEB) bağlı ve Niğde il merkezinde bulunan ilköğretim okulları ve liselerde görev yapan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları bazı değişkenler açısından karşılaştırılmıştır. Bu nedenle, yapılan bu araştırma “tarama modeli”nde (Fraenkel ve Wallen, 2009) bir araştırmadır. İlgili alanyazında tarama modeli, çok sayıda elemandan oluşan bir evrende, evren hakkında genel bir yargıya varmak amacıyla, evrenin tümü ya da ondan alınacak bir grup, örnek veya örneklem üzerinde yapılan tarama düzenlemesi olarak nitelendirilmektedir (Karasar, 2005).

Örneklem

Araştırmanın çalışma grubunu Niğde il merkezinde bulunan altı resmi ilköğretim okulunda ve dört lisede çalışmakta olan öğretmenler (n = 211) oluşturmuştur. Çalışma grubunun seçiminde, amaçsal örnekleme çeşitlerinden maksimum çeşitlilik yöntemi benimsenmiş; evrenin temsil yeteneği göz önünde bulundurularak okulların seçilmesinde üst, orta ve alt sosyo-ekonomik düzeye sahip öğrencilerin devam ettiği okullarda görev yapmakta olan öğretmenler çalışma grubuna seçilmiştir (McMillan ve Schumacher, 2006).

Araştırmaya katılan öğretmenlerin demografik nitelikleri incelendiğinde; öğretmenlerin %47.40'ının (n = 100) kadın, %52.60'ının (n = 111) ise erkek, %14.22'sinin (n = 30) 1-5 yıllık, %42.65'inin (n = 90) 6-10 yıllık, %28.44'ünün (n = 60) 11-15 yıllık ve %14.69'unun (n = 31) ise 16 yıl ve üzeri mesleki kıdeme sahip olduğu, %13.75'inin (n = 29) lisans tamamlama, %77.73'ünün (n = 164) lisans ve %8.53'ünün (n = 18) ise lisansüstü eğitim mezunu olduğu, %56.87'sinin (n = 120) ilköğretim okullarında ve %43.13'ünün (n = 91) ise liselerde görev yapıyor oldukları görülmüştür.

Veri Toplama Araçları

Araştırmada veri toplama aracı olarak “Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algı Ölçeği” kullanılmış olup, bu ölçeğe ilişkin bilgiler çalışmada verilmiştir.

Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algı Ölçeği

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarını belirleyebilmek amacıyla Kılınç (2011) tarafından geliştirilen “Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algı Ölçeği” kullanılmıştır. Çalışma grubunu farklı üniversitelerin eğitim fakültelerinde öğrenim görmekte olan 764 öğretmen adayı oluşturmuştur. Performansların beceri ve bilgi bileşenlerinden oluşması ilkesiyle; çalışma grubunun daha önceden Eğitimde Ölçme ve Değerlendirme dersini almış olmaları ve öğretmen adayı olarak belirli bir motivasyona sahip olmaları gerekliliğinden dolayı üçüncü ve dördüncü sınıf öğrencilerinden oluşmuştur. 23 özyeterlik maddesi 5'li Likert tipi ölçme maddesine (1 = hiç katılmıyorum, 2 = katılmıyorum, 3 = kararsızım, 4 = katılıyorum, 5 = tamamen katılıyorum) dönüştürülüp deneme uygulaması yapıldıktan sonra geçerlik ve güvenilirlik çalışmaları gerçekleştirilmiştir. Geçerlik ve güvenilirlik sürecinde tek bir oturum ve tek bir ölçme aracı kullanıldığı için iç tutarlılık anlamında Cronbach alfa katsayısı hesaplanmıştır. Madde analizi ve yapı geçerliğinin belirlenmesi için de açılımlayıcı ve doğrulayıcı faktör analizine başvurulmuştur. Ölçekteki maddelerin faktör yük değerlerinin .36 ile .86 arasında değiştiği görülmüştür. Ayrıca, yapılan doğrulayıcı faktör analizi sonucunda ise ölçeğin uyum indekslerinin (GFI [Goodness of Fit Index] = .94; CFI [Comparative Fit Index] = .92; NNFI [Non-normed Fit Index] = .92; RMSEA [Root Mean Error of Approximation] = .049) kabul edilebilir düzeye sahip olduğu görülmüş ve yeterli kanıt edilmiştir. Ölçeğin genel Cronbach alfa katsayısı .96 olarak hesaplanmıştır. Alt boyutlar bazında ise; birinci alt boyut (bilgi) için Cronbach alfa katsayısı .93 ve ikinci alt boyut (beceri) için Cronbach alfa katsayısı ise .95 olarak hesaplanmıştır. Toplam 23 maddeye sahip olan ölçeğin bilgi boyutu 10 (örn., ölçme ile ilgili kavramları açıklayabilirim, ölçme aracının güvenilirliğinin önemini açıklayabilirim, vb.), beceri

boyutu ise 13 maddeden (örn., madde güçlük indisini hesaplayabilirim, ölçme sonuçlarının ortalamasını hesaplayabilirim, vb.) meydana gelmektedir (Kılınç, 2011).

Verilerin Analizi

Çalışma verilerinin analizinde öncelikle, ilköğretim okulu ve lise öğretmenlerinin “Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algı Ölçeği”ne (Kılınç, 2011) verdikleri yanıtlara ilişkin betimsel istatistikler hesaplanmıştır. Ardından, öğretmenlerin kişisel ve mesleki bazı özelliklerine göre, eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının farklılaşıp farklılaşmadığı incelenmiştir. İlgili analizler öncesinde verilerin normal dağılım gösterip göstermediği Kolmogorov-Smirnov Z testi işe koşulup kontrol edilerek, verilerin normal dağılım varsayımını karşıladıkları tespit edilmiştir ($K-S_z = .968, p > .05$). Buna göre, araştırmada parametrik test istatistiklerinin kullanılmasına karar verilmiştir. Cinsiyet ve görev yapılan öğretim kurumu değişkenlerine göre yapılan karşılaştırmalarda bağımsız gruplar için t-testi, ikiden fazla grup kategorisi karşılaştırmalarında ise tek yönlü varyans (ANOVA) analizi kullanılmıştır. Anlamlı çıkan F değerleri için, ortaya çıkan farkın kaynağını belirlemek üzere çoklu karşılaştırma testlerinden Scheffe testi kullanılmıştır. Huck’a (2000) göre, izleme testleri arasında en sıkı olanı Scheffe testidir. Akbulut (2010) ise, Scheffe testinin anlamsız olan bir farkın anlamlı görünmesine engel olmak, yani I. tip hatadan kaçınmak amacıyla yapılan en güçlü testlerden biri olduğu belirtmektedir (s. 125). İfade edilen bu gerekçeyle, yapılan çalışmada ikiden fazla grup kategorisi karşılaştırmalarında ortaya çıkan anlamlılığın hangi grup(lar)dan kaynaklandığını test etmek amacıyla Scheffe testinin kullanılmasına karar verilmiştir.

BULGULAR

Araştırmanın bu bölümünde, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları ile ilgili betimsel istatistiklere, ardından cinsiyet, mesleki kıdem, eğitim durumu ve görev yapılan öğretim kademesi değişkenlerine ilişkin karşılaştırmalara yer verilmektedir.

Betimsel İstatistikler

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarına ilişkin betimsel istatistikler Tablo 1’de sunulmaktadır.

Tablo 1. Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarına İlişkin Betimsel İstatistikler

| <i>Boyut</i> | <i>n</i> | \bar{X} | <i>SS</i> | Sh_x |
|--------------|----------|-----------|-----------|--------|
| Bilgi | 211 | 3.13 | 7.05 | 0.602 |
| Beceri | 211 | 2.82 | 8.74 | 0.485 |

Tablo 1’e bakıldığında, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarına ilişkin görüşleri bilgi boyutunda $\bar{X} = 3.13$ ($SS = 7.05$), beceri boyutunda ise $\bar{X} = 2.82$ ($SS = 8.74$) şeklinde gerçekleştiği görülmektedir.

Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Bazı Değişkenlere Göre Karşılaştırılması

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının bazı değişkenlere göre incelenmesi aşağıda başlıklar altında verilmiştir.

Cinsiyet

Tablo 2’de öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının cinsiyete göre karşılaştırılması ile ilgili bağımsız gruplar *t* testi sonuçlarına yer verilmiştir.

Tablo 2. Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Cinsiyete Göre Karşılaştırılması

| Boyut | Cinsiyet | <i>n</i> | \bar{X} | <i>SS</i> | <i>sd</i> | <i>t</i> | <i>p</i> | η^2 |
|--------|----------|----------|-----------|-----------|-----------|----------|----------|----------|
| Bilgi | Erkek | 111 | 3.02 | 6.60 | 209 | -1.939 | .054 | .028 |
| | Kadın | 100 | 3.25 | 7.36 | | | | |
| Beceri | Erkek | 111 | 2.74 | 8.96 | 209 | -2.447 | .015 | .018 |
| | Kadın | 100 | 2.92 | 8.37 | | | | |

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının cinsiyet değişkenine göre karşılaştırılabilmesi için bağımsız gruplar *t* testi yapılmış ve test sonuçları ise Tablo 2’de verilmiştir. Yapılan analiz sonucunda, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarına ilişkin görüşlerinin cinsiyet değişkenine göre beceri boyutunda kadın öğretmenlerin lehine ($t_{(209)} = -2.447, p < .05$) anlamlı farklılık yarattığı, bilgi boyutunda ise cinsiyet değişkenine göre anlamlı farklılık yaratmadığı ($t_{(209)} = -1.939, p > .05$) sonucuna ulaşılmıştır. Aynı zamanda, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının beceri boyutundaki farka ilişkin etki büyüklüğünün küçük düzeyde ($\eta^2 = .018$) olduğu görülmüştür.

Mesleki Kıdem

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının mesleki kıdem değişkeni açısından karşılaştırması yapılmıştır. Bu amaçla, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının mesleki kıdem değişkeni açısından karşılaştırılması için tek yönlü ANOVA analizi yapılmış ve analiz sonuçları Tablo 3’te verilmiştir.

Tablo 3. Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Mesleki Kıdeme Göre Karşılaştırılması

| Boyut | Mesleki Kıdem | <i>n</i> | \bar{X} | <i>SS</i> | <i>sd</i> | <i>F</i> | <i>p</i> | Fark | η^2 |
|--------|---------------|----------|-----------|-----------|-----------|----------|----------|------|----------|
| Bilgi | 1-5 yıl | 30 | 3.74 | 5.32 | 3-207 | 27.94 | .000 | 1-2 | .32 |
| | 6-10 yıl | 90 | 3.33 | 6.75 | | | | 1-3 | |
| | 11-15 yıl | 60 | 2.90 | 5.44 | | | | 1-4 | |
| | 16 yıl üstü | 31 | 2.41 | 3.82 | | | | | |
| Beceri | 1-5 yıl | 30 | 3.53 | 6.33 | 3-207 | 33.00 | .000 | 1-2 | .28 |
| | 6-10 yıl | 90 | 2.90 | 8.97 | | | | 1-3 | |
| | 11-15 yıl | 60 | 2.64 | 6.70 | | | | 1-4 | |
| | 16 ve üstü | 31 | 2.26 | 3.90 | | | | | |

Tablo 3’te öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının mesleki kıdem değişkeni açısından bilgi boyutunda ($F_{(3,207)} = 27.94, p < .05$) ve beceri boyutunda ($F_{(3,207)} = 33.00, p < .05$) farklılaştığı görülmektedir. Ortaya çıkan farklılığın hangi gruplardan kaynaklandığını tespit etmek amacıyla Scheffe testi yapılmıştır. Yapılan Scheffe testi sonucunda bilgi boyutunda en yüksek düzeyde eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algısına 1-5 yıllık mesleki kıdemdeki öğretmenler ($\bar{X} = 3.74, SS = 5.32$) sahipken, en az düzeyde eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algısına ise 16 yıl ve üzeri yıllık mesleki kıdemdeki öğretmenler ($\bar{X} = 2.41, SS = 3.82$) sahiptir. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının bilgi boyutundaki bu farka ilişkin etki büyüklüğünün Cohen’in (1988) sınıflamasına göre orta düzeyde ($\eta^2 = .32$) olduğu ve toplam varyansın %32’sini açıkladığı görülmüştür. Diğer taraftan, beceri boyutunda ise en yüksek düzeyde eğitimde ölçme ve

değerlendirmeye yönelik özyeterlik algısına yine 1-5 yıllık mesleki kıdemdeki öğretmenler ($\bar{X} = 3.53$, $SS = 6.33$) sahipken, en az düzeyde eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algısına ise 16 yıl ve üzeri mesleki kıdemdeki öğretmenler ($\bar{X} = 2.26$, $SS = 3.90$) sahiptir. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının beceri boyutundaki bu farka ilişkin etki büyüklüğü incelendiğinde, etki büyüklüğünün orta düzeyde ($\eta^2 = .28$) olduğu ve toplam varyansın %28'ini açıkladığı görülmüştür.

Eğitim Durumu

Aşağıda öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları eğitim durumu değişkenine göre tek yönlü ANOVA analizi ile karşılaştırılmış ve analiz sonuçları Tablo 4'te verilmiştir.

Tablo 4. Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Eğitim Durumuna Göre Karşılaştırılması

| Boyut | Eğitim Durumu | n | \bar{X} | SS | sd | F | p | Fark | η^2 |
|--------|---------------|-----|-----------|-------|-------|--------|------|------|----------|
| Bilgi | Lisans Tam. | 29 | 2.33 | 2.22 | 2-208 | 29.829 | .000 | 1-2 | .22 |
| | Lisans | 164 | 3.22 | 5.94 | | | | 1-3 | |
| | Lisansüstü | 18 | 3.64 | 15.54 | | | | | |
| Beceri | Lisans Tam. | 29 | 2.19 | 1.50 | 2-208 | 34.636 | .000 | 1-2 | .25 |
| | Lisans | 164 | 2.85 | 6.98 | | | | 1-3 | |
| | Lisansüstü | 18 | 3.56 | 11.46 | | | | | |

Tablo 4'te öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları eğitim durumu değişkenine göre varyans (ANOVA) analizi ile karşılaştırılmıştır. Öğretmenlerinin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları arasında eğitim durumu değişkenine göre bilgi boyutunda ($F_{(2,208)} = 29.829$, $p < .05$) ve beceri boyutunda ($F_{(2,208)} = 34.636$, $p < .05$) anlamlı farklılık olduğu görülmektedir. Ortaya çıkan bu farklılığın hangi gruplardan kaynaklandığı tespit etmek amacıyla yapılan Scheffe testi sonucuna göre, ortaya çıkan bu farklılığın 1-2 ve 1-3 gruplar arasında gerçekleştiği görülmektedir. Yapılan analizde ortaya çıkan anlamlı farklılığın her iki boyutta da lisansüstü eğitimi mezunu öğretmenlerin lehine olduğu saptanmış olup, lisansüstü eğitimi mezunu öğretmenlerin bilgi boyutunda ($\bar{X} = 3.64$, $SS = 15.54$) eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları lisans mezunu öğretmenlerin ($\bar{X} = 3.22$, $SS = 5.94$) ve lisans tamamlama mezunu öğretmenlerinkine ($\bar{X} = 2.33$, $SS = 2.22$) göre anlamlı olarak daha yüksek düzeyde bulunmuştur. Beceri boyutunda ise lisansüstü eğitimi mezunu öğretmenlerin ($\bar{X} = 3.56$, $SS = 11.46$) eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları lisans mezunu öğretmenlerin ($\bar{X} = 2.85$, $SS = 6.98$) ve lisans tamamlama mezunu öğretmenlerinkine ($\bar{X} = 2.19$, $SS = 1.50$) göre anlamlı olarak daha yüksek düzeyde olduğu saptanmıştır. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının bilgi boyutundaki bu farka ilişkin etki büyüklüğünün orta düzeyde ($\eta^2 = .22$) olduğu ve toplam varyansın %22'sini açıkladığı görülmüştür. Aynı zamanda, beceri boyutundaki farka ilişkin etki büyüklüğünün de aynı şekilde orta düzeyde ($\eta^2 = .25$) olduğu ve toplam varyansın %25'ini açıkladığı tespit edilmiştir.

Görev Yapılan Öğretim Kademesi

Aşağıda öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları görev yapılan öğretim kademesi değişkenine göre bağımsız gruplar t testi ile karşılaştırılmış ve analiz sonuçları Tablo 5'te verilmiştir.

Tablo 5. Öğretmenlerin Eğitimde Ölçme ve Değerlendirmeye Yönelik Özyeterlik Algılarının Görev Yapılan Öğretim Kademesine Göre Karşılaştırılması

| Boyut | Öğretim Kademesi | n | \bar{X} | SS | sd | t | p | η^2 |
|--------|------------------|-----|-----------|------|-----|-------|------|----------|
| Bilgi | İlköğretim | 120 | 3.31 | 7.26 | 209 | 6.474 | .000 | .168 |
| | Lise | 91 | 2.66 | 3.53 | | | | |
| Beceri | İlköğretim | 120 | 3.00 | 9.01 | 209 | 6.503 | .000 | .167 |
| | Lise | 91 | 2.38 | 4.39 | | | | |

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının görev yapılan öğretim kademesi değişkenine göre karşılaştırılabilmesi için bağımsız gruplar *t* testi yapılmış ve test sonuçları Tablo 5'te verilmiştir. Yapılan analiz sonucunda, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının görev yapılan öğretim kademesi değişkenine göre gerek bilgi boyutunda ($t_{(209)} = 6.474, p < .05$) gerekse de beceri boyutunda ($t_{(209)} = 6.503, p < .05$) ilköğretim öğretmenlerinin lehine istatistiksel olarak anlamlı bir farklılık yarattığı görülmüştür. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının gerek bilgi boyutundaki ($\eta^2 = .16$), gerekse beceri boyutundaki ($\eta^2 = .16$) farka ilişkin etki büyüklüklerinin orta düzeyde oldukları ve toplam varyansın %16'sını açıkladıkları görülmüştür.

SONUÇLAR ve TARTIŞMA

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının araştırıldığı bu çalışmada elde edilen bulgulara göre, öğretmenler eğitimde ölçme ve değerlendirmeye yönelik gerek bilgi boyutunda, gerekse de beceri (uygulama) boyutunda düşük özyeterlik algısına sahiptirler. Eğitimde ölçme ve değerlendirmeye yönelik özyeterlikleri üzerine yapılan çalışmaların hemen hemen hepsinde öğretmenlerin kendilerini yetersiz olarak algıladıkları görülmüştür (örn., Haynie, 1992; Aydın, 2001; Culbertson ve Wenfan, 2003; Karaca, 2003; Çakan, 2004; Karaca, 2004; Aydın, 2005; Begtaş, 2005; Flowers ve diğerleri, 2005; Cheng, 2006; Çakır ve Çimer, 2007; Doğan, Karakaya ve Gelbal, 2007; Güven ve Eskitürk, 2007; Gömleksiz ve Kan, 2007; Volante ve Fazio, 2007; Birgin ve Gürbüz, 2008; Numanoglu ve Bayır, 2009). Bu araştırmaların aksine Pektaş (2010) ve Yavuz (2011) öğretmen adayları üzerine yapmış oldukları araştırmalarda, öğretmen adaylarının ölçme ve değerlendirme alanında kendilerini genel olarak yeterli gördükleri saptanmıştır. Araştırmada elde edilen bulguya göre, öğretmenlerin ölçme ve değerlendirmede kendilerini bilgi düzeyinde beceri düzeyine nazaran daha yeterli algıladıkları bulunmuştur. Öğretmenler, her ne kadar bilgi düzeyinde beceri düzeyine nazaran daha yüksek özyeterlik algısına sahip olsalar da, genel itibari ile her iki boyutta da özyeterlik algıları düşük olarak saptanmıştır. İlgili alanyazından elde edilen bulgular, bu araştırmanın ilgili bulgusunu destekler niteliktedir. Örneğin, Birgin (2007) yapmış olduğu araştırmada, öğretmen adaylarının ölçme ve değerlendirme teknikleri ve yaklaşımları hakkında bilgiye sahip olsalar da, bunları kullanmaya yönelik yeterliklerinin düşük olduğu sonucuna ulaşmıştır. Yapılan bu araştırmanın ilgili bulgusuna paralel olarak, Culbertson ve Wenfan (2003), Erdemir (2007), Gömleksiz ve Kan (2007) ve Okur (2008), yaptıkları araştırmalarda, öğretmenlerin mezun oldukları kurumlarından ölçme ve değerlendirmeyle ilgili yeterli bilgi düzeyine sahip olmadıklarını tespit etmişlerdir. Bununla birlikte, Culbertson ve Wenfan (2003), Cheng (2006), Çakır ve Çimer (2007), Erdemir (2007) ve Gelbal ve Kelecioğlu (2007) yaptıkları araştırmalarda, öğretmenlerin ve öğretmen adaylarının ölçme ve değerlendirmenin beceri boyutunda da özyeterlik algılarının düşük olduğu bulgusuna ulaşmışlardır. Sonuçta, eğitimde ölçme ve değerlendirmede öğretmenlerin genel olarak özyeterlik algılarının düşük olduğu, ölçme ve değerlendirmenin gerek bilgi gerekse de beceri boyutundaki özyeterlik algıları düşük olmakla birlikte, öğretmenlerin bilgi boyutunda beceri boyutuna nazaran kendilerini daha yeterli hissettikleri anlaşılmıştır. Elde edilen bu bulgu, öğretmenlerin sınıf içi süreçleri, öğretme ve öğrenmeyi değerlendirmede ölçme ve değerlendirmeye bilimsel açıdan çok fazla önem vermediklerini gösterebilir. Bu bulgu, ayrıca, öğretmenlerin mezun oldukları okullarda eğitimde ölçme ve değerlendirmeye yönelik yeterli

düzeyde bilgi ve beceri ile mezun olmadıkları ile de açıklanabilir. Ayrıca, sınıfların kalabalık oluşu, ölçme ve değerlendirme uygulamalarının fazla zaman alıyor oluşu, vb. etkenler de öğretmenlerin böylesi düşük özyeterlik algısına sahip olmalarının sebepleri arasında olabilir.

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının araştırıldığı bu çalışmada elde edilen bir başka bulguya göre, öğretmenlerin ölçme ve değerlendirmeye yönelik özyeterlik algıları cinsiyet değişkenine göre bilgi boyutunda istatistiksel olarak anlamlı farklılık göstermese de, beceri boyutunda kadın öğretmenler lehine anlamlı farklılık göstermektedir. Ayrıca, gerek erkek gerekse de kadın öğretmenlerin eğitimde ölçme ve değerlendirmenin bilgi boyutundaki genel özyeterlik algıları arasında anlamlı bir farklılık olmasa da, aritmetik ortalamalar karşılaştırıldığında bunun yine kadın öğretmenler lehine olduğu görülmüştür. Yapılan bu araştırmanın ilgili bulgusu, kadın öğretmenlerin daha titiz oldukları ve öğrenci başarısını ve öğretme sürecini değerlendirmeye daha fazla önem vermeleri ile açıklanabilir. Bu araştırmanın ilgili bulgusundan farklı olarak Pektaş (2010) yapmış olduğu çalışmada erkek öğretmen adayları lehine anlamlı farklılık saptarken, Karaca (2004) ve Yavuz (2011) yapmış oldukları çalışmalarda öğretmen adaylarının ölçme ve değerlendirme alanındaki yeterliklerine ilişkin görüşlerinde cinsiyetin anlamlı bir farklılaşma oluşturmadığını saptamıştır.

Yapılan bu çalışmada elde edilen bir başka bulguya göre, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları bilgi ve beceri boyutlarında mesleki kıdem değişkenine göre mesleki kıdemi daha az olan genç öğretmenler lehine istatistiksel olarak anlamlı farklılık göstermektedir. Araştırmada elde edilen bu bulguya göre, mesleki kıdem azaldıkça öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları gerek bilgi gerekse de beceri boyutunda artmakta, mesleki kıdemi arttıkça bu konudaki özyeterlik algıları her iki boyutta da düşmektedir. Bu bulgu, mesleki kıdemi daha az olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının mesleki kıdemi daha fazla olan meslektaşlarından daha yüksek olduğunu göstermektedir. Bu bulgu, Aydın'ın (2005) yapmış olduğu araştırmanın bulgusu ile benzerlik göstermektedir. Aydın (2005) yapmış olduğu çalışmada, yeni öğretmenlerin, yani mesleki kıdemi düşük olan öğretmenlerin ölçme ve değerlendirmede daha yaşlı meslektaşlarına göre daha etkili olduğunu saptamıştır. Benzer bir şekilde, Haynie'nin (1992) ve Yayla'nın (2011) yapmış olduğu çalışmalarda da, mesleki kıdemi daha düşük olan öğretmenlerin mesleki kıdemi yüksek olan meslektaşlarına nazaran ölçme ve değerlendirmeye yönelik daha yüksek özyeterlik algısına sahip oldukları bulunmuştur.

Araştırmada elde edilen bir diğer bulguya göre, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları bilgi ve beceri boyutlarında eğitim durumu değişkenine göre lisans ve lisansüstü eğitimi mezunu öğretmenler lehine istatistiksel olarak anlamlı farklılık saptanmıştır. Bu bulgu, eğitim durumu lisans ve lisansüstü eğitim mezunları lehine işlemekle birlikte, eğitim durumu lisansüstü eğitim olanlar gerek bilgi gerekse de beceri boyutlarında daha yüksek özyeterlik algısına sahip oldukları saptanmıştır. Bu bulgu, gerek lisans, gerekse de lisansüstü eğitimi mezunu öğretmenlerin ölçme ve değerlendirme dersini lisans tamamlama mezunu öğretmenlere nazaran daha etkili ve yoğun bir şekilde aldıklarını göstermektedir. Bir diğer ifadeyle, ölçme ve değerlendirme dersini uzun süre alanın, lisans ve lisansüstü eğitimi mezunu öğretmenlerin bu alanda kendilerini daha yeterli algılamalarında önemli bir etken olabilir. Nitekim Buldur (2009), öğretmen adaylarının alternatif ölçme ve değerlendirme yaklaşımlarına yönelik okuryazarlık düzeylerinin ve öz yeterliklerinin tespit edilmesi için deneysel bir çalışma yapmış, bu çalışmada öğretmen adaylarının alternatif ölçme ve değerlendirme yaklaşımlarına yönelik özyeterliklerini ve okuryazarlık düzeylerini geliştirmek amacıyla onlara eğitim vermiştir. Buna göre öğretmen adaylarının alternatif ölçme ve değerlendirme yaklaşımını kullanmaya yönelik özyeterlikleri uygulamalar ilerledikçe artmıştır. Benzer olarak öğretmen adaylarının okuryazarlık ön, orta ve son test puanları arasında da anlamlı farklılık tespit edilmiş ve bu yaklaşımlara yönelik okuryazarlık düzeyleri de uygulamalar ilerledikçe gelişmiştir. Araştırmadan elde edilen nitel verilere dayanılarak, öğretmen adaylarının aldıkları kuramsal ve uygulamalı eğitimin, yaptıkları gözlemlerin ve bu yaklaşımları kullanarak yaptıkları ders sunumlarının özyeterlik ve okuryazarlık düzeylerini geliştirdiği sonucuna ulaşılmıştır. Benzer bir biçimde, Gijbels ve Dochy de (2006) yaptıkları çalışmalarında, Buldur'un (2009) elde etmiş

olduğu bulgulara paralel sonuçlar elde etmişlerdir. Bu bağlamda, Gijbels ve Dochy'nin (2006) ve Buldur'un (2009) yapmış olduğu araştırmadan elde edilen bulgunun, yapılan bu çalışmanın ilgili bulgusu ile paralellik gösterdiği söylenebilir. Benzer bulguların Aydın'ın (2001), Aydın'ın (2005) ve Haynie'nin (1992) yapmış oldukları araştırmalarında da olduğu görülmüştür.

Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının araştırıldığı bu çalışmanın son bulgusuna göre, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları görev yapılan öğretim kademesi (ilköğretim ve lise) değişkenine göre bilgi ve beceri boyutlarında ilköğretim öğretmenleri lehine istatistiksel olarak anlamlı farklılık yaratmaktadır. Bu bulgu, 2005 yılında değiştirilen yeni ilköğretim programlarının daha farklı ve alternatif ölçme ve değerlendirme teknikleri ve yaklaşımları getirmesi bulgusu ile açıklanabilir. Zira, yeni ilköğretim programlarıyla öğretmenler alternatif ölçme ve değerlendirme teknikleri ile karşılaşmışlardır. Birgin ve Gürbüz'e (2008) göre, 2005 yılında uygulamaya konulan ilköğretim programları, öğretmen ve öğretmen adaylarından öğrenme ortamlarında öğrencileri alternatif ölçme ve değerlendirme etkinlikleriyle değerlendirmelerini zorunlu kılmaktadır. Özellikle, eğitim denetmelerinin (maarif müfettişlerinin) okulu yılda iki kez rehberlik ve denetim amacıyla ziyaret etmeleri ve denetimler sırasında bu hususlardan öğretmenleri sorumlu tutmaları ister istemez öğretmenleri bu konuda hizmet-içi eğitim almaya ve kendilerini yetiştirmeye yönelmiş olabilir. Bunun aksine, liselerde gerek denetimin dört - beş yılda bir olması, gerekse de bu kurumlarda görev yapan öğretmenlerin ölçme ve değerlendirmeden pek fazla sorumlu tutulmamaları, bu öğretmenlerin ölçme ve değerlendirmeye yönelik düşük özyeterlik algısına sahip olmasına sebep olmuş olabilir. Nitekim Çakan (2004) tarafından yapılan araştırmada da ilköğretim ve ortaöğretimde (lise) görev yapmakta olan öğretmenlerin yeterlilik düzeyleri arasında anlamlı bir farklılığa rastlanmıştır. Öğretmenlerin algıları buldukları öğretim kademesine göre değişmektedir. İlköğretim kademesindeki öğretmenler ortaöğretim kademesindeki öğretmenlere göre kendilerini daha yeterli görmekteyler. Yapılan bu araştırmada her ne kadar, ilköğretim ve liselerde görev yapmakta olan öğretmenler arasında ilköğretim öğretmenleri lehine anlamlı bir farklılık saptansa da, öğretmenlerin ve öğretmen adaylarının geleneksel ölçme ve değerlendirme yaklaşımlarını kullanmaya meyilli oldukları, alternatif ölçme ve değerlendirme yaklaşımları ile ilgili bilgiye sahip olsalar da bu yaklaşımları kullanmaya yönelik yeterliklerinin düşük olduğu sonuçlarına ulaşılmıştır (örn., Birgin, 2007; Çakır ve Çimer, 2007; Doğan, Karakaya ve Gelbal, 2007; Güven ve Eskiürk, 2007; Okur, 2008). Yapılan araştırmalarda, genellikle öğretmenler sıklıkla kendilerini yeterli gördükleri yöntemleri kullandıkları rapor edilmektedir (örn., Gelbal ve Kelecioğlu, 2007). Öğretmenlerin en çok klasik soru hazırlamada kendilerini yeterli hissettiklerini ve sınavlarda klasik sorulara diğer soru tiplerine göre daha çok yer verdiklerini ortaya konulmuştur (örn., Balcı ve Tekkaya, 2000). Bu durum öğretmenlerle tartışıldığında sınıfların kalabalık oluşu, ekonomik düzey, uygulanmasının ve değerlendirilmesinin kolay olması, kısa zamanda yapılabilmesi, üniversite sınav sistemi gibi nedenlerden dolayı geleneksel ölçme-değerlendirme tekniklerini kullandıkları ortaya çıkmıştır (örn., Balcı ve Tekkaya, 2000; Korkmaz ve Kaptan, 2003). Benzer bir biçimde, Güven (2001) tarafından yapılan araştırmada da öğretmenlerinin ölçme ve değerlendirmenin yapılma amaçları konusunda yeterli bilince sahip olmadıkları ve ölçme ve değerlendirmeyi daha çok öğrenciye not vermek ve başarısını belirlemek amacıyla yaptıkları ortaya çıkmıştır. Yapılan bu çalışmada ise, öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik orta düzeyde bilgi ve beceri düzeyine sahip oldukları anlaşılmıştır. Bu sonuç, öğretmenlerin eğitimde ölçme ve değerlendirme konusunda yeterli bilgi ve beceri düzeyine sahip olmadıklarının bir göstergesi niteliğindedir.

Yapılan bu araştırmanın evrene genellenebilirliği düşüktür. Zira, yapılan bu araştırma belli bir bölgedeki görece küçük bir ildeki belli sayıdaki öğretmenler üzerinde gerçekleştirilmiştir. Ayrıca, çalışmada örneklemin belirlenmesinde yalnızca merkez ilçede görev yapmakta olan ilköğretim ve lise öğretmenleri dikkate alınmış olup; köy, kasaba ve ilçe merkezlerindeki öğretmenler araştırma kapsamında değerlendirilmemiştir. Bu anlamda, benzer bir araştırmanın daha büyük illerdeki kır-kent ayrımını da gözetir bir biçimde büyük örneklem grupları ile gerçekleştirilmesi yararlı olabilir. Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının araştırıldığı bu çalışmada elde edilen bulgulara dayanarak, öncelikle öğretmenlerin eğitimde ölçme ve

değerlendirmeye yönelik gerek bilgi boyutunda gerekse de beceri boyutunda uzman kişilerce hizmet-içi eğitimden geçirilmeleri gerekmektedir. Zira yapılan bu çalışmada öğretmenlerin genel itibari ile eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları oldukça düşük bir seviyede bulunmuştur. Diğer yandan, yapılan çalışmada, lisansüstü eğitimi mezunu öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik daha yüksek özyeterlik algılarına sahip oldukları saptanmıştır. Buna paralel olarak, öğretmenlerin özellikle eğitim bilimleri alanında lisansüstü eğitime yönlendirilmeleri, onların eğitimde ölçme ve değerlendirmeye yönelik kuramsal ve pratik bilgi ve becerileri kazanmalarını sağlayabilir. Bu konuda, Milli Eğitim Bakanlığı (MEB) ile üniversitelerin eğitim fakülteleri işbirliği yaparak, gerek yüz yüze gerekse de çevrimiçi (online) eğitim seçenekleriyle öğretmenlere daha etkili ve uygulamaya dönük eğitimde ölçme ve değerlendirmeye yönelik bir eğitim süreci sağlayabilirler. Bunun yanında, öğretmenlere eğitimde ölçme ve değerlendirmenin nasıl yapıldığına ilişkin çeşitli kitaplar, istatistik ve test analiz programları ve kaynak CD'ler sağlanabilir. Bir başka yandan, çalışmada mesleki kıdemi yüksek olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algıları genç meslektaşlarına göre daha düşük bir düzeyde bulunmuştur. Buna paralel olarak, mesleki kıdemi fazla olan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik hizmet-içi eğitimden geçirilmeleri ve sınıflarında öğretme-öğrenme sürecini değerlendirmede farklı ölçme ve değerlendirme tekniklerini kullanmaya ve bunlara ilişkin gerekli olan istatistiksel işlemleri yapmaya özendirilmeleri gerekmektedir. Çalışmada, ayrıca, ilköğretim okullarında görev yapan öğretmenlerin liselerde görev yapan meslektaşlarına göre eğitimde ölçme ve değerlendirmeye yönelik daha düşük düzeyde özyeterlik algısına sahip oldukları saptanmıştır. Bu bulguya paralel olarak, liselerde görev yapan öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik uzman kişilerce hizmet-içi eğitimden geçirilerek, lisansüstü eğitime yönlendirilmeleri gerekmektedir. Burada gerek hizmet-içi eğitimlerle gerekse de lisansüstü eğitimlerle ilgili olarak, verilecek eğitimlerin mutlak suretle kuram ve uygulamayı bütünleştirmesi ve bunların öğretmenlere somut bir şekilde gösterilmesi gerektiği söylenebilir.

KAYNAKÇA

- Airasian, P. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.
- Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları*. İstanbul: İdeal Kültür Yayıncılık.
- Atılğan, H., Doğan, N. ve Kan, A. (2007). *Eğitimde ölçme ve değerlendirme* (2. baskı). Ankara: Anı Yayıncılık.
- Aydın, A. (2001). *Eğitim fakültesi mezunu olan ve olmayan öğretmenlerin ölçme ve değerlendirme yeterliliklerinin karşılaştırılmasına yönelik bir çalışma* (Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Aydın, F. (2005, Eylül). *Öğretmenlerin alternatif ölçme değerlendirme konusundaki düşünceleri ve uygulamaları*. XIV. Ulusal Eğitim Bilimleri Kongresi'nde sunulan bildiri, Pamukkale Üniversitesi Eğitim Fakültesi, Denizli.
- Balcı, E. ve Tekkaya, C. (2000). Ölçme ve değerlendirme tekniklerine yönelik bir ölçeğin geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 18, 42-50.
- Bayram, E. (2011). Öğretmenlerin ölçme ve değerlendirme yeterliklerinin incelenmesi. (Yayımlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Begtaş, A. D. (2005). *Fen öğretiminde değerlendirme etkinlikleri üzerine öğretmen görüşleri* (Yayımlanmamış yüksek lisans tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Birgin, O. (2007, Eylül). *Sınıf öğretmeni adaylarının ölçme-değerlendirme konusundaki okuryazarlık düzeylerinin incelenmesi*. XVI. Ulusal Eğitim Bilimleri Kongresi'nde sunulan bildiri, Gaziosmanpaşa Üniversitesi Eğitim Fakültesi, Tokat.
- Birgin, O. ve Gürbüz, R. (2008). Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki bilgi düzeylerinin incelenmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20, 163-181.
- Buldur, S. (2009). *Fen bilgisi öğretmen adaylarının alternatif ölçme ve değerlendirme yaklaşımlarına yönelik okuryazarlık ve öz yeterlik düzeylerinin geliştirilmesi* (Yayımlanmamış yüksek lisans tezi, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.

- Cheng, H. M. (2006). Junior secondary science teachers' understanding and practice of alternative assessment in Hong Kong: Implications for teacher professional development. *Canadian Journal of Science, Mathematics and Technology Education*, 6(3), 227-243.
- Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Culbertson, L. D., & Wenfan, Y. (2003, April). *Alternative assessment: Primary grade literacy teachers' attitudes and practices*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37, 99-114.
- Çakır, İ. ve Çimer, O. S. (2007, Kasım). *Fen ve teknoloji öğretmenlerinin alternatif ölçme değerlendirme konusundaki yeterlilikleri ve uygulamada karşılaşılan problemler*. I. Ulusal İlköğretim Kongresi'nde sunulan bildiri, Hacettepe Üniversitesi Eğitim Fakültesi, Ankara.
- Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *Journal of Educational Research*, 91, 331-343.
- Doğan, N., Karakaya, İ. ve Gelbal, S. (2007, Kasım). *İlköğretim öğretmenlerinin ölçme araçlarıyla ilgili yeterlik algıları ve bu araçları kullanma durumları*. I. Ulusal İlköğretim Kongresi'nde sunulan bildiri, Hacettepe Üniversitesi Eğitim Fakültesi, Ankara.
- Erdemir, A. Z. (2007). *İlköğretim ikinci kademe öğretmenlerinin ölçme değerlendirme tekniklerini etkin kullanabilme yeterliklerinin araştırılması* (Yayımlanmamış yüksek lisans tezi, Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Enstitüsü, Kahramanmaraş). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Ertürk, S. (1972). *Eğitimde program geliştirme*. Ankara: Yelkentepe Yayınları.
- Flowers, C., Ahlgrim-Dezell, D., Browder, D., & Spooner, F. (2005). Teachers' perceptions of alternate assessments. *Research and Practice for Persons with Severe Disabilities*, 30(2), 81-92.
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education* (7th ed.). New York: McGraw-Hill, Inc.
- Gelbal, S. ve Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, 32(4), 399-409.
- Gömleksiz, M. N. ve Kan, Ü. A. (2007, Kasım). *Sınıf öğretmenliği adaylarının yeni ilköğretim programındaki ölçme-değerlendirme yöntemlerini tanıma düzeylerine ilişkin görüşleri*. I. Ulusal İlköğretim Kongresi'nde sunulan bildiri, Hacettepe Üniversitesi Eğitim Fakültesi, Ankara.
- Güven, B. ve Eskitürk, M. (2007, Eylül). *Sınıf öğretmenlerinin ölçme ve değerlendirmede kullandıkları yöntem ve teknikler*. XVI. Ulusal Eğitim Bilimleri Kongresi'nde sunulan bildiri, Gaziosmanpaşa Üniversitesi Eğitim Fakültesi, Tokat.
- Güven, S. (2001, Haziran). *Sınıf öğretmenlerin ölçme ve değerlendirmede kullandıkları yöntem ve tekniklerin belirlenmesi*. X. Ulusal Eğitim Bilimleri Kongresi'nde sunulan bildiri, Abant İzzet Baysal Üniversitesi Eğitim Fakültesi, Bolu.
- Haynie, W. J. (1992). Post-hoc analysis of test items written by technology education teachers. *Journal of Technology of Education*, 4(1), 27-40.
- Huck, S. W. (2000). *Reading statistics and research* (3rd ed.). New York: Addison Wesley.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi* (15. baskı). Ankara: Nobel Yayın Dağıtım.
- Karaca, E. (2008). An investigation of primary and high school teachers' perception levels of efficacy of measurement and evaluation in education in Turkey. *Social Behavior and Personality*, 36(8), 1111-1122.
- Karaca, E. (2004). *Öğretmen adaylarının ölçme ve değerlendirme yeterliklerine ilişkin algıları*. Eskişehir: Anadolu Üniversitesi Eğitim Fakültesi Yayınları.
- Karaca, E. (2003). Öğretmen adaylarının ölçme ve değerlendirme yeterliklerine ilişkin likert tipi bir yeterlik algısı ölçeğinin geliştirilmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9, 179-198.
- Kılınç, M. (2011). Öğretmen adaylarının eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algı ölçeği. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 12(4), 81-93.
- Korkmaz, H. ve Kaptan, F. (2003). İlköğretim fen öğretmenlerinin portfolyoların uygulanabilirliğine yönelik güçlükler hakkındaki algıları. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 159-166.
- Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice* (7th ed.). New Jersey: John Wiley and Sons, Inc.
- Linn, R., & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). New York: Prentice-Hall.

- McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry* (6th ed.). Boston: Pearson Education Ltd.
- MEB (2015). *Milli eğitim istatistikleri: Örgün eğitim 2014-2015*. Ankara: Milli Eğitim Bakanlığı, Strateji Geliştirme Başkanlığı.
- MEB (2009). *Öğretmenlik mesleki genel yeterlikleri*. Ankara: Milli Eğitim Basımevi.
- Numanoğlu, G. ve Bayır, G. (2009). Bilgisayar öğretmen adaylarının öğretmenlik mesleği genel yeterliklerine ilişkin görüşleri. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 10(1), 197-212.
- Okur, M. (2008). *4. ve 5. sınıf öğretmenlerinin fen ve teknoloji dersinde kullanılan alternatif ölçme ve değerlendirme tekniklerine ilişkin görüşlerinin belirlenmesi* (Yayımlanmamış yüksek lisans tezi, Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Enstitüsü, Zonguldak). <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Özçelik, D. A. (1998). *Ölçme ve değerlendirme*. Ankara: ÖSYM Yayınları.
- Pektaş, S. (2010). *Öğretmen adaylarının ölçme ve değerlendirme yeterlik algılarının incelenmesi* (Yayımlanmamış yüksek lisans tezi, Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü, Bolu). <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Turgut, M. F. (1988). *Eğitimde ölçme ve değerlendirme metotları* (6. baskı). Ankara: Saydam Matbaacılık.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770.
- Yaman, S. ve Karamustafaoğlu, S. (2011). Öğretmen adaylarının ölçme ve değerlendirme alanına yönelik yeterlik algı düzeylerinin incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 44(2), 53-72.
- Yavuz, G. (2011). *Öğretmen adaylarının öğretme öğrenme süreci ve ölçme ve değerlendirme alanındaki yeterliklerine ilişkin görüşleri* (Yayımlanmamış yüksek lisans tezi, Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Mersin). <http://tez2.yok.gov.tr/adresinden edinilmiştir>.
- Yayla, R. G. (2011, April). *Fen ve teknoloji öğretmenlerinin tecrübeleriyle alternatif ölçme ve değerlendirme yaklaşımlarına yönelik özyeterlikleri arasındaki ilişki*. Paper presented at the 2nd International Conference on New Trends in Education and Their Implications, Antalya, Turkey.

EXTENDED ABSTRACT

Introduction

Teachers are required to have the necessary knowledge and skills in educational measurement and evaluation so as to evaluate teaching and learning in their classrooms (Kubiszyn and Borich, 2003). The effective realisation of educational measurement and evaluation activities is possible with teachers having the required knowledge, skills, and attitudes, in other words self-efficacy (Yavuz, 2011). The high level of self-efficacy perception in educational measurement and evaluation of teachers shows that it affects the instruction in the classroom in a considerable extent (Pektaş, 2010; Kılınç, 2011). Especially, it is seen very important that teachers have a high self-efficacy in educational measurement and evaluation. In order to practise measurement and evaluation in education effectively, teachers must develop necessary measurement tools to determine the readiness levels and learning incompetence of their students as well as identify the level of development of learning of them in the end of the curriculum. Also, teachers are required to know methods and techniques in measurement and evaluation in order to use the necessary measurement tools effectively (MEB, 2009; Yavuz, 2011). The limited or inadequate knowledge and skills in regard of measurement and evaluation techniques of teachers may create some problems when they practice them in the classroom (Çakan, 2004). Hence, it can be stated that teachers must have the necessary knowledge and skills in measurement and evaluation so as to assess the teaching and learning in their classrooms. Without having the necessary knowledge and skills in measurement and evaluation, teachers cannot evaluate the instruction effectively and maintain the teaching-learning process efficiently. The purpose of this study was to examine teachers' self-efficacy perceptions on educational measurement and evaluation with respect to gender, occupational experience, educational level, and instruction stage in which teachers work.

Method

The study was carried out in the centre of Niğde province from six elementary schools and four high schools. Teachers (n = 211), selected according to the purposive sampling method, participated in the study. The “survey method” was adopted in the research. In accordance with the purpose of the study, percentage, mean, standard deviation, independent samples *t* test, one-way ANOVA (analysis of variance) and Scheffe tests were employed in the study in order to examine the data obtained in terms of some variables (gender, occupational experience, educational level, and instruction stage). “The Scale for Measurement and Evaluation Self-Efficacy in Education” (Kılınç, 2011) was used in order to collect data for the study. For the statistical analyses SPSS 17.0 was utilised in the research.

Results

According to findings of the study, it was seen that teachers participated in the study had low level of self-efficacy perceptions on educational measurement and evaluation both in knowledge ($M = 3.13$, $SD = 7.05$) and skill ($M = 2.82$, $SD = 8.74$) sub-dimensions. On the other hand, teachers’ self-efficacy perceptions on educational measurement and evaluation were found out to be differed with respect to gender, occupational experience, educational level, and instruction stage which teachers work in variables in the study. Firstly, it was found that teachers’ self-efficacy perceptions on educational measurement and evaluation statistically differed in skill sub-dimension ($t_{(209)} = -2.447$, $p < .05$) in relation to gender variable. However, it was understood that teachers’ self-efficacy perceptions on educational measurement and evaluation statistically did not differ in knowledge sub-dimension ($t_{(209)} = -1.939$, $p > .05$) in relation to gender variable. Similarly, it was found out that teachers’ self-efficacy perceptions on educational measurement and evaluation statistically differed both in knowledge ($F_{(3,207)} = 27.94$, $p < .05$) and skill ($F_{(3,207)} = 33.00$, $p < .05$) sub-dimensions in relation to occupational experience variable in favour of teachers who had 1-5 years ($M_{\text{Knowledge}} = 3.74$, $SD = 5.32$; $M_{\text{Skill}} = 3.53$, $SD = 6.33$) of occupational experience. Thus, it was also understood that teachers had 16 and above years ($M_{\text{Knowledge}} = 2.41$, $SD = 3.82$; $M_{\text{Skill}} = 2.26$, $SD = 3.90$) of occupational experience had low level of self-efficacy perceptions on educational measurement and evaluation. According to the findings of the study in regard of educational level variable, it was found out that teachers’ self-efficacy perceptions on educational measurement and evaluation both in knowledge ($F_{(2,208)} = 29.829$, $p < .05$) and skill ($F_{(2,208)} = 34.636$, $p < .05$) sub-dimensions in favour of teachers had postgraduate level of education in the study. Lastly, according to the findings obtained in the study in terms of instruction stage where teachers work in, it was seen that teachers’ self-efficacy perceptions on educational measurement and evaluation differed significantly in regard of instruction stage where teachers work in variable in favour of teachers who worked in elementary schools both in knowledge ($t_{(209)} = 6.474$, $p < .05$) and skill ($t_{(209)} = 6.503$, $p < .05$) sub-dimensions.

Discussion

The results acquired in the study showed that teachers had low level of self-efficacy perceptions on educational measurement and evaluation both in knowledge and skill sub-dimensions. According to findings of the study, it was seen that teachers participated in the study had low level of self-efficacy perceptions on educational measurement and evaluation both in knowledge and skill sub-dimensions. In most studies in the related literature, it was seen that teachers had low levels of self-efficacy perceptions on educational measurement and evaluation (e.g., Haynie, 1992; Aydın, 2001; Culbertson and Wenfan, 2003; Karaca, 2003; Çakan, 2004; Karaca, 2004; Aydın, 2005; Begtaş, 2005; Cheng, 2006). However, it was also seen that there were some but limited number of studies that teachers had high level of self-efficacy perceptions on educational measurement and evaluation (e.g., Pektaş, 2010; Yavuz, 2011). Generally, it was reported that teachers mostly had self-efficacy perceptions on educational measurement and evaluation in knowledge dimension instead of skill dimension in many studies. On the other hand, teachers’ self-efficacy perceptions on educational measurement and evaluation were found out to be differed with respect to gender, occupational

experience, educational level, and instruction stage which teachers work in variables in the study. Firstly, it was found out that teachers' self-efficacy perceptions on educational measurement and evaluation statistically differed in skill sub-dimension in relation to gender variable. However, it was understood that teachers' self-efficacy perceptions on educational measurement and evaluation statistically did not differ in knowledge sub-dimension in relation to gender variable. Similarly, it was found out that teachers' self-efficacy perceptions on educational measurement and evaluation statistically differed both in knowledge and skill sub-dimensions in relation to occupational experience variable in favour of teachers who had 1-5 years of occupational experience. Thus, it was also understood that teachers had 16 and above years of occupational experience had low level of self-efficacy perceptions on educational measurement and evaluation. Therefore, it can be said that in light of the data acquired in the study in terms of occupational experience, younger teachers have high level of self-efficacy perceptions on educational measurement and evaluation compared to their older colleagues. On the other hand, according to the findings of the study in regard of educational level variable, it was found out teachers' self-efficacy perceptions on educational measurement and evaluation both in knowledge and skill sub-dimensions in favour of teachers had postgraduate level of education in the study. According to this finding, it was understood that teachers with postgraduate level of education had higher level of self-efficacy perceptions on educational measurement and evaluation compared to the teachers who had graduate or lower level of education. Lastly, according to the findings obtained in the study in terms of instruction stage where teachers work in, it was seen that teachers' self-efficacy perceptions on educational measurement and evaluation differed significantly in regard of instruction stage where teachers work in variable in favour of teachers who worked in elementary schools both in knowledge and skill sub-dimensions.

TEOG Ortak ve Mazeret Sınavındaki Türkçe ve Matematik Alt Testlerinin Psikometrik Özelliklerinin Karşılaştırılması*

The Comparison of Psychometric Properties of Standardised and Make Up Maths and Turkish Subtest Questions in TEOG

Abdullah Faruk KILIÇ **

Hülya KELECİOĞLU ***

Öz

Bu çalışmanın amacı 2013-2014 eğitim öğretim yılında uygulanan TEOG ortak ve mazeret sınavındaki Türkçe ve matematik alt testlerinin psikometrik özelliklerini karşılaştırmaktır. Türkçe ve matematik testlerinin ortak ve mazeret sınavları madde güçlüğü, madde ayırıcılık indeksleri, testlerin güvenilirliği ve geçerliği gibi psikometrik özellikler çevresinde karşılaştırmalı olarak incelenmiştir. Araştırma Türkçe ortak sınavı için 9773, matematik ortak sınavı için 9485, Türkçe mazeret sınavı için 2747 ve matematik mazeret sınavı için 2323 öğrencinin madde puanları üzerinden yürütülmüştür. Kapsam geçerliği çalışması için uzman görüşlerine başvurulmuştur. Bunun için araştırmacılar tarafından TEOG ortak ve mazeret sınavlarındaki soruların ölçtüğü kazanımları belirlemek ve karşılaştırmak amacıyla soru ve kazanım eşleştirme formu ile soru değerlendirme formu oluşturulmuştur. Araştırma sonucunda hem matematik hem de Türkçe ortak ve mazeret sınavlarının test ve madde istatistiklerinin manidar düzeyde farklı olduğu gözlenmiştir. Uzman görüşlerine dayalı olarak değerlendirilen Türkçe ve matematik testlerinin kapsamının ortak ve mazeret sınavları için benzer olduğu görülmüştür. Matematik testi ortak ve mazeret sınavlarındaki maddelerin aynı yapıyı ölçmediği fakat Türkçe testi ortak ve mazeret sınavlarındaki maddelerin benzer yapıyı ölçtüğü sonucuna ulaşılmıştır.

Anahtar Kelimeler: TEOG, geçerlik, güvenilirlik, faktör analizi, test ve madde istatistikleri

Abstract

The aim of this study is to compare the psychometric properties of Turkish and math subtests in common and make-up examination of TEOG applied in 2013-2014 school year. Turkish and math tests of the common and make-up examinations were investigated comparatively within the frame of psychometric properties such as item difficulty, item discrimination index, test reliability and validity. Research was carried out via item scores matrix of 9773 students for Turkish common exam, 9485 students for math common exam, 2747 students for Turkish make-up exam and 2323 students for math make-up exam. It was consulted to expert opinion for the content validity study. According to this, in order to determine and compare the acquisitions that is measured in common and make-up examinations of TEOG, question-acquisition matching form and question assessment form was prepared by researcher. As a result of research, it has been observed that Turkish and math subtests test and item statistics are different in significantly. Content validity which based on expert opinions is in similar level for Turkish and math subtests. The results of the study revealed that math tests of the common and make-up exams items didn't measure the same things in terms of construct validity. Unlike math, Turkish tests of the common and make-up exams items measure the same things in terms of construct validity.

Key Words: TEOG, validity, reliability, factor analysis, test and item statistics.

* Bu makale birinci yazar tarafından ikinci yazar danışmanlığında hazırlanan yüksek lisans tezinden türetilmiştir.

** Araş. Gör., Adıyaman Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Adıyaman-Türkiye, e-posta: afarukkilic@windowslive.com

*** Prof. Dr, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: hulyakelecioglu@gmail.com

GİRİŞ

Türkiye’de eğitimin hemen hemen her düzeyinde test uygulanmakta ve bunların sonucunda bireyler hakkında kararlar verilmektedir. Örneğin ortaokul düzeyinde Temel Eğitimden Ortaöğretime Geçiş (TEOG) sınavı, lise düzeyinde Yüksek Öğretime Geçiş Sınavı (YGS), Lisans Yerleştirme Sınavı (LYS), lisans düzeyinde Akademik Personel ve Lisansüstü Eğitim Sınavı (ALES), Yabancı Dil Sınavı (YDS), Kamu Personeli Seçme Sınavı (KPSS) gibi sınavlar uygulanmaktadır. Bu sınavların ortak özelliği ise bireylerin hayatını, yüksek derecede etkileme gücüne sahip olmalarıdır. Genç nüfusun fazla olduğu ülkemizde, bireyleri belli bir amaç için sıralamak ve bu sıralamaya göre bireyler hakkında karar vermek kaçınılmazdır.

Hazırlanan bir test, ölçülmesi planlanan davranış evreninden ölçülecek davranışların bir örneklemini sağlamaktadır. Bu nedenle testin ölçülmek istenilen davranış evrenini ne derecede iyi temsil ettiği ve testin ölçülmek istenilen davranış evrenini ne derece doğru ölçtüğü sorulması gereken temel sorulardır (Alıcı ve diğerleri., 2011). Bu soruların cevaplarına ulaşabilmek için testin geçerlik ve güvenilirlik çalışmalarının yapılması gerekmektedir.

Güvenirlik; test puanlarının aynı ölçme işlemlerinden geçen tekrarlı ölçmeler sonucunda tutarlı sonuç verme derecesi olduğu gibi aynı zamanda hatalardan arınık olma derecesidir (AERA, APA, ve NCME, 2014; Özgüven, 2011; Turgut ve Baykul, 2012). Geçerlik ise testin amaca hizmet etme derecesi olarak tanımlanmakta (Turgut ve Baykul, 2012) ve test puanları hakkında testin hazırlanma amacı doğrultusunda yapılan yorumların, hem kuramsal olarak hem de toplanan kanıtlarla desteklenme derecesi olarak düşünülmektedir (AERA ve diğerleri., 2014). Tanımlardan da anlaşılacağı üzere hem geçerlik hem de güvenilirlik bir derece olarak belirtilmektedir. Bir test için var ya da yok şeklinde ifade etmek yerine *yeterli* veya *yetersiz* olarak değerlendirilmektedir.

Bir ölçme aracından elde edilen sonuçlar üzerinden geçerlik ve güvenilirlik çalışmalarının yapılması, ölçümlerin ne kadar hatasız olduğu, testte ölçülmek istenilen yapıyla ne kadar örtüştüğü gibi ölçme sonuçları hakkında önemli kanıtlar sunmaktadır. Ulusal düzeyde uygulanan sınavlar için de geçerlik ve güvenilirlik çalışmaları yürütülerek ölçümlerin ne kadar hatasız yapıldığı, ölçülmek istenilen kapsamın ne kadar yoklandığı ve ölçülmek istenilen yapıyla ilişkisi araştırılmalıdır. Ulusal düzeyde gerçekleştirilen sınavlardan biri de 2013-2014 eğitim öğretim yılından beri Millî Eğitim Bakanlığı tarafından uygulanan Temel Eğitimden Ortaöğretime Geçiş Sınavı (TEOG)’dır. Görmez ve Coşkun’da (2015) TEOG için geçerlik ve güvenilirlik çalışmaları yapılmasının önemini vurgulamıştır.

TEOG ilk kez 2013-2014 eğitim öğretim yılında 8. sınıf düzeyinde uygulamaya konmuştur. Bireylerin başarılarını sürece yayarak değerlendirmeyi hedefleyen bir sınav sistemi olduğu belirtilmektedir. TEOG’da Türkçe, matematik, fen ve teknoloji, din kültürü ve ahlak bilgisi, T.C. inkılâp tarihi ve Atatürkçülük ile yabancı dil dersleri için Millî Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğüne her dönem ortak sınavlar yapılarak öğrenciler değerlendirilmektedir (Millî Eğitim Bakanlığı [MEB], 2015).

Her dönem yapılan ortak sınavlarda ağırlıklandırılmış puanların hesaplanmasında; Türkçe, matematik, fen ve teknoloji dersleri için dört; T.C. inkılâp tarihi ve Atatürkçülük, yabancı dil ile din kültürü ve ahlâk bilgisi için iki katsayısı o dersin puanı ile çarpılarak her bir dersin ağırlıklandırılmış ortak sınav puanı hesaplanmaktadır. Puanlama 700 tam puan üzerinden yapılmaktadır. Öğrencilerin; 6, 7 ve 8’inci sınıf yılsonu başarı puanları ile 8. sınıf ağırlıklandırılmış ortak sınav puanı toplanarak, elde edilen toplam, ikiye bölünüp merkezî sistemle öğrenci alan ortaöğretim kurumlarına yerleştirmeye esas puanın hesaplanmasında kullanılmaktadır. Bu hesaplamada ise puanlama, 500 tam puan üzerinden yapılmaktadır (MEB, 2015).

TEOG’da ortak sınavlar ile mazeret sınavları adı altında iki sınav yapılmaktadır. Ortak sınavlarda her ders için çoktan seçmeli 20 soru sorulmakta, değerlendirmede yanlış cevap sayısı doğru cevap sayısını etkilememektedir. Sınav süresi her ders için 40 dakika olarak verilmektedir. Mazeret sınavları ise; ortak sınavlara katılmayan öğrencilerin girdiği sınavlardır. Mazeret sınavına girecek öğrenciler “Millî Eğitim Bakanlığı Okul Öncesi Eğitim ve İlköğretim Kurumları yönetmeliğinin 23. maddesinde belirtilen hükümler çerçevesinde değerlendirilerek okul müdürlüğüne belirlenmektedir.

Bu öğrencilerin bilgileri, aynı gün il/ilçe millî eğitim müdürlüklerine bildirilerek okul müdürlüklerince e-okul sistemine sınavlar tamamlandıktan sonra 5 (beş) gün içerisinde giriş yapılmaktadır (MEB, 2015).

Mazereti okul müdürlüğünce uygun görülen öğrenciler için Bakanlıkça belirlenen tarih ve merkezlerde mazeret sınavı yapılarak geçerli bir mazeretten dolayı sınava giremeyen öğrencilere yeni bir sınav hakkı verilmektedir. Mazeret sınavları da ortak sınavlarda olduğu gibi her ders için 20 sorudan oluşmakta, sınav süresi yine 40 dakika olarak verilmekte ve yanlış cevap sayısı doğru cevap sayısını etkilememektedir. Şad ve Şahiner (2016) tarafından öğrenci, öğretmen ve veli görüşler, üzerinden yürütülen araştırmada mazeret sınavının olması ve puanlamada düzeltme formülünün kullanılmamasının olumlu karşılandığı gözlenmiştir.

TEOG sınavlarında belirtilen altı dersten; bir dönemde iki sınav yapılan derslerin ilk sınavı, bir dönemde üç sınav yapılan derslerin ise ikinci sınavı ortak sınav şeklinde uygulanmaktadır.

TEOG sınavlarında ortak ve mazeret sınavı olmak üzere iki sınav yaklaşık 1 ay ara ile yapılmaktadır. Bu durumda her iki sınavda uygulanan testlerin ve testlerde yer alan maddelerin madde güçlüğü, madde ayırıcılığı, maddelerin ölçtüğü kazanımlar, test sonuçlarının güvenilirlik düzeyi ve testlerin ölçtüğü yapılar gibi test ve madde özelliklerinin birbiriyle eşdeğer olup olmadığı sorunu ortaya çıkmaktadır. Bireylerin aynı şekilde değerlendirildiği, aynı ölçüt üzerinden karar verildiği iki test ve testte bulunan maddeler birbirine eşdeğer olmadığına, bir grubun dezavantajlı olması durumu ortaya çıkmaktadır. Bunun için de bu testlerin psikometrik nitelikleri incelenmeli, eşdeğerlikleri sağlanmalı ve gelecek testlerde bunlara yönelik önlemler alınmalıdır.

TEOG ile ilgili yapılan araştırmalarda kapsam geçerliğini ve sınavlardaki maddelerin hangi kazanımla ilişkili olduğunu inceleyen (Dalak, 2015; Karadeniz, Eker, ve Ulusoy, 2015; Kardeş-Birinci, 2014; Yılmaz, 2014; Yılmaz-Koçar ve Aygün, 2015; Yorgancı, 2015) çalışmalarla birlikte TEOG sistemine yönelik öğrenci ve öğretmen görüşlerine göre değerlendirilmesi ön plana çıkmaktadır (Atıla ve Özeken, 2015; Dinç, Dere, ve Koluman, 2014; Karaca, Bektaş, ve Armağan, 2015; Karaca, 2015; Zayimoğlu-Öztürk ve Aksoy, 2014). Bu araştırmaların yanında Özkan, (2015) tarafından 2013-2014 eğitim öğretim yılının I. döneminde gerçekleştirilen TEOG sınavlarından matematik ortak ve mazeret sınavlarının test eşitleme çalışması yapılmıştır. TEOG sınavlarının psikometrik özelliklerine yönelik yapılacak araştırmaların da bu sınavların geleceğini belirleme konusunda öneme sahip olduğu düşünülmektedir.

Araştırmanın Amacı

Bu çalışmanın amacı 2013-2014 eğitim öğretim yılında uygulanan TEOG ortak ve mazeret sınavındaki Türkçe ve matematik alt testlerinin psikometrik özelliklerini karşılaştırmaktır. Bu amaçla çalışmada Türkçe ve matematik alt testlerine ait ortak ve mazeret sınavlarının;

1. Test ve madde istatistikleri
2. Kapsam geçerlikleri ve
3. Yapı geçerlikleri karşılaştırmalı olarak incelenmiştir.

YÖNTEM

Araştırmanın Modeli

Bu çalışma tarama modellerinden, genel tarama modelinde yürütülen bir araştırmadır. Genel tarama modelinde çok sayıda elemandan oluşan bir evrende, evren hakkında genel bir kanıya varmak amacıyla evren üzerinde ya da örnekleminde yapılan tarama düzenlemeleridir (Karasar, 2014). Ayrıca düzeyi bakımından bu araştırma betimsel bir araştırmadır. Betimsel araştırmalarda verilen bir durum dikkatli bir şekilde tanımlamaya çalışılır (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz, ve Demirel, 2013).

Evren ve Örneklem

Araştırmanın iki farklı evreni bulunmaktadır. Bunlardan birincisi 2013-2014 eğitim öğretim yılının ikinci döneminde uygulanan TEOG Türkçe ve matematik ortak sınavına katılan 1.271.284 öğrencidir. Aynı yıl gerçekleştirilen TEOG Türkçe ve matematik mazeret sınavına katılan 2890 öğrenci araştırmanın ikinci evrenini oluşturmaktadır.

Araştırmanın örneklemini ortak sınavlara giren öğrencilerden rastgele olarak seçilen 10.000 öğrenci oluşturmaktadır. Örneklem incelenip tüm sorulara aynı cevabı veren ya da sistematik olarak cevapları tekrarlayan bireylerin cevapları temizlenerek matematik ortak sınavı için 9485, Türkçe ortak sınavı için 9773 öğrencinin verileri üzerinden analizler yapılmıştır. Mazeret sınavlarından örneklem alınmamış evren üzerinde analizler gerçekleştirilmiştir. Veri temizleme işlemi yapıldıktan sonra Türkçe mazeret sınavında 2747 öğrencinin, matematik mazeret sınavında ise 2323 öğrencinin cevapları üzerinden analizler gerçekleştirilmiştir.

Veri Toplama Araçları

Araştırmada, TEOG Türkçe ve matematik alt testlerinin ortak ve mazeret sınavlarının verilerinin yanı sıra araştırmacılar tarafından geliştirilen ve aşağıda açıklanan formlar kullanılmıştır.

Soru-Kazanım Eşleştirme Formu

Uzman görüşlerine göre TEOG ortak ve mazeret sınavlarındaki soruların ölçtüğü kazanımları belirlemek ve karşılaştırmak amacıyla Soru ve Kazanım Eşleştirme Formu oluşturulmuştur. Türkçe ve matematik derslerinin ortak ve mazeret sınav soruları bir tablo halinde alan uzmanlarına verilmiş ve soruların karşılıklarına, sorunun hangi kazanımı ölçtüğünü belirtmeleri istenmiştir. Türkçe ve matematik derslerinin kazanım listesi ise uzmanlara farklı bir formda verilmiştir.

Soru Değerlendirme Formu

Maddelerin kapsam geçerlik oranı indeksini (KGO_i) hesaplamak amacıyla Lawshe (1975) tarafından önerilen yöntemle göre alan uzmanlarının maddeleri “gerekli/önemli”, “yararlı ama önemli değil” ve “gereksiz” şeklinde sınıflandırabileceği Soru Değerlendirme Formu hazırlanmıştır.

Soru kazanım eşleştirme formu ve soru değerlendirme formu 14 matematik ve 10 Türkçe alan uzmanınca değerlendirilmiştir. Matematik alan uzmanlarından 7’si ilköğretim matematik öğretmenliği lisans mezunu; 1’i matematik eğitimi alanında yüksek lisans; 6’sı ise matematik öğretmenliği tezsiz yüksek lisans mezunudur. Türkçe alan uzmanlarından 4’ü Türkçe öğretmenliği lisans mezunu, 1’i Türkçe eğitimi yüksek lisans mezunu, 5’i ise Türk dili ve edebiyatı öğretmenliği tezsiz yüksek lisans mezunudur.

Verilerin Analizi

Türkçe ve matematik alt testlerinin ortak ve mazeret sınavları için test ve madde istatistikleri hesaplanarak; testlerin aritmetik ortalamaları arasında fark olup olmadığı bağımsız örneklem t-testi ile, madde ayırıcılık ortancaları arasında fark olup olmadığı Mann-Whitney U testi ile, testlerin ortama güçlükleri ve madde güçlükleri arasında fark olup olmadığı iki oran arasındaki farkı test etmeye yarayan z testi ile, varyansları arasında fark olup olmadığı F testi ile ve güvenilirlikleri arasında fark olup olmadığı korelasyonların test edilmesinde kullanılan z testi ile karşılaştırılmıştır. İç tutarlık anlamında güvenilirlik belirlenirken konjenerik ölçmelerde kullanılması önerilen (Yurdugül, 2006) McDonald’ın ω katsayısı hesaplanmıştır. Güvenirlik katsayısının hesaplanmasında Factor 10.3 programı, madde ayırıcılık indekslerinin hesaplanmasında ise ITEMAN 3.5 analiz programı diğer hesaplamalarda ise Office 2014 paket programı içerisinde bulunan Excel programı kullanılmıştır.

İstatistiksel sınavın yanında etki büyüklüklerinin de verilmesi araştırmaların pratikteki anlamlılığını da ortaya koyacağı ve daha nitelikli sonuçların elde edilmesini sağlayacağı için raporlanması bir gereklilik haline gelmiştir (American Educational Research Association, 2006;

American Psychological Association, 2010; Özsoy ve Özsoy, 2013). Bu çalışmada ortalamalar arasındaki etki büyüklüğünü belirlemek için Cohen's d hesaplanmıştır. Cohen's d istatistiği işaretine bakılmaksızın 0,20 küçük, 0,50 orta ve 0,80 ise büyük etki gücüne sahip olduğunu ifade etmektedir (Büyüköztürk, 2013; Cohen, 1988, 1992; Cumming, 2012). Oranlar arasındaki etki büyüklüğünü belirlemek için Cohen's h istatistiği kullanılmıştır (Cohen, 1988, 1992). Elde edilen h değerlerinin yorumu d değerlerinin yorumu gibidir. Varyanslar arasındaki farkın etki büyüklüğü yine Cohen's d istatistiği ile belirlenmiş ancak, varyansların anlamlılığı F testi ile yapıldığı için F testine yönelik olarak elde edilen formül kullanılmıştır (Thalheimer ve Cook, 2002). Bu istatistiğin yorumu da h istatistiğiyle ayıdır. Güvenirlik katsayıları arasındaki farkın etki büyüklüğünü belirlemek için Cohen's q istatistiği, madde ayırıcılık ortancaları arasındaki farkın etki büyüklüğünü belirlemek için ise r istatistiği kullanılmıştır (Cohen, 1988, 1992). Cohen's q ve Cohen's r istatistikleri 0,10 küçük, 0,30 orta ve 0,50 büyük etki büyüklüğüne sahip olduğu şeklinde yorumlanmaktadır (Cohen, 1992).

Soru kazanım eşleştirme formu verilerine göre kazanımların, öğrenme alanlarına göre dağılımı belirlenmiş ve öğrenme alanını temsil etme oranı hesaplanmıştır. Uzman görüşlerinden elde edilen Kapsam Geçerlik İndeksleri (KGİ) ve Kappa istatistiği arasında uyum olup olmadığını belirlemek için Soru Değerlendirme Formu verileri kullanılmıştır.

Maddelerin KGO_i değerlerini hesaplamak amacıyla Soru Değerlendirme Formu verileri kullanılmıştır. Alan uzmanlarının maddeleri sınıflandırması sonucunda her bir madde için kapsam geçerlik oranı elde edilmiştir (Yurdugül ve Bayrak, 2012). Kapsam geçerlik oranı her bir madde için;

$$KGO_i = \frac{N_G - \frac{N}{2}}{\frac{N}{2}}$$

şeklinde ifade edilmektedir. Burada KGO_i , i. maddenin kapsam geçerlik oranını, N_G i. madde için gerekli diyen uzman sayısını, N ise toplam uzman sayısını göstermektedir. KGO_i -1 ile +1 arasında değer almaktadır. Tüm maddeler için elde edilen KGO_i değerlerinin ortalaması ise KGİ değerini vermektedir.

Bu değerler arasında manidar bir fark olup olmadığı ise z testi ile belirlenmiştir. Bu hesaplamalarda Office 2014 paket programı içerisinde bulunan Excel programı kullanılmıştır.

Yapı geçerliğine yönelik kanıt toplamaya başlamadan önce verilerin analiz varsayımlarını sağlayıp sağlamadığı incelenmiştir. Öncelikle puan dağılımlarının normalliği incelemek amacıyla Kolmogorov Smirnov testi yapılmış bu testin sonucu manidar bulunmasına rağmen puan dağılımlarının histogram grafikleri incelenip dağılımların çarpıklık ve basıklık katsayıları gözden geçirildiğinde verilerin normal dağılıma uygun olduğu sonucuna ulaşılmıştır. K-S testinin manidar çıkmasının nedeninin örneklem büyüklüğü olduğu düşünülmüştür. Çarpıklık değerleri Türkçe testinin ortak sınavı için -0,63, mazeret sınavı için 0,35; matematik testinin ortak sınavı için 0,65, mazeret sınavı için 1,61 olduğu gözlenmiştir. Basıklık değerlerinin ise Türkçe testinin ortak sınavı için -0,58, mazeret sınavı için -0,89; matematik testinin ortak sınavı için -0,36, mazeret sınavı için 3,25 olduğu gözlenmiştir. Verilerin normal dağılımdan aşırı sapma göstermediği söylenebilir. Buna göre değişken dağılımlarının normallik sayılımasını sağladığı iddia edilebilir (George ve Mallery, 2012; West, Finch, ve Curran, 1995).

Veri temizleme işlemi sonrası veri setlerinde kayıp veri olup olmadığı incelenmiş ve veri setlerinde kayıp veri bulunmadığı gözlenmiştir.

Veri setlerinde uç değerlerin bulunup bulunmadığı Mahalanobis uzaklıkları kullanılarak incelenmiş ve uç değerlerin bulunmadığı görülmüştür. Türkçe testi ortak sınavının Mahalanobis uzaklık değerlerinin 43,06 ile 2,13; mazeret sınavının ise 33,57 ile 4,59 arasında değiştiği ve ki-kare istatistiğinin anlamlı olmadığı gözlenmiştir ($p>0,001$). Benzer şekilde matematik testi ortak sınavının Mahalanobis uzaklık değerlerinin 36,24 ile 7,39; mazeret sınavının ise 36,42 ile 3,62 arasında değiştiği ve ki-kare istatistiğinin anlamlı olmadığı gözlenmiştir ($p>0,001$). Bu değerlere göre veri setlerinde uç değerlerin bulunmadığı yorumu yapılabilir.

Veri setlerinde çoklu bağlantı problemi olup olmadığını belirlemek için Tolerans değerleri, varyans şişkinlik faktörü ve koşul indeksleri incelenmiştir. İncelemeler sonucunda Tolerans değerinin 0,01'den büyük, varyans şişkinlik faktör değeri 10'dan küçük ve koşul indeksi 30'dan küçük olduğu gözlenmiştir. Buna göre çoklu bağlantı olmadığı söylenebilir (Kline, 2011; Tabachnik ve Fidell, 2012).

Yapı geçerliklerinin hangi düzeyde olduğunu araştırmak amacıyla Türkçe ve matematik testlerinin ortak ve mazeret sınavlarının faktör yapıları açıcı ve doğrulayıcı faktör analizleri ile araştırılmıştır. Faktör analizine başlamadan önce verilerin faktörleşmeye uygun olup olmadığı test edilmiştir. Bu amaçla Kaiser-Meyer-Olkin İstatistiği ve Bartlett Küresellik Testi sonuçları incelenmiştir. KMO değerinin yorumlanmasında Kaiser ve Rice (1974) tarafından belirtilen değerler dikkate alınmıştır.

KMO ve Bartlett Küresellik Testi sonuçları incelendiğinde Türkçe ortak sınavından elde edilen verilerin KMO değeri 0,96 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=43512,92$; $sd=190$; $p=0,00$). Türkçe mazeret sınavından elde edilen verilerin KMO değeri 0,94 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=10446,33$; $sd=190$; $p=0,00$).

Matematik testi ortak sınavından elde edilen verilerin KMO değeri 0,92 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=12648,95$; $sd=190$; $p=0,00$). Matematik testi mazeret sınavından elde edilen verilerin KMO değeri 0,84 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=1722,32$; $sd=190$; $p=0,00$). Bu sonuçlara göre tüm verilerin faktör analizine uygun olduğu söylenebilir.

DFA için bu araştırmada kullanılan uyum ölçütleri Schermelleh-Engel, Moosbrugger, ve Müller (2003)'den derlenerek aşağıdaki tabloda belirtilmiştir.

Tablo 1.Uyum Ölçütleri

| Uyum ölçütleri | İyi Uyum | Kabul Edilebilir Uyum |
|----------------|----------------------------|-----------------------------|
| χ^2/sd | $0 \leq \chi^2/sd \leq 2$ | $2 \leq \chi^2/sd \leq 3$ |
| p değeri | $0,05 < p \leq 1,00$ | $0,01 \leq p \leq 0,05$ |
| RMSEA | $0 \leq RMSEA \leq 0,05$ | $0,05 \leq RMSEA \leq 0,10$ |
| SRMR | $0 \leq SRMR \leq 0,05$ | $0,05 \leq SRMR \leq 0,10$ |
| NFI | $0,95 \leq NFI \leq 1,00$ | $0,90 \leq NFI \leq 0,95$ |
| NNFI | $0,97 \leq NNFI \leq 1,00$ | $0,95 \leq NNFI \leq 0,97$ |
| CFI | $0,95 \leq CFI \leq 1,00$ | $0,90 \leq CFI \leq 0,95$ |
| GFI | $0,95 \leq GFI \leq 1,00$ | $0,90 \leq GFI \leq 0,95$ |
| AGFI | $0,90 \leq AGFI \leq 1,00$ | $0,85 \leq AGFI \leq 0,90$ |

Uyum ölçütleri incelendiğinde χ^2/sd oranının 3'ten küçük olduğu durumda mükemmel uyum, 5'ten küçük olduğu durumda orta düzey uyum olduğunu belirten kaynaklar bulunmaktadır (Kline, 2011; Meydan ve Şeşen, 2015). Fakat bu oranın büyük ya da küçük olması araştırmacının bakış açısıyla ilişkilidir. Kesin olarak tanımlanan bir standart bulunmamaktadır. Uygulamada, uyumlu olduğu bilinen modellerde bile, önerilen bu oran 3,4,5 ve hatta 5'ten büyük de çıkabilmektedir. χ^2/sd oranının yorumlanması χ^2 istatistiğinin büyük örneklerde büyük çıkma eğilimini ortadan kaldırmamaktadır. Büyük örnekler büyük χ^2 istatistiği üretmekte bu da χ^2/sd oranının büyümesine neden olmaktadır. Çünkü serbestlik derecesi örneklem büyüklüğü kullanılarak hesaplanmamaktadır (Mueller, 1996).

AFA ve DFA'yı gerçekleştirmek amacıyla veriler öncelikle rastgele iki gruba ayrılmıştır. Grubun birinde AFA için tetrakorik korelasyon katsayılarına dayalı olarak faktör analizi yapılmış diğer grupta ise DFA uygulanmıştır. AFA için Statistica 10 ve Factor 10.3, DFA için Factor 10.3 ve Lisrel 8.80 programları kullanılmıştır.

BULGULAR

Araştırma sonucunda elde edilen bulgular, sırasıyla test ve madde istatistiklerinin karşılaştırılması, kapsam geçerliklerinin karşılaştırılması ve yapı geçerlik düzeylerinin karşılaştırılması başlıklarıyla verilmiştir.

Türkçe Testlerinin Test İstatistiklerinin Karşılaştırılması

Türkçe testine ait test istatistiklerinin ortak ve mazeret sınavlarına göre farklılık gösterip göstermediği sınanmış ve sonuçlar Tablo 2’de belirtilmiştir.

Tablo 2. Türkçe Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

| Test Adı | \bar{X} | t | Cohen’s d | \bar{p} | z | Cohen’s h | S_x^2 | F | Cohen’s d |
|----------------|-----------|--------|-----------|-----------|--------|-----------|---------|--------|-----------|
| Türkçe-Ortak | 14,30 | | | 0,72 | | | 22,58 | | |
| Türkçe-Mazeret | 9,93 | 41,12* | 0,91 | 0,49 | 21,49* | 0,48 | 24,73 | 12,05* | 0,07 |

*p<0.05

Ortak ve mazeret sınavlarının aritmetik ortalamaları arasında fark ortak sınav lehine manidar bulunmuştur ($t(4257,661)=41,12$, $p<0,05$). Bu karşılaştırmaya ait etki büyüklüğü değeri ise 0,91 olarak elde edilmiştir. Buna göre yüksek derecede etki büyüklüğüne sahip olduğu ve bu farkın örneklemelerden kaynaklanmadığı söylenebilir. Aynı zamanda sınavların ortalama güçlükleri incelendiğinde Türkçe testinin ortak sınavının ortalama güçlüğü 0,72, mazeret sınavının ortalama güçlüğü ise 0,49 olduğu gözlenmiştir. Ortak sınav ortalamasının, mazeret sınavına göre daha yüksek ve ortalama güçlüklerinin manidar düzeyde farklı olduğu ($z(12519)=21,49$, $p<0,05$) ve ortalama güçlük farkının etki büyüklüğü orta düzeyde (0,47) olarak bulunmuştur. Bu bulgulara dayalı olarak mazeret sınavının öğrenciler için bir dezavantaj oluşturdu yorumu yapılabilir.

Ortak ve mazeret sınavlarının varyansları arasında mazeret sınavı lehine manidar düzeyde fark bulunmuştur ($F(1,12519)=12,05$, $p<0,05$). Bu durum ortak sınava katılan grubun mazeret sınavına katılan gruba göre daha homojen bir grup olduğu anlamına gelmektedir. Ancak etki büyüklüğü incelendiğinde Cohen’s d istatistiği 0,07 olarak bulunduğu görülmektedir. Buna göre varyanslar arasındaki farkın çok küçük bir etki büyüklüğüne sahip olduğu, yani bu farkın örneklem büyüklüğünden kaynaklanmış olabileceği söylenebilir.

Tablo 3. Türkçe Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

| Test Adı | Güvenirlilik | z | Cohen’s q | r_{jk} Ortancası | Mann-Whitney U | r |
|----------------|--------------|-------|-----------|--------------------|----------------|------|
| Türkçe-Ortak | 0,94 | | | 0,57 | | |
| Türkçe-Mazeret | 0,91 | 9,74* | 0,21 | 0,64 | 193,00 | 0,03 |

*p<0.05

Tablo 3 incelendiğinde ortak ve mazeret sınavlarının McDonald ω değerlerinin karşılaştırılması sonucu ortak sınav lehine manidar düzeyinde fark bulunduğu görülmektedir ($z(12518)=9,74$, $p<0,05$). Bulunan bu farka yönelik elde edilen etki büyüklüğü incelendiğinde 0,21 olduğu görülmektedir. 0,10 küçük, 0,30 orta derecede etki büyüklüğüne sahip olduğu yönünde yorumlanmaktadır. Bu durumda güvenilirliklerin etki büyüklüğünün orta düzeyin altında olduğu söylenebilir. Ortak ve mazeret sınavlarının güvenilirliklerinin istatistiksel olarak farklı olması ölçme sonuçlarına karışan hata miktarlarının farklı olması anlamına gelebilir. Ancak bu sonuca analize alınan veri sayılarının farklı olması da neden olabilmektedir. Bu nedenle elde edilen etki büyüklüğü değerleri, güvenilirlik katsayıları arasındaki farkta örneklem büyüklüklerinin etkili olduğunu göstermektedir.

Ortak sınav ve mazeret sınavının madde ayırıcılık indeksleri ortancası (medyan) karşılaştırıldığında anlamlı fark gözlenmemiştir ($U=193$, $z=-0,19$, $p>0,05$, $r=0,03$). Etki büyüklüğünün çok düşük (0,03) olduğu görülmektedir. Ortak sınav ve mazeret sınavının bilen ve bilmeyen bireyleri birbirinden ayırma düzeylerinin benzer olduğu söylenebilir.

Türkçe ortak ve mazeret sınavlarının test istatistikleri incelendiğinde madde ayırıcılık indeksleri ortancaları dışındaki istatistiklerinin tümünün farklılaştığı görülmektedir. İki testin paralel olabilmesi için madde sayıları, güvenilirlikleri, varyansları, ortalamalarının eşit madde tiplerinin aynı olması gerekmektedir (Gulliksen, 1950). Ayrıca aynı kapsamı ya da yapıyı ölçmeli, her iki test de içerdiği maddelerle ölçülecek yapıyı eşit şekilde ölçmelidir (Bracken ve diğerleri., 2013). Burada testlerin paralel olup olmadığına karar verilirken testler aynı gruba uygulandıktan sonra adı geçen özellikleri incelenmektedir. Bu çalışma için iki farklı testi aynı gruba uygulama imkânı olmadığından ve testi alan grupların farklı olmasına rağmen tümünün sekizinci sınıfta olduğu göz önüne alınıp örneklem büyüklüklerinin de yeterli olduğu düşünüldüğünde testlerin özelliklerinin bu denli farklılaşmaması beklenebilir. Tüm sonuçlar birlikte değerlendirildiğinde Türkçe testi ortak ve mazeret sınavlarının psikometrik özelliklerinin farklılaşması nedeniyle birbirinin alternatifi olarak kullanılmayacağı ifade edilebilir.

Türkçe Testlerinin Madde İstatistiklerinin Karşılaştırılması

Türkçe testi ortak ve mazeret sınavlarında aynı kazanımı ölçen 11 madde bulunmaktadır (Tabloda 1-11. sıradaki maddeler). Kalan 9 maddenin 7'si aynı alt öğrenme alanında yer almaktadır (Tabloda 12-17. sıradaki maddeler). Hem aynı alt öğrenme alanından olmayan hem de aynı kazanımı ölçmeyen 2 madde ise diğer 2 madde ile eşleştirilerek karşılaştırılmıştır (Tabloda 19 ve 20. sıradaki maddeler). Eşleşen maddelerin madde istatistikleri arasında fark olup olmadığı araştırılmıştır. Türkçe ortak ve mazeret sınavında bulunan maddelerin madde güçlük indeksi ve madde ayırıcılık indeksleri açısından karşılaştırılması Tablo 4'te verilmiştir.

Tablo 4. Türkçe Testi Ortak ve Mazeret Sınavlarının Madde Güçlük ve Ayırıcılık İndekslerinin Karşılaştırılması

| S. No | Ortak Sınav Madde No | Mazeret Sınavı Madde No | p_j^* | p_j^{**} | z | Cohen's h | r_{jx}^* | r_{jx}^{**} | z | Cohen's q |
|-------|----------------------|-------------------------|---------|------------|--------|-----------|------------|---------------|---------|-----------|
| 1 | 1 | 1 | 0,66 | 0,41 | 23,18* | 0,51 | 0,69 | 0,46 | 16,23* | 0,35 |
| 2 | 2 | 3 | 0,63 | 0,40 | 21,50* | 0,46 | 0,68 | 0,54 | 10,41* | 0,22 |
| 3 | 6 | 8 | 0,64 | 0,61 | 3,16* | 0,06 | 0,68 | 0,63 | 4,06* | 0,09 |
| 4 | 7 | 5 | 0,82 | 0,69 | 13,95* | 0,30 | 0,53 | 0,65 | -8,57* | -0,19 |
| 5 | 8 | 7 | 0,74 | 0,65 | 9,87* | 0,20 | 0,56 | 0,66 | -7,41* | -0,16 |
| 6 | 10 | 14 | 0,80 | 0,64 | 16,78* | 0,36 | 0,53 | 0,68 | -11,06* | -0,24 |
| 7 | 12 | 16 | 0,83 | 0,53 | 33,25* | 0,66 | 0,47 | 0,67 | -13,92* | -0,30 |
| 8 | 14 | 12 | 0,49 | 0,39 | 8,77* | 0,20 | 0,68 | 0,5 | 12,95* | 0,28 |
| 9 | 15 | 10 | 0,64 | 0,34 | 27,73* | 0,61 | 0,63 | 0,44 | 12,46* | 0,27 |
| 10 | 19 | 19 | 0,65 | 0,46 | 17,90* | 0,38 | 0,67 | 0,53 | 10,21* | 0,22 |
| 11 | 20 | 20 | 0,66 | 0,38 | 27,13* | 0,57 | 0,46 | 0,44 | 1,16 | 0,03 |
| 12 | 5 | 4 | 0,92 | 0,54 | 47,65* | 0,92 | 0,27 | 0,78 | -35,57* | -0,77 |
| 13 | 18 | 17 | 0,50 | 0,34 | 14,98* | 0,33 | 0,67 | 0,43 | 16,24* | 0,35 |
| 14 | 4 | 2 | 0,81 | 0,65 | 18,15* | 0,36 | 0,53 | 0,66 | -9,38* | -0,20 |
| 15 | 9 | 9 | 0,73 | 0,48 | 24,01* | 0,52 | 0,57 | 0,71 | -11,09* | -0,24 |
| 16 | 11 | 6 | 0,84 | 0,32 | 54,83* | 1,12 | 0,46 | 0,32 | 7,67* | 0,17 |
| 17 | 13 | 13 | 0,84 | 0,57 | 29,94* | 0,61 | 0,46 | 0,75 | -22,02* | -0,48 |
| 18 | 16 | 11 | 0,65 | 0,63 | 2,04 | 0,04 | 0,77 | 0,72 | 5,22* | 0,11 |
| 19 | 3 | 18 | 0,76 | 0,34 | 41,62* | 0,87 | 0,49 | 0,52 | -1,86 | -0,04 |
| 20 | 17 | 15 | 0,69 | 0,55 | 13,45* | 0,29 | 0,60 | 0,66 | -4,61* | -0,10 |

* $p<0,05$, * Ortak Sınav, ** Mazeret Sınavı

Tablo 4 incelendiğinde madde güçlük indekslerinin ortak sınav için 0,49 ile 0,92; mazeret sınavı için 0,32 ile 0,69 arasında değiştiği görülmektedir. Ortak sınavda çok zor ve zor madde bulunmamaktadır. Mazeret sınavında ise çok zor ve çok kolay madde yoktur. Madde güçlüğü bakımından ortak sınavda orta güçlükte 2, kolay 12 ve çok kolay 6 madde bulunurken, mazeret sınavında zor 6, orta güçlükte 8 ve kolay 6 madde bulunmaktadır.

Mazeret sınavının gruba daha zor geldiği ve soruların güçlük düzeyi dağılımının ortak sınavdaki gibi olmadığı söylenebilir. Aynı ya da benzer kazanımı ölçen maddelerin madde güçlük düzeylerinin ortak sınavdaki 16. ve mazeret sınavındaki 11. maddeler için istatistiksel olarak anlamlı bir farklılık olmadığı diğer maddelerin ise istatistiksel olarak anlamlı düzeyde farklılaştığı görülmektedir. Bu durum, testi alan bireylerin farklı güçlükteki maddelere yanıt verdiğini göstermektedir. Birbirinin yerine kullanılan testlerin madde güçlüklerinin bu kadar farklılaşması testin adalet yönünü zedeleyeceği düşünülmektedir. Madde güçlük indeksi karşılaştırmalarının etki büyüklükleri incelendiğinde ise 0,04 ile 1,12 arasında değişen etki büyüklüklerine sahip oldukları görülmektedir. Bu karşılaştırmalara göre 2 karşılaştırmanın etki büyüklüğü çok küçük, 8 karşılaştırmanın etki büyüklüğü küçük, 7 karşılaştırmanın etki büyüklüğü orta ve 3 karşılaştırmanın ise etki büyüklüğü çok yüksektir. Buna göre maddelerin yarısının etki büyüklüklerinin orta ve üzerinde olduğu söylenebilir.

Tablo 4'te madde ayırıcılık indekslerinin ortak sınav için 0,27 ile 0,77; mazeret sınavı için ise 0,32 ile 0,78 arasında değiştiği görülmektedir. Herhangi bir maddenin, teste alınabilmesi için 0,30 ve üzerinde madde ayırıcılık indeksine sahip olması gerekmektedir (Erkuş, 2014). Ortak sınav bu açıdan incelendiğinde 5. maddenin madde ayırıcılık indeksinin 0,27 olduğu görülmektedir. Diğer 19 madde ise 0,41 ve üzerinde madde ayırıcılık indeksine sahiptir. Mazeret sınavında ise tüm maddeler 0,30'un üzerinde madde ayırıcılık değerine sahiptir.

Aynı ya da benzer kazanımı ölçen maddelerin, madde ayırıcılık indeksleri karşılaştırıldığında 2 madde haricindeki tüm maddelerde manidar fark bulunmuştur. Aynı kazanımı ölçen maddelerin büyük çoğunluğu madde ayırıcılık indeksi olarak farklılaşmasına rağmen testlerin madde ayırıcılık indeksi ortancaları farklılaşmamaktadır. Etki büyüklükleri işaretleri dikkate alınmadan incelendiğinde 0,03 ile 0,77 arasında değiştiği görülmektedir. 13 karşılaştırmanın küçük, 5'inin orta ve 2'sinin ise yüksek etki gücüne sahip olduğu söylenebilir.

Türkçe Testlerinin Kapsam Geçerlik Oranı İndekslerinin Karşılaştırılması

Tablo 5'te alt öğrenme alanlarına göre öğretim programında bulunan kazanım yüzdesi, ortak sınav ve mazeret sınavında bu alt öğrenme alanından kaç madde bulunduğu ve bu maddelerin yüzdesi verilmiştir.

Tablo 5. Türkçe Testi Ortak ve Mazeret Sınavında Bulunan Maddelerin Alt Öğrenme Alanına Göre Dağılımı

| Alt Öğrenme Alanı | Öğretim Programındaki Kazanım Yüzdesi (%) | Ortak Sınavda Bulunan Madde Sayısı | Ortak Sınavda Bulunan Madde Yüzdesi | Mazeret Sınavında Bulunan Madde Sayısı | Mazeret Sınavında Bulunan Madde Yüzdesi |
|---|---|------------------------------------|-------------------------------------|--|---|
| Anlatım Bozukluklarını Belirleme ve Düzeltme | 3,57 | 0 | 0 | 0 | 0 |
| Okuduğu Metni Değerlendirme | 3,57 | 0 | 0 | 0 | 0 |
| Yazım ve Noktalama Kurallarını Uygulama | 3,57 | 0 | 0 | 0 | 0 |
| Fiilimsiler İle İlgili Bilgi ve Kuralları Kavrama ve Uygulama | 7,14 | 2 | 10 | 1 | 5 |
| Söz Varlığını Zenginleştirme | 7,14 | 3 | 15 | 2 | 10 |
| Cümleyle İlgili Bilgi ve Kuralları Kavrama ve Uygulama | 19,64 | 4 | 20 | 5 | 25 |
| Okuduğu Metni Anlama ve Çözümleme | 55,36 | 11 | 55 | 12 | 60 |
| Toplam | 100 | 20 | 100 | 20 | 100 |

Tablo 5 incelendiğinde en çok kazanıma sahip olan alt öğrenme alanı okuduğu metni anlama ve çözümleme oluğu görülmektedir. Ortak sınavda toplam kazanımların yaklaşık %11'ine karşılık gelen anlatım bozukluklarını belirleme ve düzeltme, okuduğu metni değerlendirme ile yazım ve noktalama kurallarını uygulama alt öğrenme alanlarına ait madde bulunmamaktadır. En çok madde ise %55,36 kazanım yüzdesine sahip okuduğu metni anlama ve çözümleme alt öğrenme alanında bulunmaktadır. Alt öğrenme alanlarında bulunan kazanımların sayısı ve yüzdeleriyle birlikte ortak sınavdaki madde sayısı da azalmaktadır. Fakat %7,14 ağırlığa sahip olan fiilimsiler ile ilgili bilgi ve kuralları kavrama ve uygulama, söz varlığını zenginleştirme alt öğrenme alanlarından eşit miktarda madde bulunmamaktadır. Burada maddelerin sadece tek bir kazanımı ölçmediği diğer kazanımlarla da ilişkili olabileceği göz önünde bulundurulmalıdır. Ortak sınavda, 7 alt öğrenme alanından 4'üne ait madde bulunduğundan ortak sınavın alt öğrenme alanlarını %57,1 oranında temsil ettiği söylenebilir.

Mazeret sınavında da ortak sınavla benzer şekilde toplam kazanımların yaklaşık %11'ine karşılık gelen anlatım bozukluklarını belirleme ve düzeltme, okuduğu metni değerlendirme ile yazım ve noktalama kurallarını uygulama alt öğrenme alanlarına ait madde bulunmamaktadır. En çok madde ise okuduğu metni anlama ve çözümleme alt öğrenme alanından bulunmaktadır. Alt öğrenme alanlarında bulunan kazanımların sayısı ve yüzdelerine göre mazeret sınavındaki madde sayısı da azalmaktadır. Mazeret sınavında toplam 7 alt öğrenme alanından 4'üne ait madde bulunduğundan mazeret sınavının alt öğrenme alanlarını %57,1 oranında temsil ettiği söylenebilir.

Ortak ve mazeret sınavları için alt öğrenme alanlarında kazanım sayıları eşit olmasa da aynı alt öğrenme alanlarından maddeler bulunmakta farklı alt öğrenme alanlarından madde bulunmamaktadır. Bu açıdan bakıldığında benzer kapsamı ölçtükleri söylenebilir.

Ortak ve mazeret sınavlarının KGİ'leri karşılaştırılmış, ayrıca Yurdugül ve Bayrak'ın (2012) da belirttiği gibi Kappa istatistiği madde bazındaki KGİ'ye göre daha tutarlı sonuçlar ürettiğinden Kappa istatistiği de hesaplanmış ve sonuçlar Tablo 6'da verilmiştir.

Tablo 6. Türkçe Testi Ortak ve Mazeret Sınavları KGİ ve Kappa İstatistikleri Karşılaştırılması

| Ortak Sınav KGİ | Mazeret Sınavı KGİ | z | Cohen's h | Ortak Sınav Kappa | Mazeret Sınavı Kappa | z | Cohen's h |
|-----------------|--------------------|------|-----------|-------------------|----------------------|-------|-----------|
| 0,50 | 0,49 | 0,05 | 0,02 | 0,57 | 0,58 | -0,01 | -0,02 |

* $\alpha = 0,05$ düzeyinde manidardır

Tablo 6 incelendiğinde KGİ ve Kappa istatistiklerinin karşılaştırılması sonucunda KGİ ($z=0,05$, $p>0,05$) ve Kappa ($z= -0,01$, $p>0,05$) istatistikleri manidar düzeyde farklılaşmamıştır. Etki büyüklükleri incelendiğinde ise sırasıyla 0,02 ve -0,02 olduğu görülmektedir. Buna göre etki büyüklüğü yok denecek kadar azdır. Türkçe ortak ve mazeret sınavlarının kapsamlarının KGİ ve Fleiss Kappa katsayısı sonuçlarına göre benzer olduğu söylenebilir.

Türkçe Testlerinin Yapı Geçerlik Düzeylerinin Araştırılması

Türkçe Testi Ortak Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Türkçe testi ortak sınavının verileri üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak gerçekleştirilen faktör analizi sonucunda, özdeğeri 1'in üzerinde olan bir faktör bulunmuştur. Bu faktörün özdeğeri 10,88, açıklanan varyans oranı %54,40 olduğu gözlenmiştir. Birinci faktörün açıklanan varyans oranının %54,40 olduğu, öz değeri 1'in üzerinde sadece 1 faktör olması Türkçe testi ortak sınavının genel bir faktöre sahip olabileceğini düşündürmektedir. Tablo 7'de maddelerin faktör yükleri verilmiştir.

Tablo 7. Türkçe Testi Ortak Sınav Maddelerinin Faktör Yükleri

| Madde Numarası | Faktör Yüğü |
|-----------------------|-------------|
| 1 | 0,74 |
| 2 | 0,71 |
| 3 | 0,69 |
| 4 | 0,83 |
| 5 | 0,87 |
| 6 | 0,73 |
| 7 | 0,84 |
| 8 | 0,71 |
| 9 | 0,68 |
| 10 | 0,77 |
| 11 | 0,85 |
| 12 | 0,81 |
| 13 | 0,82 |
| 14 | 0,66 |
| 15 | 0,65 |
| 16 | 0,81 |
| 17 | 0,66 |
| 18 | 0,63 |
| 19 | 0,69 |
| 20 | 0,52 |
| Özdeğer | 10,88 |
| Açıklanan Varyans (%) | 54,40 |

Tablo 7 incelendiğinde faktör yük değerlerinin 0,52 (20. madde) ile 0,87 (5. madde) arasında deęiştii görülmektedir. 0,40'ın altında faktör yüküne sahip herhangi bir madde bulunmamaktadır. Açıklanan toplam varyans oranı ise %54,40'tır. Bu sonuçlar birlikte değerlendirildiğinde Türkçe testi ortak sınavının tek faktörlü bir yapıda olduđu söylenebilir. Bu sonuçlar literatürdeki diđer çalışmalarla da benzerlik göstermektedir (Anıl ve diđerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010).

20 maddeden oluşan Türkçe testi ortak sınavı için tek faktörlü bir ölçme modeli tanımlanmış ve bu model doğrulayıcı faktör analizi ile test edilmiştir. Analiz sonucunda göstergeler incelendiğinde faktör yük değerlerinin $\lambda=0,30$ 'un üzerinde olduđu ve 0,39 ile 0,77 arasında deęiştii, özgül varyansların ise tüm deęerlerde $\epsilon<0,90$ olduđu görülmüştür.

Tanımlanan tek faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 8'de sunulmuştur.

Tablo 8. Türkçe Testi Ortak Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

| Uyum İndeksleri | İstatistik Deęeri |
|-----------------|---------------------------|
| χ^2 | 5047,21 (p=0,000, sd=170) |
| χ^2/sd | 29,68 |
| GFI | 0,906 |
| AGFI | 0,884 |
| CFI | 0,977 |
| NFI | 0,976 |
| NNFI | 0,974 |
| RMSEA | 0,077 |
| RMR | 0,038 |
| Standardize RMR | 0,038 |

Tablo 8 incelendiğinde $\chi^2/sd=29,68$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistikleri sırasıyla 0,906, 0,884 ve 0,977 olduğu görülmektedir. NFI ve NNFI istatistikleri ise 0,976 ve 0,974 olarak bulunmuştur. RMSEA değeri 0,077, RMR ve standardize RMR değerleri ise 0,05'in altındadır. RMR ve SRMR indekslerinin 0,05'in altında olması mükemmel, 0,08'in altında olması iyi uyuma işaret etmektedir (Brown, 2015). Bu değerlere göre Türkçe testi ortak sınavının 20 maddelik tek faktörlü yapısı, DFA sonucunda doğrulandığı gözlenmiştir.

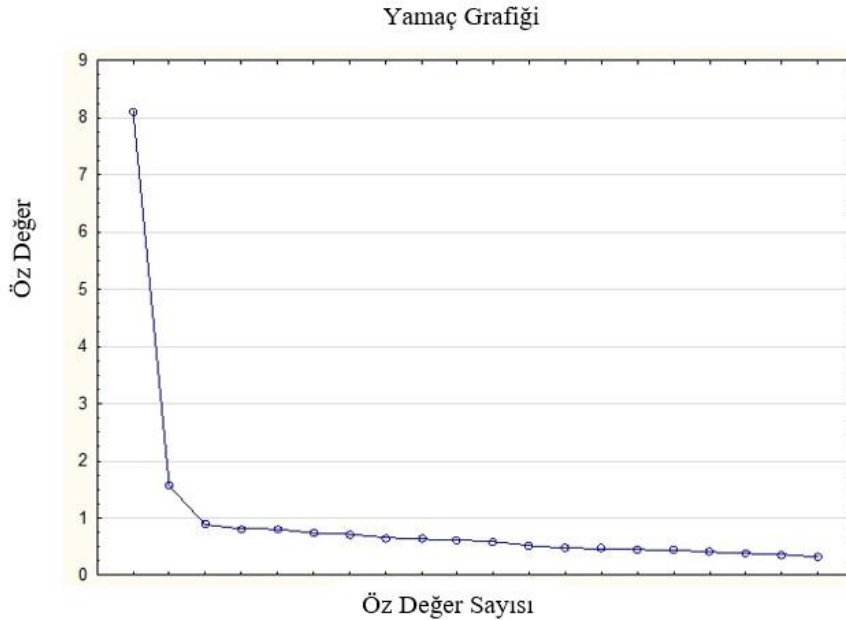
Türkçe Testi Mazeret Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Türkçe testi mazeret sınavından elde edilen verilerin üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak yürütülen faktör analizi sonucunda, döndürme uygulanmadan elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 9'da verilmiştir.

Tablo 9. Türkçe Testi Mazeret Sınavının AFA Sonuçları

| Faktörler | Faktör Özdeğeri | Açıklanan Varyans (%) | Toplam Varyans (%) |
|-----------|-----------------|-----------------------|--------------------|
| 1 | 8,19 | 40,97 | 40,97 |
| 2 | 1,63 | 8,19 | 49,16 |

Tablo 9 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde faktör öz değerlerine sahip 2 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %40,97, ikinci faktörde açıklanan varyansın %8,19 olduğu görülmektedir. Ayrıca özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 1'de verilmiştir.



Şekil 1. Türkçe Testi Mazeret Sınavına Ait Yamaç Grafiği

Faktör yapısına karar vermek için Şekil 1'de verilen yamaç grafiği incelendiğinde birinci faktörden sonra ani bir düşüş yaşandığı sonrasında ise grafiğin plato yaptığı görülmektedir.

Faktör analizi sonuçları genel olarak değerlendirildiğinde; birinci faktörün açıklanan varyans oranının %40,97 olması, öz değer grafiğinde birinci faktörden sonra ani bir düşüşün yaşanması testin genel bir faktöre sahip olabileceğini düşündürmektedir. Veriler tek faktörle sınırlandırılıp analiz tekrarlandığında Tablo 10'daki sonuç elde edilmiştir.

Tablo 10. Türkçe Testi Mazeret Sınavının Tek Faktörlü AFA Sonuçları

| Madde No | Faktör Yüğü |
|-----------------------|-------------|
| 1 | 0,48 |
| 2 | 0,72 |
| 3 | 0,59 |
| 4 | 0,78 |
| 5 | 0,73 |
| 6 | 0,47 |
| 7 | 0,71 |
| 8 | 0,67 |
| 9 | 0,71 |
| 10 | 0,51 |
| 11 | 0,73 |
| 12 | 0,57 |
| 13 | 0,77 |
| 14 | 0,71 |
| 15 | 0,66 |
| 16 | 0,67 |
| 17 | 0,52 |
| 18 | 0,65 |
| 19 | 0,53 |
| 20 | 0,46 |
| Özdeğer | 8,19 |
| Açıklanan Varyans (%) | 40,97 |

Tablo 10 incelendiğinde tek faktörlü yapıda faktör yük değerlerinin 0,40 (6. madde) ile 0,78 (4. madde) arasında değiştiği görülmektedir. 0,30'un altında faktör yüküne sahip herhangi bir madde bulunmamaktadır. Ayrıca açıklanan toplam varyans %40,52 olduğundan Türkçe testi mazeret sınavının tek faktörlü bir yapıda olduğu düşünülmüştür. Bu sonuçlar literatürdeki diğer çalışmalarla da benzerlik göstermektedir (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010).

Türkçe testi mazeret sınavının faktör yapısına ilişkin tek faktörlü bir ölçme modeli tanımlanmış ve bu model doğrulayıcı faktör analizi ile test edilmiştir. Analiz sonucunda oluşturulan ölçme modeli incelendiğinde, göstergelere ilişkin faktör yük değerlerinin $\lambda=0,30$ 'un üzerinde olduğu ve 0,34 ile 0,76 arasında değiştiği, özgül varyansların ise tüm değerlerde $\epsilon<0,90$ olduğu görülmüştür.

Tanımlanan tek faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 11'de sunulmuştur.

Tablo 11. Türkçe Testi Mazeret Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

| Uyum İndeksleri | İstatistik Değeri |
|-----------------|---------------------------|
| χ^2 | 2087,47 (p=0,000, sd=170) |
| χ^2/sd | 12,27 |
| GFI | 0,867 |
| AGFI | 0,851 |
| CFI | 0,956 |
| NFI | 0,951 |
| NNFI | 0,950 |
| RMSEA | 0,091 |
| RMR | 0,058 |
| Standardize RMR | 0,058 |

Tablo 11 incelendiğinde $\chi^2/sd=12,27$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistikleri sırasıyla 0,867, 0,851 ve 0,956 olduğu görülmektedir. GFI 0,90 ve üzerinde olduğu durumda kabul edilebilir uyum olduğu düşünülmektedir. Fakat GFI değeri örneklem büyüklüğünden etkilenen bir istatistik olduğu için 0,867 olarak bulunmuştur. Buna rağmen diğer tüm değerler uyumlu olduğunu gösterdiği için diğer istatistiklerde incelenmiştir. NFI ve NNFI istatistikleri ise 0,951 ve 0,950 olarak bulunmuştur. RMSEA değeri 0,091, RMR ve standardize RMR değerleri ise 0,058 olarak bulunmuştur. Bu değerlere göre Türkçe testi mazeret sınavının 20 maddelik tek faktörlü yapısı DFA sonucunda doğrulanmıştır.

Türkçe testi ortak ve mazeret sınavlarının yapı geçerlik düzeylerini incelemek amacıyla AFA ve DFA gerçekleştirilmiş ve her iki sınav için de tek faktörlü bir yapı tanımlanmıştır. Bu yapılarda tüm maddeler faktörlerine yüklenmiş ve faktör yükü 0,30'un altında herhangi bir madde olmamıştır. Bu durumda yapı geçerliği açısından Türkçe testi ortak ve mazeret sınavlarının birbirine yakın düzeyde olduğunu söylemek mümkündür.

Matematik Testlerinin Test İstatistiklerinin Karşılaştırılması

Matematik testi ortak ve mazeret sınavları için test istatistiklerinin farklılık gösterip göstermediği sınıanmış ve sonuçlar Tablo 12'de belirtilmiştir.

Tablo 12. Matematik Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

| Test Adı | \bar{X} | t | Cohen's d | \bar{p} | z | Cohen's h | S_x | F | Cohen's d |
|-------------------|-----------|--------|-----------|-----------|--------|-----------|-------|---------|-----------|
| Matematik-Ortak | 9,48 | | | 0,47 | | | 19,46 | | |
| Matematik-Mazeret | 5,89 | 42,68* | 1,28 | | 15,40* | 0,35 | | 401,41* | 0,46 |
| | | | | 0,30 | | | 11,63 | | |

*p<0.05

Matematik ortak ve mazeret sınavlarının aritmetik ortalamaları arasında fark, ortak sınav lehine manidar bulunmuştur ($t(4431,75)=42,68$, $p<0,05$). Etki büyüklüğü 1,28 olarak bulunduğu görülmektedir. Buna göre ortalamalar arasındaki farkın çok yüksek düzeyde etki büyüklüğüne sahip olduğu, bu farkın örneklem büyüklüğünden kaynaklanmadığı söylenebilir. Aynı zamanda sınavların ortalama güçlükleri incelendiğinde matematik testi ortak sınavının ortalama güçlüğü 0,47, mazeret sınavının ortalama güçlüğü ise 0,30 olduğu gözlenmiştir. Matematik testinin ortak sınavının ortalaması, mazeret sınavına göre daha yüksek olduğu ve ortalama güçlüklerinin manidar düzey farklılaşması ($z(11806)=15,40$, $p<0,05$), mazeret sınavının öğrenciler için bir dezavantaj oluşturduğunu düşündürmektedir. Ortalama güçlük karşılaştırmasına yönelik olarak elde edilen etki büyüklüğü 0,35 olarak bulunmuştur. Buna göre orta düzeye yakın bir etki büyüklüğüne sahip olduğu söylenebilir.

Ortak ve mazeret sınavlarının varyansları arasında ortak sınav lehine manidar düzeyde fark bulunmuştur ($F(1,11807)=401,41$, $p<0,05$). Bu durum ortak sınava katılan grubun mazeret sınavına katılan gruba göre daha homojen bir grup olduğu anlamına gelmektedir. Varyansların karşılaştırmasına yönelik olarak elde edilen etki büyüklüğü değeri 0,46 olarak bulunmuştur. Buna göre orta düzeyde bir etki büyüklüğüne sahip olduğu söylenebilir.

Tablo 13. Matematik Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

| Test Adı | Güvenirlilik | z | Cohen's q | r_{jx} Ortancası | Mann-Withney U | r |
|-------------------|--------------|--------|-----------|--------------------|----------------|------|
| Matematik-Ortak | 0,88 | | | 0,57 | | |
| Matematik-Mazeret | 0,81 | 11,94* | 0,25 | 0,32 | 62,50* | 0,58 |

*p<0.05

Tablo 13’de görüldüğü üzere sınavların McDonald ω değerleri karşılaştırıldığında ortak sınav lehine manidar fark bulunmuştur ($z(11807)=11,94$, $p<0,05$). Etki büyüklüğü incelendiğinde ise Cohen’s q istatistiği 0,25 olduğu görülmektedir. Buna göre orta düzeyde bir etki büyüklüğüne sahip olduğu, farkın gözlenmesinde örneklem büyüklüğünün etkisinin az olduğu söylenebilir.

Testlerin madde ayırıcılık indeksi ortancaları (medyan) kıyaslandığında ortak sınav lehine manidar fark bulunmuştur ($U=62,50$, $z=-3,72$, $p<0,05$, $r=0,58$). Etki büyüklüğü de incelendiğinde yüksek düzeyde etki büyüklüğüne sahip olduğu görülmektedir. Bu durum testlerin, testte ölçülen kazanımlara sahip olan ve olmayan bireyleri birbirinden ayırma düzeylerinin farklı olduğu anlamına gelmektedir.

Tüm bu karşılaştırmalar incelendiğinde matematik ortak ve mazeret sınavlarının birbirinin alternatifi olarak kullanılmamasının daha uygun olacağı söylenebilir.

Matematik Testlerinin Madde İstatistiklerinin Karşılaştırılması

Matematik ortak ve mazeret sınavlarındaki 13 maddenin (Tablo 14’te 1-13. sıradaki maddeler) aynı kazanımı ölçtüğü, kalan 7 maddenin ise farklı kazanımları ölçmekle beraber bu 7 maddenin 5’inin (Tablo 14’te 14-18. sıradaki maddeler) aynı alt öğrenme alanında bulunduğu ve konu olarak eşleştiği gözlenmiştir. Tabloda 14’te 19. sırada yer alan ortak sınavdaki madde üçgenler alt öğrenme alanındayken mazeret sınavındaki madde ise üçgenlerde ölçme alt öğrenme alanındadır. Birbirine çok yakın alt öğrenme alanlarından olduğu için bu maddeler birbiriyle eşleştirilmiştir. Kalan 1 madde (Tablo 14’te 20. sıradaki maddeler) ise farklı öğrenme alanlarındadır. Buna göre maddeler eşleştirilmiş ve eşleşen maddelerin madde istatistikleri arasında fark olup olmadığı araştırılmıştır. Matematik ortak ve mazeret sınavında bulunan maddelerin madde güçlükleri ve madde ayırıcılık indeksleri açısından karşılaştırılması Tablo 14’te verilmiştir.

Tablo 14. Matematik Testi Ortak ve Mazeret Sınavlarının Madde Güçlük ve Ayırıcılık İndekslerinin Karşılaştırılması

| S.No | Ortak Sınav Madde No | Mazeret Sınavı Soru No | p_j^* | p_j^{**} | z | Cohen’s h | r_{jx}^* | r_{jx}^{**} | z | Cohen’s q |
|------|----------------------|------------------------|---------|------------|--------|-------------|------------|---------------|--------|-------------|
| 1 | 6 | 11 | 0,65 | 0,37 | 24,52* | 0,57 | 0,62 | 0,46 | 9,83* | 0,23 |
| 2 | 11 | 9 | 0,61 | 0,29 | 28,49* | 0,66 | 0,57 | 0,39 | 10,18* | 0,24 |
| 3 | 13 | 2 | 0,52 | 0,35 | 14,84* | 0,34 | 0,68 | 0,28 | 23,38* | 0,54 |
| 4 | 7 | 12 | 0,39 | 0,43 | -3,20* | -0,08 | 0,42 | 0,30 | 5,97* | 0,14 |
| 5 | 10 | 8 | 0,49 | 0,24 | 21,92* | 0,53 | 0,56 | 0,28 | 14,90* | 0,35 |
| 6 | 3 | 5 | 0,83 | 0,36 | 46,33* | 1,00 | 0,36 | 0,51 | -8,02* | -0,19 |
| 7 | 17 | 16 | 0,40 | 0,38 | 2,09 | 0,04 | 0,62 | 0,31 | 17,46* | 0,40 |
| 8 | 8 | 14 | 0,50 | 0,25 | 21,64* | 0,52 | 0,75 | 0,34 | 26,72* | 0,62 |
| 9 | 18 | 18 | 0,41 | 0,29 | 11,29* | 0,25 | 0,71 | 0,31 | 24,46* | 0,57 |
| 10 | 19 | 19 | 0,32 | 0,32 | 0,22 | 0,00 | 0,63 | 0,42 | 12,68* | 0,29 |
| 11 | 20 | 20 | 0,33 | 0,22 | 10,41* | 0,25 | 0,20 | 0,17 | 1,34 | 0,03 |
| 12 | 4 | 6 | 0,44 | 0,28 | 14,44* | 0,34 | 0,47 | 0,29 | 9,13* | 0,21 |
| 13 | 5 | 15 | 0,71 | 0,33 | 34,16* | 0,78 | 0,55 | 0,33 | 11,90* | 0,28 |
| 14 | 1 | 3 | 0,45 | 0,23 | 19,01* | 0,47 | 0,60 | 0,31 | 16,09* | 0,37 |
| 15 | 9 | 10 | 0,37 | 0,15 | 20,11* | 0,51 | 0,60 | 0,29 | 17,04* | 0,39 |
| 16 | 12 | 13 | 0,55 | 0,37 | 15,67* | 0,36 | 0,76 | 0,37 | 26,24* | 0,61 |
| 17 | 15 | 7 | 0,33 | 0,27 | 5,53* | 0,13 | 0,30 | 0,34 | -1,92 | -0,04 |
| 18 | 16 | 17 | 0,36 | 0,24 | 10,64* | 0,26 | 0,46 | 0,23 | 11,36* | 0,26 |
| 19 | 2 | 4 | 0,31 | 0,31 | 0,09 | 0,00 | 0,50 | 0,37 | 6,95* | 0,16 |
| 20 | 14 | 1 | 0,51 | 0,24 | 23,18* | 0,57 | 0,26 | 0,32 | -2,83* | -0,07 |

* $p<0,05$, * Ortak Sınav, ** Mazeret Sınavı

Tablo 14 incelendiğinde madde güçlük indekslerinin ortak sınav için 0,31 ile 0,83; mazeret sınavı için 0,15 ile 0,43 arasında değiştiği görülmektedir. Ortak sınavda çok zor madde bulunmamaktadır. Mazeret sınavında ise kolay ve çok kolay madde yoktur. Madde güçlüğü bakımından ortak sınavda

zor 8, orta güçlükte 8, kolay 3 ve çok kolay 1 madde bulunurken, mazeret sınavında çok zor 1, zor 18, ve orta güçlükte 1 madde bulunmaktadır.

Mazeret sınavının gruba daha zor geldiği ve soruların güçlük düzeyi dağılımının ortak sınavdaki gibi olmadığı söylenebilir. Aynı ya da benzer kazanımı ölçen maddelerin madde güçlük düzeyleri bakımından 7., 10. ve 19. sıradaki maddeler için istatistiksel olarak anlamlı düzeyde farklılık olmadığı diğer maddelerin ise istatistiksel olarak anlamlı düzeyde farklılaştığı görülmektedir. Bu durum, testi alan bireylerin farklı güçlükteki maddelere yanıt verdiğini göstermektedir. Birbirinin yerine kullanılan testlerin madde güçlüklerinin bu kadar farklılaşması testin adalet yönünü zedeleyeceği düşünülmektedir. Madde güçlük indeksi karşılaştırmalarının etki büyüklükleri incelendiğinde ise -0,08 ile 1,00 arasında değiştiği görülmektedir. Buna göre; 4 karşılaştırmanın etki büyüklüğü çok küçük, 7 karşılaştırmanın etki büyüklüğü küçük, 7 karşılaştırmanın etki büyüklüğü orta ve 2 karşılaştırmanın ise etki büyüklüğü çok yüksektir. Karşılaştırmaların yaklaşık yarısının etki büyüklüklerinin orta ve üzerinde olduğu söylenebilir.

Tablo 14'te ortak sınavın madde ayırıcılık indekslerinin 0,20 ile 0,76 arasında değiştiği görülmektedir. 20. ve 14. maddelere ait madde ayırıcılık indeksleri sırasıyla 0,20 ve 0,26 olarak bulunmuştur. Diğer 18 madde ise 0,30 ve üzerinde madde ayırıcılık indeksine sahiptir. Mazeret sınavının madde ayırıcılık indeksleri incelendiğinde 0,17 ile 0,51 arasında değiştiği görülmektedir. Mazeret sınavında madde ayırıcılık indeksi 0,20'nin altında olan 1 madde (20. madde), 0,20-0,30 arasında olan 6 madde (2., 6., 8., 10., 12. ve 17. maddeler) bulunmaktadır.

Aynı ya da benzer kazanımı ölçen maddelerin madde ayırıcılık indeksleri karşılaştırıldığında, 2 madde haricindeki tüm maddelerde $\alpha=0,05$ düzeyinde manidar bulunmuştur. Ortak sınavın madde ayırıcılık indeksleri daha yüksek olduğu için ortak sınavın mazeret sınavında göre bu psikometrik özellik açısından daha uygun bir test olduğu söylenebilir. Etki büyüklükleri incelendiğinde etki büyüklüğünün -0,19 ile 0,62 arasında değiştiği Tablo 14'te görülmektedir. Buna göre karşılaştırmalarının büyük çoğunluğunun etki büyüklüğü orta ve altında iken 4 adet karşılaştırmanın yüksek düzeyde etki büyüklüğüne sahip olduğu gözlenmiştir.

Matematik Testlerinin Kapsam Geçerlik Oranı İndekslerinin Karşılaştırılması

Tablo 15'te alt öğrenme alanlarına göre öğretim programında bulunan kazanım yüzdesi, ortak sınav ve mazeret sınavında bu alt öğrenme alanından kaç madde bulunduğu ve bu maddelerin yüzdesi verilmiştir.

Tablo 15. Matematik Ortak ve Mazeret Sınavlarında Bulunan Maddelerin Alt Öğrenme Alanına Göre Dağılımı

| Alt Öğrenme Alanı | Öğretim Programındaki Kazanım Yüzdesi | Ortak Sınavda Bulunan Madde Sayısı | Ortak Sınavda Bulunan Madde Yüzdesi | Mazeret Sınavında Bulunan Madde Sayısı | Mazeret Sınavında Bulunan Madde Yüzdesi |
|------------------------------------|---------------------------------------|------------------------------------|-------------------------------------|--|---|
| Örüntü ve Süslemeler | 1,92 | 0 | 0 | 0 | 0 |
| Tablo ve Grafikler | 1,92 | 0 | 0 | 0 | 0 |
| Olasılık Çeşitleri | 1,92 | 0 | 0 | 0 | 0 |
| Gerçek sayılar | 1,92 | 0 | 0 | 1 | 5 |
| Merkezi eğilim ve yayılma ölçüleri | 3,85 | 0 | 0 | 0 | 0 |
| Örüntüler ve İlişkiler | 3,85 | 1 | 5 | 1 | 5 |
| Dönüşüm Geometrisi | 3,85 | 1 | 5 | 1 | 5 |
| Olası Durumları Belirleme | 3,85 | 1 | 5 | 1 | 5 |
| Olay Çeşitleri | 3,85 | 1 | 5 | 1 | 5 |
| Üslü Sayılar | 5,77 | 1 | 5 | 1 | 5 |
| Köklü Sayılar | 7,69 | 1 | 5 | 1 | 5 |
| Üçgenlerde Ölçme | 7,69 | 2 | 10 | 3 | 15 |
| Cebirsel İfadeler | 9,62 | 2 | 10 | 2 | 10 |
| Geometrik Cisimler | 11,54 | 2 | 10 | 2 | 10 |
| Denklemler | 13,46 | 4 | 20 | 4 | 20 |
| Üçgenler | 17,31 | 4 | 20 | 2 | 10 |
| Toplam | 100 | 20 | 100 | 20 | 100 |

Tablo 15 incelendiğinde en çok kazanıma ve en çok ders saatine sahip olan alt öğrenme alanı üçgenlerdir. Toplam kazanımın %13,5'ine karşılık gelen örüntü ve süslemeler, tablo ve grafikler, olasılık çeşitleri, gerçek sayılar ve merkezi eğilim ve yayılma ölçüleri alt öğrenme alanlarına ait ortak sınavda madde bulunmamaktadır. En çok madde ise %17,31 kazanım yüzdesine sahip üçgenler alt öğrenme alanında bulunmaktadır. Denklemler alt öğrenme alanında toplam kazanımın %9,62'si bulunmasına rağmen ortak sınavda bu öğrenme alanına ait 4 madde bulunmaktadır. Geometrik cisimler alt öğrenme alanında ise 2 madde vardır. Oysaki üçgenler alt öğrenme alanından sonra en çok kazanım bulunan öğrenme alanı geometrik cisimler alt öğrenme alanıdır. Burada maddelerin sadece tek bir kazanımı ölçmediği diğer kazanımlarla da ilişkili olabileceği göz önünde bulundurulmalıdır. Ortak sınavda toplam 16 alt öğrenme alanından 11'ine ait madde bulunduğu ortak sınavın alt öğrenme alanlarını %68,75 oranında temsil ettiği söylenebilir.

Toplam kazanımın %9,61'ine karşılık gelen örüntü ve süslemeler, tablo ve grafikler, olasılık çeşitleri ve merkezi eğilim ve yayılma ölçüleri alt öğrenme alanlarından mazeret sınavında madde bulunmamaktadır. Mazeret sınavında ortak sınava benzer şekilde denklemler alt öğrenme alanından 4 madde bulunmaktadır. Fakat ortak sınavda üçgenler alt öğrenme alanından 4 madde varken mazeret sınavında 2 madde bulunmaktadır. Mazeret sınavında toplam 16 alt öğrenme alanından 12'sine ait madde bulunduğu mazeret sınavının alt öğrenme alanlarını %75 oranında temsil ettiği söylenebilir.

Mazeret sınavında, gerçek sayılar alt öğrenme alanından 1 madde bulunmasına rağmen ortak sınavda bu alt öğrenme alanından madde bulunmamaktadır. Üçgenlerde ölçme alt öğrenme alanından ortak sınavda 2, mazeret sınavında ise 3 madde bulunmaktadır. Üçgenler alt öğrenme alanından ise ortak sınavda 4, mazeret sınavında ise 2 madde vardır. Diğer alt öğrenme alanlarından eşit sayıda madde bulunmaktadır.

Ortak ve mazeret sınavlarının KGİ'leri ve Kappa istatistikleri karşılaştırılmış ve sonuçlar Tablo 16'da verilmiştir.

Tablo 16. Matematik Ortak ve Mazeret Sınavları KGİ ve Kappa İstatistikleri Karşılaştırılması

| Ortak Sınav KGİ | Mazeret Sınavı KGİ | z | Cohen's h | Ortak Sınav Kappa | Mazeret Sınavı Kappa | z | Cohen's h |
|--------------------|-----------------------|------|-----------|----------------------|----------------------------|-------|--------------|
| 0,46 | 0,54 | 0,42 | 0,16 | 0,69 | 0,75 | -0,29 | -0,13 |

* $\alpha = 0,05$ düzeyinde manidardır.

Matematik ortak ve mazeret sınavlarına ilişkin hesaplanan KGİ ($z=0,42$, $p>0,05$) ve Kappa istatistikleri ($z= -0,29$, $p>0,05$) arasında anlamlı farklılıklar gözlenmemiştir. Etki büyüklükleri de incelendiğinde çok küçük oldukları görülmektedir. Matematik ortak ve matematik mazeret sınavlarının KGİ ve Fleiss Kappa katsayısı açısından benzer olduğu söylenebilir. Yılmaz-Koğar ve Aygün, (2015)'de 2013-2014 eğitim öğretim yılında uygulanan TEOG sınavlarından matematik alt testinin ortak sınavını iki dönem içinde incelemiş ve kapsam geçerliği için gerekli ölçütü sağladığını belirtmiştir.

Matematik ortak ve mazeret sınavlarının benzer kapsamı ölçmesine rağmen madde güçlük ve ayrıricılık indeksleri gibi madde istatistikleri bakımından farklılaştığı görülmektedir. Benzer kapsamı ölçen testlerin, madde istatistikleri açısından farklılaşması testin uygulandığı grupların birbirinden farklı özelliklerde olmasından kaynaklanabileceği düşünülmektedir. Madde ve test istatistikleri gruba bağımlı istatistikler olduğu için aynı kapsamı ölçen maddelerde de farklılaşabilmektedir.

Matematik Testlerinin Yapı Geçerlik Düzeylerinin Araştırılması

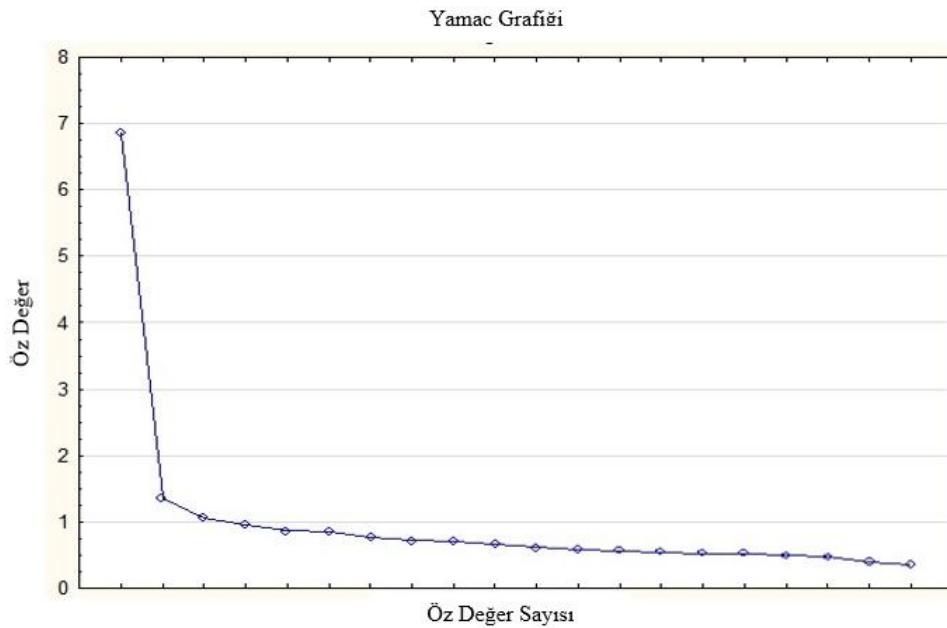
Matematik Testi Ortak Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Matematik testi ortak sınavının verileri üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak gerçekleştirilen faktör analizi sonucunda, elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 17'de verilmiştir.

Tablo 17. Matematik Ortak Sınavının AFA Sonuçları

| Faktörler | Faktör Özdeğerleri | Açıklanan Varyans (%) | Toplam Varyans (%) |
|-----------|--------------------|-----------------------|--------------------|
| 1 | 6,95 | 34,75 | 34,75 |
| 2 | 1,35 | 6,79 | 41,54 |
| 3 | 1,10 | 5,50 | 47,04 |

Tablo 17 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde öz değere sahip 3 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %34,75 olduğu ve üç faktörün tamamının %47,04 oranında toplam varyansı açıkladığı görülmektedir. Faktörlerin öz değerleri birinci faktör için 6,95, ikinci faktör için 1,35 ve üçüncü faktör için 1,10 olarak bulunmuştur. Ayrıca özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 2’de verilmiştir.



Şekil 2. Matematik Testi Ortak Sınavına Ait Yamaç Grafiği

Faktör yapısına karar vermeden önce Şekil 2’de verilen yamaç grafiği de incelenmiş birinci faktörden sonra ani ve ivmeli bir düşüş olduğu fakat diğer faktörlerden sonra grafiğin yatay olarak devam ettiği ve ani bir düşüş olmadığı gözlenmiştir.

Faktör analizi sonuçları uzman görüşlerine göre de değerlendirilmiş ve matematik testinin cebir ve geometri olarak iki yapıda olabileceği sonucuna varılmıştır. Matematik ortak sınavı için faktör analizi iki faktörle sınırlandırılarak faktör analizi tekrarlanmış ve eğik (oblique) döndürme yöntemlerinden promax yöntemi kullanılarak faktörler döndürülmüştür.

Uzman görüşleri doğrultusunda cebir ve geometri olarak iki faktörlü bir yapıda olduğu düşünülen Matematik ortak sınavındaki maddelerin iki faktöre göre dağılımı Tablo 18’de verilmiştir.

Tablo 18. Matematik Ortak Sınavı 2 Faktörlü, Döndürülmüş AFA için Maddelerin Faktörlere Göre Dağılımı

| Madde No | Cebir | Geometri |
|-----------------------|---------------|---------------|
| 1 | 0,27 | 0,43* |
| 2 | 0,09 | 0,61** |
| 3 | 0,74** | -0,23 |
| 4 | 0,40* | 0,11 |
| 5 | 0,82* | -0,23 |
| 6 | 0,82* | -0,17 |
| 7 | -0,14 | 0,71** |
| 8 | 0,70** | 0,09 |
| 9 | 0,30 | 0,45* |
| 10 | 0,30 | 0,31** |
| 11 | 0,70** | -0,15 |
| 12 | 0,74* | 0,03 |
| 13 | 0,61* | 0,10 |
| 14 | 0,25* | -0,08 |
| 15 | -0,24 | 0,67* |
| 16 | 0,05 | 0,56** |
| 17 | 0,41** | 0,35 |
| 18 | 0,55** | 0,25 |
| 19 | 0,45** | 0,36 |
| 20 | -0,33 | 0,60** |
| Özdeğer | 6,75 | 1,36 |
| Açıklanan Varyans (%) | 34,75 | 6,79 |

*Uzman görüşüne göre diğer faktörde yer alan fakat analiz sonucunda mevcut faktöre yük veren madde.

** Faktör yükü ve uzman görüşüne göre mevcut faktörde yer alan madde.

Tablo 18 incelendiğinde ortak sınavda bulunan 20 maddeden 12'si cebir faktöründe, 8'i geometri faktöründe yer almaktadır. Bu şekilde oluşturulan yapı toplam varyansın %41,54'ünü açıklamaktadır. Matematik ortak sınavı uzman görüşleri doğrultusunda incelendiğinde cebir konu alanıyla ilgili olan 9 madde, (1., 3., 8., 9., 11., 15., 17., 18. ve 19. maddeler) geometri konu alanıyla ilgili 11 madde (2., 4., 5., 6., 7., 10., 12., 13., 14., 16. ve 20. madde) olduğu görüşüne ulaşılmıştır. Bu duruma göre FA sonuçları incelendiğinde 1., 9. ve 15. maddenin cebir konu alanıyla ilgili olmasına rağmen geometri alanını ölçüyor gibi görünmektedir. Benzer şekilde 4., 5., 6., 12., 13., 14. maddeler geometri konu alanıyla ilgili olmasına rağmen cebir konu alanını ölçüyor gibi görünmektedir. Bu durum testin yapı geçerliğini zedeleyici bir etki yapmaktadır. Literatürdeki diğer çalışmalar incelendiğinde genellikle bu tür sınavlarda tek faktörlü yapılar elde edilmiştir. Ancak bu çalışmalarda faktörlere yük vermeyen maddeler analiz dışı bırakılmıştır (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010). Fakat burada uzman görüşleri ve faktör yükleri dikkate alınarak 2 faktörlü bir yapı tanımlanmıştır. Tanımlanan bu yapı DFA ile test edilmiştir. Analiz sonucu tanımlanan bu ölçme modeli incelendiğinde göstergelere ilişkin faktör yük değerlerinin 14. ve 20. maddeler hariç $\lambda=0,30$ 'un üzerinde olduğu ve 0,31 ile 0,74 arasında değiştiği, özgül varyanslara bakıldığında genellikle $\epsilon<0,90$ olduğu ancak 14. ve 20. maddeler için 0,90'ın üzerinde olduğu görülmüştür. Tanımlanan 2 faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 19'da sunulmuştur.

Tablo 19. Matematik Ortak Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

| Uyum İndeksleri | İstatistik Değeri |
|-----------------|--|
| χ^2 | 2772,38 (p=0,000, s _d =169) |
| χ^2/sd | 16,404 |
| GFI | 0,944 |
| AGFI | 0,931 |
| CFI | 0,970 |
| NFI | 0,968 |
| NNFI | 0,966 |
| RMSEA | 0,057 |
| RMR | 0,038 |
| Standardize RMR | 0,038 |

Tablo 19 incelendiğinde $\chi^2/sd=16,404$ olarak bulunmuştur. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistiklerinin sırasıyla 0,944, 0,931, 0,970 olduğu görülmektedir. NFI ve NNFI istatistikleri ise 0,968 ve 0,966 olarak bulunmuştur. RMSEA, RMR ve standardize RMR değerleri 0,057 ve 0,038'dir. Bu değerlere göre matematik ortak sınavının 20 maddelik iki faktörlü yapısı uyumlu görülmüştür. Fakat 2 maddenin ilişki katsayısının 0,30'un altında olması maddelerin faktörlerle ilişkilerinin çok düşük düzeyde olduğu anlamına gelmektedir. Bu durum da yapı geçerlik düzeyini düşürücü bir etki yapmaktadır.

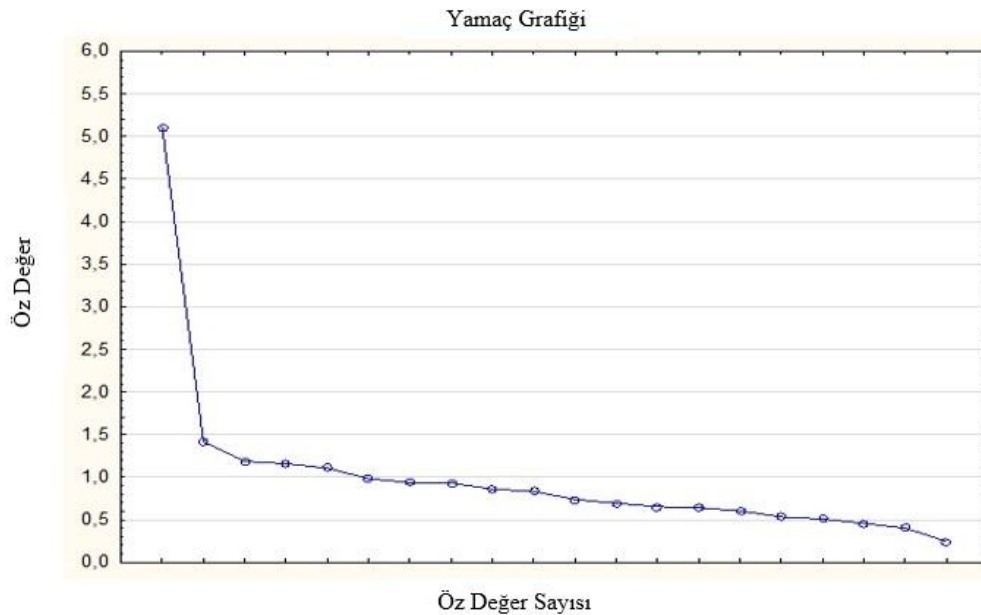
Matematik Mazeret Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Matematik mazeret sınavının faktör yapısını belirleyebilmek amacıyla temel bileşenler faktör çıkarma yöntemi kullanılarak AFA gerçekleştirilmiş ve döndürme uygulanmadan elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 20'de verilmiştir.

Tablo 20. Matematik Mazeret Sınavının AFA Sonuçları

| Faktörler | Faktör Özdeğerleri | Açıklanan Varyans (%) | Toplam Varyans (%) |
|-----------|--------------------|-----------------------|--------------------|
| 1 | 5,11 | 25,55 | 25,55 |
| 2 | 1,42 | 7,08 | 32,63 |
| 3 | 1,19 | 5,93 | 38,55 |
| 4 | 1,16 | 5,79 | 44,34 |
| 5 | 1,11 | 5,53 | 49,87 |

Tablo 20 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde faktör öz değerlerine sahip 5 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %25,55 olduğu ve beş faktörün tamamının %49,87 oranında toplam varyansı açıkladığı görülmektedir. Faktörlerin öz değerleri birinci faktör için 5,11, ikinci faktör için 1,42, üçüncü faktör için 1,19, dördüncü faktör için 1,16, beşinci faktör için 1,11 olarak bulunmuştur. Özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 3'te verilmiştir.



Şekil 3. Matematik Testi Mazeret Sınavına Ait Yamaç Grafiği

Şekil 3 incelendiğinde birinci faktörden sonra ani ve ivmeli bir düşüş olduğu fakat diğer faktörlerden sonra grafiğin yatay olarak devam ettiği ve ani bir düşüş olmadığı görülmektedir.

Faktör analizi sonuçları uzman görüşlerine göre de değerlendirilmiş ve matematik testinin cebir ve geometri gibi iki faktörlü bir yapıda olabileceği sonucuna varılmıştır. Matematik mazeret sınavı için faktör analizi iki faktörle sınırlandırılarak faktör analizi tekrarlanmış ve eğik (oblique) döndürme yöntemlerinden promax döndürme yöntemi kullanılarak faktörler döndürülmüştür.

Uzman görüşleri doğrultusunda cebir ve geometri olarak iki faktörlü bir yapıda olduğu düşünülen Matematik mazeret sınavındaki maddelerin iki faktöre göre dağılımı Tablo 21’de verilmiştir

Tablo 21. Matematik Mazeret Sınavının 2 Faktörlü ve Döndürülmüş AFA Sonuçları

| Madde No | Cebir | Geometri |
|-----------------------|-------------|-------------|
| 1 | 0,58 | 0,15 |
| 2 | -0,06 | 0,34 |
| 3 | 0,55 | 0,01 |
| 4 | 0,38 | 0,17 |
| 5 | 0,28 | 0,45 |
| 6 | 0,07 | 0,47 |
| 7 | 0,63 | -0,08 |
| 8 | 0,44 | -0,03 |
| 9 | 0,60 | 0,15 |
| 10 | 0,70 | 0,17 |
| 11 | 0,23 | 0,48 |
| 12 | -0,12 | 0,56 |
| 13 | -0,09 | 0,60 |
| 14 | 0,43 | 0,23 |
| 15 | -0,13 | 0,62 |
| 16 | 0,45 | -0,07 |
| 17 | 0,68 | -0,45 |
| 18 | 0,19 | 0,24 |
| 19 | 0,49 | 0,13 |
| 20 | 0,47 | -0,35 |
| Özdeğer | 5,11 | 1,42 |
| Açıklanan Varyans (%) | 25,55 | 7,08 |

*Uzman görüşüne göre diğer faktörde yer alan fakat analiz sonucunda mevcut faktöre yük veren madde.

** Faktör yükü ve uzman görüşüne göre mevcut faktörde yer alan madde.

Tablo 21 incelendiğinde 7 maddenin geometri, 13 maddenin ise cebir faktöründe yer aldığı görülmektedir. Matematik mazeret sınavı uzman görüşleri doğrultusunda incelendiğinde cebir konu alanıyla ilgili olan 12 madde (1., 3., 5., 7., 9., 10., 14., 16., 17., 18., 19. ve 20. maddeler) geometri konu alanıyla ilgili 8 madde (2., 4., 6., 8., 11., 12., 13., ve 15. madde) olduğu sonucuna ulaşılmıştır. Bu duruma göre FA sonuçları incelendiğinde 4. ve 8. maddeler geometri konu alanıyla ilgili olmasına rağmen cebir konu alanıyla ilişki ölçüm yapıyormuş gibi görünmektedir. Benzer şekilde cebir konu alanına ait olan 5. madde ise geometri konusu gibi algılandığı görülmektedir. Bu durum mazeret sınavının yapı geçerliğini zedeleyici bir etki yapmaktadır.

Literatürdeki diğer çalışmalar incelendiğinde genellikle tek faktörlü yapılar elde edilmiştir. (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010). Fakat bu çalışmada madde yükleri, yamaç grafiği, açıklanan varyans oranları ve uzman görüşleri dikkate alınarak 2 faktörlü bir yapı tanımlanmıştır. Tanımlanan bu yapı DFA ile test edilmiş, faktör yük değerleri ile maddeler arasındaki ilişki katsayılarının 0,04 ile 0,73 arasında olduğu görülmüştür. 2. ,6. 12., 16. ve 20. maddelerin faktörleriyle ilişki katsayıları 0,30’un altında olduğu, özgül varyansların ise genellikle 0,90 altında olduğu fakat 2. ,6. 12., 16. 17. ve 20. maddelerin 0,90’ın üzerinde olduğu gözlenmiştir.

Tanımlanan 2 faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 22’de sunulmuştur.

Tablo 22. Matematik Mazeret Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

| Uyum İndeksleri | İstatistik Değeri |
|-----------------|---------------------------------------|
| χ^2 | 892,01 (p=0,000, s _d =169) |
| χ^2/sd | 5,278 |
| GFI | 0,929 |
| AGFI | 0,912 |
| CFI | 0,905 |
| NFI | 0,890 |
| NNFI | 0,893 |
| RMSEA | 0,060 |
| RMR | 0,048 |
| Standardize RMR | 0,048 |

Tablo 22 incelendiğinde $\chi^2/sd=5,278$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Bunun yanında model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistiklerinin sırasıyla 0,929, 0,912 ve 0,905 olduğu gözlenmiştir. NFI ve NNFI istatistikleri ise 0,890 ve 0,893 olarak bulunmuştur. RMSEA 0,060 olup, RMR ve standardize RMR değerleri ise 0,048 olarak bulunmuştur. Uyum katsayıları genel olarak incelendiğinde kurulan modelin uyumlu olduğu söylenebilir.

Matematik ortak ve mazeret sınavlarının yapı geçerlik düzeylerini ortaya koymak için AFA ve DFA sonucunda elde edilen yapılar, testlerin birbirinin alternatif olarak kullanılmamasının daha uygun olacağını göstermektedir. Testlerde uzman görüşleri ile cebir ve geometri olmak üzere iki faktörlü bir yapı oluşturulmuştur. Fakat her iki sınav içinde geometri ve cebirle ilişkili olup faktör analizi sonucu kendi faktörlerine yüklenmeyen maddeler olduğu gözlenmiştir. Geometri öğrenme alanıyla ilişki olan maddenin cebir öğrenme alanında bir maddeymiş gibi yüklenmesi, maddenin farklı bir yapıyı ölçtüğüne işaret etmektedir. Bu durum yapı geçerlik düzeylerinin düşmesine neden olmaktadır.

SONUÇLAR ve TARTIŞMA

Bu araştırma sonucunda aşağıdaki sonuçlara ulaşılmıştır.

1. 2013-2014 eğitim öğretim yılında uygulanan TEOG Türkçe ve matematik testleri ortak ve mazeret sınavlarının madde ve test istatistikleri karşılaştırılmış ve bu istatistiklerin manidar düzeyde farklı olduğu sonucuna ulaşılmıştır ($\alpha=0,05$). Literatür incelendiğinde Özkan (2015) tarafından 2013-2014 eğitim öğretim yılı I. döneminde gerçekleştirilen matematik ortak ve mazeret sınavlarının test eşitleme çalışması yapılmış ve her iki sınavın güvenilirlik ve ortalama güçlüklerinin eşit olduğu, ortalamaları arasında manidar farklılık bulunmasına karşın, testlerin varyanslarının eşit olduğu belirtilmiştir. Bu çalışmada ise ortak ve mazeret sınavlarının test ve madde istatistiklerinin farklılaştığı sonucuna ulaşılmıştır. Testlerin uygulanan dönemlerinin farklı olmasının böyle bir sonucu doğurmuş olabileceği düşünülmektedir. Bu araştırma sonucuna göre test ve madde istatistiklerinin farklılaşması durumunun sınavların fırsat eşitliği yönünü zedelediği söylenebilir.

2. 2013-2014 eğitim öğretim yılında uygulanan TEOG Türkçe ve matematik testleri ortak ve mazeret sınavlarının KGİ ve Kappa istatistikleri arasında anlamlı fark bulunmamıştır. Bu nedenle ortak ve mazeret sınavlarının uzman görüşlerine dayalı olarak değerlendirilen kapsam geçerliklerinin benzer düzeyde olduğu sonucuna ulaşılmıştır. Genel olarak bu sonuç TEOG alt testleri için literatürdeki diğer bulgularla da benzerlik göstermektedir (Karadeniz ve diğerleri., 2015; Kardeş-Birinci, 2014;

Yılmaz-Koğar ve Aygün, 2015). Benzer kapsamı ölçtüğü gözlenen testlerin, test ve madde istatistiklerinin farklılaştığı da göz önüne alındığında testi alan grupların farklı olmasının bu duruma neden olabileceği düşünülmüştür. Ancak etki büyüklükleri incelendiğinde etki büyüklüğü yüksek olan farklılıkların da bulunduğu gözlenmiştir.

3. Türkçe ortak ve mazeret sınavları için AFA ve DFA sonucunda tanımlanan modeller doğrulanmıştır. Türkçe testi ortak ve mazeret sınavlarının yapı geçerlik düzeyine yönelik toplanan ipuçları değerlendirildiğinde bu sınavların yapı geçerlik düzeylerinin bu amaçla kullanılacak bir sınav için yeterli olduğu düşünülmektedir.

Matematik ortak ve mazeret sınavları için AFA sonucunda, uzman görüşleri de değerlendirilerek cebir ve geometri olmak üzere 2 faktörlü bir yapı tanımlanmıştır. DFA ile tanımlanan bu ölçme modeli test edilmiş ve tanımlanan ölçme modelinin uyumlu olduğu gözlenmiştir. Ancak ortak sınavda ve mazeret sınavında da yapı geçerliğini zedeleyici durumlar gözlenmiştir. Uzman görüşlerine göre cebir faktöründe olması gereken bazı maddeler geometri faktörüne yüklenmiştir. Bu durumun ortaya çıkmasında testi alan bireylerin testteki bu maddeleri tam olarak anlayamamış olmasının da rolü olabileceği düşünülmektedir. Bu durum ortak sınav ve mazeret sınavının yapı geçerlik düzeyini zedeleyici bir etki yapmıştır. Literatürdeki benzer araştırmalarda tek faktörlü yapılar elde edilmiş ve faktöre yeterince yüklenmeyen maddeler analizlerden çıkarılmıştır (Anıl ve diğerleri., 2010; Örs, 2010). Ayrıca matematik ortak sınavına bulunan 14. maddenin ve matematik mazeret sınavında bulunan 18. maddenin herhangi bir faktöre yük vermemesi maddenin tam olarak anlaşılabilir olabileceğini düşündürmektedir. Madde incelendiğinde öğrencilerin maddeyi anlamakta güçlük çekebileceği sonucuna ulaşılmıştır. Matematik ortak ve mazeret sınavları üzerinde gerçekleştirilen AFA ve DFA sonucunda bu testlerin birbiri yerine kullanılmamasının daha uygun olacağı görüşü benimsenmiştir.

Araştırma sonuçları değerlendirildiğinde ortak ve mazeret sınavlarının psikometrik özelliklerinin benzer düzeyde hazırlanması için, bu sınavların test ve madde istatistikleri uzman görüşleri doğrultusunda tahmin edilerek düzenlenmesi, ortak ve mazeret sınavlarının sonuçlarının birbirinin yerine kullanılabilmesi için test eşitleme çalışmasının yapılması önerilmektedir. Bunun için sınavların uygulama süreci, test eşitlemeye olanak verecek biçimde yürütülmelidir. Testler arasındaki süre birkaç hafta olacak şekilde düzenlenerek sorular açıklanmayıp ortak maddeler teste eklenebilir. Böylece test eşitleme çalışmalarına imkân verilebilir. Faktör analizi sonucunda matematik testi ortak ve mazeret sınavları için farklı faktörlere yük veren maddeler olduğu gözlenmiştir. Örneğin faktör analizi sonucunda cebir faktöründe çıkan bir madde uzman görüşüne göre geometri faktöründe yer almakta ya da tam tersi bir durum söz konusu olmaktadır. Maddeler, öğrenciler tarafından tam olarak anlaşılmadığı için bu durumun oluşabileceği düşünülmektedir. Bu nedenle öğrencilerin seviyeleri de göz önünde bulundurularak maddeler yazılırken öğrencilerin maddeyi tam olarak anlayabilecekleri şekilde yazılmasına dikkat edilmelidir.

KAYNAKÇA

- AERA, APA, ve NCME. (2014). *Standarts for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Alıcı, D., Başol, G., Çakan, M., Kan, A., Karaca, E., Özbek, Ö. Y., ve Yaşar, M. (2011). *Eğitimde ölçme ve değerlendirme*. (S. Tekindal, Ed.). Ankara: Pegem Akademi.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications: American Educational Research Association. *Educational Researcher*, 35(6), 33–40. doi.org/10.3102/0013189X035006033
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6. Baskı). Washington, DC: American Psychological Association. doi.org/10.1017/CBO9781107415324.004
- Anıl, D., ve Güzeller, C. O. (2011). Seviye belirleme sınavı fen ve teknoloji alt testi ile diğer alt testler arasındaki ilişkinin yol analizi ile incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 11(1), 1–10.

- Anıl, D., Güzeller, C. O., Çokluk, Ö., ve Şekercioğlu, G. (2010). Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub- test. *Procedia - Social and Behavioral Sciences*, 2(2), 5292–5298. doi.org/10.1016/j.sbspro.2010.03.863
- Atıla, M. E., ve Özekten, Ö. F. (2015). Temel eğitimden ortaöğretime geçiş sınavı: Fen bilimleri öğretmenleri ne düşünüyor? *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 34(1), 124–140. doi.org/10.7822/omuefd.34.1.7
- Bracken, B. A., Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., ve Rodriguez, M. C. (2013). *APA handbook of testing and assesment in psychology*. (K. F. Geisinger, Ed.) (1. Cilt). Washington, DC: American Psychological Association.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. Baskı). New York: The Guilford Press.
- Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (18. Baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2013). *Bilimsel araştırma yöntemleri* (14. Baskı). Ankara: Pegem Akademi.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. *Journal of Chemical Information and Modeling* (2. Baskı, Vol. 53). New York: Lawrence Erlbaum Associates, Inc. http://doi.org/10.1017/CBO9781107415324.004
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge. doi.org/10.1037/a0028079
- Dalak, O. (2015). *TEOG sınav soruları ile 8. sınıf öğretim programlarındaki ilgili kazanımları yenilenmiş Bloom Taksonomisine göre incelenmesi*. (Yüksek lisans tezi, Gaziantep Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Gaziantep). http://tez2.yok.gov.tr/ adresinden edinilmiştir.
- Dinç, E., Dere, İ., ve Koluman, S. (2014). Kademeler arası geçiş uygulamalarına yönelik görüş ve deneyimler. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 7(17), 397–423. dx.doi.org/10.14520/adyusbd.761
- Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (2. Baskı). Ankara: Pegem Akademi.
- George, D., ve Mallery, P. (2012). *IBM SPSS Statistics 19 Step by Step: A simple guide and reference*. *Journal of Chemical Information and Modeling* (12. Ed., Vol. 53). Boston: Pearson.
- Görmez, M., ve Coşkun, İ. (2015). *1. yılında temel eğitimden ortaöğretime geçiş reformunun değerlendirilmesi*. Ankara:SETA.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley ve Sons Inc.
- Kaiser, H. F., ve Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. doi.org/10.1177/001316447403400115
- Karaca, M. (2015). *TEOG sınavlarında soru sorulan ve sorulmayan fen ve teknoloji dersi konularına ilişkin öğrenci ve öğretmen bakış açıları*. (Yüksek lisans tezi, Erciyes Üniversitesi, İlköğretim Anabilim Dalı, Kayseri). http://tez2.yok.gov.tr/ adresinden edinilmiştir.
- Karaca, M., Bektaş, O., ve Armağan, F. Ö. (2015). 8. sınıf öğrencilerinin merkezi sınavlarda sorulmayan fen bilimleri konularına yönelik görüşleri. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 63–86. http://doi.org/10.17152/gefad.99679
- Karadeniz, O., Eker, C., ve Ulusoy, M. (2015). TEOG sınavlarındaki T.C. İnkılap Tarihi ve Atatürkçülük dersine ait soruların kazanım temelli olarak değerlendirilmesi. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, 6(18), 115–134.
- Karasar, N. (2014). *Bilimsel araştırma yöntemi* (26. Baskı). Ankara: Nobel Yayınları.
- Kardeş-Birinci, D. (2014). Merkezi sistem ortak sınavlarda ilk deneyim: Matematik dersi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 3(2), 8–16.
- Kline, R. B. (2011). *Principles and practise of structural equating modeling*. (T. G. Press, Ed.) (3. Baskı). New York.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- MEB. (2015). 2014-2015 öğretim yılı ortak sınavlar e-kılavuzu, [Çevrim-İçi: http://oges.meb.gov.tr/meb_iys_dosyalar/2014_11/04053521_ortaksnavlareklavuz20142015.pdf], erişim tarihi: 20 Mart 2015
- Meydan, C. H., ve Şeşen, H. (2015). *Yapısal eşitlik modellemesi* (2. Baskı). Ankara: Detay Yayıncılık.
- Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS. Design* (Vol. 102). New York: Springer Science ve Business Media, LLC. http://doi.org/10.1016/j.peva.2007.06.006

- Örs, S. (2010). *İlköğretim 6., 7. ve 8. sınıf Seviye Belirleme Sınavı fen ve teknoloji alt testlerinin faktör yapılarının belirlenmesi*. (Yüksek lisans tezi, Ankara Üniversitesi, Ölçme ve Değerlendirme Anabilim Dalı, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Özgül, İ. E. (2011). *Psikolojik testler*. Ankara: PDREM Yayınları.
- Özkan, M. (2015). *TEOG kapsamında uygulanan matematik alt testi ile matematik mazeret alt testinin istatistiksel eşitliğinin sınanması*. (Yayınlanmamış yüksek lisans tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Özsoy, S., ve Özsoy, G. (2013). Eğitim araştırmalarında etki büyüklüğünün raporlanması. *İlköğretim Online*, 12(2), 334–346.
- Schermelleh-Engel, K., Moosbrugger, H., ve Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Şad, S. N., ve Şahiner, Y. K. (2016). Temel eğitimden ortaöğretime geçiş (TEOG) sistemine ilişkin öğrenci, öğretmen ve veli görüşleri. *İlköğretim Online*, 15(1), 53–76.
- Tabachnik, B. G., ve Fidell, L. S. (2012). *Using multivariate statistics* (6. Baskı). Boston: Pearson.
- Thalheimer, W., ve Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. <http://doi.org/10.1113/jphysiol.2004.078915>
- Turgut, M. F., ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Ankara: Pegem Akademi.
- West, S. G., Finch, J. F., ve Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: SAGE.
- Yılmaz, G. (2014). *SBS ve TEOG sınavındaki söz varlığımızla ilgili Türkçe sorularının 2005 Türkçe Öğretim Programı'ndaki sözcük varlığımızla ilgili kazanımları ölçme yeterlilik düzeyinin öğretmen görüşlerine göre incelenmesi*. (Yüksek lisans tezi, Erzincan Üniversitesi, Türkçe Eğitimi Anabilim Dalı, Erzincan). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yılmaz-Koçar, E., ve Aygün, B. (2015). Temel eğitimden orta öğretime geçiş sınavı (TEOG)'nın matematik temel alanına ait testlerin kapsam geçerliğinin incelenmesi. *Pegem Eğitim ve Öğretim Dergisi*, 5(5), 667–680. <http://doi.org/10.14527/pegegog.2015.036>
- Yorgancı, O. K. (2015). *Sekizinci sınıf Türkçe dersi ortak sınavı sorularının öğretim programına göre değerlendirilmesi*. (Yüksek lisans tezi, Gazi Üniversitesi, Türkçe Eğitimi Anabilim Dalı, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yurdugül, H. (2006). Parelel, eşdeğer ve konjenetik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15–37.
- Yurdugül, H., ve Bayrak, F. (2012). Ölçek geliştirme çalışmalarında kapsam geçerlik ölçüleri: Kapsam geçerlik indeksi ve Kappa istatistiğinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, (Özel Sayı 2), 264–271.
- Zayimoğlu-Öztürk, F., ve Aksoy, H. (2014). Temel eğitimden ortaöğretime geçiş modelinin 8 . sınıf öğrenci görüşlerine göre değerlendirilmesi (Ordu ili örneği). *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 33(2), 439–454. <http://doi.org/10.7822/omuefd.33.2.8>

EXTENDED ABSTRACT

Introduction

Tests are applied almost at all levels of education system in Turkey and as a result of these tests decisions made about individuals. Transition from basic to secondary education exam (TEOG) at lower secondary level, transition to higher education exam at upper secondary level (YGS), undergraduate placement exam (LYS) can be given as example for the exams applied in nation-wide. Common characteristics of these exams are being highly determinant for the life of individuals. For the tests having significant power of influence, it is necessary to collect evidence for validity and reliability.

Analyzing the validity and reliability of a test gives off important evidence so collecting these data especially for the national exams are necessary. Transition from basic to secondary education exam which has been applied since 2013-2014 school year is one of these exams. This examination system aims to evaluate the achievements of individuals throughout the education duration (MEB, 2015). Transition from basic to secondary education exam contains Turkish, mathematics, science and

technology, education of religion and ethics, TR History of Revolution and Kemalism and foreign language course. This exam is applied by Ministry of National Education.

Common and make up exams are carried out within the scope of TEOG exams. Individuals who couldn't attend common exam and having a valid excuse can attend make up exams instead of common exams on a date determined by the Ministry of National Education. On that case it is come out as a matter whether tests applied in both exams and their items are identical or not. In these two tests individuals are evaluated in the same way and decisions made through the same criterion, if the items aren't identical this cause disadvantage for one of the groups. To be able overcome this risk psychometric properties of both tests should be investigated, equivalence of both tests should be provided and necessary precautions should be taken for similar test to be conducted.

In this study the psychometric properties of Turkish and math subtests in common and make-up examination of TEOG have been compared.

Method

Sample of the research is consisted of randomly selected 10.000 students attended common exams. Data obtained from the individuals in the sample analyzed and data with systematically repeated answers and same answers to all questions was omitted. Research was carried out via item scores matrix of 9773 students for Turkish common exam, 9485 students for math common exam, 2747 students for Turkish make-up exam and 2323 students for math make-up exam.

It was consulted to expert opinion for the content validity study. According to this, in order to determine and compare the acquisitions that is measured in common and make-up examinations of TEOG, question-acquisition matching form and question assessment form was prepared by researchers.

To conduct analysis Statistica 10, Factor 10.3, Lisrel 8.80, ITEMAN 3.5 and Excel 2016 package programs were used.

Results and Discussion

As a result of research, item and test statistics of the common and make up exams have been compared and it has been observed that test and item statistics are different in significantly ($\alpha=0,05$). Because of this situation, it can be said that fairness of exams has been damaged. Common and make up exams content validity which based on expert opinions is in similar level.

To evaluate clues related construct validity, it is believed that Turkish common and make up exams construct validity level is adequate for using this purpose. But the results of the study revealed that math's common and make-up exams items don't measure the same things in terms of construct validity.

Therefore, common and make-up examinations are needed to be prepared in similar level by investigating the test and item statistics of them through expert predictions. Besides, it can be conducted test equating studies for this exams. For this purpose, common exam questions shouldn't announce until make up exams being applied.

A Note on the Use of Item Parceling in Structural Equation Modeling with Missing Data*

Yapısal Eşitlik Modellemesinde Kayıp Verilerin Madde Parsellemenin Kullanımı Üzerine Bir Not

Fatih Orçan**

Yanyun Yang***

Abstract

Item parceling procedure may be applied to alleviate some difficulties in analysis with missing data and/or nonnormal data in structural equation modeling. A simulation study was conducted to investigate how item parceling behaves under various conditions in structural equation model with missing and nonnormal distributed data. Design factors included missing mechanism, percentage of missingness, distribution of item data, and sample size. Results showed that analysis conducted at the parcel level yielded lower model rejection rates than analysis based on the individual items, and the patterns were consistent across missing mechanism, percentage of missing, and distribution of item data. In addition, parcel-level analyses resulted in comparable parameter estimates to item-level analyses.

Keywords: Item parceling, missing data, nonnormal distribution, structural equation model, SEM

Öz

Yapısal eşitlik modellerinde madde parselleme prosedürü kayıp veri ve/veya normal dağılım olmadığı durumlarda karşılaşılan zorlukları aztmakta kullanılabilir. Madde parsellemenin kayıp ve normal dağılmayan verilerin olması durumunda yapısal eşitlik modelinde nasıl davranacağı bir simülasyon çalışması ile araştırılmıştır. Dizayn faktörleri kayıp mekanizması, kayıp veri yüzdesi, madde veri dağılımları ve örneklem büyüklüğünü içermektedir. Sonçlara göre, parsel seviyesinde yapılan analizler madde seviyesinde yapılan analizlere oranla daha az model reddedilmesine sebep olmuş ve kayıp mekanizması, kayıp veri yüzdesi ve madde veri dağılımlarında benzer davranmıştır. Ayrıca, parsel seviyesindeki analizler madde seviyesindeki analizlere karşılaştırılabilir parametre tahminleriyle sonuçlanmıştır.

Anahtar Kelimeler: Madde parselleme, kayıp veriler, normal olmayan dağılım, yapısal eşitlik modellemesi, YEM

INTRODUCTION AND LITERATURE REVIEW

Structural equation modeling (SEM) has been frequently used in empirical data analysis to examine hypothesized relationships among a set of variables. A commonly used estimation method in SEM, maximum likelihood (ML), requires the sample size be sufficiently large and observed variables be multivariate normally distributed. Violation of these assumptions results in inaccurate model chi-square statistic, fit indices, parameter estimates, and standard errors associated with parameter estimates (e.g., Bollen, 1989; Curran, West, & Finch, 1996; Chou, Bentler, & Satorra, 1991), and the degree of bias tends to increase as the model complexity increases. Alternatively, estimation methods that do not require assumptions as restrictive as ML may be applied. One alternative is robust maximum likelihood (MLR) estimation method. MLR corrects for the positive bias in model chi-square statistic and the negative bias of standard errors associated with parameter estimates. Another alternative is weighted least square (WLS). WLS does not require variables be multivariate normally distributed. However, WLS requires very large sample sizes even when the data are

* This study was conducted as part the first author's dissertation

** Asst. Prof., Karadeniz Technical University, College of Education, Trabzon – Turkey, fatihorcan@ktu.edu.tr

*** Assoc. Prof. Florida State University, College of Education, Tallahassee – FL, yyang3@admin.fsu.edu

multivariate normally distributed (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992) and the performance of WLS becomes worse as the complexity of the model increases (Muthén & Kaplan, 1992). When the purpose of the SEM analysis is not to examine the psychometric properties of individual items, instead, it is to investigate the relationship among latent factors, a parceling technique may be applied to reduce the model complexity, particularly when the analysis involves small sample size and the number of indicators per factor is large (e.g., Sterba & MacCallum, 2010).

Parceling is referred to as a procedure for computing sums or average scores across multiple items. The variables based on the sum or average (called parcels) instead of the individual items are then used as indicators of latent factors in the SEM analysis (Bandalos 2002, 2008; Little, Cunningham, Shahar, & Widaman, 2002; Sass & Smith, 2006; Sterba, 2011; Sterba & MacCallum, 2010; Yang, Nay, & Hoyle, 2010). Bandalos and Finney (2001) reviewed the use of item parceling in five journals published in 1989-1994. They found that about 20% of empirical studies (62 out of 317) used some kinds of item parceling techniques; and the percentage varied from 9% in *Journal of Marketing Research* to 60% in *Journal of Educational Measurement*. A much higher percentage was found in the review of three psychology journals published in 1996-1999 (Plummer, 2000); among 102 articles used structural equation modeling, about 50% of which involved some kinds of parceling techniques in the analysis. Use of parcels is appealing in that it reduces model complexity, reduces the requirements on sample sizes, reduces influences of individual items' systematic errors on the model estimation, helps reach optimal reliability, increases model convergence rate, and increases the model fit when the dimensionality of the items is known (Bandalos, 2002; Little et al, 2002; Matsunaga, 2008; Meade & Kroustalis, 2005; Nasser & Takahashi, 2003; Nasser & Wisenbaker, 2003; Plummer, 2000; Sass & Smith, 2006; Yang et al., 2010). Sass and Smith (2006) showed analytically that use of parcels does not lead to a bias estimate of the structural relationship among latent factors when model assumptions are met and items are unidimensional. On the other hand, there are some arguments against the use of parceling techniques. For example, using parcels in the analysis may blur the dimensionality of original measures and produce biased estimates of model parameters (Bandalos, 2002; Matsunaga, 2008). It has been recommended that parceling be used thoughtfully so that the drawback of such a use is minimized. Parceling is most beneficial when the analysis is conducted based on a small sample size and the relationship between items and the underlying latent factor is not strong (Sterba & MacCallum, 2010), on the other hand, when the dimensionality of the items is not clear, parceling should not be used (Bandalos & Finney, 2001; Little et al., 2002; Meade & Kroustalis, 2006).

For unidimensional measures, items can be assigned to parcels either randomly or purposively (Little et al., 2002; Matsunaga, 2008; Sterba & MacCallum, 2010). In this study, we use the following procedure to allocate items to parcels although results from current simulation study may be also valid for other item allocation methods (Sterba & MacCallum, 2010): First, a factor analysis is to conduct on the items to be parceled. Second, the obtained factor loadings are used to allocate items such that the sums of loadings are as equivalent as possible across parcels. Third, an SEM analysis is conducted based on parcel scores. From the classical test score theory perspective, the approach is preferred because parcels tend to be essentially tau-equivalent and thus maximize reliability of the scale based on the parcels, consequently, lead to the least bias estimates of structural coefficients among latent factors. This approach has been adopted in previous studies, however, in these studies the parcels are created based on the magnitude of loadings in population. In other words, the assignment of items to parcels is the same across all samples. Sterba and MacCallum (2010) argued that assigning the same items to parcels across all datasets do not take sampling error into consideration because the estimates of loadings may vary across samples. In our study, we conducted factor analysis on the items to be parceled for each sample, sorted items based on the magnitude of the loadings, and then assigned items to parcels. Consequently, the assignment of items to parcels varies from sample to sample while the sums of loadings are as equivalent as possible across parcels for each sample.

When the distributional assumptions underlying ML are violated, an applied researcher might face the decision of analyzing the model based on individual items using alternative estimation methods (e.g., MLR, WLS) or creating parcels and then analyzing the model based on the parcel scores. If item-level analysis using alternative estimation methods yields no worse results than those from parcel-level analysis, then adopting a parceling technique becomes an unnecessary work, not mentioning that there have been many debates regarding the use of parceling (Little, et al., 2002). As we are aware of, no previous study has compared model results (e.g., model rejection rates based on chi-square statistic, accuracy of estimated structural coefficients among latent factors) between item-level analysis and parcel-level analysis using alternative estimation methods when the distributional assumptions underlying ML are violated. This is one of the purposes of current study.

The presence of missing data creates a potential problem in SEM analysis. Although some other types of missing mechanisms have been discussed in the literature, most widely discussed missing mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Among numerous techniques proposed to handle missing data (see Peugh & Enders, 2004), maximum likelihood is the most commonly used in empirical studies in social sciences and is considered the best approach in general (Enders & Bandalos, 2001); when the missing data are present, it is named as full information maximum likelihood (FIML). This method uses all the available information to estimate the model (Acock, 2005; Chen & Astebro, 2003). The likelihood function based on the available observed variables is first computed for each observation in the sample. The individual likelihoods are then summed to give the likelihood of the whole sample (Enders & Bandalos, 2001). The multivariate normality assumption plays an essential role in the FIML estimation under MAR (Allison, 2002; Chen & Astebro, 2003; Enders, 2004). FIML produces accurate model-data fits when data are multivariate normal (Enders, 2001). Researchers also found that FIML showed unbiased parameter estimates under MCAR and MAR (Enders & Bandalos, 2001).

In an empirical study involving relatively small sample size, large number of measurement indicators, and the presence of missing data, a parceling technique may be applied to obviate some of these difficulties in SEM analysis, when the purpose of the study is to examine the hypothesized relationship among latent factors, but not the psychometric properties of individual items. Schafer and Graham (2002) raised a similar idea but labeled it in an ambiguous manner. They suggested averaging scores across a subset of items when multiple items are available in measuring the same/similar construct (i.e., the same latent variable). The parcels are then used as indicators of latent factors in SEM analysis. They labeled this method as “case-by-case item deletion” or “ipsative mean imputation”. This method has been applied in empirical studies (e.g., Achenbach, Bernstein, & Dumenci, 2005; Signorella, & Cooper, 2011; Yoder, Snell, & Tobias, 2012), however, “its properties remained largely unstudied” (Schafer & Graham, 2002, p. 158).

Examination of the effects of item parceling on model fit indices and parameter estimates has mainly been conducted under the conditions when the model assumptions are met. Specifically, the analysis model is consistent with the data generation model (Finney & DiStefano, 2006), item scores are continuously and multivariate normally distributed, and no missing data are present. A few studies focused on categorical item scores and misspecified model (Bandalos, 2002; Bandalos, 2008). However, as we are aware of, none of the studies examined item parceling techniques for data with missing values, and compared the performance of parcel-level analysis to item-level analysis using estimation methods other than ML. The purpose of this simulation study is to investigate how item parceling behaves under various conditions in SEM with missing and nonnormal distributed data via a simulation study. Results based on the parcels are compared with those based on the individual items. For both parcel-level and item-level analysis, both maximum likelihood and robust maximum likelihood estimation methods are applied.

METHODS

Data were generated based on a structural equation model as shown in Figure 1. Specifically, the model consisted of three latent factors measured by 21 items. The first factor, F1, was measured by 15 items (item1 - item15); both the second and the third factors, F2 and F3, were measured by three items (item16 – item18 and item19 - item21, respectively). The factor loading was .70 from items 16-21. The factor loadings associated with F1 varied across items and were .40, .60, and .80. The variance of uniqueness for each item was fixed as one minus the squared loading. The path coefficient from F1 to F2 (F1→F2) and F1 to F3 (F1→F3) was .40 and .60, respectively. The covariance between the disturbances of F2 and F3 (F2 ↔F3) was .50. Based on this model, we evaluated the performance of parcel level analysis by manipulating four design factors in the simulation study: missing mechanism, percentage of missingness, degree of nonnormality of item scores, and sample size.

Design Factors for Data Generation

- *Missing mechanisms.* Three missing mechanisms were considered: MCAR, MAR, and MNAR). The results from these conditions were compared to the corresponding conditions with no missing data.
- *Percentage of missingness.* Three levels of percentage of missingness were considered: 10%, 20%, and 40%. Previous studies have showed that percent of missingness had an effect on parameter estimates for analysis based on individual items (Davey & Savla, 2005; Enders, 2001).
- *Distribution of item scores.* Three types of distribution of item scores were considered: (1) normal distribution; (2) moderate skewness and low kurtosis ($Sk = 1$, and $K = 1.5$); and (3) high skewness and high kurtosis ($Sk = 1.75$, and $K = 3.75$). Nonnormality was only applied to items 1-15 with the same population skewness and kurtosis. Items 16 through 21 were distributed normally in all generation conditions.
- *Sample sizes.* Three different sample sizes were considered: 100, 300, and 1000. These sample sizes were chosen to represent a range of small to large sample sizes in SEM analysis.

These four design factors created a total of 90 conditions for data generation. Among them, 81 were formed from the conditions with missing data (3 patterns of missingness × 3 percentages of missingness × 3 types of distributions × 3 sample sizes), and 9 were formed from the conditions with no missing data (3 types of distributions × 3 sample sizes). For each condition, 2000 data sets were generated.

Data Generation Procedure

Data were generated in R (version 2.13.2) based on the model shown in Figure 1. First, the correlation matrix among factors was obtained based on the parameters specified in the model. Second, 24 random variables were generated each with a mean of zero and standard deviation of one, three of which represented factor scores and the other 21 represented item uniqueness. Factors were then converted to have multivariate normal distribution with the intended correlations using the Cholesky decomposition method. Third, the observed item scores were obtained as a weighted linear combination of the factor score and item uniqueness such that (e.g., Bernstein & Teng, 1989):

$$X_{ij} = \lambda_i * F_j + \left(\sqrt{1 - \lambda_i^2} \right) * E_{ij}$$

where X_{ij} is an observed score on item i for individual j , F_j is the factor score for person j , λ_i indicates the factor loading for the item i , and E_{ij} indicates item uniqueness. For conditions with nonnormally distributed data, Fleishman's power transformation method (Fleishman, 1978) was applied to normally distributed data to obtain observed scores with the predefined skewness and kurtosis.

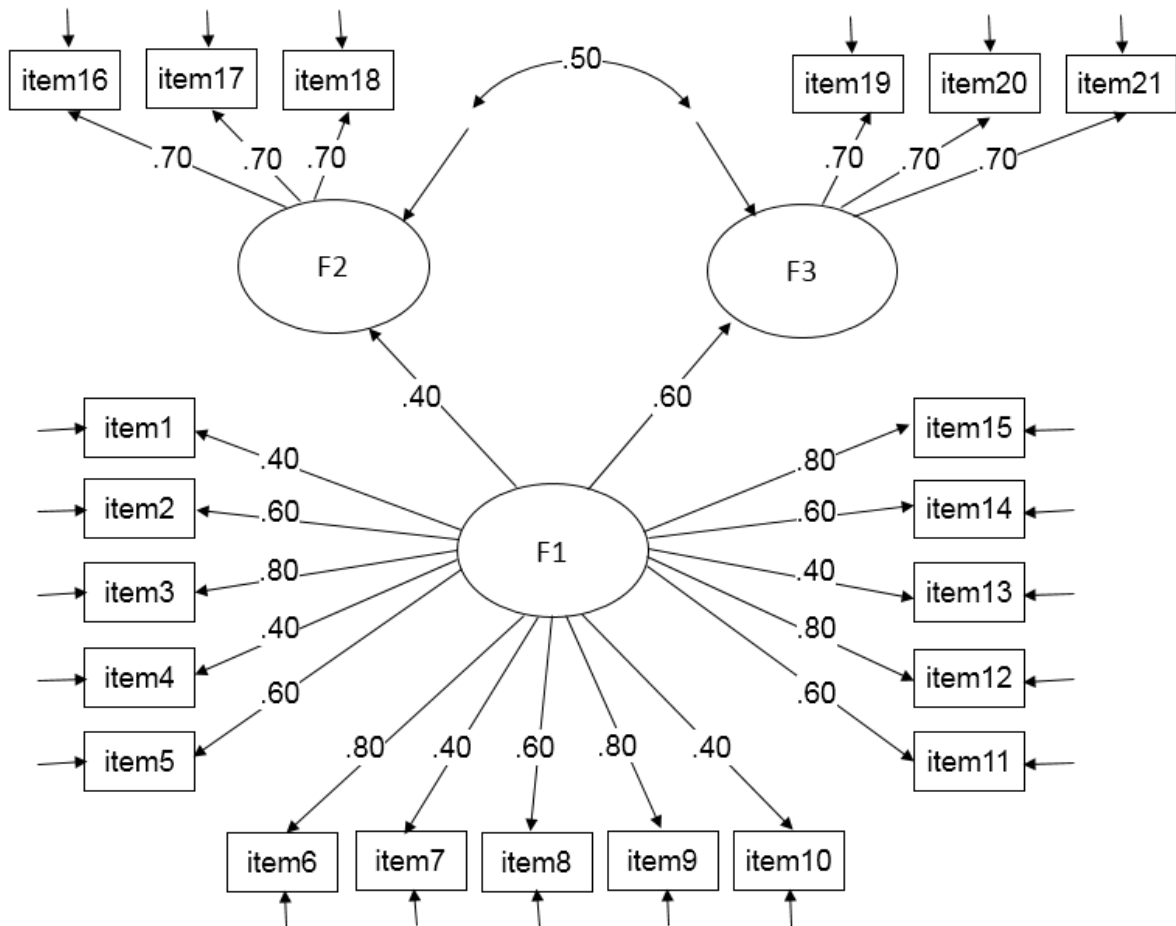


Figure 1. Models for Data Generation

The following rules were applied to create missing data. Only six items have missing values: items 1, 2, 3, 6, 7, and 8. To create data that are MCAR, randomly selected cases on the specified items were removed. To create data that are MAR, missingness on item1 was related to the values of item4. That is, item4 were sorted from the smallest to the largest, and then the cases with the lowest values on the item4 (e.g., 10%) were assigned being missing on item1 with a probability of .90. The rest of the values (highest 90%) were assigned being missing with a probability of .10. The same procedure was applied to the other five items; specifically, the missingness on the item2, 3, 6, 7, and 8 was related to the values on the item5, 16, 11, 12, and 13, respectively. For conditions with MNAR, missingness on variables was related to the variables themselves.

Data Analysis

Mplus 6.1 (Muthén&Muthén, 1998-2008) was used for the data analysis. For each dataset, two different models were considered. One was based on item level and the other was based on parcel level. All analysis models were considered correctly specified given that the analysis model was consistent with the data generation model. To create parcel factorial parceling technique was used. The factorial algorithm technique decomposes “item specific components” and combines them within different parcels (Matsunaga, 2008). For the analysis based on parcel scores, a CFA model was first conducted on items 1 to 15. The items were sorted based on the magnitudes of the loadings (labeled as 1st to 15th from the largest to the smallest). Three parcels were then created such that the 1st parcel contained items with the order of 1st, 6th, 7th, 12th, and 13th, the 2nd parcel consisted of items with the order of 2nd, 5th, 8th, 11th, and 14th, and the 3rd parcel comprised the rest of the five items. The mean across the five (or available variables if there were missing values) items was computed as the parcel score. The items associated with F2 and F3 were not parceled. For both the item and the parcel level analyses, ML and MLR were used for model estimation.

Analysis of Outcome Variables

For each condition, overall model-data fit was evaluated based on the chi-square test, Comparative Fit Index (CFI; Bentler, 1990) and Root Mean Square Error of Approximation (RMSEA; Steiger& Lind, 1980). For the chi-square test, rejection rate based on the nominal level of .05 was reported. CFI values larger than .95 and RMSEA smaller than .08 were considered reasonable model fits (Hu & Bentler, 1999). For parameter estimates, direct effects among factors, that is, from F1 to F2 (F1→ F2) and from F1 to F3 (F1→ F3), and the covariance between the disturbances of F2 and F3 (F2 ↔ F3) were evaluated. Relative bias was computed for both point estimates and standard errors associated with the parameter estimates as:

$$\text{Relative Bias} = \frac{\hat{\theta} - \theta}{\theta} \times 100\%,$$

where $\hat{\theta}$ and θ indicate the mean of estimates and the population parameter, respectively. The true standard errors were approximated by using the standard deviations of parameter estimates based on the corresponding conditions without missing data (Bandalos, 2006). Hoogland and Boomsma (1998) suggested biases (in absolute value) smaller than 5% for point estimates and 10% for standard errors to be acceptable.

RESULTS

For each condition, model convergence, rejection rates based on the chi-square test, and rejection rates based on fit indices were reported. Parameter estimates and their standard errors were examined by computing relative biases. There were no inadmissible solutions across all conditions except those conditions with a sample size of 100 and 40% of missing data when analyses were conducted on the item level. For these conditions, 53% to 72% of replications encountered inadmissible solutions and the rates of inadmissible solutions from ML estimation method were similar to those from MLR estimation method. Replications with inadmissible solutions were excluded from further analyses. Parcel-level analysis did not encounter any inadmissible solutions. In the following section, selected results were reported.

Model Rejection Rates Based on CFI and RMSEA

CFI greater than .95 and RMSEA smaller than .08 indicate good model-data fit (Hu & Bentler, 1999). Table 2 report percentages of replications with CFI smaller than .95 and RMSEA greater than .08 for conditions with sample size of 100. Results for conditions with sample size of 300 and 1000 were not provided because the percentage was zero or closed to zero for these conditions.

Table 2. Percentage of Replications with CFI Smaller Than .95 and RMSEA Greater than .08 for Conditions with Sample Size of 100

| Sample Size | Skewness/Kurtosis | No Missing | MCAR | | | MAR | | | MNAR | | |
|--------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | | 10% | 20% | 40% | 10% | 20% | 40% | 10% | 20% | 40% |
| CFI | | | | | | | | | | | |
| Item | 0/0 | 18 (23) | 23 (30) | 30 (37) | 80 (52) | 24 (31) | 32 (40) | 78 (58) | 28 (34) | 36 (45) | 83 (60) |
| | 1/1.5 | 38 (43) | 44 (50) | 54 (63) | 90 (73) | 45 (51) | 52 (61) | 86 (75) | 86 (53) | 47 (65) | 89 (77) |
| | 1.75/3.75 | 78 (77) | 82 (83) | 86 (87) | 98 (91) | 81 (81) | 84 (86) | 97 (93) | 79 (81) | 84 (88) | 97 (93) |
| Parcel | 0/0 | 1 (2) | 1 (2) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | 2 (3) | 2 (2) | 2 (3) | 2 (3) |
| | 1/1.5 | 2 (3) | 2 (3) | 4 (5) | 4 (6) | 2 (3) | 4 (5) | 4 (5) | 2 (4) | 3 (5) | 4 (5) |
| | 1.75/3.75 | 3 (4) | 3 (5) | 3 (5) | 3 (5) | 3 (4) | 3 (5) | 3 (5) | 3 (5) | 3 (5) | 3 (5) |
| RMSEA | | | | | | | | | | | |
| Item | 0/0 | 0 (0) | 0 (0) | 0 (0) | 12 (0) | 0 (0) | 0 (0) | 14 (0) | 0 (0) | 0 (0) | 11 (0) |
| | 1/1.5 | 0 (0) | 0 (0) | 0 (0) | 19 (1) | 0 (0) | 0 (0) | 18 (1) | 0 (0) | 0 (0) | 17 (0) |
| | 1.75/3.75 | 0 (1) | 0 (1) | 1 (2) | 43 (5) | 0 (1) | 1 (2) | 36 (5) | 0 (1) | 1 (2) | 33 (5) |
| Parcel | 0/0 | 4 (5) | 3 (5) | 4 (5) | 3 (5) | 4 (5) | 4 (6) | 4 (6) | 3 (5) | 4 (5) | 4 (6) |
| | 1/1.5 | 5 (7) | 4 (6) | 5 (7) | 5 (8) | 4 (7) | 5 (7) | 5 (7) | 4 (6) | 5 (7) | 4 (7) |
| | 1.75/3.75 | 4 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) | 3 (6) |

Note. Percentages from robust maximum likelihood estimation method are in parentheses.

Consistent with the findings based on the model chi-square test, when the analysis was conducted at the item level, CFI tended to demonstrate misfit when the sample size was 100 (18%-97% of models had CFI<.95). Models were more likely to demonstrate misfit when larger percentage of data were missing and the distribution of data became more skewed. As sample size reached to 300, less than 9% of models had CFI<.95 and the percentage was nearly zero for most of the conditions. The pattern was similar across missing mechanisms. When the analysis was conducted at the parcel level, less than 6% of replications had CFI<.95 for all conditions and the percentages were nearly zero when sample size reached to 300, regardless of the degree of nonnormality, missing mechanism, and percentage of missingness.

The findings from RMSEA were slightly different. When the analysis was conducted at the item level data, less than 2% of models showed RMSEA>.08 when the sample size was 100, unless 40%

of data in the sample were missing; for these conditions, 11%-43% of models yielded $RMSEA > .08$. However, the percentage was less than 5% when MLR was applied. As sample size reached to 300, nearly 0% of models showed $RMSEA > .08$. When the analysis was conducted at parcel level, 3%-8% of models yielded $RMSEA > .08$ and the percentages were relatively stable across missing mechanisms, percentage of data being missing and the degree of nonnormality. In addition, results from MLR and ML were comparable.

Parameter Estimates and Standard Errors

Table 3 reports relative bias of estimates for the parameter $F1 \rightarrow F2$ from MLR. Parameter estimates from ML were identical to those from MLR and thus were not provided.

Table 3. Relative Bias (%) of the Direct Effect from F1 to F2

| Sample Size | Level | No Missing | MCAR | | | MAR | | | MNAR | | |
|-------------------------------|--------|------------|------|-----|-----|-----|-----|-----|------|-----|-----|
| | | | 10% | 20% | 40% | 10% | 20% | 40% | 10% | 20% | 40% |
| Skewness=0 & Kurtosis=0 | | | | | | | | | | | |
| 100 | Item | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| | Parcel | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 300 | Item | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Parcel | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1000 | Item | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Parcel | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Skewness=1 & Kurtosis=1.5 | | | | | | | | | | | |
| 100 | Item | -3 | -3 | -2 | -3 | -3 | -2 | -1 | -3 | -2 | -2 |
| | Parcel | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| 300 | Item | -2 | -2 | -1 | -1 | -2 | -1 | -1 | -2 | -1 | -2 |
| | Parcel | -2 | -2 | -1 | -1 | -2 | -1 | -1 | -2 | -1 | -1 |
| 1000 | Item | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | -2 |
| | Parcel | -1 | -1 | -2 | -2 | -1 | -1 | -1 | -1 | -2 | -1 |
| Skewness=1.75 & Kurtosis=3.75 | | | | | | | | | | | |
| 100 | Item | -11 | -11 | -12 | -13 | -11 | -12 | -11 | -11 | -11 | -11 |
| | Parcel | -10 | -10 | -10 | -10 | -10 | -9 | -9 | -9 | -10 | -9 |
| 300 | Item | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 |
| | Parcel | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 |
| 1000 | Item | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 | -11 |
| | Parcel | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -9 | -8 |

Results showed that relative bias of parameter estimates was mainly attributable to the degree of data nonnormality, while missing mechanisms, percentage of missingness, and sample size did not appear to influence the bias of the parameter estimates. When the data were normally distributed or with skewness of 1 and kurtosis of 1.5, the relative bias was smaller than 5% for all conditions. When the skewness and kurtosis increased to 1.75 and 3.75, parameter estimates tended to be negatively biased with the relative bias be greater than 5% (in the range of -9% to -11%). Parcel-level analysis and item-level analysis yielded very similar degrees of relative bias. Although not shown in the table, similar findings were obtained for the other two parameters $F1 \rightarrow F3$ and $F2 \leftrightarrow F3$, except that the

relative bias for $F2 \leftrightarrow F3$ tended to be positive when the data had skewness of 1.75 and kurtosis of 3.75.

Standard errors of parameter estimates were also examined. The detailed results from standard errors are available upon requests. In summary, relative bias of standard errors were similar across different missing mechanism and percentage of missingness. The absolute value of relative bias was smaller than 5% for all conditions, which was smaller than the suggested cutoff of 10% for being nontrivial (Hoogland&Boomsma, 1998). In addition, parcel-level analysis and item-level analysis yielded very similar degree of relative bias.

DISCUSSION AND CONCLUSION

Applied researchers have been using item parceling techniques in their empirical data analysis, particularly when the analysis involves relatively small sample sizes. It has been shown that item parceling helps reduce model complexity, avoid violation of normality assumptions, and obtain better model-data fit, among other benefits (e.g., Little et al., 2002; Yang et al., 2010).

It may also help obviate some difficulties in analysis when missing data are present. However, how parcel-level analysis behaves when the analysis involves with missing data and/or nonnormally distributed data has not been examined. In this study, we examined the performance of parcel-level analysis under various conditions of missing and nonnormally distributed data. Results from parcel-level analysis were compared to those from item-level analysis. Robust maximum likelihood (MLR) estimation method has been recommended when data demonstrate nonnormality. However, the use of MLR has not been discussed in the literature of parceling techniques, we thus analyzed data using both ML and MLR estimation methods.

Based on the results from this simulation study, we offered four reasons for advocating the use of parceling in SEM when the analysis involves a small sample size. The smallest sample size manipulated in this study was 100, which yielded a ratio of sample size to the number of observed variables being slightly less than 5:1. This ratio can be viewed as small in SEM analysis. First, because parceling reduces model complexity, parcel-level analysis is less likely than item-level analysis to encounter estimation difficulties when the sample size is small. Second, we found that model rejection rates based on RMSEA was around the idea of 5% when the analysis was conducted at the parcel level. The chi-square test and CFI were low when the analysis was conducted at the parcel level, while the rejection rates were too high for most of the conditions when the analysis was conducted at the item level, unless sample size was large (300 or 1000). Because the analysis model was considered correctly specified in our study, parcel-level analysis yields more reasonable empirical Type I error rates for chi-square test. These two findings are consistent with previous studies (e.g., Little et al., 2002). The current study adds to this existing literature that such an advantage becomes more obvious as the percentage of data being missing increases under all three missing mechanisms: MCAR, MAR, and MNAR. Third, although MLR corrected to certain degree for the inflated model chi-square when the data were nonnormally distributed, MLR from the item-level analysis still resulted in very high Type I error rates. On the contrary, the model rejection rates based on chi-square and fit indices were reasonable when the analysis was conducted at the parcel level. Four, parcel-level analysis and item-level analysis yielded similar estimates and standard errors for structural coefficients among latent factors. This finding itself does not support the use of parceling. This was consistent with the literature where it was indicated that parceling may not be appealing under optimal conditions (Matsunaga, 2008). However, parameter estimates tend not to be interpreted when the model and data show misfit. Instead, additional parameters will be added to the model in an attempt to improve model-data fit. In other words, item-level analysis is more likely to result in an over-parameterized model, particularly when the sample size is small.

Similar to other simulation studies, one should be cautious when generalizing these conclusions to other situations. First, all the generated data are continuous. However, in practice, data are often

categorical. Future research may consider categorical incomplete data. Second, the analysis model was consistent with the data generation model. The performance of fit indices may be different and the superiority of parceling might not hold when the model is misspecified. Future research may consider misspecified models to investigate the performance of item parceling with missing data. Finally, only a limited number of levels were considered for each design factor. Based on the findings from the current study, it may be worth considering increasing the number of variables with missing values and/or the percentage of missingness. The ratio of the number of variables with missing values to the total number of variables was only .29 (= 6/21) in this study. The largest percentage of missingness was 40%. Consequently, only 11% (= 29% × 40%), at maximum, of the data were missing. Some other simulation studies have considered much higher percentage of missingness. For example, Davey and Savla (2005) and Allison (2003) included conditions with 95% and 90% missingness, respectively. Although having 95% or 90% of missing data is unlikely to encounter in an empirical study, including conditions with percentages of missingness higher than what this study had is worth considering.

REFERENCES

- Achenbach, T. M., Bernstein, A. & Dumenci, L. (2005). DSM-oriented scales and statistically based syndromes for ages 18 to 59: Linking taxonomic paradigms to facilitate multitaxonomic approaches. *Journal of personality assessment*, 84, 49-63. doi:10.1207/s15327752jpa8401_10
- Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028. doi:10.1111/j.1741-3737.2005.00191.x
- Allison, P. D. (2002). *Missing data*. California: Sage Publications.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112, 545-557. doi:10.1037/0021-843X.112.4.545
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9, 78-102. doi:10.1207/S15328007SEM0901_5
- Bandalos, D. L. (2006). *The Use of Monte Carlo Studies in Structural Equation Modeling Research*. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 385-426). Greenwich, Connecticut: Information Age Publishing.
- Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling*, 15, 211-240. doi:10.1080/10705510801922340
- Bandalos, D.L., & Finney, S. J. (2001). *Item parceling issues in structural equation modeling*. In G.A. Marcoulides & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp.269-293). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bentler, P.M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107, 238-246. doi:10.1037/0033-2909.107.2.238
- Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, 105, 467-477. doi:10.1037/0033-2909.105.3.467
- Bollen, K.A. (1989) *Structural equations with latent variables*, Wiley, New York.
- Chen, G., & Astebro, T. (2003). How to deal with missing categorical data: Test of a simple Bayesian method. *Organizational Research Methods*, 6, 309-327. doi:10.1177/1094428103254672
- Chou, C.P., Bentler, P.M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357. doi:10.1111/j.2044-8317.1991.tb00966.x
- Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29. doi:10.1037/1082-989X.1.1.16
- Davey, A., & Savla, J. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling*, 12(4), 578-597. doi:10.1207/s15328007sem1204_4
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6, 352-370. doi:10.1037/1082-989X.6.4.352

- Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*, 64, 419-436. doi:10.1177/0013164403261050
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 3, 430-457. doi:10.1207/S15328007SEM0803_5
- Finney, S. J., & DiStefano, C. (2006). *Non-Normal and Categorical Data in Structural Equation Modeling*. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 269-314). Greenwich, Connecticut: Information Age Publishing.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532. doi:10.1007/BF02293811
- Hoogland, J. J. & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research*, 26, 329-367. doi:10.1177/0049124198026003003
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. doi:10.1080/10705519909540118
- Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173. doi:10.1207/S15328007SEM0902_1
- Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures*, 2, 260-293. doi:10.1080/19312450802458935
- Meade, A. W., & Kroustalis, C. M. (2005, April). *Problems with item parceling for confirmatory factor analysis tests of measurement invariance of factor loadings*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30
- Muthén, B., & Muthén, L. (1998-2008). *Mplus user's guide (6th edition)*. Los Angeles, CA: Muthén&Muthén.
- Nasser, F., & Takahashi, T. (2003). The effect of using item parcels on Ad Hoc goodness-of-fit indexes in confirmatory factor analysis : An example using Sarason's reactions to tests. *Applied Measurement in Education*, 16, 75-97. doi:10.1207/S15324818AME1601_4
- Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement*, 63, 729-757. doi:10.1177/0013164403258228
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525-556. doi:10.3102/00346543074004525
- Plummer, B. A. (2000). *To parcel or not to parcel: The effects of item parceling in confirmatory factor analysis*. Unpublished doctoral dissertation, University of Rhode Island.
- Sass, D. A., & Smith, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling*, 13, 566-586. doi:10.1207/s15328007sem1304_4
- Schafer, J. L., & Graham, J. W. (2002). Missing data : Our view of the state of the art. *Psychological Methods*, 7, 147-177. doi:10.1037/1082-989X.7.2.147
- Signorella, M. L., & Cooper, J. E. (2011). Relationship suggestions from self-help books: *Gender stereotyping, preferences, and context effects*. *Sex Roles*, 65, 371-382. doi:10.1007/s11199-011-0023-4
- Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Sterba, S.K. (2011). Implications of parcel-allocation variability for comparing fit of item-solutions and parcelsolutions. *Structural Equation Modeling*, 18, 554-577.
- Sterba, S.K. & MacCallum, R.C. (2010). *Variability in parameter estimates and model fit across repeated allocations of items to parcels*. *Multivariate Behavioral Research*, 45, 322-358.

- Yang, C., Nay, S., & Hoyle, R. H. (2010). Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. *Applied Psychological Measurement*, 43, 122-142. doi:10.1177/0146621609338592
- Yoder, J. D., Snell, A. F., & Tobias, A. (2012). Balancing multicultural competence with social justice: Feminist beliefs and optimal psychological functioning. *The Counseling Psychologist*, 40, 1101-1132. doi:10.1177/001100001142629

UZUN ÖZET

Giriş

Yapısal Eşitlik Modellemesi (YEM) bir grup değişken arasındaki varsayılan ilişkilerin ampirik veriler kullanılarak test edilmesi için sıklıkla kullanılır. YEM’de yoğun olarak kullanılan maksimum olabilirlik (ML) tahmin metodu örneklem büyüklüğünün yüksek ve gözlenen değişkenlerin normal dağılmasını varsaymaktadır. Bu varsayımların sağlanmaması ki-kare istatistiği, uyum ideksleri, parameter tahmini ve parameter tahmininin standart hatalarının hatalı olmasına sebep olmaktadır (Bollen, 1989; Curran, West, & Finch, 1996; Chou, Bentler, & Satorra, 1991). Ayrıca bu yanlış modelin karmaşıklığı arttıkça artma eğilimi göstermektedir. Varsayımların sağlanmaması durumunda ML’ye alternatif olarak sağlam maksimum olabilirlik (MLR) tahmin metodu kullanılabilir. MLR ki-kare değerlerindeki pozitif yanlılığı ve standart hatalardaki negatif yanlılığı düzeltmektedir.

YEM’in amacının tekil maddelerin özelliklerini test etmekten ziyade gizil değişkenler arasındaki ilişkiyi incelemek olduğunda, özellikle örneklem büyüklüğünün yetersiz olduğu ve her bir gizil değişkenin gösterge sayısının fazla olması durumunda, parsel tekniklerinden biri kullanılabilir (Sterba & MacCallum, 2010). Parselleme birçok maddenin toplam veya ortalama puanlarının hesaplanması olarak tanımlanan bir prosedür olarak tanımlanır. Parsel olarak tanımlanan bu yeni değişkenler bireysel maddeler yerine YEM analizinde gizil değişkenlerin göstergesi olarak kullanılır (Bandalos 2002, 2008; Little, Cunningham, Shahar, & Widaman, 2002; Sass & Smith, 2006; Sterba, 2011; Sterba & MacCallum, 2010; Yang, Nay, & Hoyle, 2010). Parsel kullanmak model karmaşıklığını azalttığı, örneklem gereksinimin azalttığı, bireysel maddelerden kaynaklanan sistematik hataların model üzerindeki etkisini azalttığı, yüksek güvenilirliğe yardımcı olduğundan, modellerin yakınsama oranlarını arttırdığından ve maddelerin boyutları bilindiğinde model veri uyumunu arttırdığından dolayı caziptir. Sass ve Smith (2006) model varsayımlarının sağlanması ve maddelerin tek boyutlu olduğu durumlarda parsellemenin faktörler arasındaki yapısal ilişki parametrelerinin tahmininde yanlılığa sebep olmadığını göstermişlerdir.

Kayıp verilerin varlığı YEM analizlerinde muhtemel problemler oluşturmaktadır. Literatürde en yaygın tartışılan kayıp veri türleri *tamamen rastgele kayıp* (MCAR), *rastgele kayıp* (MAR) ve *rastgele olmayan kayıp* (MNAR) olarak tanımlanabilir. Kayıp verileri ele alan bir çok teknikten sosyal bilimlerdeki ampirik analizlerde en çok kullanılan maksimum olabilirlik metodudur (Enders & Bandalos, 2001). Kayıp verilerin olması durumunda bu metod tam bilgi maksimum olabilirlik (FIML) olarak adlandırılır. Araştırmacılar FIML metodunun verilerin MCAR ve MAR olması durumunda yansız parametre tahmini yaptığını bulmuşlardır (Enders & Bandalos, 2001).

Örneklem büyüklüğünün düşük olması, modelin karmaşık olması ve verilerde kayıpların olması ve araştırmanın amacının maddeler arasındaki ilişkiden ziyade faktörler arasındaki ilişkiyi incelemek olduğunda, ampirik çalışmalarda parselleme tekniği, YEM analizlerindeki bazı zorlukları gidermek için kullanılabilir. Schafer and Graham (2002) buna benzer fikirleri farklı şekilde adlandırarak (örn. “*case-by-case item deletion*” veya “*ipsative mean imputation*”) ileri sürmüştür.

Buna bağlı olarak, çalışmanın amacı, parsellemenin kayıp ve normal olmayan verilerle YEM analizlerinde nasıl davranacağını simülasyon aracılığıyla araştırmaktır. Parsellemeye dayalı sonuçlar bireysel maddelerle oluşturulan modellerle karşılaştırılmıştır.

Metod

Veriler şekil 1’de gösterilen modele göre üretilmiştir. Faktör yükleri şekilde gösterildiği gibidir. Her bir maddenin hata terimi ise bir eksi yükün karesi olarak sabitlenmiştir. Veri üretimini için kullanılan dizayn faktörleri şu şekildedir:

- Kayıp mekanizması: Üç farklı mekanizma dikkate alınmıştır; MCAR, MAR ve MNAR.
- Kayıp yüzdesi: Üç farklı yüzde seviyesi dikkate alınmıştır; 10%, 20% ve 40%.
- Madde puanlarının dağılımı: Üç farklı dağılım dikkate alınmıştır; (1) normal dağılım; (2) orta seviye çarpık ve düşük basıklık ($Sk = 1$, ve $K=1.5$); (3) yüksek çarpıklık ve yüksek basıklık ($Sk = 1.75$ ve $K=3.75$). Sadece 1-15 arasındaki maddelerde non-normallik kullanılmıştır.
- Örneklem büyüklüğü: Üç farklı örneklem büyüklüğü (SS) dikkate alınmıştır; 100, 300, 1000.

Bu dizayn faktörler kullanılarak, toplam 90 farklı durum oluşturulmuştur. Herbir durum için 2000 veri seti, R (versiyon 2.13.2) programında üretilmiştir. İlk olarak faktörler arasındaki korelasyon elde edilmiştir. Daha sonra ortalama sıfır ve standart sapma bir olacak şekilde rastgele 24 değişken üretilmiştir. Cholesky ayrıştırma metodu kullanılarak faktör puanları çoklu normal dağılacak şekilde dönüştürülmüştür. Üçüncü olarak, gözlenen madde puanları faktörlerin ve hata terimlerinin doğrusal bir kombinasyonu olarak elde edilmiştir (örn. Bernstein & Teng, 1989). Sabit çarpıklık ve basıklık değerleri için Fleishman’ın tekniği kullanılarak gözlenen değerler oluşturulmuştur (Fleishman, 1978). Verilerin MCAR olması durumunda rastgele seçilen değerler değişkenlerden silinmiştir. MAR olması durumunda ise madde 1, 2, 3, 6, 7 ve 8’in değerleri sırasıyla 4, 5, 16, 11, 12 ve 13’ün değerlerine göre silinmiştir. MNAR olması durumunda ise değişkenlerin değerleri kendi değerlerine göre silinmiştir.

Mplus 6.1 (Muthén & Muthén, 1998-2008) programı kullanılarak herbir veri seti madde ve parsel seviyelerinde test edilmiştir. Parcel oluşturmak için factorial parselleme tekniği kullanılmış (Matsunaga, 2008). Herbir model ise ML ve MLR tekniği altında test edilmiştir. Her bir durum için genel model veri uyumu ki-kare testi, karşılaştırmalı uyum indeksi (CFI) ve kök ortalama kare yaklaşım hatasına (RMSEA) dayanarak Hu ve Bentler’in (1999) kriterleriyle karşılaştırılmıştır. Ayrıca nokta tahmini ve standart hatası için göreceli yanlışlık değerleri hesaplanmıştır.

Sonuçlar ve Tartışma

Her bir durum için model yakınsaması ve model reddetme oranları rapor edilmiştir. Tablo 1 ki-kare ye dayalı model reddetme oranlarını göstermektedir. Model doğru tanımlanmış olduğundan red oranının 5% olması beklenir. Madde seviyesindeki ML’ye dayalı oranlar 5% ‘den oldukça yüksektir. Bu oran SS=1000 olması durumunda azalmaktadır. Bu reddetme oranı kayıp veri yüzdesi ile artmaktadır. Aynı şartlar altında parsel seviyesindeki modellerin reddetme oranları daha düşüktür ve dağılıma kayıp mekanizmasına ve kayıp yüzdesine göre göreceli olarak sabittir.

Sonuçlara göre, tahmin yanlışlığı çoğunlukla verilerin normal dağılmamasından kaynaklanmaktadır. Kayıp oranı ve mekanizması ve örneklem büyüklüğünün yanlışlığı etkilediği söylenemez. Parsel ve madde analizleri benzer parametre yanlışlıkları ve parametre tahmininin standart hata yanlışlığı göstermiştir.

Parselleme teknikleri ampirik çalışmada kullanılmaktadır, özellikle örneklem büyüklüğünün düşük olması durumunda. Parsellemenin model karmaşıklığını azalttığı ve daha iyi bir model-veri uyumu sağladığı gibi bazı yararları vardır. Bu çalışmada parselleme tekniğinin kayıp ve normal olmayan verilerdeki performansı incelenmiştir.

Sonuçlara göre, parsel seviyesindeki modellerde daha az tahmin zorluğu olmuştur, özellikle küçük örneklem büyüklüğünde. Örneklem büyüklüğünün yüksek olması durumlarından başka, parsel seviyesinde ki-kare ve CFI’ya dayalı model reddetme oranı yüksek fakat RMSEA değerleri .05

civarındadır. Ayrıca, madde seviyesinde model reddetme oranları MLR'nin kullanılması durumunda da parsel seviyesindeki değerlerden yüksektir. Son olarak, madde ve parsel seviyesinde modeller YEM parametreleri ve bunların standart hataları bakımından benzer sonuçlar vermiştir. Bu çalışmada sürekli değişkenler üzerinde ve doğru tanımlanmış modeller üzerinden simülasyon yapılmıştır. Daha sonraki çalışmalarda doğru tanımlanmamış maddeler veya kategorik değişkenler için bir simülasyon yapılabilir.

Araştırmacıların Nicel Veri Analizinde Karşılaştıkları Güçlüklerin İkili Karşılaştırmalar Yoluyla İncelenmesi

An Analysis of Researchers' Difficulties in Quantitative Data Analysis with the Use of Pairwise Comparisons

Mustafa İLHAN *

Öz

Bu çalışmada, Thurstone'un ikili karşılaştırmalar yöntemi kullanılarak araştırmacıların nicel veri analizinde karşılaştıkları güçlükler incelenmiştir. Araştırma, alan eğitimi ve eğitim bilimleri alanında çalışan 110 katılımcı üzerinde yürütülmüştür. Çalışmada, nicel veri analizi sürecinde takip edilen işlemler altı başlıkta ifade edilmiş ve bu altı başlık ikili karşılaştırmalar halinde katılımcılara sunulmuştur. Dolayısıyla, ölçme aracında nicel veri analizi sırasında takip edilen altı işlem basamağının ikili kombinasyonlarından oluşan 15 madde yer almıştır. Katılımcılardan nicel veri analizinde yaşadıkları güçlükleri dikkate alarak, karşılaştırılan iki işlemden hangisinde daha fazla zorlandıklarını belirtmeleri istenmiştir. Ölçekleme çalışması sonucunda, *i*) araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların test edilmesi araştırmacıların nicel veri analizinde en fazla güçlük yaşadıkları işlem olarak tespit edilmiştir. Bunu sırasıyla; *ii*) araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi, *iii*) analiz sonuçlarının yorumlanması, *iv*) analiz sonuçlarının raporlaştırılması, *v*) araştırma problemine uygun olan ve varsayımlarının karşılandığı belirlenen istatistiksel tekniğin bir paket program aracılığıyla gerçekleştirilmesi ve son olarak *vi*) veri dosyasının oluşturulup düzenlenmesi işlemleri takip etmiştir.

Anahtar Kelimeler: Nicel veri analizi, istatistik, ölçekleme teknikleri, ikili karşılaştırmalarla ölçekleme

Abstract

This study analyzed the difficulties that researchers have with quantitative data analysis using Thurstone's pairwise comparison method. The study sample included 110 participants working in educational sciences and teacher education for specific subject areas. The study presented the operations performed in quantitative data analysis under six headings, which were presented to the participants as pairwise comparisons. Thus, the study included 15 items consisting of the pair combinations of the six operational stages followed in quantitative data analysis. The participants were asked to consider their difficulties in quantitative data analysis and identify with which of the two operations they had the greatest difficulty. The scaling study found that the researchers' greatest difficulty in quantitative data analysis was: *i*) testing the assumptions associated with the statistical techniques that are determined to be suitable for the study. Respectively, this is followed by: *ii*) determining the suitable statistical techniques for the research problem, the study design and the variables in the study, *iii*) interpreting the analysis results, *iv*) preparing a report on the analysis results, *v*) applying statistical technique that is determined to be suitable for the research problem and confirm the assumptions using the package program, and finally, *vi*) creating and arranging the data file.

Key Words: Quantitative data analysis, statistics, scaling techniques, scaling the pairwise comparison

GİRİŞ

Bilimsel araştırma sürecinin en önemli adımlarından biri verilerin analiz edilmesidir (Bayram, 2009). Çünkü çalışma kapsamında toplanan verilerin araştırma problemi bağlamında ne ifade ettiği veri analizi ile ortaya konulmaktadır. Bu anlamda veri analizi, ham verilere anlam kazandırma süreci olarak görülebilir (Altunışık, Coşkun, Bayraktaroglu ve Yıldırım, 2012). Veri analizi; veri kümesinin düzenlenmesi, betimsel istatistiklerin hesaplanması, hipotez testlerinin yapılması, korelasyon ve

* Yrd. Doç. Dr., Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, İlköğretim Bölümü, Diyarbakır-Türkiye.
e-posta: mustafailhan21@gmail.com

regresyon ile diğer istatistiksel işlemlerin gerçekleştirilmesini kapsamaktadır. Sıralanan istatistiksel işlemler, bundan 25-30 yıl öncesine kadar elle yapılıyor ya da basit denebilecek hesap makineleriyle yürütülüyordu. Bu durum geniş kapsamlı araştırmaların yapılamamasına, yapılanların uzun zaman almasına ve bazen de araştırma hatalarına neden olabilmekteydi (Baykul ve Güzeller, 2013). Bilgisayarın bilimsel araştırmalarda kullanılmaya başlaması ve istatistiksel işlemleri yapmaya hizmet eden paket programların geliştirilmesiyle birlikte, veri analizine ilişkin söz konusu güçlükler büyük ölçüde aşılmıştır (Field, 2009). Geçmişte yapılması mümkün olmayan ya da çok uzun sürede yapılabilen istatistiksel analizler günümüzde bilgisayarlar ve paket programlar aracılığıyla kolayca gerçekleştirilebilir hale gelmiştir (Kalaycı, 2010; Tan, 2016). Bununla birlikte, veri analizi sürecinde paket programlardan yararlanabilmek için araştırmacının bir takım yeterliliklere sahip olması gerekmektedir. Bu yeterlilikler, istatistik bilgisi ve paket programlar bilgisi şeklinde iki başlıkta ele alınmaktadır (Sipahi, Yurtkoru ve Çinko, 2010).

Araştırmacının; *i*) araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklere karar verebilmesi (Akbulut, 2010; Büyüköztürk, 2010), *ii*) bu istatistiksel tekniklere ilişkin varsayımların karşılanıp karşılanmadığını belirleyebilmesi (Seçer, 2013), *iii*) analiz sonuçlarını nasıl yorumlayacağını ve raporlayacağını bilmesi paket programları kullanabilmek için sahip olması gereken istatistiksel yeterliliklerdir (Aziz, 2011). Bir araştırmacı bir istatistikçi olmak zorunda değildir; fakat en azından araştırmasında hangi analizi niçin yaptığını ve yaptığı analizleri nasıl yorumlayacağını bilecek kadar istatistiğe hâkim olmak durumundadır. Daha açık bir anlatımla, çalışma kapsamında topladığı verileri doğru bir şekilde analiz edebilmesi, araştırmacının istatistik teknikleri konusunda yetişmiş olmasını gerektirmektedir (Erkuş, 2011). Ancak, istatistik konusunda yeterli bilgiye sahip olması araştırmacının veri analizinde paket programları etkin bir şekilde kullanabilmesi için tek başına yeterli değildir. Gerçekleştirilecek istatistiksel analize uygun bilgisayar programını seçebilmesi (Creswell, 2012), araştırma verilerini bu programın öngördüğü doğrultuda düzenleyebilmesi ve programdaki komutları yerine getirebilmesi araştırmacının veri analizinde paket programlardan yararlanabilmesinin diğer ön şartlarıdır (Pallant, 2005).

Araştırmacının yeterli istatistik bilgisine sahip olmaması ya da paket programları kullanmada sorun yaşaması; araştırma problemine uygun olmayan istatistiksel teknikleri kullanmasına, gerekli olan analizleri yapamamasına, analiz sonuçlarını yanlış ya da eksik yorumlamasına yol açabilmektedir. Bu durum araştırmacıyı sınırlayan ve bilimsel çalışmaların geçerliği ile güvenilirliğini tehdit eden önemli bir faktördür (Baykul ve Güzeller, 2013). Literatürdeki çok sayıda çalışmada, araştırmacıların veri analizi konusundaki eksiklikleri, bilimsel araştırmaların niteliğini olumsuz yönde etkileyen bir özellik olarak belirlenmiştir. Örneğin, Kabaca ve Erdoğan (2007) fen bilimleri, bilgisayar ve matematik eğitimi alanlarındaki tez çalışmalarını istatistiksel açıdan inceledikleri araştırmada; *a*) birden fazla bağımlı değişken içeren çalışmalarda çok değişkenli varyans analizinin yerine, tek değişkenli istatistiklerin kullanıldığı *b*) deneysel çalışmalarda ortak değişkenin kontrol altına alınmasına olanak tanıyan kovaryans analizinin yerine *t*-testinin uygulandığı ve *c*) parametrik testlerin kullanımı ile ilgili hatalar yapıldığı sonucuna ulaşmışlardır. Aynı şekilde; Evrekli, İnel, Deniz ve Balım'ın (2011) fen eğitimi alanındaki lisansüstü tezlerdeki istatistiksel sorunları inceledikleri araştırmada, tezlerin birçoğunda parametrik ya da non-parametrik yöntem seçiminin yanlış yapıldığı ve analizlerin veri durumu göz önünde bulundurulmadan gerçekleştirildiği saptanmıştır. Bilimsel araştırmalarda yapılan istatistiksel hatalara örnek olabilecek bir diğer çalışma da Ozan ve Köse (2014) tarafından yapılmıştır. Ozan ve Köse'nin (2014), eğitim programları ve öğretim alanındaki araştırma eğilimlerini inceledikleri çalışmada, araştırmacıların çalışma problemine uygun istatistiksel teknikler yerine daha iyi bildikleri analizlere yöneldikleri tespit edilmiştir. Söz gelimi; deneysel çalışmalarda kovaryans analizinin kullanılması daha uygun iken *t*-testinin kullanıldığı, birden fazla bağımlı değişken içeren çalışmalarda çok değişkenli istatistiklere başvurulması daha doğru iken; tek değişkenli analizlerin uygulandığı belirlenmiştir. Araştırmaların geçerlik ve güvenilirliğinin artırılıp, bilimsel niteliklerinin yükseltilebilmesi için veri analizine ilişkin söz konusu hataların giderilmesi oldukça önemlidir. Veri analizindeki hataların giderilmesi ise araştırmacıların bu konudaki yeterliliklerinin geliştirilmesiyle mümkün olabilir. Araştırmacıların veri analizinde yaşadıkları güçlüklerin ortaya konulması, bu konudaki yeterliliklerinin geliştirilmesi için

neler yapılması gerektiği hakkında ipuçları verebilir. Dolayısıyla, veri analizinde yaşanan güçlüklerin belirlenmesine yönelik bir çalışmanın alanyazına kazandırılması önem arz etmektedir.

Araştırmanın Amacı

Bu çalışmada, alan eğitimi ve eğitim bilimleri alanında çalışan araştırmacıların nicel veri analizinde karşılaştıkları güçlüklerin ikili karşılaştırmalar yoluyla incelenmesi amaçlanmıştır. Veri analizi bilimsel araştırmaların temel basamaklarından biri olmasına karşın, literatürde araştırmacıların bu süreçte karşılaştıkları güçlükleri belirlemeye yönelik bir çalışmaya rastlanmamıştır. Bu bakımdan çalışmanın alanyazına katkı sağlayacağı düşünülmektedir. Çalışma sonucunda araştırmacıların veri analizini gerçekleştirirken daha çok hangi işlemlerde zorlandıkları tespit edileceğinden, lisans ve lisansüstü düzeyde okutulan istatistik ve veri analizi ile ilgili derslerin öğretim programları düzenlenirken araştırmada ulaşılan bulgulardan yararlanılabileceği tahmin edilmektedir. Bunun yanı sıra; istatistik, veri analizi ve bilimsel araştırma yöntemleri ile ilgili kitap hazırlayan araştırmacıların da çalışmada ulaşılabilecek sonuçlardan istifade edebileceklerine inanılmaktadır.

YÖNTEM

Bu bölümde araştırma modeli, çalışma grubu, veri toplama aracı, veri toplama süreci ve verilerin analizi hakkındaki bilgilere yer verilmiştir.

Araştırma Modeli

Nicel veri analizinde karşılaşılan güçlüklerin belirlenmesinin amaçlandığı bu araştırma, durum tespitine yönelik bir tarama çalışmasıdır. Tarama modelindeki araştırmalar, mevcut durumun herhangi bir müdahalede bulunulmaksızın olduğu gibi yansıtılması esasına dayanır (Karasar, 2009).

Çalışma Grubu

Araştırma, alan eğitimi ve eğitim bilimleri alanında çalışan 50'si (%45.50) kadın ve 60'ı (%54.50) erkek toplam 110 katılımcı üzerinde yürütülmüştür. Katılımcıların yaşları 23 ile 53 arasında değişmekte olup grubun yaş ortalaması 30.65'tir. Çalışma grubunun eğitim düzeylerine göre dağılımı; 12 (%10.90) yüksek lisans öğrencisi, 1 (%0.9) yüksek lisans mezunu, 82 (%74.5) doktora öğrencisi ve 15 (13.6) doktora mezunu şeklindedir. Çalışma grubunda, alan eğitiminden 52 (%47.30) katılımcı yer alırken, eğitim bilimleri alanından 58 (%52.70) katılımcı bulunmaktadır.

Veri Toplama Aracı ve Verilerin Toplanması

Araştırmada ikili karşılaştırmalar ölçekleme yöntemi kullanılmış ve veri toplama aracı bu ölçekleme yöntemine uygun olarak hazırlanmıştır. Bu doğrultuda, nicel veri analizi sürecinde takip edilen basamaklar altı başlıkta ifade edilmiştir. Sözü edilen başlıklar Tablo 1'de gösterilmiştir. Tablo 1'de de görüldüğü üzere, veri analizinde takip edilen bu altı işlem basamağı için U1, U2, U3, U4, U5 ve U6 gösterimleri kullanılmıştır. Bu gösterimlerdeki U harfi *uyarıcı* ifadesini temsil etmektedir.

Tablo 1. Nicel veri analizinde takip edilen basamaklar

| | |
|----|---|
| U1 | Veri dosyasının oluşturulması ve düzenlenmesi |
| U2 | Araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi |
| U3 | Araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların test edilmesi |
| U4 | Araştırma problemine uygun olan ve varsayımlarının karşılandığı belirlenen istatistiksel tekniğin bir paket program aracılığıyla gerçekleştirilmesi |
| U5 | Analiz sonuçlarının yorumlanması |
| U6 | Analiz sonuçlarının raporlaştırılması |

Tablo 1'deki altı başlık ikili karşılaştırmalar halinde katılımcılara sunulmuştur. Dolayısıyla, ölçme aracında nicel veri analizinde takip edilen altı işlem basamağının ikili kombinasyonlarından oluşan 15 madde yer almıştır. Ayrıca, katılımcıların demografik özellikleri hakkında bilgi edinmek için ölçme aracına cinsiyet, yaş, eğitim düzeyi ve lisansüstü eğitim görülen alan değişkenleri eklenmiştir. Araştırmanın verileri 2015 yılı Kasım ayı içerisinde toplanmıştır. Veri toplama aracı katılımcıların bir kısmına yüz yüze uygulanmış, bir kısmına ise elektronik posta yoluyla gönderilmiştir. Katılımcılardan nicel veri analizinde yaşadıkları güçlükleri dikkate alarak, karşılaştırılan iki işlemden hangisinde daha fazla zorlandıklarını belirtmeleri istenmiştir. Veri toplama aracının başına, sunulan ikililerden mutlaka birinin tercih edilmesi gerektiğine dair bir uyarı eklenmiştir. Buna rağmen, yanıtız bırakılan ikili karşılaştırma içeren ölçme araçları kapsam dışı tutulmuştur.

Veri Analizi

Araştırma kapsamında toplanan veriler, Thurstone'un karşılaştırmalı yargı kanununun beşinci hal denklemi ile tam veri matrisinden ölçeklenmiştir. Bunun için öncelikle, katılımcıların nicel veri analizinde takip edilen basamaklara ilişkin yaptıkları karşılaştırmalara ait frekans değerleri hesaplanmıştır. Hesaplanan bu değerler kullanılarak frekans matrisi oluşturulmuş ve sonrasında frekans matrisindeki her bir değer toplam katılımcı sayısına bölünerek oranlar matrisi elde edilmiştir. Ardından, oranlar matrisinin elemanlarına karşılık gelen standart Z puanları belirlenmiş ve böylece birim normal sapmalar matrisine (Z matrisi) ulaşılmıştır (Erkuş, 2012). Z matrisinde, esas köşegene simetrik olan elemanlarının mutlak değerce eşit ve işaretçe zıt olması gerekir (Turgut ve Baykul, 1992). Çalışmada bu özellikten yararlanılarak, Z matrisinin düzenlenmesinde bir hata yapıp yapılmadığı kontrol edilmiştir. Bu işlemi takiben birim normal sapmalar matrisinin en alt satırına, her bir sütuna ait değerlerin ortalamasına karşılık gelen bir satır eklenerek ölçek değerleri bulunmuştur. Ölçek değerleri için en küçük eleman sıfır olacak şekilde bir öteleme yapıldıktan sonra, sıfıra daha yakın olan uyarıcılar katılımcıların nicel veri analizinde daha fazla güçlük yaşadıkları basamaklar olarak yorumlanmıştır.

Ölçekleme çalışmasının son aşamasında, kestirilen ölçek değerlerinin iç tutarlılığına bakılmıştır. Bu amaçla gözlenen oranlar ile beklenen/teorik oranların uyum düzeyi incelenmiştir. Uyum düzeyinin değerlendirilmesinde ilk olarak ölçek değerlerinin ortalama hatası hesaplanmıştır. Turgut ve Baykul'a (1992) göre, hesaplanan ortalama hatanın küçük olması ölçek değerlerinin güvenilir olduğuna işaret ederken; yüksek olması katılımcı yargılarının güvenilir olmadığı anlamına gelmektedir. Ortalama hata değeri, gözlenen ve beklenen oranlar arasındaki uyum hakkında bilgi verse de bu konudaki nihai karar için Ki Kare testinin uygulanması gerekmektedir. Matematiksel ölçekleme modelinin ampirik veriler ile uygunluk derecesi Ki Kare Testi ile saptanmaktadır (Turgut ve Baykul, 1992). Bu doğrultuda, çalışmada hesaplanan Ki Kare değerinin ilgili serbestlik derecesindeki Ki Kare değerinden küçük olması, gözlenen oranlar ile beklenen oranlar arasında anlamlı bir fark bulunmadığına ve ölçek değerlerinin iç tutarlılığının sağlandığına yönelik bir kanıt olarak kabul edilmiştir. Çalışmada, veri analizi ile ilgili bütün hesaplamalar Microsoft Office Excel programında yapılmıştır.

BULGULAR

Bu bölümde araştırmadan elde edilen bulgulara yer verilmiştir. İlk olarak, her bir uyarıcıya ait frekans değerleri belirlenmiştir. Bu frekans değerlerinden oluşan frekans matrisi Tablo 2'de sunulmuştur. Frekans matrisinin i . satır ile j . sütuna karşılık gelen eleman (x_{ij}); satırdaki uyarıcıyı sütundaki uyarıcıya göre daha güç bir işlem olarak gören katılımcı sayısını yansıtmaktadır. Örneğin (U1, U2) değeri; 20 katılımcının bir numaralı uyarıcıyı (veri dosyasının oluşturulması ve düzenlenmesi), iki numaralı uyarıcıya (araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi) kıyasla daha zor bir işlem olarak gördüklerine işaret etmektedir. Tablo 2'den anlaşılacağı üzere, frekans matrisinde esas köşegene göre simetrik olan elemanların toplamı çalışmadaki katılımcı sayısına eşittir.

İkinci adımda, frekans matrisinin her bir hücresindeki değer çalışmadaki toplam katılımcı sayısına bölünerek oranlar matrisi oluşturulmuştur. Oranlar matrisi, frekans matrisi ile birlikte Tablo 2’de sunulmuştur. Oranlar matrisinin esas köşegene göre simetrik olan elemanlarının toplamı bire eşittir. Frekans ve oranlar matrisindeki hücre değerlerine bakıldığında, U1’i diğer uyarıcılara göre daha zor bir işlem olarak gören katılımcı sayısının oldukça az olduğu dikkat çekmektedir. Yine frekans ve oranlar matrisindeki değerler, veri analizinde takip edilen basamaklardan hangisi ile karşılaştırılırsa karşılaştırılsın U3’ün araştırmacılar tarafından daha zor bir işlem olarak görüldüğünü yansıtmaktadır. Bu durum, araştırmacıların nicel veri analizinde en az zorlandıkları işlemin *veri dosyasının oluşturulması ve düzenlenmesi*; en çok zorlandıkları işlemin ise *araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların test edilmesi* olduğunu düşündürmektedir.

Tablo 2. Nicel veri analizinde karşılaşılan güçlüklerle ilişkin frekans matrisi, oranlar matrisi ve birim normal sapmalar matrisi

| | | U1 | U2 | U3 | U4 | U5 | U6 |
|-------------------------------|-----------------|--------|---------|---------|---------|---------|---------|
| Frekans Matrisi | U1 | | 20 | 18 | 28 | 25 | 27 |
| | U2 | 90 | | 50 | 74 | 54 | 66 |
| | U3 | 92 | 60 | | 71 | 59 | 63 |
| | U4 | 82 | 36 | 39 | | 43 | 49 |
| | U5 | 85 | 56 | 51 | 67 | | 65 |
| | U6 | 83 | 44 | 47 | 61 | 45 | |
| | U6 | 83 | 44 | 47 | 61 | 45 | |
| Oranlar Matrisi | U1 | | 0.1818 | 0.1636 | 0.2545 | 0.2273 | 0.2455 |
| | U2 | 0.8182 | | 0.4545 | 0.6727 | 0.4909 | 0.6000 |
| | U3 | 0.8364 | 0.5455 | | 0.6455 | 0.5364 | 0.5727 |
| | U4 | 0.7455 | 0.3273 | 0.3545 | | 0.3909 | 0.4455 |
| | U5 | 0.7727 | 0.5091 | 0.4636 | 0.6091 | | 0.5909 |
| | U6 | 0.7545 | 0.4000 | 0.4273 | 0.5545 | 0.4091 | |
| | U6 | 0.7545 | 0.4000 | 0.4273 | 0.5545 | 0.4091 | |
| Birim Normal Sapmalar Matrisi | U1 | | -0.9085 | -0.9796 | -0.6603 | -0.7479 | -0.6889 |
| | U2 | 0.9085 | | -0.1142 | 0.4475 | -0.0228 | 0.2533 |
| | U3 | 0.9796 | 0.1142 | | 0.3731 | 0.0913 | 0.1833 |
| | U4 | 0.6603 | -0.4475 | -0.3731 | | -0.2770 | -0.1372 |
| | U5 | 0.7479 | 0.0228 | -0.0913 | 0.2770 | | 0.2299 |
| | U6 | 0.6889 | -0.2533 | -0.1833 | 0.1372 | -0.2299 | |
| | ΣZ_j | 3.9851 | -1.4723 | -1.7415 | 0.5744 | -1.1862 | -0.1595 |
| | Z_{j_ort} | 0.6642 | -0.2454 | -0.2902 | 0.0957 | -0.1977 | -0.0266 |
| | Ölçek (S_j) | 0.9544 | 0.0449 | 0.0000 | 0.3860 | 0.0925 | 0.2637 |

Üçüncü adımda, oranlar matrisindeki hücre değerlerine karşılık gelen z değerleri belirlenerek birim normal sapmalar matrisine (Z matrisi) ulaşılmıştır. Z matrisi de frekans ve oranlar matrisi ile birlikte Tablo 2’de gösterilmiştir. Z matrisinin elemanlarının esas köşegene göre mutlak değerce eşit ve işaret yönünden zıt olması gerekir. Tablo 2’deki hücre elemanları bu gerekliliği karşıladığından, Z matrisinin oluşturulmasında bir hata yapılmadığı söylenebilir. Tablo 2’deki ΣZ_j satırı, her bir sütundaki elemanların toplamına eşittir. ΣZ_j satırındaki değerlerin karşılaştırılan uyarıcı sayısına bölünmesiyle Z_{j_ort} satırındaki değerler hesaplanmıştır. Tablo 2’ye bakıldığında, Z_{j_ort} satırındaki en küçük değer -0.2902 olduğu görülmektedir. Eksenin başlangıç noktasını sıfıra kaydırmak için tüm Z_{j_ort} değerlerine 0.2902 eklenmiştir. Böylece ölçek değerlerine ulaşılmıştır. Ölçek değerleri küçük olan uyarıcılar, katılımcıların nicel veri analizinde daha fazla güçlük yaşadıkları işlem adımlarını

temsil etmektedir. Buna göre, katılımcıların nicel veri analizinde yaşadıkları güçlükler Tablo 3'teki gibi sıralanmaktadır.

Tablo 3. Nicel veri analizinde karşılaşılan güçlüklerle ilişkin ölçek değerleri ve sıralamalar

| Nicel veri analizinde karşılaşılan güçlükler | Ölçek Değerleri | Sıralama |
|--|-----------------|----------|
| U1 - Veri dosyasının oluşturulması ve düzenlenmesi | 0.9544 | 6 |
| U2 - Araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi | 0.0449 | 2 |
| U3 - Araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların test edilmesi | 0.0000 | 1 |
| U4 - Araştırma problemine uygun olan ve varsayımlarının karşılandığı belirlenen istatistiksel tekniğin bir paket program aracılığıyla gerçekleştirilmesi | 0.3860 | 5 |
| U5 - Analiz sonuçlarının yorumlanması | 0.0925 | 3 |
| U6 - Analiz sonuçlarının raporlaştırılması | 0.2637 | 4 |

Tablo 3'e göre, katılımcılar en fazla *araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların testinde* zorlanmaktadır. Bunu sırasıyla; araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi, analiz sonuçlarının yorumlanması, analiz sonuçlarının raporlaştırılması, araştırma problemine uygun olan ve varsayımlarının karşılandığı belirlenen istatistiksel tekniğin bir paket program aracılığıyla gerçekleştirilmesi ve son olarak veri dosyasının oluşturulup düzenlenmesi izlemektedir.

Kestirilen Ölçek Değerlerinin İç Tutarlılığının Saptanması

Katılımcıların ikili karşılaştırmaları yaparken dikkatli davranıp davranmadıklarının test edilmesi için ulaşılan sonuçların güvenilirliğinin belirlenmesi gerekir (Nartgün, 2006). Ölçek değerlerinin güvenilirliği gözlenen ve beklenen oranlar arasındaki uyumdan hareketle test edilmektedir (Tekindal, 2009). Dolayısıyla güvenilirlik çalışması kapsamında yapılması gereken ilk işlem beklenen oranlar matrisinin oluşturulmasıdır. Beklenen oranlar matrisi, beklenen birim normal sapmalar matrisindeki (Z' matrisi) değerlerden hesaplanmaktadır. Z' matrisi Tablo 4'te sunulmuştur. Bu matrisin esas köşegenin bir tarafındaki elemanların yazılması yeterlidir. Simetri yardımıyla diğer taraftaki elemanlar kolaylıkla bulunabilmektedir. Z' matrisinin i . satır ile j . sütuna karşılık gelen elemanı $z'_{jk} = U_j - U_k$ eşitliği yardımıyla hesaplanmaktadır (Turgut ve Baykul, 1992). Örneğin; $z'_{U_1 U_2} = U_1 - U_2 = 0.9544 - 0.0449 = 0.9095$ olarak bulunur.

Tablo 4. Nicel veri analizinde karşılaşılan güçlüklerle ilişkin beklenen birim normal sapmalar matrisi, beklenen oranlar matrisi ve hatalar matrisi

| | U1 (0.9544) | U2 (0.0449) | U3 (0.0000) | U4 (0.3860) | U5 (0.0925) | U6 (0.2637) |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| U1 (0.9544) | 0.0000 | | | | | |
| U2 (0.0449) | 0.9095 | 0.0000 | | | | |
| U3 (0.0000) | 0.9544 | 0.0449 | 0.0000 | | | |
| U4 (0.3860) | 0.5684 | -0.3411 | -0.3860 | 0.0000 | | |
| U5 (0.0925) | 0.8619 | -0.0476 | -0.0925 | 0.2935 | 0.0000 | |
| U6 (0.2637) | 0.6907 | -0.2188 | -0.2637 | 0.1223 | -0.1712 | 0.0000 |

| | | U1 | U2 | U3 | U4 | U5 | U6 |
|--------------------------|--------|--------|--------|--------|--------|--------|----|
| Beklenen Oranlar Matrisi | U1 | | | | | | |
| | U2 | 0.8185 | | | | | |
| | U3 | 0.8301 | 0.5179 | | | | |
| | U4 | 0.7151 | 0.3665 | 0.3497 | | | |
| | U5 | 0.8056 | 0.4810 | 0.4632 | 0.6154 | | |
| | U6 | 0.7551 | 0.4134 | 0.3960 | 0.5487 | 0.4321 | |
| Hatalar Matrisi | U1 | | | | | | |
| | U2 | 0.0003 | | | | | |
| | U3 | 0.0063 | 0.0276 | | | | |
| | U4 | 0.0304 | 0.0392 | 0.0048 | | | |
| | U5 | 0.0329 | 0.0281 | 0.0004 | 0.0063 | | |
| | U6 | 0.0006 | 0.0134 | 0.0313 | 0.0058 | 0.0230 | |
| | Toplam | 0.0705 | 0.1083 | 0.0365 | 0.0122 | 0.0230 | |

Z' matrisinin ardından bu matristeki değerlerden yararlanılarak beklenen oranlar matrisi oluşturulmuştur. Bu amaçla, Excel'de standart normal dağılım işlevi kullanılmış ve Z' matrisinin elemanlarına karşılık gelen oran değerleri bulunmuştur. Bulunan bu değerlerden oluşan beklenen oranlar matrisi, beklenen birim normal sapmalar matrisi ile birlikte Tablo 4'te sunulmuştur. Beklenen oranlar matrisi oluşturulduktan sonra, gözlenen oranlar ile beklenen oranlar arasındaki farkların mutlak değerleri alınarak hatalar matrisine ulaşılmıştır. Tablo 4'teki hatalar matrisinin en alt satırında, sütun toplamları yer almaktadır. Bu sütun toplamlarının toplamı alınıp tablodaki eleman sayısına bölüldüğünde, ortalama hata değeri elde edilmektedir. K uyarıcı sayısı olmak üzere tablodaki eleman sayısı $k(k-1)$ ifadesine eşittir. Buna göre; sütun toplamlarının toplamı 0.2504 ve tablodaki eleman sayısı 30 olup, ortalama hata değeri $0.2504/30 = 0.008$ 'dir. Kabul edilebilir hata oranının 0.05 olduğu bir noktada hesaplanan hata değerinin 0.05'ten küçük olması katılımcılar tarafından yapılan ikili karşılaştırmaların iç tutarlılığının yüksek olduğu şeklinde yorumlanmaktadır (Nartgün, 2006). Dolayısıyla, çalışmada hesaplanan ölçek değerlerinin güvenilir olduğu söylenebilir.

Matematiksel Modelin Ampirik Verilere Uygunluğunun Test Edilmesi

Ortalama hata değerinin hesaplanması, ölçek değerlerinin iç tutarlılığı hakkında fikir vermektedir. Ancak ölçek değerlerinin güvenilirliğinin daha net bir biçimde ortaya konulabilmesi için elde edilen ölçek değerleri (matematiksel model) ile ampirik veriler arasındaki farkın anlamlılığının test edilmesi gerekir. Dönüştürülmüş gözlenen oranlar matrisi, dönüştürülmüş beklenen oranlar matrisi ile bu iki matrisinin elemanlarının farkına karşılık gelen farklar matrisinin oluşturulması ve sonrasında Ki Kare değerinin hesaplanması bu kapsamda yapılması gereken işlemlerdir. Ki Kare değeri, farklar matrisindeki elemanların kareleri toplamının $821/N$ ($N =$ çalışmadaki katılımcı sayısı) ifadesine bölünmesiyle hesaplanmaktadır. Hesaplanan Ki Kare değerinin, k uyarıcı sayısı olmak üzere, $(k-1)(k-2)/2$ serbestlik derecesindeki Ki Kare değerinden küçük olması, ölçek değerlerinin güvenilir olduğuna ilişkin bir kanıt olarak değerlendirilmektedir (Turgut ve Baykul, 1992).

Tablo 5. Dönüştürülmüş gözlenen oranlar matrisi, dönüştürülmüş beklenen oranlar matrisi ve farklar matrisi

| | | U1 | U2 | U3 | U4 | U5 | U6 |
|--|----|-------|-------|-------|-------|-------|----|
| Dönüştürülmüş Gözlenen Oranlar Matrisi | U1 | | | | | | |
| | U2 | 64.75 | | | | | |
| | U3 | 66.11 | 47.64 | | | | |
| | U4 | 59.74 | 34.88 | 36.57 | | | |
| | U5 | 61.55 | 45.52 | 42.94 | 51.30 | | |
| | U6 | 60.33 | 39.23 | 40.80 | 48.16 | 39.76 | |

| | | U1 | U2 | U3 | U4 | U5 | U6 |
|--|----|-------|-------|-------|-------|-------|----|
| Dönüştürülmüş Beklenen Oranlar Matrisi | U1 | | | | | | |
| | U2 | 64.82 | | | | | |
| | U3 | 65.65 | 46.03 | | | | |
| | U4 | 57.73 | 37.29 | 36.27 | | | |
| | U5 | 63.87 | 43.91 | 42.88 | 51.65 | | |
| | U6 | 60.33 | 39.99 | 39.00 | 47.81 | 41.09 | |
| Farklar Matrisi | U1 | | | | | | |
| | U2 | -0.07 | | | | | |
| | U3 | 0.46 | 1.61 | | | | |
| | U4 | 2.01 | -2.41 | 0.30 | | | |
| | U5 | -2.32 | -1.61 | 0.06 | -0.35 | | |
| | U6 | 0.00 | -0.76 | 1.80 | 0.35 | -1.33 | |

Güvenirlilik çalışması için oluşturulan dönüştürülmüş gözlenen oranlar matrisi, dönüştürülmüş beklenen oranlar matrisi ve farklar matrisi Tablo 5'te sunulmuştur. Farklar matrisindeki elemanların kareleri alınıp toplandığında, 26.56 değeri elde edilmiştir. Bu değer, 821/110'a bölünmesiyle Ki Kare değeri 3.56 olarak bulunur [$\chi^2_{\text{hesaplanan}}=3.56$]. Araştırmadaki serbestlik derecesi 10 olup, bu serbestlik derecesi için .05 düzeyindeki kritik Ki Kare değeri 18.307'dir [$\chi^2_{\text{kritik}}(0.05,10)=18.307$]. Hesaplanan Ki Kare değeri kritik Ki Kare değerini aşmadığından, matematiksel modelin ampirik verilere uygun ve ölçek değerlerinin güvenilir olduğu söylenebilir.

TARTIŞMA VE SONUÇ

Bilimsel çalışmalarda anlamlı sonuçlara ulaşabilmesi, araştırmacıların çalışmaya uygun istatistiksel teknikleri kullanmasına ve bu teknikleri hatasız bir şekilde uygulayıp yorumlamasına bağlıdır (Bayram, 2012). Bundan dolayı, bilimsel araştırmaların geçerlik ve güvenilirliğinin artırılması araştırmacıların veri analizi ile ilgili yeterliklerinin geliştirilmesiyle yakından ilgilidir. Araştırmacıların veri analizi sırasında karşılaştıkları güçlüklerin belirlenmesi, bu konudaki yeterliklerinin geliştirilebilmesi için bir ön koşul niteliği taşımaktadır. Çünkü Kaptan'ın (1998) da ifade ettiği gibi tüm bilimsel çalışmaların başlangıç noktasını mevcut durumun belirlenmesi oluşturmaktadır. Bu noktadan hareketle, çalışmada ikili karşılaştırmalarla ölçekleme yöntemi kullanılarak araştırmacıların nicel veri analizinde yaşadıkları güçlükler incelenmiştir.

Çalışmadan elde edilen bulgulara göre, *araştırmaya uygun olduğuna karar verilen istatistiksel tekniklere ilişkin varsayımların test edilmesi* araştırmacıların nicel veri analizi sırasında en fazla zorlandıkları işlemidir. Bu bulgu, literatürdeki araştırmalar ile örtüşmektedir. Örneğin; Yim, Nahm, Han ve Park (2010) tarafından yapılan araştırmada, non-parametrik verilerde parametrik testlerin kullanımı araştırmalarda en sık yapılan istatistiksel hatalardan biri olarak belirlenmiştir. Yücel Toy ve Güneri Tosunoğlu (2007) tarafından yapılan araştırmada ulaşılan sonuçlar da aynı doğrultudadır. Yücel Toy ve Güneri Tosunoğlu (2007), sosyal bilimler alanında yapılan araştırmalardaki istatistiksel hataları incelemiş ve çalışmaların birçoğunda istatistiksel analizlerin normallik ve homojenlik gibi varsayımların test edilmeden gerçekleştirildiği sonucuna ulaşmıştır. Ergün ve Çilingir'in (2013) ilköğretim bölümünde yapılan lisansüstü tezleri incelediği araştırmada ulaşılan sonuçlar da bu çalışmadan elde edilen bulgular ile paralellik göstermektedir. Ergün ve Çilingir (2013) Ondokuz Mayıs Üniversitesi örneğinde gerçekleştirdikleri çalışmada, 47 yüksek lisans tezini incelemiş ve inceledikleri tezlerin %38'inde verilerin normalliği hakkında bilgi verilmediğini tespit etmişlerdir. Araştırmacıların istatistiksel tekniklere ilişkin varsayımların testinde zorluk yaşaması, yaptıkları çalışmalarda bu varsayımların test edilmesine yönelik işlemlere yer vermemesi şeklinde bir sonucu ortaya çıkarmış olabilir. Dolayısıyla, literatürdeki çalışmaların, araştırmadan elde edilen bulguları desteklediği söylenebilir. Araştırmacıların istatistiksel tekniklere ilişkin varsayımların testinde yaşadıkları güçlükleri aşabilmeleri için istatistik ve veri analizi ile ilgili derslerde ve

kitaplarda bu konuya daha fazla yer verilmesi önerilebilir. Ayrıca istatistiksel analizler uygulanmadan önce, bu analizlerle ilgili varsayımların test edilmesinin araştırmadan elde edilen bulguların doğruluğu üzerinde son derece etkili olduğu vurgulanıp, araştırmacılar bu konuya daha fazla ağırlık vermeleri hususunda motive edilmelidir.

Çalışmadan elde edilen bulgular, araştırmacıların nicel veri analizinde en fazla zorlandıkları ikinci işlemin *araştırma problemi ve deseni ile çalışmadaki değişkenlere uygun istatistiksel tekniklerin belirlenmesi* olduğunu göstermiştir. IV. İstatistik Kongresi'nde düzenlenen istatistik alanında yapılan çalışmalarda çok sık rastlanan hatalar ve bilimsel etik konulu çalıştay raporu bu araştırma bulgusu ile aynı eksendedir. Bahsi geçen çalıştayda, verinin yapısına uygun olmayan analiz tekniklerinin kullanılması bilimsel çalışmalarda sıklıkla karşılaşılan istatistik hatalardan biri olarak rapor edilmiştir (Akt., Yücel Toy ve Güneri Tosunoğlu, 2007). Yim vd. (2010) tarafından yapılan çalışmanın sonuçlarına göre de verilere uygun olmayan analiz yöntemlerinin kullanılması bilimsel çalışmalarda karşılaşılan temel istatistik hatalarından biridir. Yine Çelik, Gedik, Karaman, Demirel ve Göktaş'ın (2014) eğitim araştırmaları yayınlayan bilimsel dergilere gönderilen araştırmalardaki istatistik hatalarını ve bu hataların araştırmanın reddedilmesine etkilerini inceledikleri çalışmada, kullanılan istatistiksel yöntemlerin araştırma sorularına veya toplanan verilere uygun olmaması bilimsel araştırmaların reddedilmesine sebebiyet veren en önemli faktörlerden biri olarak tespit edilmiştir. İstatistik ve veri analizi ilgili derslerde; araştırma problemi, araştırmada kullanılan model, çalışmadaki değişkenlerin sayısı ve düzeyleri (sınıflama, sıralama, aralık ya da oran ölçeği) ile analizde kullanılacak istatistiksel teknik arasındaki ilişki üzerinde daha fazla durulması, araştırmacıların çalışmaya uygun olan istatistiksel tekniği belirlemede yaşadıkları güçlüğün üstesinden gelmelerine yardımcı olabilir.

Çalışmada ulaşılan sonuçlara göre; araştırmacıların nicel veri analizinde yaşadıkları güçlükler arasında bir sıralama yapıldığında, üçüncü sırada *analiz sonuçlarının yorumlanması*, dördüncü sırada ise *analiz sonuçlarının raporlaştırılması* yer almaktadır. Buna göre; nicel veri analizine ilişkin bu iki basamak araştırmacıların çalışmaya uygun analizin belirlenmesi ve bu analize ilişkin varsayımların test edilmesine kıyasla daha az zorlandıkları işlemlerdir. Araştırmacılar çalışmaya uygun analiz tekniklerini belirleyip, varsayımları karşılanan istatistiksel tekniği bir paket program aracılığıyla gerçekleştirdikten sonra; analiz sonuçlarını literatürde aynı analizi kullanan benzer çalışmalardan hareketle yorumlayıp raporlaştırabilmektedir. Bunun yanında birçok veri analizi kitabında farklı istatistiksel tekniklere ilişkin analiz sonuçlarının nasıl yorumlanması ve raporlaştırılması gerektiğine dair detaylı açıklamalar bulunmaktadır. Oysa araştırmacıların istatistik kitaplarında ya da literatürdeki mevcut çalışmalarda doğrudan kendi araştırmadaki problem, desen ve değişkenler ile örtüşen bir örneğe rastlama olasılığı oldukça düşüktür. Buna bağlı olarak, analiz sonuçlarının yorumlanması ve raporlaştırılması araştırmacıların uygun analizin belirlenmesi ve bu analizin varsayımlarının test edilmesine göre daha rahat yaptıkları işlemler olabilmektedir.

Çalışmada ulaşılan bulgulara göre, *araştırma problemine uygun olan ve varsayımlarının karşılandığı belirlenen istatistiksel tekniğin bir paket program aracılığıyla gerçekleştirilmesi* nicel veri analizinde karşılaşılan güçlükler arasında beşinci sırada yer bulmuştur. Son sırada ise, *veri dosyasının oluşturulması ve düzenlenmesi* işlemi yer almıştır. Bu iki işlem araştırmacının nicel veri analizini gerçekleştirebilmesi için sahip olması gereken paket programlar bilgisini içermektedir. Araştırmanın giriş kısmında da belirtildiği gibi, istatistik bilgisi ve paket programlar bilgisi nicel verilerin analizi için gerekli olan iki yeterliliklerdir. Çalışmadan elde edilen bulgular; araştırmacıların istatistik bilgisini gerektiren işlemlerde, paket program bilgisini gerektiren işlemlere göre daha fazla zorlandıklarını yansıtmaktadır. Günümüzde bilgisayar kullanımının yaygınlaşması ve bireylerin bilgisayar kullanma yeterliliklerinin giderek yükselmesi, araştırmacıların paket program kullanma gibi veri analizinin teknik beceri gerektiren kısmında daha az zorlanmalarına kaynaklık eden bir faktör olabilir. Ancak, istatistik bilgisi olmadan paket program kullanımına yönelik teknik bilgi tek başına bir anlam ifade etmemektedir. Nitekim Erkuş (2011) yalnızca teknik anlamdaki paket program kullanma becerisinin veri analizi için yeterli olmadığını, değişkenlerin yapısı ve yapılacak işlemler bilinmeden gerçekleştirilecek istatistiksel analizlerin hatalı olacağını dile getirmiştir. Buna göre, paket program bilgisinden önce iyi bir istatistik bilgisine sahip olmaları araştırmacıların nicel veri analizinde sorun

yaşamamaları adına bir zorunluluk olarak ifade edilebilir. Bu kapsamda, lisansüstü eğitim programlarında bilgisayar destekli veri analizi dersini alacak araştırmacılarda istatistik dersini başarıyla tamamlamış olma ölçütünün aranması yerinde bir uygulama olacaktır.

İleri Araştırmalara Yönelik Öneriler

Bu çalışmada, Thurstone'un ikili karşılaştırmalar yöntemi kullanılarak araştırmacıların nicel veri analizinde karşılaştıkları güçlükler incelenmiştir. Alanyazında araştırmacıların nicel veri analizi sırasında daha çok hangi işlemlerde zorlandıklarını ortaya koyabilecek bir çalışmaya rastlanmamıştır. Bu yönüyle, çalışmanın özgün olduğu ve literatüre katkı sağlayacağı söylenebilir. Öte yandan, araştırmacının bir takım sınırlılıklarının bulunduğu ve bu sınırlılıkların aşılabilmesi için konu ile ilgili başka çalışmalara ihtiyaç duyulduğu unutulmamalıdır. İlk olarak, bu çalışmada araştırmacıların nicel veri analizinde karşılaştıkları güçlükler nicel yöntemlerle incelenmiştir. İleri çalışmalarda araştırmacılar ile veri analizinin hangi işlem basamağında neden zorlandıklarına ilişkin görüşmeler yapılması konu ile ilgili literatürün olgunlaşmasına katkı sağlayabilir. Bundan dolayı, ileri araştırmalara yönelik getirilebilecek ilk öneri benzer bir çalışmanın nitel olarak yürütülmesidir. İkinci olarak bu çalışmada, araştırmacıların veri analizinde karşılaştıkları güçlükler yalnızca nicel çalışmalar ile sınırlandırılmıştır. Bu kapsamda, benzer bir çalışmanın araştırmacıların nitel veri analizinde karşılaştıkları güçlüklerin tespitine yönelik olarak yapılması önerilebilir.

KAYNAKÇA

- Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları*. İstanbul: İdeal Kültür Yayıncılık.
- Altunışık, R., Coşkun, R., Bayraktaroğlu, S., & Yıldırım, E. (2012). *Sosyal bilimlerde araştırma yöntemleri SPSS uygulamalı*. Sakarya: Sakarya Yayıncılık.
- Aziz, A. (2011). *Sosyal bilimlerde araştırma yöntem ve teknikleri*. Ankara: Nobel Yayın Dağıtım.
- Baykul, Y., & Güzeller, C.O.(2013). *Sosyal bilimler için istatistik-SPSS Uygulamalı*. Ankara: Pegem Akademi Yayıncılık.
- Bayram, N. (2009). *Sosyal bilimlerde SPSS ile veri analizi*. Bursa: Ezgi Kitabevi.
- Bayram, N. (2012). *Veri analizi - Excel ve SPSS uygulamalarıyla birlikte*. Ankara: Siyasal Kitabevi.
- Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Ankara: PegemA Yayınları.
- Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Boston: Pearson.
- Çelik, E., Gedik, N., Karaman, G., Demirel, T., & Göktaş, Y. (2014). Eğitim makalelerinde karşılaşılan hatalar ve bunların reddedilmeye etkileri. *Scientometrics*, 98(3),1837-1853.
- Ergün, M., & Çilingir, F. (2013, Mayıs). *İlköğretim bölümünde yapılan lisansüstü tezlerin incelenmesi: Ondokuz Mayıs Üniversitesi örneği*. VI. Ulusal Lisansüstü Eğitim Sempozyumu'nda sunulmuş sözlü bildiri, Sakarya, Türkiye.
- Erkuş, A. (2011). *Davranış bilimleri için bilimsel araştırma süreci*. Ankara: Seçkin Yayıncılık.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-I*. Ankara: Pegem Akademi Yayıncılık.
- Evrekli, E., İnel, D., Deniz, H., & Balım, A.G. (2011). Fen eğitimi alanındaki lisansüstü tezlerdeki yöntemsel ve istatistiksel sorunlar. *İlköğretim Online*, 10(1), 206-218.
- Field, A. (2009). *Discovering Statics Using SPSS*. London: SAGE Publications Ltd.
- Kabaca, T., & Erdoğan, Y. (2007). Fen bilimleri ve matematik eğitimi alanlarındaki tez çalışmalarının istatistiksel açıdan incelenmesi. *Pamukkale Eğitim Fakültesi Dergisi*, 22(2), 54-64.
- Kalaycı, Ş. (Ed.) (2010). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. Ankara: Asil Yayın Dağıtım.
- Kaptan, S. (1998). *Bilimsel araştırma ve istatistik teknikleri*. Ankara: Tekışık Web Ofset Tesisleri.
- Karasar, N. (2009). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Nartgün, Z. (2006). Öğretmenlik meslek bilgisi derslerinin önem düzeyinin ikili karşılaştırmalarla ölçeklenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 6 (2), 161-176.
- Ozan, C., & Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, 4(1), 116-136.
- Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows*. Australia: Australian Copyright.
- Sipahi, B., Yurtkoru, S., & Çinko, M. (2010). *Sosyal bilimlerde SPSS'le veri analizi*. İstanbul: Beta Basım Yayım Dağıtım.
- Seçer, İ. (2013). *SPSS ve LISREL ile pratik veri analizi: Analiz ve raporlaştırma*. Ankara: Anı Yayıncılık.

- Tan, Ş. (2016). *SPSS ve Excel uygulamalı temel istatistik-I*. Ankara: Pegem Akademi Yayıncılık.
- Tekindal, S. (2009). *Duyuşsal özelliklerin ölçülmesi için araç oluşturma*. Ankara: Pegem Akademi Yayıncılık.
- Turgut, M.F., & Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM Yayınları.
- Yim, K.H., Nahm, F.S., Han, K.A., & Park, S.Y. (2010). Analysis of statistical methods and errors in the articles published in the Korean journal of pain. *Korean Journal Pain*, 23(1), 35-41.
- Yücel Toy, B., & Güneri Tosunoğlu, N. (2007). Sosyal bilimler alanındaki araştırmalarda bilimsel araştırma süreci, istatistiksel teknikler ve yapılan hatalar. *Gazi Üniversitesi Ticaret ve Turizm Eğitim Fakültesi Dergisi*, 1, 1-20.

EXTENDED ABSTRACT

Introduction

This study aimed to analyze the difficulties faced by teacher education for specific areas and educational sciences researchers in quantitative data analysis using pairwise comparison. Although data analysis is an essential step in scientific research, there are no studies in the relevant literature to determine the difficulties that researchers have in this process. Thus, this study will contribute to the relevant literature. Since the study results determine the operations with which the researchers have the most difficulty, the study findings will help to organize the statistics and data analysis curricula in undergraduate and graduate programs. Moreover, the researchers who write books about statistics, data analysis and scientific research methods will also benefit from the study results.

Method

This is a survey of case determination. The study sample included 110 participants of whom 50 (45.5%) were females, and 60 (54.5%) were males. They work in teacher education for specific areas and educational sciences. The study used pairwise comparison as the scaling method, and a data collection tool was prepared for this scaling. With this purpose, the study followed six steps in quantitative data analysis: *i*) creating and arranging the data file, *ii*) determining the suitable statistical techniques for the research problem, the study design and the variables in the study, *iii*) testing the assumptions associated with the statistical techniques that are determined to be suitable for the research, *iv*) applying the statistical technique determined to be suitable for the research problem and provide assumptions that are confirmed using a package program, *v*) interpreting the analysis results, and *vi*) preparing a report on the analysis results. These six steps were presented to the participants as pairwise comparisons. Thus, the study included 15 items consisting of the pair combinations of the six operational stages followed in quantitative data analysis. The study data were collected in November 2015. The participants were asked to consider their difficulties in quantitative data analysis and identify with which of the two operations they had the greatest difficulty. The researchers added a warning at the beginning of the data collection tool saying that one of the two given operations had to be selected. The pair comparisons that were not answered despite this warning were excluded from the study.

The collected data were scaled on the complete data matrix using the fifth case equation of Thurstone's law of comparative judgment. For this scaling, first the study computed the frequency values of the comparisons made by the participants about the steps of the quantitative data analysis. Then, the study created a frequency matrix using these values, and divided each value in the frequency matrix by the number of participants, which generated a rates matrix. After this operation, the study determined the standard Z-scores for the elements of the rates matrix and derived the unit normal deviations matrix. Then, the study found the scale values by adding a line to the bottom line of the unit normal deviations matrix corresponding to the values of each column. The researchers also made a shift to bring the smallest element of scale values to zero, and it was interpreted that the stimuli that were closer to zero were the steps with which participants had more difficulty in quantitative data analysis.

Results and Discussion

According to the study findings, the operation with which the participants had the greatest difficulty was testing the assumptions associated with the statistical techniques that are suitable for the research. This operation was followed by determining the suitable statistical techniques for the research problem, the study design and the variables in the study, interpreting the analysis results and preparing a report on the analysis results. Accordingly, the operations which the researchers have the least difficulty with were interpreting the analysis results and preparing a report on them compared to the determination of the analysis that is suitable for the study and testing the assumptions associated with this analysis. After the researchers determined the analysis that was suitable for the study and applied the statistical technique using a package program, they interpreted the analysis results and prepared a report on them based on relevant studies using the same analysis. In addition, many books on data analysis include detailed explanations about how to interpret and report analysis results from different statistical techniques. However, finding an example which perfectly matches with problems, designs and variables in their study in the statistics books or the available studies in the relevant literature is very unlikely. Thus, researchers interpret and report the analysis results more easily than they determine the suitable analysis and test its assumptions.

According to the study findings, the fifth difficulty researchers had in quantitative data analysis was applying the statistical technique that was suitable for the research problem and determined to confirm the assumptions using a package program. The last of the six operations the researchers had difficulty with was the creation and arrangement of the data file. These two operations include the package program information that a researcher should have to carry out the quantitative data analysis. As stated in the introduction, statistical information and package programs information are the two competencies required for quantitative data analysis. The study findings revealed that researchers had more difficulty in operations requiring statistical information than the operations requiring package programs information.

Öğrencilerin Sosyoekonomik Düzeylerinin TEOG Matematik Başarısına Etkisi*

Effect of Socio-Economic Status on Student's TEOG Mathematics Achievement*

Sinan YAVUZ**

Metin ODABAŞ***

Atilla Özdemir****

Öz

Bu çalışmada Temel Eğitimden Ortaöğretime Geçiş Sistemi (TEOG) ortak sınavından elde edilen matematik başarıları ile öğrencilerin sosyoekonomik düzeyleri (SED) arasındaki ilişki incelenmiştir. Araştırmada 2013 yılında TEOG sınavına giren öğrencilerin matematik puanları bağımlı değişken ve SED bağımsız değişken olarak kullanılmıştır. Araştırma kapsamında kurulan hiyerarşik doğrusal (lineer) modelde öğrencilerin SED ve TEOG matematik başarı puanları birinci düzey, okulların ortalama SED değerleri ise ikinci düzey değişkenler olarak alınmıştır. Araştırmada Ankara ili Altındağ ilçesi sınırları içerisinde bulunan 32 ortaokuldan toplam 1194 öğrencinin SED ve 2013 yılı birinci dönem TEOG ortak sınavı matematik başarı puanları kullanılmıştır. Ayrıca araştırmada öğrencilerin okul düzeyinde ortalama matematik başarıları ve SED değerleri belirlenerek matematik başarılarındaki değişkenlik incelenmiştir. Okulların ortalama matematik başarıları arasındaki fark istatistiksel olarak anlamlı bulunmuştur. Okullara ait ortalama SED değerleri ortalama matematik başarıları üzerindeki etkisinin istatistiksel olarak anlamlı olmadığı tespit edilmiştir.

Anahtar Kelimeler: Hiyerarşik doğrusal model, TEOG ortak sınavı matematik başarıları, sosyoekonomik düzey

Abstract

In this study, the relationship between TEOG mathematics achievement and social economic status (SES) was examined. In the research, student's 2013 first semester TEOG mathematics achievement was used as dependent variable and SES is used as independent variable. In the context of the research, SES of students and math achievement scores were used as a first level, the average SES of school-level variable is used as the second level in hierarchic linear model (HLM). In the research, 32 secondary schools, which are located in Altındağ district within the boundaries of Ankara, and 1194 students participated. SES and 2013 TEOG student mathematics achievement of the test scores were used. The difference between average mathematics achievements of the schools was found as statistically significant. It is determined that the effect of mean SES of the schools on mathematics achievement was not statistically significant.

Key Words: Hierarchical linear models, TEOG exam mathematics achievement, socio-economic status

GİRİŞ

Temel eğitim Milli Eğitim'in genel amaçları ve temel ilkeleri doğrultusunda, "Her Türk çocuğunun iyi birer yurttaş olabilmesi için, gerekli temel bilgi, beceri, davranış ve alışkanlık kazanmasını, millî ahlak anlayışına uygun olarak yetişmesini, ilgi, yeti ve yetenekleri doğrultusunda hayata ve bir üst öğrenime hazırlanmasını" sağlamayı amaçlamaktadır (MEB Temel Kanunu, 2014). Dolayısıyla, öğrencilerin ortaöğretim ve üniversitedeki akademik başarıları için de temel oluşturmaktadır.

Bireylerin ortaöğretim başarılarına etki eden çeşitli değişkenler bulunmaktadır. Araştırmacılar uzun yıllardır öğrencilerin başarılarını etkileyen faktörler üzerinde çalışmalar yapmışlardır. Gelbal (2008) araştırmasında MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı tarafından 2005 yılında sekizinci sınıflara uygulanan Öğrenci Başarı Belirleme Sınavı Türkçe test başarılarını etkileyen

* Bu çalışma IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde (9-13 Haziran 2014) sözlü bildiri olarak sunulmuştur

** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, İletişim: sinanyavuz@hacettepe.edu.tr

*** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, İletişim: m.odabas@hacettepe.edu.tr

****Dr, Milli Eğitim Bakanlığı, İletişim: atimaths06@gmail.com

birçok bağımsız değişkenin etkisini ayrı ayrı incelemiştir. Bu araştırmaya göre öğrencinin ev içindeki eğitim olanakları, ailenin sosyoekonomik düzeyi (SED), çevresel etkenler ve okuldaki öğrenme-öğretme olanaklarının öğrencinin Türkçe başarıları düzeyine etkisi olduğu sonucuna ulaşmıştır.

Ailelerin sosyoekonomik düzeylerinin, öğrencilerin eğitimsel kazanımları elde etmesinde ve akademik başarıları üzerinde önemli bir belirleyici faktör olduğu düşünülmektedir. Öksüzler ve Sürekçi'nin 2010'da Bandırma'da Ortaöğretim Kurumları Sınavı'na (OKS) girmiş ve 9. sınıfa devam eden öğrenciler üzerinde yaptıkları çalışmada öğrenci başarılarını etkileyen faktörlerin başında ailenin gelir ve eğitim düzeyinin geldiği sonucuna ulaşmışlardır. Bunun yanında kitap okuma ve planlı çalışma alışkanlığı ile dershaneye gitme faktörlerinin de OKS başarıları üzerinde olumlu etkiye sahip olduğunu göstermişlerdir. Öğrencilerin sosyoekonomik düzeylerinin başarıya olan etkisinin incelendiği diğer bir çalışmada Barr (2015), 8000 öğrenci ile 2010 ve 2012 yıllarında geliştirilen testler üzerinden boylamsal bir çalışma yaparak ailenin sosyoekonomik düzeyinin öğrencinin akademik başarıları üzerinde önemli bir etkisinin olduğunu belirlemiştir. Acar (2013), yaptığı çalışmada SED değerlerine benzer değişken olan olanaklar değişkeninin öğrencilerin Türkçe başarılarını yordadığı sonucuna ulaşmıştır. Yine farklı bir çalışmada öğrencilerin çevre bilinci ve çevre akademik başarıları ile SED'leri arasında anlamlı farklılık bulunmuştur (Uzun & Sağlam, 2005). Çiftçi ve Çeçen (2009) okuduğunu anlama kazanımlarıyla ilgili olarak, öğrencilerin kavrama, uygulama, analiz ve değerlendirme basamaklarındaki başarılarında SED açısından anlamlı farklılık bulunduğunu tespit etmişlerdir.

SED akademik başarının yanında bireylerin tamamlamış olduğu eğitim düzeyinin üzerinde de etkili olmaktadır (Duncan ve Magnusan, 2011). SED için çok farklı tanımlamalar yapılmıştır. Genel anlamda Mueller ve Parsel (1981), SED'i, bir birey ya da bir aile için gelir, sosyal statü, eğitim gibi niteliklerin birleşiminin erişimindeki hiyerarşik durumu olarak açıklamıştır. Yüksek SED'e sahip olan öğrencilerin eğitimsel imkân ve fırsatlara sahip olmada düşük SED'e sahip olan öğrencilere kıyasla daha avantajlı durumda oldukları söylenebilmektedir. Literatür incelendiğinde SED akademik başarının açıklanmasında önemli bir yordayıcı olarak kullanılmaktadır (Sammons, West & Hind 1997; Thomas, Sammons, Mortimore & Smees, 1997).

Yeni Zelanda hükümeti tarafından 2008 yılında hazırlanan ve TIMMS matematik başarılarının 1994 ve 2006 yılları arasındaki değişiminin incelendiği 12 yıllık süreci kapsayan raporda, yüksek SED'e sahip olan öğrencilerin düşük olanlara göre daha yüksek matematik başarıları elde ettiği tespit edilmiştir (Caygil ve Kirkham, 2008). Güneybatı Nijerya'da 36 farklı okuldan 1722 öğrenci üzerinde yapılan diğer bir çalışmada yine benzer bir sonuçla, yüksek SED'e sahip olan öğrencilerin düşük SED'e sahip olanlardan daha yüksek matematik başarılarına sahip olduğu görülmüştür (Olatunde, 2010).

Yılmaz Fındık ve Kavak (2013) sosyoekonomik açıdan dezavantajlı öğrencilerin PISA sınavlarında genellikle düşük başarı gösterdiği ve yeterli düzeylerinin de oldukça alt seviyelerde olduğu sonucuna ulaşmıştır. Sarier (2010) PISA ve ortaöğretime giriş sınavlarındaki öğrenci başarılarını incelemiş ve sosyoekonomik farklılıkların başarıda anlamlı farklılıklar oluşturabileceğini belirtmiştir.

Arı (2007), ilkökul dördüncü sınıf öğrencileri üzerinde yaptığı araştırmada Türkçe ve matematik başarılarına etki eden çeşitli faktörleri incelemiş ve SED yükseldikçe Türkçe ve matematik derslerindeki başarının da yükseldiğini tespit etmiştir. Yine benzer bir çalışmada dört ve beşinci sınıflarda eğitim gören öğrencilerde uzunluk, alan ve hacim ölçüleri konularında SED'i yüksek olanların daha başarılı olduğu sonucuna ulaşılmıştır (Kültür, Kaplan & Kaplan, 2002).

Genel kaniya bakıldığında SED ile akademik başarı arasında pozitif bir ilişkinin olduğu düşünülmekle birlikte, alanyazında bu durumun farklı şekilde olduğunu, yani SED'in akademik başarı üzerinde etkili olmadığını veya negatif etkisinin olduğunu raporlayan çalışmalara da rastlanmıştır. White, 1982'deki yapmış olduğu meta-analiz çalışmasında akademik akademik başarı ile SED arasındaki ilişkinin incelendiği yaklaşık 200 makaleyi taramış, sonuç olarak da aralarında zayıf bir ilişki olduğunu tespit etmiştir. Diğer bir meta-analiz çalışmasında ise 1990-2000

yılları arasında Amerika Birleşik Devletleri'nde yapılmış olan 58 çalışma taranmıştır. Bu çalışmalarda elde edilen korelasyonların ortalaması 0,29 olarak hesaplanmıştır (Sirin, 2005). Kurul Tural (2002) benzer bir şekilde öğrencilerin SED geçmişleri kontrol edildiğinde, öğrenci başına yapılan harcama ile öğrenci başarısı arasında zayıf bir ilişki olduğundan bahsetmiştir. Hernandez (2014), ABD'nin Florida eyaletinde devlet okulunda eğitim gören ortaokul öğrencileri üzerinde yaptığı çalışmada, öğrencilerin matematik puanları ile SED'leri arasında manidar ve negatif yönde bir ilişki tespit etmiştir. SED öğrencinin çeşitli eğitimsel materyallere ve fırsatlara ulaşımını kolaylaştırırsa da, SED-akademik başarı ilişkisi üzerinde öğrencinin derse karşı tutumu, motivasyonu, çocuk-ebeveyn etkileşimi, ailenin beklentilerinin öğrenci üzerindeki etkisi, öğrenci ve aile sağlığı gibi değişkenler önemli etkilere sahiptir (Brooks-Gunn & Duncan, 1997). Özellikle sağlık sorunları doğrudan öğrencilerin bilişsel gelişimini sınırlayarak akademik yeteneklerini etkileyebilmektedir (Heckman, 2007).

2000'lerden önce yapılan çalışmalarda SED ile matematik başarısı arasındaki ilişkinin incelenmesi korelasyonel tekniklerle sınırlı kalmıştır. Genel olarak bu çalışmalarda SED ve matematik başarı arasında pozitif korelasyon elde edilmiştir (Jurdak, 2009). Ma ve Klinger (2000) 6. sınıf öğrencileri üzerinde yapmış oldukları çalışmada, SED ve matematik başarısı arasındaki ilişkinin hesaplanması için hiyerarşik modellemeyi kullanmışlar ve SED'i anlamlı bir yordayıcı olarak belirlemişlerdir.

SED gibi değişkenler öğrencilerin akademik başarılarının yordanmasında belirleyici olarak kullanılmaktadır. Akademik başarı olarak bazı çalışmalarda PISA, TIMMS gibi uluslararası verilerle çalışılırken (Sarier 2010; Yılmaz Fındık ve Yüksek 2013; Marks 2006), bazı çalışmalarda ise (2013 yılında uygulamaya geçen) Temel Eğitimden Orta Öğretime Geçiş Sınavı (TEOG) ve Yükseköğretime Geçiş Sınavı (YGS) gibi ulusal sınavların verileri kullanılmaktadır (Cayhan ve Akın 2015; Yılmaz Koğar ve Aygün 2015; Doğan ve Demir 2015). Ülkemizde okullarda verilen öğretimin ölçülmesinde gerek öğretmen yapımı, gerek standart testler olarak çeşitli sınav uygulamaları gerçekleştirilmektedir. 2013 yılında MEB tarafından yayınlanan genelge ile TEOG ortak sınav uygulamaları 2013 – 2014 Eğitim Öğretim yılında itibaren 8. sınıflar için başlamıştır (Tebliğler Dergisi, 2013).

TEOG 2013 yılı güz döneminde gerçekleştirilen sınavının birinci oturumu, Türkiye genelinde ve yurt dışında 14 bin 407 sınav merkezi ve 86 bin 25 salonda gerçekleştirilmiştir (Basın Açıklaması, MEB, 2013). TEOG ortak sınavı her biri 20 soru içeren 6 alt testten (Türkçe, matematik, din kültürü, fen ve teknoloji, inkılap tarihi ve yabancı dil) oluşmaktadır. Öğrencilerin bu testlerin her birinden aldıkları puanlar yüz üzerinden hesaplanarak dönem içerisindeki 2. sınav yerine kullanılmaktadır. Bu çalışmada, bu sınavdan elde edilmiş ve yüz puan üzerinden hesaplanmış olan matematik başarı puanları kullanılmıştır.

Araştırmanın Amacı

Yapılan alanyazın incelemeleri sırasında ülkemizde SED değişkeninin başarı üzerindeki etkisini hiyerarşik modelleme kurgusuyla inceleyen çalışmaların sınırlı olduğu tespit edilmiştir. Bu çalışmada öncelikle okullar arası başarı farklılaşmasının olup olmadığı araştırılmış, daha sonra ise başarı farkının SED tarafından açıklanıp açıklanamayacağı incelenmiştir.

Sosyal bilimler alanında yapılan çalışmalarda, birçok veri hiyerarşik yapıya sahiptir, diğer bir ifadeyle veriler farklı düzeylerde yuvalanmışlardır (Atar, 2010). Bu çalışmada da farklı düzeylerde yuvalanmış veriler kullanılmıştır. Çalışma kapsamında aşağıdaki problemlere cevaplar aranmaktadır;

1. Ortaokulların ortalama TEOG matematik başarıları puanları arasında anlamlı farklılık var mıdır?
2. Ortalama SED değeri yüksek olan okulların düşük olan okullara göre TEOG matematik başarısı daha yüksek midir?

YÖNTEM

Çalışma Grubu

Bu araştırmada Ankara ili Altındağ ilçesinde 32 ortaokulda bulunan ve TEOG sınavına giren 1194 öğrenciye ait veri kullanılmıştır. Tablo 1’de örneklemdaki öğrencilere ait betimsel istatistikler gösterilmiştir.

Tablo 1. Cinsiyete Dair Betimsel İstatistikler

| | Frekans | Yüzde |
|--------|---------|-------|
| Erkek | 574 | 48,1 |
| Kız | 620 | 51,9 |
| Toplam | 1194 | 100,0 |

Veri Toplama Yöntemleri

Araştırmanın verileri Hacettepe Üniversitesi ve Altındağ İlçe Milli Eğitim Müdürlüğü’nün imzalamış olduğu protokol çerçevesinde elde edilmiştir. Çalışmada kullanılan bağımlı değişkene ait veri, TEOG sınavı matematik testi başarısıdır. Öğrencilere ait demografik özellikler “e-okul portalı” kullanılarak elde edilmiştir. Çalışmada bağımsız değişken olarak kullanılan öğrencilere ait SED değerleri için, ailenin aylık geliri 5 kategoriye, oturulan evin kendilerine ait olup olmaması durumu, evin ısınmasının ne ile sağlandığı, öğrencinin kendisine ait odasının olup olmaması değişkenleri ise 2’şer kategoriye sahiptir. LISREL 8.80 programı ile bir yol analizi yapılarak, elde edilen yol katsayıları ile SED değerleri hesaplanmıştır. Alanyazınına göre SED değerleri, diğer değişkenler ile tanımlanmıştır (Bielderma, de Greef, Krijnen ve van der Schans 2015; Konopack, McAuley, 2012).

Tablo 2. SED değişkeni oluşturulmasında kullanılan değişkenlere ait yol katsayıları

| Değişkenler | Yol Katsayısı |
|--|---------------|
| Oturduğu evin kendine ait olup olmaması | 0.34 |
| Kendine ait odasının bulunup bulunmaması | 0.65 |
| Evin ısınma şekli | 0.67 |
| Ailenin gelir durumu | 0.56 |

Tablo 2 incelendiğinde, bu araştırma için öğrencilerin evinin ısınma şekli ve kendine ait odasının bulunup bulunmaması SED değişkeninin elde edilmesinde yaklaşık aynı etkiye sahipken, bunu ailenin gelir durumu ve oturduğu evin kendine ait olup olmaması izlemektedir.

Veri Çözümleme Teknikleri

Raudenbush ve Byrk’a (1992) göre eğitim sistemi hiyerarşik (aşamalı) bir yapıya sahip olduğundan (öğrenciler okul içinde yuvalanmış), araştırmacılar hem öğrenci hem de okul özelliklerini incelemek durumundadırlar. Alanyazında öğrenci ve okul düzeyinde başarıya etki eden faktörlerin incelendiği çeşitli çalışmalar da bulunmaktadır (Atar ve Atar, 2012; Braun, Jenkins & Grigg, 2006). Bu çalışmaların bir kısmında farklı hiyerarşik düzeylere ait veriler kullanılmıştır. Farklı düzeye sahip, hiyerarşik bir yapı bulunduran ve iç içe geçmiş verilerin kullanılmasında hiyerarşik doğrusal (lineer) modeller (HLM) tercih edilmektedir. HLM, sıra düzenli, farklı düzeylerdeki iç içe geçmiş verilerin analizlerine imkân sağlayan istatistiksel bir modeldir. Ayrıca HLM, hiyerarşik yapıda olan yordayıcı

değişkenlerin yordanan değişkenler üzerindeki değişkenliğinin incelenmesine olanak sağlayan ve en küçük kareler regresyonun farklı bir şeklidir (Woltman, Feldtain, Mackay & Rocchi, 2012).

Verilerin Analizi

Bu araştırmada belirlenmiş öğrenci ve okul özelliklerine ait değişkenler Tablo 3'te verilmiştir. SED değişkeninin başarı üzerindeki etkisini inceleme amaçlı elde edilen veriler iki farklı düzeyde toplanmıştır. Farklı düzeylerde toplanmış olan bu verilerin incelenmesinde HLM kullanılmıştır. Düzey-1'de öğrencinin TEOG ortak sınavından almış olduğu matematik başarı puanı, öğrencinin oturduğu evin kentine ait olup olmaması, kendine ait odasının bulunup bulunmaması, evin ısınma şekli ve ailenin gelir durumu değişkenleri bulunmaktadır. Düzey-2'de ise okullara ait TEOG ortak sınav matematik başarı puan ortalaması ve SED değişkeni yer almaktadır. HLM analizleri HLM 7 programı ile gerçekleştirilmiştir. Ayrıca verinin HLM 7 programına uygun hale getirilmesinde ve varsayımların tespitinde SPSS 21 programı kullanılarak düzey-1 üzerinden düzey-2'ye ait okulların ortalama SED değerleri elde edilmiştir. Araştırma kapsamında belirlenen problemlere karşılık rastgele etkiler ANOVA modeli ve rastgele katsayılar regresyon modeli analizleri yapılmıştır.

Tablo 3. Araştırmanın Düzey Değişkenleri

| Öğrenci Düzeyi (Düzey-1) | Okul Düzeyi (Düzey-2) |
|--|--|
| Öğrenci TEOG ortak sınav matematik başarı puan | Okul TEOG ortak sınav matematik başarı puan ortalaması |
| Oturduğu evin kendine ait olup olmaması | SED |
| Kendine ait odasının bulunup bulunmaması | |
| Evin ısınma şekli | |
| Ailenin gelir durumu | |

Model-1: Tek-Yönlü Varyans Analizi Rastgele Etkiler Modeli

Bu model, iki düzeyli en basit modeldir. Birinci ve ikinci düzeydeki TEOG matematik başarı puanlarının varyansını açıklayacak yordayıcı değişkenleri içermez. Bu yüzden farklı bir değişken eklemekten genel matematik başarı ortalaması ile matematik başarısında gözlenen farklılıkların ne kadarının bireysel farklılık ve ne kadarının okul ortamından kaynaklandığı kestirilmektedir.

Birinci düzey denklemi:

$$TEOG_{ij} = \beta_{0j} + r_{ij}$$

İkinci düzey denklemi:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$TEOG_{ij}$: j okulundaki i. öğrencinin matematik başarı puanı

β_{0j} : j okulunun matematik başarı puanının ortalaması

r_{ij} : j okulundaki i öğrencisinin matematik başarısının j okulunun matematik başarı ortalamasından farkı (birinci düzey hata terimi)

γ_{00} : Tüm okullara ait matematik başarı puanının genel ortalaması

u_{0j} : j okulunun matematik başarı ortalamasının genel matematik başarı ortalamasından farkı (ikinci düzey hata terimi) (Raudenbush ve Bryk, 2002).

Model-2: Rastgele-Katsayılar Regresyon Modeli

Bu modelde öğrenciler arasındaki bireysel farklılığın SED'den kaynaklanan kısmını açıklamak için birinci düzey, okullar arasındaki farklılığı ve bunun etkilerini açıklamak içinse ikinci düzey denklem kurulmuştur.

Birinci düzey denklemi:

$$TEOG_{ij} = \beta_{0j} + \beta_{SEDj} (SED_{ij} - SED_{ORTj}) + r_{ij}$$

İkinci düzey denklemi:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{SEDj} = \gamma_{SEDO} + u_{SEDj}$$

$TEOG_{ij}$: j okulundaki i. öğrencinin matematik başarı puanı

β_{0j} : j okulunun düzeltilmiş (adjusted) matematik başarı puanının ortalaması (kesişim katsayısı)

β_{SEDj} : j okulunda SED'in matematik başarısı üzerindeki etkisi (eğim katsayısı)

SED_{ij} : j okulundaki i. öğrencinin sosyoekonomik düzeyi

SED_ORT_j : j okulunun sosyoekonomik düzey ortalaması

γ_{00} : düzeltilmiş genel matematik başarı ortalaması

u_{0j} : j okulunun düzeltilmiş matematik başarı ortalamasının düzeltilmiş genel matematik başarı ortalamasından farkı

γ_{SEDO} : SED eğim katsayısına ait genel ortalama

u_{SEDj} : j okulunun SED ortalamasının genel SED ortalamasından farkı (Raudenbush ve Bryk, 2002).

BULGULAR

Model-1: Tek-Yönlü Varyans Analizi Rastgele Etkiler Modeli

Tek-yönlü varyans analizi rastgele etkiler modelinin birinci düzeyine ait güvenilirlik katsayısı 0.763 olarak kestirilmiştir. Bu modelin analiz sonuçları Tablo 4'te verilmiştir. Bu modelde, öğrencilerin TEOG matematik başarılarının okul ortalamasından farklarının varyans bileşeni yaklaşık 50.43, okul ortalamalarının genel ortalamadan farklarının varyans bileşeni ise yaklaşık 507.24 olarak kestirilmiştir (Tablo 4). Okullar arası değişkenliğin tahmini değerinin (u_{0j}) anlamlı olarak sıfırdan büyük çıkması ($p < 0.001$) okulların TEOG matematik başarı ortalamalarında farklılık olduğu, diğer bir ifade ile her okulun aynı ortalama başarıya sahip olmadığı şeklinde yorumlanabilir. Okul ortalamaları için %95 olasılıklı değer aralığı aşağıdaki eşitlik ile hesaplandığında (1), okul ortalamalarının %95'inin 28.13 – 55.97 puan aralığında olması beklenir. Bu değerler dikkate alındığında, okulların ortalama başarıları arasında önemli farklar olduğu söylenebilir.

Tablo 4. Model-1: Tek Yönlü Varyans Analizi Rastgele Etkiler Modeli Analiz Sonuçları

| Sabit etki | Katsayı | Standart Hata | p |
|--|----------------|--------------------|--------|
| Genel TEOG matematik başarısı, γ_{00} | 42.05 | 1.44 | <0.001 |
| Rastgele Etki | Standart Sapma | Varyans Bileşenler | p |
| 2. Düzey hata terimi, u_{0j} | 7.10 | 50.43 | <0.001 |
| 1. Düzey hata terimi, r_{ij} | 22.52 | 507.24 | |

$$\gamma_{00} \pm (1.96)(\sqrt{\sigma_{u_{0j}}}) = 42.05 \pm (1.96)(\sqrt{50.43}) = 42.05 \pm (13.92) = 28.13 - 55.97 \quad (1)$$

Sınıflar arası korelasyon, Tablo 4'teki düzey-2 varyansının toplam varyansa bölünmesi ile yaklaşık 0.09 olarak hesaplanmıştır ($50.43/(50.43+507.24)$). Bu sonuç, matematik başarısında gözlenen farklılıkların yaklaşık %9'luk kısmının okullar arasındaki ortalama matematik başarısındaki farklılıktan, yaklaşık %91'lik kısmının ise öğrenciler arasındaki bireysel farklılıklardan kaynaklandığını göstermektedir.

Model-2: Rastgele-Katsayılar Regresyon Modeli

Rastgele-katsayılar regresyon modelinin birinci düzeyine ait güvenilirlik katsayısı 0.757 olarak kestirilmiştir.

Tablo 5. Model-2: Rastgele-Katsayılar Regresyon Modeli Analiz Sonuçları

| Sabit etki | Katsayı | Standart Hata | p |
|--|----------------|--------------------|--------|
| Genel TEOG matematik başarısı, γ_{00} | 42.06 | 1.44 | <0.001 |
| Ortalama SED etkisi, γ_{SEDO} | -0.52 | 0.80 | 0.523 |
| Rastgele Etki | Standart Sapma | Varyans Bileşenler | p |

| | | | |
|--------------------------------|-------|--------|--------|
| 2. Düzey hata terimi, u_{0i} | 7.11 | 50.55 | <0.001 |
| SED etkisi, u_{SEDi} | 1.91 | 3.65 | 0.254 |
| 1. Düzey hata terimi, r_{ij} | 22.46 | 504.42 | |

Tablo 5’te görüleceği üzere SED değerinin ortalama matematik başarıları üzerine etkisinin istatistiksel olarak anlamlı olmadığı görülmüştür ($p>0.05$). Varyans bileşenlerinin son kestirimleri incelendiğinde, öğrenci düzeyinde SED etkisi katıldıktan sonra öğrencilerin genel matematik başarılarının varyans bileşeni 504.42 olarak kestirilmiştir. Tablo 5’teki öğrenci düzeyi varyans değeri ile karşılaştırıldığında yakın sonuca ulaşıldığı ve SED değerlerinin öğrencilerin matematik başarılarındaki varyansı açıklamakta etkili olmadığı sonucuna ulaşılmıştır. Okul düzeyinde ortalama SED kontrol altına alındıktan sonra okulların ortalama matematik başarılarının varyans bileşeni (artık varyans) 50.55 olarak kestirilmiştir. Tablo 5’teki okul düzeyi varyans değeri ile karşılaştırıldığında yine yakın bir sonuca ulaşılmıştır. SED değişkeninin modele eklenmesinin matematik başarı puanları üzerinde anlamlı bir etkisi olmamıştır.

SONUÇLAR ve TARTIŞMA

Yapılan HLM analizleri ile tek-yönlü varyans analizi rastgele etkiler modelinin (Model 1) anlamlı sonuçlar verdiği, rastgele-katsayılar regresyon modelinin (Model 2) ise istatistiksel olarak anlamlı sonuçlar vermediği tespit edilmiştir. Tek-yönlü varyans analizi rastgele etkiler modelinden yola çıkarak, öğrenciler arasında ve okullar arasında TEOG matematik başarı puanlarının anlamlı olarak farklılaştığı sonucuna ulaşılmıştır. Birinci alt probleme cevap olarak; ortaokulların ortalama matematik başarıları puanları arasında anlamlı farklılık vardır. Bu farklılığın sebebi incelendiğinde, %91’inin öğrencilerden kaynaklandığı, yalnızca %9’unun okullardan kaynaklandığı bulunmuştur. Sonuç eğitimin çıktıları açısından incelendiğinde, eğitim sistemi içerisinde öğrencilerin başarılarının okullardan değil bireysel farklılıklardan kaynaklanması istenir.

T.C. Anayasası’nın 42. maddesi olan “Eğitim ve öğretim hakkı ve ödevi” ve eşit haklar ilkesi dikkate alındığında, her Türk vatandaşının eşit eğitim imkânlarına sahip olması ve okullardan kaynaklanan başarı farkının yüksek olmaması beklenir. Elde edilen sonuçlar bu doğrultuda, öğrenci başarılarındaki temel farklılığın yine öğrenciye ait özelliklerden kaynaklandığının göstergesidir. Çalışma grubunun bulunduğu Ankara ili Altındağ ilçesi için yorum yapılacak olursa, Altındağ ilçesindeki okulların başarı ortalamalarının farklı olduğu, fakat bu farklılığın çok büyük bir bölümün (%91) öğrenciden, ancak küçük bir kısmının (%9) okullardan kaynaklandığı sonucuna ulaşılmıştır.

Rastgele-katsayılar regresyon modelinin kurulması ile ikinci alt problem cevaplanmıştır. Alanyazına bakıldığında, çoğunlukla SED’in akademik başarı üzerinde etkili olduğu sonuçlarıyla karşılaştırılmıştır. Fakat bu çalışmada uygulamalı olarak SED’in öğrenciler ve okullar arasındaki TEOG matematik başarı farkını açıklamakta yetersiz kaldığı bulunmuştur. Ankara ili Altındağ ilçesindeki öğrencilerin SED’leri matematik başarılarında anlamlı bir farklılığa yol açmamıştır.

Araştırmanın bulguları doğrultusunda, öğrencinin oturduğu evin kendine ait olup olmaması, kendine ait odasının bulunup bulunmaması, evin ısınma şekli ve ailenin gelir durumunun, öğrencinin matematik başarıları üzerinde anlamlı bir etkiye sahip olmadığı tespit edilmiştir.

Gerek bazı çalışmalarda doğrudan matematik dersine ilişkin başarı yordanmaması (White, 1982; Sirin, 2005), gerekse SED değerlerinin farklı çalışmalarda farklı değişkenlerden (ebeveyn eğitim durumu, kardeş sayısı, vb.) elde ediliyor olması, anlamlı farklılık gösteren çalışmalara benzerlik göstermeyen bir sonuç çıkmasının sebebi olarak düşünülmektedir. Ayrıca bu çalışma, alanyazında SED’in matematik başarıları üzerinde anlamlı etkisi olmadığını ortaya koyan çalışmaları (Kurul Tural, 2002; Hernandez, 2014) destekler nitelikte sonuç verirken, anlamlı etkisi bulunan çalışmaların (Caygil ve Kirkham, 2008; Olatunde, 2010; Yılmaz Fındık ve Kavak, 2013; Sarier, 2010; Arı, 2007; Kültür vd. 2002) aksine bir sonuç ortaya koymuştur.

Öneriler

Elde edilen bulgular incelendiğinde, ailelerin matematik başarısı üzerinde etkisi bulunmayan (öğrencinin oturduğu evin kendine ait olup olmaması, kendine ait odasının bulunup bulunmaması, evin ısınma şekli ve ailenin gelir durumu) değişkenler yerine daha farklı bağımsız değişkenleri manipüle ederek başarıyı arttırmaları önerilmektedir. Bundan sonra yapılacak çalışmalarda farklı bölgelerden, okullardan veya farklı kademedeki eğitim kurumlarından örneklemeler seçilerek SED'in matematik ile beraber diğer derslerin başarısı üzerindeki etkileri incelenebilir. SED bu çalışma kapsamında yukarıda belirtilen 4 farklı kategorik değişkenin gizil değişkeni kullanılarak elde edilmiştir. Öğrencilerin SED puanları elde edilirken daha farklı değişkenlerin kullanılması önerilmektedir.

KAYNAKÇA

- Acar, M. (2013). *Öğrenci Başarılarının Belirlenmesi Sınavında Türkçe Dersi Başarısının Öğrenci ve Okul Özellikleri ile İlişkisinin Hiyerarşik Lineer Model ile Analizi* (Doktora Tezi) Ankara Üniversitesi / Eğitim Bilimleri Enstitüsü, Ankara.
- Arı, A. (2007). Öğrencilerin okul başarısını etkileyen çeşitli faktörlerin incelenmesi. *Milli Eğitim Dergisi*, 176, 169-179.
- Atar, H. Y., & Atar, B. (2012). Türk Eğitim Reformunun Öğrencilerin TIMSS 2007 Fen Başarılarına Etkisinin İncelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(4), 2621-2636.
- Burcu, A. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2).
- Barr, A. B. (2015). Family socioeconomic status, family health, and changes in students' math achievement across high school: A mediational model. *Social Science & Medicine*, 140, 27-34.
- Bielderman, A., de Greef, M. H. G., Krijnen, W. P., & van der Schans, C. P. (2015). Relationship between socioeconomic status and quality of life in older adults: a path analysis. *Quality of Life Research*, 24(7), 1697-1705.
- Braun, H., Jenkins, F., & Grigg, W. (2006). A Closer Look at Charter Schools Using Hierarchical Linear Modeling. NCES 2006-460. *National Center for Education Statistics*.
- Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The future of children*, 55-71.
- Caygill, R., & Kirkham, S. (2008). *Mathematics: trends in year 5 mathematics achievement 1994 to 2006: New Zealand results from three cycles of the trends in international mathematics and science study (TIMSS)*. Comparative Education Research Unit, Research Division, Ministry of Education.
- Cayhan, C., & Erhan, A. (2016). TEOG sınavı Türkçe dersi sorularının Türkçe dersi öğretim programındaki kazanımlar açısından değerlendirilmesi. *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (04).
- Çiftçi, Ö., & Çeçen, M. A. (2009). Beşinci sınıf öğrencilerinin okuduğunu anlama kazanımlarıyla ilgili bilişsel becerilere ulaşma düzeylerinin incelenmesi. *Kastamonu Eğitim Dergisi*, 17(2), 637-648.
- Doğan, E., & Demir, S. B. (2015). Examination of the Relation between Teog Score and School Success in Terms of Various Variables. *Journal of Education and Training Studies*, 3(5), 113-121.
- Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. *Whither opportunity*, 47-70.
- Fındık, L. Y., & Kavak, Y. (2013). Türkiye'deki sosyo-ekonomik açıdan dezavantajlı öğrencilerin PISA 2009 başarılarının değerlendirilmesi. *Kuram ve Uygulamada Eğitim Yönetimi*, 2(2), 249-273.
- Gelbal, S. (2008). Sekizinci sınıf öğrencilerinin sosyoekonomik özelliklerinin Türkçe başarısı üzerinde etkisi. *Eğitim ve Bilim*, 33(150).
- Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences*, 104(33), 13250-13255.
- Hernandez, M. (2014). *The Relationship between Mathematics Achievement and Socio-Economic Status*, *Ejournal of Educational Policy*, Spring.
- Jurdak, M. (2009). *Toward equity in quality in mathematics education*. New York: Springer.
- Konopack, J. F., & McAuley, E. (2012). Efficacy-mediated effects of spirituality and physical activity on quality of life: A path analysis. *Health and quality of life outcomes*, 10(1), 1.
- Kültür, M. N., Kaplan, A., & Kaplan, N. (2002). İlköğretim okulları 4. ve 5. sınıflarda uzunluk, alan ve hacim ölçüleri konularının öğretiminin değerlendirilmesi. *Kastamonu Eğitim Dergisi*, 10(2), 297-308.
- Ma, X., & Klinger, D. A. (2000). Hierarchical linear modelling of student and school effects on academic achievement. *Canadian Journal of Education/Revue canadienne de l'éducation*, 41-55.
- Marks, G. N. (2006). Are between-and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*, 48(1), 21-40.
- MEB, (2014). MEB Temel Kanunu, 24.02.2016 tarihinde http://mevzuat.meb.gov.tr/html/temelkanun_0/temelkanun_0.html adresinden ulaşılmıştır.

- MEB, (2013). e-Kılavuz. 24.10.2014 tarihinde http://www.meb.gov.tr/sinavlar/dokumanlar/2013/kilavuz/2013_OGES_Klvz.pdf adresinden ulaşılmıştır.
- MEB, (2015). Ortaöğretim Kurumları Yönetmeliği. 01.03.2016 tarihinde <http://www.resmigazete.gov.tr/eskiler/2013/09/20130907-4.htm> adresinden ulaşılmıştır.
- Olatunde, Y. P. (2010). Socio-economic background and mathematics achievement of students in some selected senior secondary schools in southwestern Nigeria. *Pakistan Journal of Social Sciences*, 7(1), 23-27.
- Öksüzler, O., & Sürekçi, D. (2010). İlköğretimde başarıyı etkileyen faktörler: Bir sıralı lojit yaklaşımı. *Finans Politik & Ekonomik Yorumlar*, 47(543), 93-102.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical Linear and Nonlinear Modeling*. Illinois: Scientific Software International.
- Sarıer, Y. (2010). Ortaöğretime giriş sınavları (OKS-SBS) ve PISA sonuçları ışığında eğitimde fırsat eşitliğinin değerlendirilmesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 11(3), 107-129.
- Sammons, P., West, A., & Hind, A. (1997). Accounting for variations in pupil attainment at the end of Key Stage 1. *British Educational Research Journal*, 23(4), 489-511.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3), 417-453.
- Tebliğler Dergisi, (2013). 2773584, MEB: Ankara, 24.10.2014 tarihinde http://mevzuat.meb.gov.tr/html/temelegitortaoigrgecis/temegiortagecis_1.html adresinden ulaşılmıştır.
- Tural, N. K. (2002). Öğrenci başarısında etkili okul değişkenleri ve eğitimde verimlilik. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 35(1-2), 39-54.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Differential secondary school effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal*, 23(4), 451-469.
- Wei, X., Lenz, K. B., & Blackorby, J. (2012). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, 0741932512448253.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological bulletin*, 91(3), 461.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69.
- Koğar, E. Y., & Aygün, B. (2015). Temel eğitimden orta öğretime geçiş sınavı (TEOG)'nın matematik temel alanına ait testlerin kapsam geçerliğinin incelenmesi. *Pegem Eğitim ve Öğretim Dergisi*, 5(5), 667-680.
- Uzun, N., & Sağlam, N. (2005). Sosyo-ekonomik durumun çevre bilinci ve çevre akademik başarısı üzerine etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 29(29), 194-202.

EXTENDED ABSTRACT

Introduction

Primary education provides a basis for individual's secondary and university academic success. The main purpose of primary education is "To be a good citizen of every Turkish child, providing necessary knowledge, skills, attitudes and habits, to grow in line with the national morality, preparing students in the line of their interests, talents and capabilities for the life and to prepare them for the higher education".

Several factors are involved in determining the primary education success of individuals. To keep a high level of the individual's school performance, it is necessary to ensure the well-organized learning environment. According to the Gelbal (2008) educational opportunities in the student's house, the family's socioeconomic status, environmental factors, learning and teaching facilities in school are shown as educational environments. According to the study of Öksüzler and Sürekçi in 2010, they reached the conclusion that student's family income and education level are the most important factors affecting student success.

Researchers are investigating the factors that affect the success of students for many years. In the research of a meta-analysis, White (1982) scan about 200 articles and examined the relationship between academic success and socio-economic status (SES). As a result, he found that there is a weak relationship between them. In another meta-analysis study, 58 studies were scanned in the years 1990-2000, in the United States. The average of the correlations obtained in this study between the variables was calculated to be 0.29 (Sirin, 2005).

Various test procedures are performed each year to measure the education given in schools. In 2013, the Ministry of Education (MEB) circular issued about the transition to Secondary Education, common practice exam (TEOG). This common exam was started to use 2013-2014 school year for 8th classes.

In this study, determinants of SES were accepted as family's home ownership, student's room ownership, house warming type and family income. Effect of socio-economic status on mathematics achievement has been examined by establishing hierarchical linear models. In the research it is examined whether there is a statistical difference between mathematics achievements of schools.

The purpose of this study was to determine the effects of students' SES on TEOG mathematics achievement and to investigate that on the school level with Hierarchical Linear Models.

In the studies of social sciences, many of data has a hierarchical structure, in other words, the data have different nested levels (Atar, 2010). The data in this study was nested in different level. In the first level of the data determined as students' level, and the second level of the data determined as school level. With the help of hierarchical linear modeling, these data were compared between each other and internally.

In the context of the research, the answers to the questions are sought.

- 1- Is there a significant difference between school's mathematics achievements?
- 2- Is the mathematics success of schools, which have high mean SES, higher than schools with low mean SES?

Method

In this study, the relationship between multilevel independent variables (student level and school level SES) and dependent variable which is mathematics achievement was investigated with hierarchical linear modeling.

This study aims to describe a situation might happen in the past or today. Moreover, conditions that belong to individuals or objects are not changed or affected.

The population of the research is 8th class students which are from 32 elementary schools in Altındağ District in Ankara are 5499. The sample of the research is 1194 students chosen from all schools by simple random sampling method.

Research data were obtained with the help of the protocol, which was signed between Hacettepe University and Altındağ National Education Directorate. The dependent variables of the data used in this study was obtained from TEOG math exam. Demographic characteristics of the students were obtained from the e-school portal. For the independent variable SES values, path analysis was conducted with the variables family's home ownership, student's room ownership, house warming shape and family income. With the help of path coefficients, SES values were calculated.

In the examination of the relationship between the mathematics achievement and features at the student and the school level, two-level HLM was used. HLM analysis was carried out by the program HLM 7. Also for ensuring data compliance to HLM analysis, SPSS 21 program was used to aggregate second level mean SES values from the first level values. For the response to the problems identified within the study, random effects ANOVA model analysis and as dependent variable, average model analysis was carried.

Results and Discussion

Reliability coefficient for model 1 was found as 0,756. This is an indicator that most of variability of the second level is between the level 2 units. Predictions of the two models are close to each other. This is indication of the provision of basic assumptions (Raudenbush, Bryk, Cheong, Congdon and Toit, 2011). On the Model – 1, p value for SES was statistically significant ($p < 0,05$). There are significant differences between averages of schools TEOG mathematics scores. According to the Model – 2 results, effect of school mean SES on the mathematics achievement was examined and p value for SES was not statistically significant. ($p > 0,05$). The effect of average SES value on the average mathematics achievement was not significant. In conclusion, according to the HLM analysis, there were significant results for model 1 and so there was a significant difference on TEOG mathematics score between schools. Model 2 found as not significance. In other words, the effect of average SES value on the average mathematics achievement was not statistically significant. Data were obtained from only one district in Ankara city, variety of SES may not be appearing. The districts which are known as in the heterogeneous structure can be chosen for the following studies.

R Programı ve Microsoft Excel Kullanılarak İkili Karşılaştırma Yöntemi Uygulaması

Pair-Wise Comparasion Method Application Via R Project and Microsoft Excel

Cem Oktay GÜZELLER* Mehmet Taha ESER** Gökhan AKSU***

Öz

Bu araştırmada üniversite öğrencilerinin okul dışında gerçekleştirdikleri boş zaman aktivitelerinin değerlendirilmesine yönelik araştırmacılar tarafından “Boş Zamanı Değerlendirmeye İlişkin İkili Karşılaştırmalar Ölçeği” geliştirilmiştir. Ölçme aracından elde edilen sonuçların Thurstone tarafından geliştirilen karşılaştırmalı yargılar kanununa göre R programı ile analiz etmek ve ilgili bulguları frekans yöntemiyle excel ortamında elle gerçekleştirilmiş yöntemin analiz sonuçları ile karşılaştırmak amaçlanmıştır. Bu amaç kapsamında öncelikle ikili karşılaştırmalar yönteminin gerçekleştirilebilmesi için ölçme aracı geliştirilmiştir. Araç, 8x8 kare matris formatında oluşturulmuş ve uzman görüşleri ile öğrencilerden alınan dönütler sonucunda ölçme aracına son hali verilmiştir. Veriler 2015 – 2016 Eğitim Yılı Güz Yarıyılında bir fakültede öğrenim gören 306 öğrenciden elde edilmiştir. Elde edilen bulgulara göre hem R programı ile gerçekleştirilen hem de excel kullanılarak gerçekleştirilen hesaplamaların birbirine çok benzer olduğu görülmüştür. Hem R programı hem de excel ile gerçekleştirilen ölçekleme işlemi sonucunda öğrencilerin boş zamanlarında en fazla arkadaşlarıyla vakit geçirdikleri ve sosyal aktiviteleri tercih ettikleri, kitap okumayı ve bilgisayar başında vakit geçirmeyi ise en son tercih ettikleri belirlenmiştir.

Anahtar Kelimeler: Ölçekleme, İkili karşılaştırma, R Programı, Boş zaman etkinliği

Abstract

“Pairwise Comparison For Spending Leisure Time” was developed in this study by the researches for the evaluation of leisure time activities of university students out of the university. This study indented to analyze the results obtained by the measurement instruments through the Program R according to the comparative judgement law developed by Thurstone. Within the scope of this purpose, a measurement scale was developed to realize first the pairwise comparisons method. The tool was created in the format of 8x8 square matrix and finalized according to the expert views and the feedbacks from the students. The data was obtained from 306 students studying at a faculty in the spring term of the academic year 2015-2016. According to the findings, the calculations by both the Program R and the Microsoft Excel were found to be very similar. After the scaling by both the Program R and the Microsoft Excel, it was determined that the students spent most of their leisure time with friends and preferred to have social activities while reading books and spending time at the computer were the least preferred activities.

Key Words: Scaling, Pairwise comparison method, R Programming, Leisure activity

GİRİŞ

İkili karşılaştırmalar yöntemi, psikolojide kullanılan ölçekleme tekniklerinden bir tanesidir (Turgut ve Baykul, 1992). İkili karşılaştırma yönteminde ölçek yapısı için tasarım ve prensip Thurstone’un çalışmasına (1927) dayanmaktadır. Thurstone, tutumların ölçülebileceğini ileri sürerken, tek seferde iki uyaran arasında basit karşılaştırmalara dayalı bir grup uyarının ölçeklenmesi için kullanılabilecek olan bir işlem ve model geliştirmiştir. Thurstone sürecin formülüzasyonunu ‘Karşılaştırmalı Yargılar

*Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: cguzeller@gmail.com

**Uzman, Türkiye Belediyeler Birliği, Ankara-Türkiye, e-posta: tahaeser@gmail.com

***Öğr. Gör., Adnan Menderes Üniversitesi, Aydın MYO, Aydın-Türkiye, e-posta: gokhanaksu1983@hotmail.com

Kanunu” olarak adlandırmıştır. Örneğin, bir kişinin çeşitli ağırlıklarda beş cismin algılanan ağırlıklarını ölçmek istediğini varsayalım. Bireyler, cisimlerin ağırlıklarını ikili olarak karşılaştırarak veriler elde edilebilir ve algılanan ağırlıkların ölçek değerlerini tahmin etmek için karşılaştırmalı bir şekilde karara varabilirler. Bu, cisimlerin fiziksel ağırlığına algısal olarak denktir. Ölçek yerleri her uyarının lehine muhakemelerin oranlarından çıkarılır (Heldsinger ve Humphry, 2010).Thurstone’un ikili karşılaştırma modeli Rasch modelinin bir öncülüdür (Andrich, 1978a).

Thurstone tipi ölçekleme söz konusu olduğunda birkaç yöntem göze çarpmaktadır. İkili karşılaştırma yöntemi bunlardan bir tanesidir. İkili karşılaştırma yönteminde de diğer yöntemlerde olduğu gibi yargıcı kararlarına gerek duyulmaktadır. Ölçekleme işlemini gerçekleştirirken, araştırmacıların geliştirdiği ‘‘Rekreasyon Etkinliklerinin Değerlendirilmesine İlişkin İkili Karşılaştırmalar Ölçeği’’nden elde edilen sonuçlar kullanılmıştır.

Boş zaman, bireyin hem şahsı, hem de diğer bireyleri göz önünde bulundurarak bütün güçlüklerden uzaklaştığı, kendi öz iradesini kullanarak seçtiği bir faaliyetle uğraştığı, bireyin özgür bir şekilde kullanabildiği, yaratıcı ve manevi değerlerin kaynağı olarak görülen zamandır (Godbey ve Stanley, 1976; Hacıoğlu ve diğ., 2003). Boş zamanı değerlendirme (recreation) kavramı, bireylerin boş zamanlarında gerçekleştirdiği etkinlikleri merkeze alan bir kavramdır. Boş zamanları değerlendirme sadece bireyler için bir ihtiyaç, verimli bir hayata sahip olma şekli ve manevi anlamda rahatlama sağlayan bir yol değildir, aynı zamanda bireyler boş zamanlarını değerlendirerek kendini ifade etme ve kişiliğini geliştirme fırsatına sahip olur (Karaküçük, 1999; Terzioğlu ve Yazıcı, 2003; Gökmen ve diğerleri, 1985).

Zamanın zengin ve etkili bir şekilde değerlendirilmesi önemli bir konudur. Yanlış ve verimsiz zaman yönetimi sonucu meydana gelen sıkıntılar bireyi ve tüm toplumu olumsuz bir şekilde etkilemektedir (Jennen ve Uhlebeck, 2004). Ülkemizde boş zaman genellikle iyi bir şekilde değerlendirilememekte okuma düzeyinin düşük olması vb. sebepler düşünüldüğünde özellikle de gençlere yeterince iyi alışkanlıklar kazandırılmadığı gözlemlenmektedir (Karaküçük, 1995; Karaküçük ve Başaran, 2000). Bireylerin farklı yaşam tarzları ve alışkanlıkları göz önünde bulundurulduğunda değişik boş zaman değerlendirme şekillerinin olduğu görülmektedir. Teknolojinin ilerlemesi ile birlikte kitle iletişim araçları bireyler için ortak boş zaman değerlendirme aktivitelerinin oluşmasında ve küresel bir ortam oluşmasında büyük rol oynamıştır (Pekince, 2012; Aydın,2000).

Rekreasyon değerlendirme tercihlerinin birçok araştırmaya konu olması (Kim, 2008; Yin ve diğ., 2009; Trottier ve diğ., 2002; Terzioğlu ve Yazıcı, 2003) ve dolayısıyla boş zamanları değerlendirmeye ilişkin bir ölçek geliştirirken konu ile ilgili geniş bir literatürü inceleyebilme imkanı, ilgili konunun hala popülerliğini koruması, literatür incelendiğinde araştırmacılar tarafından boş zamanları değerlendirme tercihlerine yönelik kapsamlı bir ikili karşılaştırma ölçeğine rastlanmamış olması sonucunda ilgili ölçek oluşturulmuş ve araştırma kapsamında kullanılmıştır.

Araştırma kapsamında geliştirilen ölçekten elde edilen ikili karşılaştırma sonuçları kullanılarak frekans matrisi ile oranlar matrisi oluşturulmuş ve bu matrisler kullanılarak excel ve R programları vasıtasıyla Thurstone V. Hal ve III. Hal İkili Karşılaştırma Uygulamaları gerçekleştirilmiştir.

Araştırmanın Amacı

Alanyazında ikili karşılaştırmalar yöntemine ilişkin çalışmalar oldukça yaygındır (Anıl ve Güler, 2006; Nartgün, 2006; Acar Güvendir ve Özer Özkan, 2013; Öztürk ve diğ., 2011; Özer ve Acar, 2011; Altun ve Gelbal, 2014). Ancak, yapılan bu çalışmalar incelendiğinde, tüm çalışmaların excel ortamında formüllerin elle işlenerek frekans temelli bir şekilde gerçekleştirildiği görülmektedir. Bu durum, hem uzun hem zaman alıcı hem de hataya daha açık bir hesaplama ortamının oluşmasına sebebiyet vermektedir. Ancak R gibi, içerisinde ikili karşılaştırmalara dair istatistiksel hesaplamaların gerçekleştirilebileceği bir paket tarafından desteklenen açık kaynak kodlu yazılımların bu amaçla kullanılması hatayı minimum seviyeye indirmektedir. Bir analizi birçok değer için tekrar etme, veriyi işleme ve programlama ile benzer problemlere hızlı çözüm bulma vb. özellikler göz önünde bulundurulduğunda excel yerine R programını kullanmanın daha avantajlı

olduğu göze çarpmaktadır (Crawley, 2007; Heidberger ve Neuwirth, 2009). İkili karşılaştırmalar yöntemine dair Thurstone V. Hal Çalışması R programı ile gerçekleştirilebilmektedir. Excel ile gerçekleştirilen ikili karşılaştırmalar yönteminde excel yerine R programı kullanılarak ikili karşılaştırmaya dair ölçeklemenin daha kolay daha rahat ve daha sorunsuz bir şekilde gerçekleştirilmesi göz önünde bulundurulduğunda, konu ile ilgili gelecekte gerçekleştirilecek olan çalışmalarda R programının kullanım sıklığının artacağı düşünülmektedir. Bu kapsamda çalışma oldukça önemlidir.

YÖNTEM

Çalışma fakültelerde görevli öğretim üyeleri ve öğrencilerle yapılan görüşmeler sonucunda belirlenmiş boş zaman aktivitelerinden frekans değeri en yüksek sekiz farklı etkinliğin öğrenci yargılarına göre kendi içinde ikili karşılaştırma esasına dayanmaktadır. Thurstone (1927) tarafından geliştirilen karşılaştırmalı yargılar kanunu temel alınarak yapılan analizlerde öğrenciler her bir aktiviteyi ikili karşılaştırarak hangi aktiviteyi daha çekici buluyorlarsa ilgili matriste üstün olan aktiviteye 1, diğerine 0 puan vererek kodlama yapmışlardır. Bu bilgiler kullanılarak ikili karşılaştırma sonucunda aktivitelerin ölçek değerleri elde edilmiştir. Bu araştırmanın amacı, farklı programlarla gerçekleştirilen analiz sonuçlarını karşılaştırarak, analiz sonuçları arasındaki benzer ve farklı yönleri ortaya çıkarmak olduğundan çalışmanın yöntemi nicel araştırma yöntemlerinden karşılaştırmalı çalışmalar modeli altında değerlendirilmelidir (Mills, Bunt ve Brujin, 2006).

Çalışma Grubu

Araştırmanın çalışma grubunu Aydın ilinde yer alan bir üniversitenin farklı bölümlerinde öğrenim gören öğrenciler oluşturmaktadır. Uygulama 2015–2016 eğitim-öğretim yılı güz yarısında yapılmıştır. Çalışma gönüllülük esasına dayalı olarak, uygulamaya katılmak isteyen öğrenciler ile gerçekleştirilmiştir. Tabakalı örnekleme yöntemiyle belirlenen 350 öğrenciye uygulama formu dağıtılmıştır. Dağıtılan formlardan 44 tanesinin eksik ve hatalı doldurulması sebebiyle değerlendirmeye 306 form alınmıştır.

Veri Toplama Araçları

Öğrencilerin belirlenen 8 farklı boş zaman aktivitesinden ikili karşılaştırma yaparak hangisini daha fazla tercih ettiklerini belirlemek amacıyla 8x8 türünde kare matris oluşturulmuştur. Matrisin alt ve üst köşegen matrisinde benzer karşılaştırmaları iki defa yapmamak amacıyla alt köşegen matris kullanılarak karşılaştırma yapılmaları istenmiştir (Ek-1). Geliştirilen ölçme aracında yer alacak aktiviteleri belirlemek amacıyla sayısal ve sözel ağırlıklı bölümlerde görev yapan öğretim elemanları ve öğretim üyelerinden oluşan 10 kişilik bir ekipten uzman görüşü alınmıştır. Uzman görüşleri ve sonrasında öğrencilerle yapılan görüşmeler doğrultusunda frekans değeri en yüksek 8 aktivite (spor yapmak, kitap okumak, sosyal aktivite, arkadaşlarla vakit geçirmek, bilgisayar karşısında vakit geçirmek, telefon uygulamaları ile vakit geçirmek, tv seyretmek, uyumak) matris üzerine yerleştirilerek uygulama formuna son şekli verilmiştir. Matrisin 8x8 türünde kare matris olması, alt üçgen üzerinde işaretleme yapılması ve toplam gözenek sayısının 32 olması sebebiyle her hücreye en az 5 veri girmesi ($32 \times 5 = 160$) amacıyla katılımcı sayısı yüksek tutulmuştur.

Verilerin Analizi

Verilerin analizinde psikolojik özelliklerin ölçeklenmesinde yaygın olarak kullanılan tekniklerinden ikili karşılaştırma yöntemiyle ölçekleme tekniğinden yararlanılmıştır (Thurstone, 1927; Turgut ve Baykul, 1992). Thurstone (1927) tarafından geliştirilen karşılaştırmalı yargılar kanunu temel alınarak yapılan analizlerde öğrenciler her bir aktiviteyi ikili karşılaştırarak hangi aktiviteyi daha çok tercih ediyorlarsa bunu ilgili matriste üstün olan aktivite 1, diğerine 0 puan vererek kodlama yapmışlardır.

Bu bilgiler kullanılarak ikili karşılaştırma yapılan aktivitelerin ölçek değerleri elde edilmiştir. Araştırmada elde edilen verilerin analizine R programı için V. Hal, Excel için V. ve III. Hal Yöntemiyle ölçekleme gerçekleştirilerek sonuçlar tutarlılık açısından incelenmiştir. Ölçeklemede ikili karşılaştırmada kullanılan temel formüllerden yararlanılmıştır (Turgut ve Baykul, 1992). Öğrencilerin vermiş oldukları yanıtlarda çelişkili üçlüler olup olmadığı ki kare değerleri ile kontrol edilerek yanıtlarında çelişkili üçlü bulunan öğrencilere ait anket sonuçları analiz kapsamına alınmamıştır.

R programlama dili, ilk defa Yeni Zelanda’ da yer alan Aucland Üniversitesi İstatistik Bölümü’ nden Ross Ihaka ve Robert Gentleman tarafından oluşturulmuştur. Sonraları dünyanın farklı bölgelerindeki araştırmacılardan meydana gelen bir grup R programlama dilinin gelişimine yardımcı olmuştur. Bu gruba, 1997 yılında ‘‘R Core Team’’ ismi verilmiş ve R programlama dili bu grupta yer alan araştırmacıların da katkılarıyla günümüzdeki halini almıştır.

R programı, profesyonel veya ticari desteğe tabi, kapalı kutu yazılımlardan oluşan bir program değildir."R" dili için geliştirilmiş, dile yeni özellikler kazandıran birçok paket mevcuttur. Bu paketlerin çok büyük bir kısmı çevrim içi ortamda "CRAN" adı verilen bir "merkez" üzerinde bulunmaktadır. Araştırma kapsamında gerçekleştirilen ikili karşılaştırma yöntemi ile ölçekleme analizini içeren algoritmalar, ‘‘Psych’’ paketi içerisinde yer almaktadır.

Psych paketi, psikometri alanında kullanılan ve istatistiksel hesaplamaları içeren bir pakettir. İçerisinde güvenilirliğe, geçerliğe, klasik test teorisine, madde tepki kuramına ve ölçeklemeye ilişkin analizlerin gerçekleştirilebileceği küçük paketler yer almaktadır.

R programı kullanılarak ikili karşılaştırmalara dair hesaplama yapmak için, öncelikle ‘‘Psych’’ paketi R sistemine yüklenmelidir. Psych paketi içerisinde, ikili karşılaştırmalar yöntemi için hazırlanmış ‘‘vegetables’’ isimli örnek bir veri seti yer almaktadır.Bu veri seti, ilgili analiz konusunda deneyim sahibi olmak için kullanılabilir.R programı üzerinde ikili karşılaştırmalar yöntemi ile analiz gerçekleştirmek için <http://personality-project.org/r/html/thurstone.html> uzantılı linkten gereken kısa ve öz bilgi elde edilebilir. İlgili paket `install.packages("psych")` komutu ile sisteme yüklendikten sonra paketin `library(psych)` komutu ile program üzerinde çalışmaya hazır hale gelmesi sağlanmalıdır. Program kullanılarak ilgili analizin gerçekleştirilmesi için daha sonra çalışma için kullanılacak olan veri seti program üzerinde çalıştırılmalı ve ikili karşılaştırma yöntemi ile ölçekleme analizi gerçekleştirilmelidir.

BULGULAR

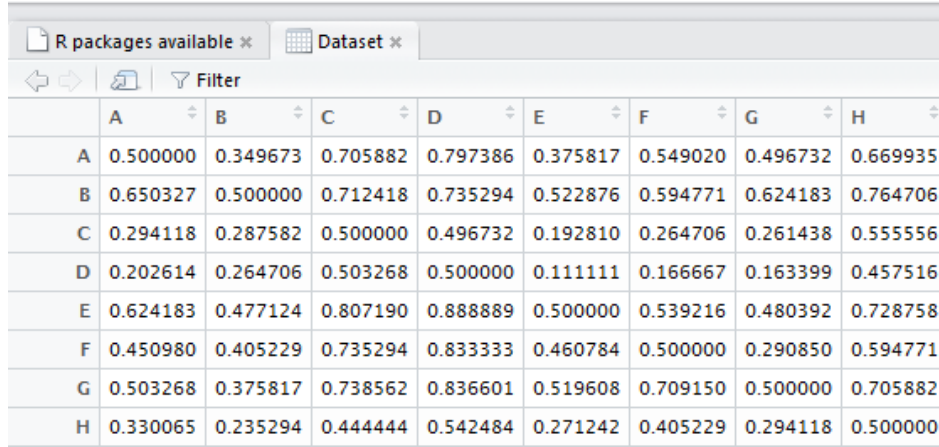
R Programı Kullanılarak III. Hal Ölçekleme Çalışması

R programı kullanılarak V. Hal Ölçekleme Çalışması gerçekleştirmek için oranlar matrisini kullanmak yeterli olmaktadır. Oranlar matrisi oluşturulurken, excel ile gerçekleştirilen hesaplamada köşegen matrisini oluşturan hücreler boş bırakılırken, R programı için oluşturulması gereken oranlar matrisinde köşegen matrisini oluşturan hücrelere 0,5 değeri yazılmalıdır. Tablo 1’ de konu ile ilgili R programı için hazırlanmış oranlar matrisi görülmektedir.

Tablo 1. R Programı için Oranlar Matrisi

| | A | B | C | D | E | F | G | H |
|---|----------|----------|----------|----------|----------|----------|----------|----------|
| A | 0,500000 | 0,349673 | 0,705882 | 0,797386 | 0,375817 | 0,549020 | 0,496732 | 0,669935 |
| B | 0,650327 | 0,500000 | 0,712418 | 0,735294 | 0,522876 | 0,594771 | 0,624183 | 0,764706 |
| C | 0,294118 | 0,287582 | 0,500000 | 0,496732 | 0,192810 | 0,264706 | 0,261438 | 0,555556 |
| D | 0,202614 | 0,264706 | 0,503268 | 0,500000 | 0,111111 | 0,166667 | 0,163399 | 0,457516 |
| E | 0,624183 | 0,477124 | 0,807190 | 0,888889 | 0,500000 | 0,539216 | 0,480392 | 0,728758 |
| F | 0,450980 | 0,405229 | 0,735294 | 0,833333 | 0,460784 | 0,500000 | 0,290850 | 0,594771 |
| G | 0,503268 | 0,375817 | 0,738562 | 0,836601 | 0,519608 | 0,709150 | 0,500000 | 0,705882 |
| H | 0,330065 | 0,235294 | 0,444444 | 0,542484 | 0,271242 | 0,405229 | 0,294118 | 0,500000 |

Oranlar matrisi oluşturulduktan sonra, matris R programı üzerinde çalıştırılmalıdır. Matris R programı üzerinde çalıştırıldıktan sonra, "thurstone(Dataset)" komutu ile analiz sonuçlandırılır. Parantez içerisinde yer alan "Dataset" yerine araştırmada kullanılacak olan veri setinin ismi yazılmalıdır. Araştırma için ilgili veri setine "Dataset" ismi verilmiştir. Şekil 1' de veri setinin program üzerinde görüntülenmiş hali yer almaktadır.



| | A | B | C | D | E | F | G | H |
|---|----------|----------|----------|----------|----------|----------|----------|----------|
| A | 0.500000 | 0.349673 | 0.705882 | 0.797386 | 0.375817 | 0.549020 | 0.496732 | 0.669935 |
| B | 0.650327 | 0.500000 | 0.712418 | 0.735294 | 0.522876 | 0.594771 | 0.624183 | 0.764706 |
| C | 0.294118 | 0.287582 | 0.500000 | 0.496732 | 0.192810 | 0.264706 | 0.261438 | 0.555556 |
| D | 0.202614 | 0.264706 | 0.503268 | 0.500000 | 0.111111 | 0.166667 | 0.163399 | 0.457516 |
| E | 0.624183 | 0.477124 | 0.807190 | 0.888889 | 0.500000 | 0.539216 | 0.480392 | 0.728758 |
| F | 0.450980 | 0.405229 | 0.735294 | 0.833333 | 0.460784 | 0.500000 | 0.290850 | 0.594771 |
| G | 0.503268 | 0.375817 | 0.738562 | 0.836601 | 0.519608 | 0.709150 | 0.500000 | 0.705882 |
| H | 0.330065 | 0.235294 | 0.444444 | 0.542484 | 0.271242 | 0.405229 | 0.294118 | 0.500000 |

Şekil 1. Veri Setinin R Programındaki Görünümü

Veri seti programa okutulduktan sonra;install packages("psych")komutu ile sisteme yüklenen ve daha sonra library(psych) komutu ile program üzerinde çalışmaya hazır hale getirilen "psych" paketi içerisinde yer alan "thurstone" komutu ile analizin gerçekleştirilmesi için Thurstone (Dataset) komutu program ekranına girilir ve "Enter" tuşuna basılarak analiz çıktısı elde edilir.

```
Thurstonian scale (case 5) scale values
Call: thurstone(x = Dataset)
  A    B    C    D    E    F    G    H
0.22 0.01 0.76 0.97 0.00 0.27 0.07 0.70

Goodness of fit of model 0.99
```

Şekil 2.R Programına İlişkin V. Hal Ölçkleme Çalışması Çıktısı

Şekil 2' de, R programında gerçekleştirilen Thurstone V. Hal ölçkleme sonuçları görülmektedir. V. Hal denklemi yardımıyla elde edilen ölçek değerleri incelendiğinde sıralama anlamında en çok tercih edilen aktivitelerin sırasıyla Arkadaşlarla vakit geçirme (D) ile Sosyal Aktiviteler (C) ve Uyku (H) olduğu, en az tercih edilen aktivitelerin ise sırasıyla Bilgisayar karşısında vakit geçirme (E), Kitap okuma (B) ve TV Seyretme (G) olduğu görülmektedir.

Excel Kullanılarak III. Hal Ölçkleme Çalışması

Çalışmanın bu bölümünde, ikili karşılaştırmalar sonucu elde edilen veri setinin Excel kullanılarak izlenen III. Hal Ölçkleme süreci ve bulgular konusunda bilgi verilmiştir. Bunun için ilk olarak, öğrencilerin ikili karşılaştırma sonucu verdikleri tepkilere ilişkin frekans değerleri belirlenmiş ve sonuçlar Tablo 2'de verilmiştir.

Tablo 2. Boş Zamanları Değerlendirme Yöntemlerine İlişkin Frekans Matrisi (F)

| | A | B | C | D | E | F | G | H |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| A | | 107 | 216 | 244 | 115 | 168 | 152 | 205 |
| B | 199 | | 218 | 225 | 160 | 182 | 191 | 234 |
| C | 90 | 88 | | 152 | 59 | 81 | 80 | 170 |
| D | 62 | 81 | 154 | | 34 | 51 | 50 | 140 |
| E | 191 | 146 | 247 | 272 | | 165 | 147 | 223 |
| F | 138 | 124 | 225 | 255 | 141 | | 89 | 182 |
| G | 154 | 115 | 226 | 256 | 159 | 217 | | 216 |
| H | 101 | 72 | 136 | 166 | 83 | 124 | 90 | |

Tablo 2’de verilen frekans matrisi öğrencilerin tercih ettikleri aktiviteleri tercih edilme sıklıklarına göre oluşturulmuştur. Tabloda spor yapmak (A), kitap okumak (B), sosyal aktivite (C), arkadaşlarla vakit geçirme (D), bilgisayar karşısında vakit geçirme (E), telefon uygulamaları ile vakit geçirmek (F), TV seyretmek (G) ve uyumak (H) olmak üzere toplam 8 boş zaman aktivitesi ikili karşılaştırılmıştır. Tabloda görüldüğü üzere B aktivitesi ile A aktivitesi karşılaştırıldığında çalışma grubunda bulunan 306 öğrencinin 199’u B aktivitesini tercih etmelerine rağmen 107’si A aktivitesini tercih etmişlerdir. Frekans matrisi tüm aktivitelerin ikili karşılaştırılmasına dayanan sonuçlardan oluşturulmuştur. Frekans matrisinde esas köşegen üzerindeki hücrelerde her aktivitenin kendisiyle karşılaştırma yapıldığı anlamına geldiğinden bu hücreler boş bırakılmıştır. Frekans matrisi oluşturulduktan sonra her bir hücrede yer alan değerlerin toplam öğrenci sayısı olan 306’ya bölünmesiyle oranlar matrisi elde edilmiştir. Oranlar matrisine ilişkin sonuçlar Tablo 3’te verilmiştir.

Tablo 3. Boş Zamanları Değerlendirme Yöntemlerine İlişkin Oranlar Matrisi (P)

| | A | B | C | D | E | F | G | H |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| A | | ,35 | ,71 | ,80 | ,36 | ,55 | ,50 | ,67 |
| B | ,65 | | ,71 | ,74 | ,52 | ,59 | ,62 | ,76 |
| C | ,29 | ,28 | | ,50 | ,19 | ,26 | ,26 | ,56 |
| D | ,20 | ,26 | ,50 | | ,11 | ,17 | ,16 | ,46 |
| E | ,62 | ,47 | ,81 | ,89 | | ,54 | ,48 | ,73 |
| F | ,45 | ,41 | ,74 | ,83 | ,46 | | ,29 | ,59 |
| G | ,50 | ,38 | ,74 | ,84 | ,52 | ,71 | | ,71 |
| H | ,33 | ,24 | ,44 | ,54 | ,27 | ,41 | ,29 | |

Tablo 3’te verilen matrisde esas köşegene simetrik olan değerlerin toplamının 1 olduğu görülmektedir. Oranlar matrisindeki hücrelerde yer alan değerlere karşılık gelen birim normal değerler (z değerleri) elde edilmiştir. Z Matrisinde sütun toplamları bulunduktan sonra bu değerler sekiz farklı etkinlik olduğundan 8’e bölünerek S_j değerleri elde edilmiştir. Eksenin başlangıç noktasını sıfır yapmak amacıyla her bir S_j değerine bu değerlerin mutlak değeri olan ,376 eklenerek her bir aktivitenin ölçek değeri belirlenmiştir. Her bir aktiviteye ilişkin ölçek değerleri Şekil 3’te gösterilmiştir.



Şekil 3.V. Hal Yöntemiyle Ölçkleme

Şekil 2 ve şekil 3 karşılaştırmalı olarak incelendiğinde, R programı ile elde edilen V. Hal yöntemine ilişkin sonuçlar ile Excel kullanılarak elde edilen V. Hal yöntemine ilişkin sonuçların neredeyse tamamıyla benzerlik gösterdiği görülmektedir.

V. Hal Yöntemiyle gerçekleştirilen ölçekleme işlemi sonucunda ortalama hata değeri 0,329 ve hesaplanan χ^2 değerinin 21 serbestlik derecesinde %95 güven derecesinde beklenen değerden büyük olması sebebiyle ve verilerin uyum ölçütlerini karşılamaması nedeniyle psikometrik yapı da göz önünde bulundurularak III. Hal denklemi yardımıyla ölçekleme yapılmasına karar verilmiştir. İki yöntem arasındaki fark ise ölçekleme yapılırken V. Hal denkleminde ayırt etme yargılarının varyansı eşit kabul edilirken III. Hal denklemi ayırt etme yargıları varyansını eşit kabul etmemekte ve her bir yargı için tek tek varyansları hesaplaması sebebiyle kullanılmaktadır (Turgut ve Baykul, 1992). Çalışmada III. Hal denklemiyle elde edilen ölçekleme sonuçları Şekil 4'te gösterilmiştir.



Şekil 4. III. Hal Yöntemiyle Ölçekleme

Hem Excel ve R ile gerçekleştirilen V. Hal hem de Excel ile gerçekleştirilen III. Hal yöntemi yardımıyla yapılan ölçekleme çalışmaları sonucunda boş zaman aktivitelerinin sıralarında büyük bir değişiklik olmadığı görülmüştür. Çalışmada V. Hal denklemi yardımıyla elde edilen ve III. Hal denklemi yardımıyla elde edilen ölçek değerleri karşılaştırıldığında sıralama anlamında en çok tercih edilen aktivite Arkadaşlarla vakit geçirme (D) ile Sosyal Aktiviteler (C) ve Uyku (H) olduğu görülmektedir. Ancak en az tercih edilen aktivitelere bakıldığında V. hal denkleminde sırasıyla Bilgisayar karşısında vakit geçirme (E), Kitap okuma (B) ve TV Seyretme (G) olmasına karşın III. hal denkleminde sırasıyla Kitap okuma (B), Bilgisayar karşısında vakit geçirme (E) ve TV Seyretme (G) olduğu görülmektedir. Hem V. hal hem de III. Hal yönteminde tüm aktivitelerin ölçek üzerindeki yerleri aynı olmasına karşın sadece en az tercih edilen Kitap okuma ve TV Seyretme aktivitelerinin yerlerinin değiştiği görülmektedir. Ölçek değeri aralıklarına bakıldığında III. Hal denklemi yardımıyla elde edilen ölçek değeri aralıklarının V. Hal denkleminde göre biraz daha genişlediği görülmüştür. Bu sonuç Aksu ve Doğan (2015) tarafından yapılan çalışmanın sonuçlarıyla benzerlik göstermektedir. Bunun yanında III. Hal denklemi yardımıyla en az tercih edilen aktivitelerin daha iyi ayrılabilirdiği belirlenmiştir. Bu sonuçlara göre öğrenci görüşlerine göre en çok tercih edilenden en az tercih edilene doğru sıralama değerleri Tablo 4'de hem V. Hal hem de III. Hal Yöntemiyle elde edilen ölçek değerleri için gösterilmiştir.

Tablo 4.V. ve III. Hal Yöntemiyle Elde Edilen Değerlerin Karşılaştırılması

| V. HAL YÖNTEMİ | | | III. HAL YÖNTEMİ | |
|----------------|--------------------------|--------------|--------------------------|--------------|
| Sıra No | Aktivite | Ölçek Değeri | Aktivite | Ölçek Değeri |
| 1 | Arkadaşlar (D) | 0,96 | Arkadaşlar (D) | 1,44 |
| 2 | Sosyal aktivite (C) | 0,76 | Sosyal aktivite (C) | 1,36 |
| 3 | Uyku (H) | 0,69 | Uyku (H) | 1,14 |
| 4 | Telefon uygulamaları (F) | 0,27 | Telefon uygulamaları (F) | 0,72 |
| 5 | Spor yapmak (A) | 0,22 | Spor yapmak (A) | 0,63 |
| 6 | TV seyretmek (G) | 0,06 | TV seyretmek (G) | 0,46 |
| 7 | Kitap okumak (B) | 0,01 | Bilgisayar (E) | 0,38 |
| 8 | Bilgisayar (E) | 0,00 | Kitap okumak (B) | 0,00 |

Tablo 4 incelendiğinde en çok tercih edilen ilk altı aktivitenin sıralamada yerleri aynı kalmasına karşın kitap okuma ve bilgisayar başında vakit geçirme aktivitelerinin sıra değerlerinde değişiklik olduğu görülmektedir.

SONUÇLAR ve TARTIŞMA

Bu çalışmada üniversite öğrencilerinin boş zamanlarını değerlendirmeleri bakımından en çok tercih ettikleri 8 farklı boş zaman aktivitesinin farklı programlarla gerçekleştirilen analiz sonuçlarını karşılaştırarak, analiz sonuçları arasındaki benzer ve farklı yönleri ortaya çıkarmak amaçlanmıştır. Yargıcı kararlarına dayalı ölçekleme tekniklerinden biri olan ikili karşılaştırmalar yöntemiyle öğrencilerin en çok tercih ettiklerini düşündükleri yöntemleri karşılaştırmaları istenmiştir (Kan, 2008).

Elde edilen sonuçlara göre, Thurstone V. Hal Ölçeklemesinin R programı ile gerçekleştirilen analiz sonuçlarının, Excel ile gerçekleştirilen analiz sonuçları ile çok benzer olduğu görülmektedir. Excel ile gerçekleştirilen V. Hal Ölçekleme Hesaplamalarının el yordamı ile gerçekleştirilmesi, fazla zaman alması ve hataya açık olması sebebi ile bu hesaplamalar R programı ile daha kısa sürede ve analiz sürecinin içerisine bireysel hata karışma ihtimali en düşük düzeyde olacak şekilde gerçekleştirilebilmektedir. R programının açık kaynak kodlu olması, ücretsiz bir yazılım olması, Excel ile oluşturulan veri setlerinin program içerisine aktarılabilmesi gibi özellikleri de ikili karşılaştırmalar ölçekleme yöntemi konusunda R programının rahatlıkla excel yerine kullanılabilmesi için geçerli sebepler olarak düşünülebilir.

Yargıcı tepkilerine dayalı ölçekleme yöntemleri düşünüldüğünde, R programı ile sadece Thurstone V. Hal Ölçekleme Yöntemi gerçekleştirilebilmektedir. Gelecekte gerçekleştirilecek olan araştırmalarda, R programının açık kaynak kod özelliği göz önünde bulundurularak Thurstone III. Hal Ölçekleme yöntemine dair kodlar yazılabilir ve sonuç olarak bu ölçekleme yöntemi R programı ile gerçekleştirilebilir ve excel ile elde edilen III. Hale ilişkin ölçek değerleri ile R programı ile elde edilen III. Hale ilişkin ölçek değerleri elde edilebilir.

R programı ile gerçekleştirilen Thurstone V. Hal Ölçekleme işlemi sonucunda ölçek değerlerinin yanında goodness of fit indice (uyum iyiliği indeksi) elde edilmektedir. Uyum iyiliği indeksinin elde edilmesi, kullandığımız yöntemin yanlılanabilir olduğunu göstermekte, yani uyum iyiliği indeksinin alacağı değere göre ilgili çalışmaya devam edip edilmeyeceği konusunda bilgi sahibi olunmaktadır.

Elde edilen sonuçlar, üniversite öğrencilerinin rekreasyon değerlendirme tercihleri açısından incelendiğinde ilgili sekiz etkinlik en çok tercih edilenden en az tercih edilene arkadaşlarla vakit geçirme, sosyal aktiviteler, uyku, tv seyretme, kitap okuma ve bilgisayar karşısında vakit çevirme şeklinde sıralanmıştır. Üniversite öğrencilerinin kitap okuma alışkanlığı açısından yoksun olduklarını göstermektedir. Üniversite öğrencilerinin okuma alışkanlığı elde edebilmesi için üniversite kütüphaneleri öğrenciler için çekici bir hale getirilmeli ve öğrencilerin de tercihleri göz önünde bulundurularak çeşitli kitapların sayısı artırılmalıdır. Üniversitelerde yer alan faaliyetler öğrencileri okumaya özendirerek şekilde düzenlenmelidir. Öğrencilerin ilgi ve ihtiyaçları göz önünde bulundurulduğunda, okuma ortamları tasarlanırken öğrencilerin ilgi ve ihtiyaçları da göz önünde bulundurulmalıdır. Sekiz etkinliğin sıralamadaki yeri bir bütün olarak düşünüldüğünde; üniversite öğrencilerinin yoksun bir rekreatif etkinlik tercihi anlayışının olduğu söylenebilir.

KAYNAKÇA

Acar Güvendir, M. ve Özer Özkan, Y. (2013). İki Ölçekleme Yönteminin Karşılaştırılması: İkili karşılaştırma ve sıralama yargıları [A Comparison of Two scaling Methods: Pair Wise Comparison and Rank-Order Judgments Scaling]. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research*. 3 (1), 105–119. [Online]: [http:// http://ebad-jesr.com/](http://http://ebad-jesr.com/) adresinden 11 Ocak 2016 tarihinde indirilmiştir.

- Aksu, G. ve Doğan, N. (2015). Öğretim Yöntem ve Tekniklerinin Öğrenci Görüşlerine Göre İkili Karşılaştırma Yöntemiyle Ölçeklenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6 (2), 194-206
- Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Altun, A. ve Gelbal S. (2014). Öğretmenlerinin Kullandıkları Ölçme ve Değerlendirme Yöntem veya Araçlarının İkili Karşılaştırma Yöntemiyle Belirlenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2 (2), 200-209.
- Anıl, D. ve Güler, N. (2006). İkili karşılaştırma Yöntemi ile Ölçekleme Çalışmasına Bir Örnek. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 30, 30-36.
- Aydın, M. (2000). *Kurumlar Sosyolojisi*. Ankara: Vadi Yayınları.
- Baykul, Y. ve Turgut, M. F. (1992). *Ölçekleme Teknikleri*. Ankara: Meteksan Anonim Şirketi
- Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language*. Wadsworth & Brooks/Cole.
- Crawley, M. J. (2007). *The R book*. Chichester, West Sussex, England: Wiley.
- Demirel, Ö. (2011). *Öğretme Sanatı*. Ankara: Pegem Akademi.
- Ergani, K. (2010). *İlköğretim 4. ve 5. Sınıf Sosyal Bilgiler Dersi Öğretim Yöntem ve Teknikleri İle Materyal Kullanımına İlişkin Öğretmen Görüşleri*. Yayınlanmış Yüksek Lisans Tezi, Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü, İlköğretim Anabilim Dalı, Kütahya.
- Gökçe, F. (1994). Eğitimde Denetimin Amaç ve İlkeleri, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 10, 73-78.
- Godbey, G., Stanley, P. (1976). *Leisure Studies and Services: An Overview*. Pennsylvania: W.B. Saundersco.
- Gönen, S., ve Kocakaya, S. (2006). Fizik Öğretmenlerinin Hizmet İçi Eğitimler Üzerine Görüşlerinin Değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 19, 37-44,
- Hacıoğlu, N., Gökdeniz, A., Dinç, Y. (2003). *Boş zaman ve Rekreasyon Yönetimi*. Ankara: Detay Yayınları.
- Heidberger, R.M., Neuwirth, E. (2009). R Through Excel: A Spreadsheet interface for Statistics, Data Analysis, and Graphics. Springer, New York.
- Heldsinger, S., Humphry, S., (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, 37(2).
- Hesapcioğlu, M. (2011). *Öğretim İlke ve Yöntemleri*. Ankara: Nobel Yayınları.
- İşman, A. ve Eskicumalı, A. (1999). *Eğitimde Planlama ve Değerlendirme*. Adapazarı: Değişim Yayınları.
- Jennen, C., Uhlebeck, G. (2004). Exercise and life satisfaction: Fitness complimentary strategies in the prevention and rehabilitation of illness. *Evidence-based Complimentary and Alternative Medicine*, 1(2), 147-165.
- Jungst, S., Licklider, B. & Wiersema, J. (2003). Providing support for faculty who wish to shift to a learning-centered paradigm in their higher education classrooms. *The Journal of Scholarship of Teaching and Learning*, 3, 69-81.
- Ishiyama, J.T., McClure, M., H. ve Amico, J. (1999) Critical thinking disposition and locus of control as predictors of evaluations of teaching strategies. *College Student Journal*, 33 (2), 10-44.
- Kan, A. (2008). Psikolojik Değişkenleri Ölçmek İçin Kullanılan Ölçekleme Yaklaşımları Üzerine Bir Karşılaştırma. *Eğitimde Kuram ve Uygulama*, 4 (1), 2-18.
- Kayabaşı, Ş. ve Erdoğan, Y. A. (2002). *Öğretimi Planlama ve Değerlendirme*. Ankara: Anı Yayınevi.
- Kahyaoglu, M ve Yangın, S. (2007). İlköğretim Öğretmen Adaylarının Mesleki Öz Yeterliklerine İlişkin Görüşleri. *Kastamonu Eğitim Dergisi*, 15, 83-105.
- Karaağaçlı, M. (2005). *Öğretimde Yöntemler ve Yaklaşımlar*. Ankara: Pelikan Yayıncılık.
- Karaküçük S., Başaran Z. (2000). Stresle Başa Çıkma da Rekreasyon Faktörü. *Gazi Beden Eğitimi ve Spor Bilimleri Dergisi*. Ankara.
- Karaküçük S. (1995). *Rekreasyon (Boş Zamanları Değerlendirme)*. Seren Matbaacılık. Ankara.
- Karaküçük, S. (1999). *Rekreasyon Boş Zamanları Değerlendirme*. Ankara: Bağırğan Yayımevi.
- Kılıç, C. (2010). *İlköğretim Okullarında (Devlet-Özel) ve Dershanelerde Görev Yapan Fen ve Teknoloji Öğretmenlerinin Kullandıkları Öğretim Yöntem ve Teknikleriyle İlgili Öğrenci Görüşleri*, Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, İlköğretim Bölümü Fen Bilgisi Öğretmenliği Bilim Dalı, Ankara.
- Kim, B. B. (2008). Research update: Perceiving leisure. *Parks & Recreation*, 43(5), 24-27
- Küçükahmet, L. (1983). *Öğretim İlke ve Yöntemleri*. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Yayınları No: 124.
- Küçükahmet, L. (1995). *Öğretim İlke ve Yöntemleri*. Ankara: Gazi Büro Kitabevi.
- Larson R. W. & Richards, M. H. (1991). Daily companionship in late childhood and early adolescence: Changing developmental contexts. *Child Development*, 62, 284-300

- Marbach-Ad, G., Seal, O. & Sokolove, P. (2001). Student attitudes and recommendations on active learning. *Journal of College Science Teaching*, JO, 434-438.
- Mills, M., Bunt, G. G. ve Brujin, J. (2006). Comparative research persistent problems and promising solutions. *International Sociology*, 21 (5), 619-631.
- Murphy, B. C., Eisenberg, N., Fabes, R. A., Shepard, S. ve Guthrie, I. K. (1999). Consistency and change in children's emotionality and regulation: A longitudinal study. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, 45, 413-444.
- Nartgün, Z. (2006). Öğretmenlik Meslek Bilgisi Derslerinin Önem Düzeyinin İkili Karşılaştırmalarla Ölçeklenmesi. *A.İ.B.Ü. Eğitim Fakültesi Dergisi*, 6 (2), 161-176.
- Önen, F., Mertoğlu, H., Saka, M. ve Gürdal, A. (2009). Hizmet içi Eğitimin Öğretmenlerin Öğretim Yöntem ve Tekniklerine İlişkin Bilgilerine Etkisi: ÖPYEP Örneği. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 10 (3), 9-23.
- Önen, F., Saka, M., Erdem, A., Uzal, G. ve Gürdal, A., (2008). Hizmet İçi Eğitime Katılan Fen Bilgisi Öğretmenlerinin Öğretim Tekniklerine İlişkin Bilgilerindeki Değişimin Tespiti: Tekirdağ Örneği. *KEFAD*, 9 (1), 45-57
- Özden, Y. (2000). *Öğrenme ve Öğretmen*. Pegem A Yayıncılık, Ankara.
- Özer, Y. ve Acar, M. (2011). Öğretmenlik Mesleği Genel Yeterlikleri Üzerine İkili Karşılaştırma Yöntemiyle Bir Ölçekleme Çalışması. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 3 (40), 89-101.
- Öztürk, N., Özdemir, S. ve Gelbal, S. (2011). İki farklı ölçekleme yaklaşımından elde edilen ölçek değerleri tutarlılığının incelenmesi. 20. Ulusal Eğitim Bilimleri Kurultayı. 8-10 Eylül 2011. Burdur.
- Peker, M. ve Mirasyedioğlu, Ş. (2003). Lise 2. Sınıf Öğrencilerinin Matematik Dersine Yönelik Tutumları ve Başarıları Arasındaki İlişki. *Pamukkale Eğitim Fakültesi Dergisi*, 2 (14), 157-166.
- Revelle, W. (2009) *An Introduction to Psychometric Theory with Applications in R*. Springer.
- Saban, A. (2004). *Öğrenme Öğretmen Süreci Yeni Teori ve Yaklaşımlar*. Nobel Yayınevi, Ankara.
- Tan, Ş., Kayabaşı, Y. ve Erdoğan, A. (2002). *Öğretimi Planlama ve Değerlendirme (3. Baskı)*, Ankara: Anı Yayıncılık.
- Taşpınar, M. (2005). *Kuramdan Uygulamaya Öğretim Yöntemleri*. Ankara: Nobel Yayınevi.
- TDK (2014). Türk Dil Kurumu Genel Türkçe Sözlük. Erişim Tarihi: 25.05.2014.
- Terzioğlu, A. & Yazıcı, M. (2003). Üniversite Öğrencilerinin Boş Zamanlarını Değerlendirme Anlayış ve Alışkanlıkları (Atatürk Üniversitesi örneği). *Erzincan Eğitim Fakültesi Dergisi*, 5(2), 1-31.
- Thurstone, L. L. (1927). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 33, 529-54.
- Thurstone, L. L. (1928). *The Measurement of Values*. Chicago, USA: The University of Chicago Press.
- Thurstone, L. L. (1959). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.
- Tohumcu, T. (2004). *Adıyaman Merkez İlköğretim Okulları 5. Sınıf Öğrencilerinin Matematik Dersindeki Başarıları ile Bu Öğrencilerin Sınıf Öğretmenlerinin Öğretim Yöntemleri Arasındaki İlişkinin İncelenmesi*, Yayınlanmış Yüksek Lisans Tezi, İnönü Üniversitesi Sosyal Bilimler Enstitüsü, Malatya.
- Trottier, A. N., Brown, G. T., Hobson, S. J. G. & Miller, W. (2002). Reliability and validity of the Leisure Satisfaction Scale (LSS-short form) and adolescent leisure interest profile (ALIP). *Occupational Therapy International*, 9(2), 131-144.
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçekleme Teknikleri*. Ankara: ÖSYM Yayınları.
- Uysal, A. (2010). *Sınıf Öğretmenlerinin 2009 Hayat Bilgisi Öğretim Programında Belirtilen Strateji, Yöntem ve Teknikleri Uygulamadaki Yeterlilik Düzeylerinin Belirlenmesi*. Yayınlanmış Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Vural, B. (2006). *Eğitim-Öğretimde Planlama-Ölçme ve Stratejiler*. İstanbul: Hayat Yayıncılık.
- Yeşilyurt, E. (2013). Öğretmenlerin Öğretim Yöntemlerini Kullanma Amaçları ve Karşılaştıkları Sorunlar. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 17 (1), 163-188.
- Yıldız, S. (2008). *Özel Eğitim Sınıflarında Çalışan Sınıf Öğretmenlerinin Matematik Öğretiminde Kullanılan Öğretim Yöntemlerine İlişkin Görüşlerin Değerlendirilmesi*. Yayınlanmış Yüksek Lisans Tezi, Selçuk Üniversitesi Sosyal Bilimler Enstitüsü, Konya
- Yin, Z., Katims, D. S. ve Zapata, J. T. (1999). Participation in leisure activities and involvement in delinquency by mexican american adolescents. *Hispanic Journal of Behavioral Sciences*, 21(2), 170-185.

EXTENDED ABSTRACT

Introduction

The design and principle for the scale structure in the paired comparison method is based on the study of Thurstone (1927, 1959). Thurstone suggested that attitudes could be measured and he developed a process and model which could be used for the scaling of a group of stimuli based on simple comparisons between two stimuli at one time. Thurstone named the formulation of the process as “Law of Comparative Judgement”. Several methods stand out with respect to Thurstone-type scaling. The paired comparison method is one of them. Decisions of judges are required in the paired comparison method like in the other methods. This study attempted to determine the level of preference of the leisure time activities of the university students. Leisure time refers to the time which is regarded as the source of creative and moral values where an individual moves away from all difficulties considering both himself/herself and other individuals and deals with an activity chosen by own will (Godbey and Stanley, 1976; Hacıoğlu, Gökdeniz and Dinç 2003). The concept of recreation centres around the activities carried out by persons in their leisure times. Making use of leisure time is not only a need of individuals, a method of having an efficient life and a way of moral relief but also an opportunity for individuals for expressing themselves and developing their personalities through recreation (Karaküçük, 1999; Terzioğlu and Yazıcı, 2003; Gökmen and diğerleri 1985). It is essential to make use of time through an enhanced and effective way. The problems which occurs through wrong and inefficient time management have a negative effect on the individual and society as whole (Jennen and Uhlebeck, 2004). Leisure time is not used generally well in our country and it is observed that young people don't have good habits due to various reasons like low level of reading etc. Considering the different life styles and habits of individuals, we see that they have various ways of spending leisure time. Following the technological advancements, mass communication played great role in the creation of common leisure time activities and a common global environment for individuals (Pekince, 2012; Aydın, 2000). The R Project was used in the study in order to compare the scaling results of the paired comparison to the results obtained through the formula suggested by Thurstone.

Method

The analyses were based on the law of comparative judgements developed by Thurstone (1927) and students conducted paired comparison for each activity and coded the attractive activities by giving 1 point to the favourable activity and 0 point to the other. They obtained the scale values of the activities after paired comparison using this information. The study group of this research consists of the faculty students studying at different departments of a university in the Aydın province. The application was conducted in the fall semester of the academic year 2015-2016. The study was conducted with the students based on voluntariness. Application forms were distributed to 350 students who were determined by the stratified sampling method. 306 forms were taken for evaluation while 44 forms were incomplete and wrongly completed. In order to determine the activities to be included in the measurement tool, expert opinions were taken from a team of 10 people consisting of lecturers and instructors who work in the numerical and verbal departments. The application form was finalized by placing on the matrix the 8 activities with the highest frequency value in line with the expert views and talks made with students. Paired comparison scaling technique was used in the data analysis which is one of the most common techniques in scaling the psychological qualities (Thurstone, 1927; Turgut and Baykul, 1992). The analyses were based on the law of comparative judgements developed by Thurstone (1927) and students conducted paired comparison for each activity and coded the attractive activities by giving 1 point to the favourable activity and 0 point to the other. For the analysis of the obtained data, the results were examined for consistency by conducting scaling through the case V. method for the R project and through the case V. and III. methods for the Excel.

Results and Discussion

Ratios matrix was used to conduct the case III scaling study by using the R project. When we look at the case V scaling results, the most preferred activities were respectively spending time with friends (D), social activities (C) and sleeping (H) while the least preferred activities were spending time on computer (E), reading books (B) and watching TV (G). In addition, as a result of the case III scaling process followed by the Excel program, the most preferred activities were respectively spending time with friends (D), social activities (C) and sleeping (H) while the least preferred activities were spending time on computer (E), reading books (B) and watching TV (G). These results obtained by the R project are similar to almost all of the results obtained by the Excel program. As a result of the scaling works done by both the case V method and case III method, no big difference was seen in the order of leisure time activities. When the scale values of obtained by the case V equation and case III equation were compared, the most preferred activities were respectively spending time with friends (D), social activities (C) and sleeping (H). However the least preferred activities in the case V equation were respectively spending time on computer (E), reading books (B) and watching TV (G) while these activities in the case III equation were respectively reading books (B), spending time on computer (E) and watching TV (G). According to the obtained results, the analysis results done by the R project were very similar with the analysis results obtained with the Excel program. The case III scaling calculations with the excel program are done gropingly, take more time and are open to error, therefore, these calculations can be done with the R project in a shorter time and with the smallest possibility to have individual errors. When the obtained results are examined with respect to the recreation evaluation preferences of the university students, the eight activities rank from the most preferred to the least preferred as spending time with friends, social activities, sleeping, watching TV, reading books and spending time on computer. It is determined that the university students don't have reading habits. University libraries should be made attractive for students and number of books should be increased in line with the student preferences so that the university students may gain reading habits. The activities within universities should be arranged to encourage students for reading. Reading environments should be designed taking the interests and needs of students into consideration. Considering the eight activities as a whole in the ranking, it can be said that the university students had a poor understanding of recreational activity preference.

EK-1. Veri Toplama Aracı

| | 1. SPOR YAPMAK | 2. KİTAP OKUMAK | 3. SOSYAL AKTİVİTE (Tiyatro, Sinema, vb.) | 4. ARKADAŞLARLA VAKİT GEÇİRMEK | 5. BİLGİSAYAR KARŞISINDA VAKİT GEÇİRMEK | 6. TELEFON UYGULAMALARI İLE VAKİT GEÇİRMEK (Oyun, Facebook, Twitter, vb.) | 7. TV SEYRETMEK | 8. UYUMAK |
|--|----------------|-----------------|---|--------------------------------|---|---|-----------------|-----------|
| 1. SPOR YAPMAK | | | | | | | | |
| 2. KİTAP OKUMAK | | | | | | | | |
| 3. SOSYAL AKTİVİTE (Tiyatro, Sinema, vb.) | | | | | | | | |
| 4. ARKADAŞLARLA VAKİT GEÇİRMEK | | | | | | | | |
| 5. BİLGİSAYAR KARŞISINDA VAKİT GEÇİRMEK | | | | | | | | |
| 6. TELEFON UYGULAMALARI İLE VAKİT GEÇİRMEK | | | | | | | | |
| 7. TV SEYRETMEK | | :) | | | | | | |
| 8. UYUMAK | | | | | | | | |

NOT: Yukarıdaki matriste yatayda (sıra) ve dikeyde (sütun) boş zaman etkinlikleri bulunmaktadır. Sizden istenilen her bir etkinliği ikili karşılaştırma yapmanızdır. Daha açık bir ifadeyle sıra ve sütunda yer alan iki etkinlikten birini tercih etmeniz gerekseydi hangisini seçerdiniz? Örneğin 7. Sıradaki TV Seyretmek ile 2. sütundaki Kitap Okumak etkinliğini karşılaştıran bir öğrenci iki seçeneğin kesiştiği noktaya (tabloda gülüncük olan hücre) eğer yataydaki etkinliği (TV Seyretmek) daha çok tercih ediyorsa "TV"; eğer dikeydeki etkinliği (Kitap Okumak) daha çok tercih ediyor ise "Kitap" yazması gerekmektedir. Benzer şekilde boş olan tüm hücreler için aynı işlemi yapmanız gerekmektedir. Ayrıca siyah alanlara işaretleme yapmayınız.

Öğretmenlerin Yükseköğrenim Alanları ile Öğrencilerin Akademik Başarısı Arasındaki İlişki*

Relationship between Higher Education Area of the Teachers and Academic Achievement of Students

İlkay ABAZAOĞLU **

Murat YATAĞAN ***

Ahmet ARİFOĞLU ****

Öz

Bu araştırmada, “Uluslararası Matematik ve Fen Eğilimleri Araştırması-2011” (TIMSS 2011) uygulamasına katılan Singapur, Güney Kore, Japonya, Çin-Tayvan, Finlandiya, Slovenya, İngiltere, Türkiye, Malezya ve Makedonya verileri kullanılarak “Sekizinci sınıf öğrencilerinin fen bilgisi ve matematik başarıları, öğretmenlerin yükseköğrenimde aldıkları eğitime göre nasıl bir değişim göstermektedir?” sorusuna yanıt aranmıştır. Araştırma kapsamında incelenen ülkelerde fen bilgisi ve matematik öğretmenlerinin yükseköğrenimde hangi alanlarda aldıkları eğitimin, öğrencilerin fen bilgisi ve matematikteki akademik başarılarını istatistiksel olarak anlamlı etkilediği ve akademik başarıyı artırıp artırmadığı tespit edilmeye çalışılmıştır. Bu çalışmanın veri kaynakları TIMSS 2011 uygulamasından elde edilmiştir. Veriler SPSS tabanlı çalışan HLM analiz programı ile analiz edilmiştir. Bu çerçevede matematik ve fen bilgisi öğretmenlerinin yükseköğrenimde aldıkları alan eğitimlerinin, öğrencilerin fen bilgisi ve matematik başarısına etkisi Hiyerarşik Lineer Modelleme (HLM) yöntemi ile belirlenmeye çalışılmıştır. Yükseköğreniminde biyoloji, fizik ve kimya alanlarında eğitim alan fen bilgisi öğretmenlerinin öğrencilerinin akademik fen başarı puanları, bu alanlarda eğitim almayan öğretmenlere göre daha yüksek bulunmuştur. Yükseköğreniminde matematik, yer bilimleri ve diğer olarak ifade edilen alanlarda eğitim alan matematik öğretmenlerinin öğrencilerinin akademik matematik başarı puanları, bu alanlarda eğitim almayan öğretmenlere göre daha yüksek çıkmıştır.

Anahtar Kelimeler: TIMSS 2011, öğrenci başarısı, fen başarısı, matematik başarısı, öğretmen, yükseköğrenim.

Abstract

This study is aimed to find out relation between courses attended by teachers in undergraduate and 8th grade students' success in science and mathematics in Singapore, South Korea, Japan, China-Taiwan, Finland, Slovenia, England, Turkey, Malaysia and Macedonia in TIMSS 2011. The study has been focused on which participated courses have statistically significant effect on students' academic success. Data for analysis has been gathered from TIMSS 2011 results. SPSS and HLM (SPSS based program) have been used for correlation and regression analysis of data. Hierarchical Linear Modeling has been used to find out effect of course content, related with field of science and mathematics' teachers, on students' science and mathematics success. Result indicated that students of science teachers who attended biology, physics and chemistry course in undergraduate education have higher science achievement scores that teachers who didn't attend these courses. Furthermore, students of mathematics teachers who attended mathematics, earth science and other courses have higher mathematics achievement score than teachers who didn't attend these courses.

Key Words: TIMSS 2011, student achievement, science achievement, mathematics achievement, teacher, higher education.

* Bu çalışma, iSER 2014 World Conference, October 29 –November 2, Kapadokya, Turkey (2014) konferansında bildirilerek sunulmuştur.

** Dr., MEB, Ankara-Mamak BİLSEM, Ankara- Türkiye, e-posta: ilkayabaza@hotmail.com

*** Dr., Milli Eğitim Bakanlığı, Ankara- Türkiye, e-posta:myatagan@gmail.com

**** Eğitim. Uzm., MEB Hayat Boyu Öğrenim Genel Müdürlüğü, Ankara- Türkiye, e-posta:axarif@yahoo.com

GİRİŞ

Eğitim kalitesinin en önemli faktörünün öğretmen kalitesi olduğu kabul edilen bir düşüncedir. Bu nedenle etkili öğretmenlerde bulunması gereken özellikler; öğretmenlerin hizmet öncesi ve sonrasında yapılan yetiştirme programları ve profillerine ilişkin literatürde çok sayıda araştırma bulunmaktadır (Wayne ve Youngs, 2003; Tondeur, Valcke, ve Van Braak, 2008). Alexander (2000) öğretmenleri kültürlere göre değişiklik gösteren mantıksal, psikolojik ve özellikle de etik özelliklerine göre değerlendirmenin psikometrik sorunlar yüzünden pek mümkün olmadığını belirtmiştir. Bu nedenle öğrenci başarısına bakarak öğretmen özelliklerini değerlendirme eğilimi önem kazanmaktadır (Zuzovsky, 2009). Bu anlamda özellikle uluslararası düzeyde gerçekleştirilen öğrenci başarısını ölçmeyi amaçlayan araştırmaların pek çoğunda PISA, TIMSS ve TALIS gibi çalışmalarda öğretmenlerle ilgili sorular da yer almaktadır. Bu durum öğretmen özelliklerinin öğrenci başarısına etkisini araştırmaya olanak sağlamaktadır.

Literatürde öğretmenin eğitimi ve tecrübesi gibi özelliklerinin öğrencinin öğrenme ve başarısıyla ilişkili olduğunu ortaya çıkaran çok sayıda uluslararası araştırmalar bulunmaktadır (Greenwald, Hedges, ve Laine 1996; Nye, Konstantopoulos, ve Hedges 2004; Rivkin, Hanushek, ve Kain 2005). Bu nedenle Wayne ve Youngs'ın (2003) da belirttiği gibi ilköğretimi iyileştirme yolları arayan araştırmacılar ve politika yapıcılar için sağlıklı bir yaklaşım da öğretmenlere odaklanmaktır. Wayne ve Youngs bu önerilerini gerekçelendirirken hem algının hem de deneysel araştırmaların öğrenci başarısının büyük ölçüde eğitimlerine giren öğretmenlere bağlı olduğunu vurgulayarak öğretmenlerin sistemin en temel kaynağı olduklarını savunmuşlardır.

Literatürde öğretmenlerle ilgili nitelikler; öğretmen özellikleri (Wayne ve Youngs, 2003; Tondeur, Valcke, ve Van Braak, 2008), öğretmen kalitesi (Rowan, Correnti, ve Miller 2002; Akiba, LeTendre, ve Scribner, 2007), öğretmen profilleri (Coults ve Lewin, 2002; Cruickshank, 2001; Scott-Kassner, 1999); ve öğretmen yeterlikleri (Zuzovsky, 2009; Croninger, Rice, Rathbun, ve Nishio, 2007) gibi farklı başlıklar altında ele alınmıştır. Ancak bu terminolojilerin içeriklerinin öğretmenin sertifika ve mezuniyet durumu ve sahip oldukları deneyim gibi özellikleri ortak bir şekilde kapsadığı için benzer tanımlara sahip oldukları söylenebilir.

Öğrencilerin fen ve matematik alanındaki akademik başarılarının yükselmesinde en önemli unsurların başında öğretmen faktörünün geldiği; öğretmenlerin hizmet öncesi eğitimleri ile hizmetiçi mesleki gelişimlerin öğrencilerin akademik başarısına olan pozitif etkisini birçok araştırma sonuçları göstermektedir (Goldhaber ve Brewer, 1998; Rowan, Chiang ve Miller, 1997; Goldhaber, 2002). Öğrencilerin akademik başarısını etkileyen diğer önemli faktörlerden birisi de öğretmenlerin yükseköğretimde aldıkları eğitimin içeriğidir. Bu nedenle öğretmenlerinin yükseköğretimde aldıkları eğitimlere odaklanmak önemli bir politika alanıdır.

Akiba ve arkadaşları (2007), çalışmalarında TIMSS 2003 uygulamasına katılan 46 ülkenin matematik ve fen verilerini kullanarak düşük ve yüksek sosyo-ekonomik düzeyden gelen öğrencilerin başarılı öğretmenlere erişimlerini araştırmışlardır. Araştırmalarında öğretmen kalitesinin ölçülebilir ve çeşitli kültürel bağlamlar arasında da ortak anlam taşıyan özelliklerine odaklanmışlardır. Araştırma sonucunda başarılı öğretmenler için ortak özellikler olarak öğretmenlerin sertifika donanımı, ana branşının matematik ya da matematik öğretmenliği alanlarında olması ve öğretmenlik deneyimleri üç yıl ve üstü olarak belirlemişlerdir.

Goldhaber ve Brewer (1998), fen ve matematik öğretmenlerinin kendi alanlarında lisans ya da yüksek lisans derecesi tamamlamış olmasının öğrencilerinin bu alanlardaki test başarılarına olumlu etkisi olduğunu ifade etmişlerdir. Rowan, Chiang ve Miller'in (1997) sadece matematik öğretmenlerinin matematik alanında derece sahibi olmasının öğrenci matematik başarısına etkisini inceledikleri araştırmada benzer bir sonuç elde etmişlerdir.

Goldhaber (2002) makalesinde, öğretmenlerin akademik hazırlığı (eğitim ve kurslar) kapsamında gerçekleştirilen araştırmaların en azından matematik ve fen alanları için öğrenci başarısını olumlu etkileyen sonuçlara yol açtığını savunmuştur. Matematik ve fen dışında başka bir alanda sahip olunan yükseköğretim derecesinin ise öğrenci başarısını etkilemediğini de bulgularına eklemiştir.

Decker, Mayer ve Glazerman'ın (2004) ve Kane, Rockoff ve Staiger'in (2006) araştırmaları sertifikasız öğretmenin öğrenci kazanımlarında önemli ya da sistematik bir dezavantaja yol açmadığına işaret etmektedir. Diğer taraftan Clotfelter, Ladd, ve Vigdor'un (2007), Darling-Hammond, Berry, ve Thoreson'ın (2001), Darling-Hammond, Holtzman, Gatlin, ve Heilig'in (2005) ve Goldhaber ve Brewer'in (2000) araştırmaları öğretmenlik sertifikası bulunan öğretmenlerin bulunmayanlara oranla öğrenme kazanımını daha yüksek oranda teşvik ettiklerini belirtmiştir. Goldhaber ve Anthony'nin (2004) boylamsal araştırmasında, sertifikası olmayan öğretmenlerin öğrencilerinin başarısının, sertifika aldıktan sonra arttuğunu tespit etmiştir.

Eberts ve Stone'un (1984), Murnane ve Phillips'in (1975), Darling-Hammond'un (2000) ve Rowan, Correnti ve Miller'in (2002) araştırmaları öğretmenin alanında veya başka bir alanda yüksek lisans derecesine sahip olmasının öğrenci başarısı üzerinde olumsuz bir etki meydana getirdiği sonucuna ulaşmıştır. Benzer bir çalışmada, Goldhaber ve Brewer (1998, 2000), Amerika'da ulusal verilere dayanarak gerçekleştirdiği araştırmada öğretmenlerin matematik alanında yüksek lisansa sahip olmasının lise öğrencilerinin matematik başarısını arttırdığına yönelik sonuçlar elde etmiştir. Bu çalışma ile öğretmenin sahip olduğu en yüksek eğitim düzeyinin değil, yüksek lisans ya da doktora derecelerinin hangi alanda olduğunun önemini ortaya koymaya çalışılmıştır.

Bu araştırma, 2011 yılında gerçekleştirilen Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS 2011)'e katılan Türkiye ile beraber 10 ülke verisi kullanılarak gerçekleştirilmiştir. Ülkeler TIMSS araştırmasına; matematik ve fen öğretim programlarının okullardaki uygulama süreçlerinin etkililiği belirlemek, uluslararası düzeyde kendi performanslarını görmek ve verilen eğitimin gelişimine yönelik olarak diğer ülke uygulamaları ile karşılaştırmalar yapabilmek amacı ile katılmaktadır. TIMSS sonuçlarına göre ülkeler, eğitim sistemlerinin işleyişinde yaşanan temel problemleri belirlemekte ve kaliteli eğitim için yapılması gereken politika, strateji ve planları somut kanıtlara dayalı ve diğer ülkelerin iyi uygulamaları çerçevesinde analiz ederek kendi sistemlerine adapte edebilmektedirler.

Bu çalışmada TIMSS 2011'e göre ilköğretim sekizinci sınıf öğrencilerinin fen ve matematik alanlarındaki akademik başarısı ile fen ve matematik öğretmenlerinin yükseköğrenimde almış oldukları alan eğitimleri arasında nasıl bir ilişki olduğu incelenmiştir. Bu çerçevede öğretmenlerin yükseköğrenim alanlarının, öğrencilerin akademik başarısına etkisi Hiyerarşik Lineer Modelleme (HLM) yöntemi ile analiz edilerek belirlenmeye çalışılmıştır. Çalışmaya dahil olan öğretmenlerin alan eğitimlerinde hangi dersleri almış oldukları Ek 1'de yer almaktadır. Öğretmenlerin Bu çerçevede, öğretmenlerin yükseköğrenimlerinde almış oldukları alan eğitiminin öğrenci başarısı ile ilişkisinin ve etki derecesinin belirlenmesi, eğitim politikacılarının ve yapımcılarının öğretmenlere yönelik hizmet öncesi ve hizmetiçi mesleki gelişimlerinde uygulanacak programlar için belirleyici bir kaynak ve döküman olacaktır.

Araştırma kapsamında "Fen bilgisi ve matematik öğretmenlerinin yükseköğrenimde almış oldukları alan eğitimleri öğrencilerin akademik anlamda fen ve matematik başarılarında nasıl bir değişime yol açmaktadır?" sorusuna yanıt aranmıştır.

YÖNTEM

Çalışmanın amacı TIMSS 2011 uygulamasında sekizinci sınıf öğrencilerinin fen ve matematik başarıları ile fen ve matematik öğretmenlerinin yükseköğrenimde almış oldukları alan eğitimleri arasında nasıl bir ilişki olduğunu incelemek ve karşılaştırmalar yapmaktır. Bu çerçevede matematik ve fen bilgisi öğretmenlerinin yükseköğrenimde almış oldukları alan eğitiminin öğrencilerin fen ve matematik başarılarına etkisi Hiyerarşik Lineer Modelleme (HLM) yöntemi ile analiz edilerek belirlenmeye çalışılmıştır. Bu çalışma betimsel modelde tasarlanmış nicel bir çalışmadır.

Geniş ölçekli ölçme ve değerlendirme çalışmalarından birisi olan Uluslararası Fen ve Matematik Eğilimleri Araştırması-TIMSS (Trends in International Mathematics and Science Study) ilk defa 1995 yılında gerçekleştirilmiştir. IEA (International Association for the Evaluation of Educational

Achievement) (<http://www.iea.nl>) tarafından yürütülen TIMSS çalışması, her 4 yılda bir yapılmakta olup; 4. ve 8. sınıf düzeyindeki öğrencilerin matematik ve fen alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesi amaçlanmaktadır.

Araştırma sorusu kapsamında TIMSS 2011 fen ve matematik öğretmeni anket uygulamasında; “Yüksek öğreniminde hangi ana dal ya da alanlarda eğitim aldınız?” sorusu öğretmenlere yöneltilmiştir. Bu soru maddesi öğretmen anketinde Şekil 1’de sunulduğu gibi yer almaktadır.

Yüksek öğreniminde hangi anadal ya da alanlarda eğitim aldınız?

Her satırda **bir** daireyi işaretleyiniz.

| | Evet | Hayır |
|---------------------------------------|-----------------------|-----------------------|
| a) Matematik | <input type="radio"/> | <input type="radio"/> |
| b) Biyoloji | <input type="radio"/> | <input type="radio"/> |
| c) Fizik | <input type="radio"/> | <input type="radio"/> |
| d) Kimya | <input type="radio"/> | <input type="radio"/> |
| e) Yer Bilimleri | <input type="radio"/> | <input type="radio"/> |
| f) Matematik Öğretmenliği | <input type="radio"/> | <input type="radio"/> |
| g) Fen Bilgisi/Bilimleri Öğretmenliği | <input type="radio"/> | <input type="radio"/> |
| h) Eğitim (Genel) | <input type="radio"/> | <input type="radio"/> |
| i) Diğer | <input type="radio"/> | <input type="radio"/> |

Şekil 1. TIMSS 2011 Öğretmen anketinde yer alan araştırma maddesi

Şekil 1’de yer alan sorulara verilen öğretmen cevapları ikinci düzey öğretmen değişkenleri olarak araştırmada incelenmiştir. Bu araştırmanın amacı öğretmen ve öğrenci düzeyi değişkenleri arasındaki ilişkiyi incelemek olmadığı için bu araştırmada her iki düzeyden değişkenleri karıştıran bir hiyerarşik modelleme gerçekleştirilmemiştir. Modelin oluşturulabilmesi için birinci düzey değişken olarak öğrenci anketinden dört tane öğrenci özelliği araştırma kapsamında incelenmiştir. Bu değişkenler; öğrencinin cinsiyeti, evde bulunan kitap sayısı, evde bilgisayar olma durumu ve evde internet olma durumu olarak seçilmiştir. Bu değişkenler öğrencilerin genel olarak sahip oldukları ve somut olarak doğruluğu açıklanabilen değişkenler olduğu için seçilmiştir.

Örneklem

Araştırmanın örneklemini Tablo 1’de yer alan 10 ülkenin ilköğretim sekizinci sınıf öğrencileri ve fen ve matematik öğretmenlerinden oluşmaktadır.

Tablo 1. Çalışmanın Örneklemini Oluşturan Ülkelerin Başarı Puanları, Öğretmen Sayıları ve Öğrenci Sayıları

| | Fen Başarı Puanı | Matematik Başarı Puanı | Fen Bilgisi Öğretmen Sayısı | Matematik Öğretmen Sayısı | Örneklem Öğrenci Sayısı |
|------------|------------------|------------------------|-----------------------------|---------------------------|-------------------------|
| Singapur | 590 | 611 | 165 | 165 | 5.927 |
| G. Kore | 560 | 613 | 150 | 150 | 5.166 |
| Japonya | 558 | 570 | 138 | 138 | 4.414 |
| Çin-Tayvan | 564 | 609 | 150 | 150 | 5.042 |
| Finlandiya | 552 | 514 | 145 | 145 | 4.266 |
| Slovenya | 543 | 505 | 186 | 186 | 4.415 |
| İngiltere | 533 | 507 | 118 | 118 | 3.842 |
| Türkiye | 483 | 452 | 239 | 239 | 6.928 |
| Malezya | 426 | 440 | 180 | 180 | 5.733 |
| Makedonya | 407 | 426 | 150 | 150 | 4.062 |

Raudenbush ve Bryk'in savunduğu gibi (2002) Hiyerarşik Lineer Modelleme (HLM) çoklu analiz yöntemi regresyon katsayılarının standart hatalarını yansız kestirebilmesi ve kümelenmiş grupların birbirine etkisini açıklayabilmesi nedeniyle, TIMSS 2011 verilerinin yapısına uygundur. TIMSS 2011 uygulamasında öğrenciler sınıf/öğretmen içinde, sınıflar/öğretmenler okul içinde ve okullar ülke içinde iç içe geçmiş yapıya sahiptir. Bunun gibi katmanlı yapı özelliği gösteren çalışmalarda belirlenen değişkenlerce diğer değişkenler açıklanabilir (Arnold, 1992).

Veri Toplama Yöntemleri

Araştırmada, IEA'nın gerçekleştirdiği TIMSS 2011 uygulamasına katılan sekizinci sınıf öğrencilerine uygulanan fen ve matematik başarı testinin sonuçları, öğrenci anketleri, fen ve matematik öğretmeni anketlerine verilen cevaplar veri olarak kullanılmıştır. Bu veriler IEA'nın veri merkezi Data Processing Center'den elde edilmiştir (rms.iea-dpc.org).

Çalışmaya kaynaklık eden TIMSS 2011 uygulamasında; katılımcı her ülkeden öncelikle ülke evrenini temsil edecek belirli sayıda okul seçilmekte, sonrasında ise her okuldan tesadüfi yöntemle bir ve/ya iki şube belirlenmektedir. Belirlenen şubedeki tüm öğrencilere başarı testi ve öğrenci anketi; şubelerin matematik ve fen bilgisi öğretmenlerine yönelik öğretmen anketleri ve seçilen okulun yöneticisine ise okul anketi uygulanmaktadır.

Verilerin Analizi

Bu araştırmada verilerin analizi, SPSS programı ve Raudenbush, Bryk ve Congdon (2004) tarafından geliştirilen HLM 7.00 programı ile yapılmıştır. Hiyerarşik Lineer Modelleme (HLM) kümelenmiş verilerin çok katmanlı veri yapılarını belirgin bir şekilde açıklamaktadır. Dolayısıyla regresyon katsayılarını yansız bir şekilde hesaplayabilmektedir (Bryk ve Raudenbush, 1992). TIMSS'te kullanılan örnekleme yöntemi ve oluşturulan veri seti iç içe geçtiği için TIMSS'in veri yapısı hiyerarşik, katmanlı bir özellik göstermektedir. Yani öğrenciler sınıf/öğretmen, öğretmenler ise okulun içinde olması nedeniyle verilerin analizinde Hiyerarşik Lineer Modelleme tekniği tercih edilmiştir.

Çalışmada fen bilgisi ve matematik öğretmenleri için ayrı ayrı oluşturulan iki düzeyli Hiyerarşik Lineer Modellemede (HLM) aşağıdaki üç aşama izlenmiştir: İlk olarak öğretmen ve öğrenci düzeyinin yer aldığı boş model (Kestiricisiz ANOVA modeli) oluşturulmuştur. Bu model 1. düzeyde öğrenci ve 2. düzeyde öğretmen düzeyi olmak üzere iki alt model içermiştir. Bu modelde elde edilen birinci düzey varyans değerleri ve ikinci düzey varyans değerleri ile öğrenci başarılarının ne kadarının öğrenciler arası farklılıklardan, ne kadarının öğretmenler arası farklılıklardan meydana geldiği tespit edilmeye çalışılmıştır. İkinci olarak modelde yalnızca 1. düzey öğrenci değişkenlerinin yer aldığı rastgele etkiler modeli kurulmuştur. Bu modelde her bir öğrencinin fen başarısı modelin 1. düzeyinde yer alan öğrenci değişkenleri; öğrencinin cinsiyeti, evde bulunan kitap sayısı, evde bilgisayar olma durumu ve evde internet olma durumu fonksiyonları olarak kestirilmeye çalışılmış ve öğrencilerin başarı puanlarına etkisi tespit edilmeye çalışılmıştır. Birinci düzey varyans değerlerine bakılarak öğrenci başarısının ne kadarının birinci düzey değişkenlere göre açıklanabildiği oranına bakılmıştır. Son olarak modelde yalnızca 2. düzey öğretmen değişkenlerinin yer aldığı koşullu model kurulmuştur. Koşullu modelde ise öğretmen özelliklerinin öğrenci başarısına etkisinin fonksiyonu için, 1. düzeyde alınan değişkenler modelden çıkarılmış ve 2. düzeyde ele alınan öğretmen değişkenleri kullanılarak öğretmen değişkenlerinin (öğretmenlerin yükseköğrenimde aldıkları alan eğitiminin) öğrenci başarısına etkisi açıklanmaya çalışılmış ve öğrencilerin başarı puanlarına etkisi tespit edilmeye çalışılmıştır. İkinci düzey varyans değerlerine bakılarak öğrenci başarısının ne kadarının ikinci düzey değişkenlere göre açıklanabildiği oranına bakılmıştır.

Boş Model (Null Model: One-way Random Effects ANOVA Model in HLM)

Boş modele göre, her bireyin bağımlı değişken üzerindeki puanı 3 elemandan oluşmaktadır: genel ortalama (γ_{00}), küme ortalamasının genel ortalamadan sapması (u_{0j}), ve bireyin puanının ait olduğu küme ortalaması hata payı (r_{ij}). Küme ortalamasının genel ortalama hata payı (u_{0j}) değeri, j kümesi içindeki tüm gözlemler için aynı olduğundan, aynı kümeden gözlemlerin bağımlılığının modellenmesini sağlar (Raudenbush ve Bryk, 2002). Yani birinci seviye modelindeki (β_{1j}) sınıfın fen başarı ortalaması tüm öğretmenler/sınıflar için sifira eşitlenmiştir. Boş modelin birinci seviyedeki eşitliği aşağıdaki gibidir:

Düzy-1 (Öğrenci Düzeyi): $Y_{ij}(FENBAS_{ij}) = \beta_{0j} + r_{ij}$

Birinci seviye modelindeki öğrenci düzeyi her hatanın r_{ij} , ortalaması sıfır olacak şekilde sabitlenir ve birinci seviye varyansının " σ^2 " olduğu varsayılarak normal olarak dağıldığı kabul edilir. Bu model her birinci seviye birimi çıktısını sadece bir ikinci seviye parametresi tarafından kestirir ki bu araştırma için bu parametre β_{0j} , öğretmenin sınıfının tahmin edilen ortalama fen başarı puanıdır. Bu durumda tesadüfi etkili tek yönlü ANOVA modelinin ikinci seviyedeki eşitliği aşağıdaki gibi bulunur:

Düzy-2 (Öğretmen Düzeyi): $\beta_{0j} = \gamma_{00} + u_{0j}$

Bu eşitlikte γ_{00} tüm sınıflar dolayısıyla tüm öğretmenler için fen başarılarının genel ortalamasıdır, ikinci seviye hata terimi olarak adlandırılan u_{0j} , sınıfın hata puanı ise öğretmen birimiyle ilişkilendirilen ortalamasının ve varyansının sıfır olduğu varsayılan tesadüfi etki olarak varyansının " τ_{00} " olduğu varsayılır. Yukarıdaki ANOVA modeli birinci ve ikinci seviye eşitlikler birleştirilince aşağıdaki birleştirilmiş tek eşitlik denklemi elde edilir:

Birleştirilmiş Model: $Y_{ij}(FENBAS_{ij}) = \gamma_{00} + u_{0j} + r_{ij}$

Burada,

- Y_{ij} : her bir öğrencinin başarı puanını,
- β_{0j} : j. sınıfın fen başarı ortalamasını,
- r_{ij} : j. okuldaki i. öğrencinin hata puanını,
- γ_{00} : sınıfların fen başarı puanı ortalamalarını,
- u_{0j} : j. sınıftaki hata puanını göstermektedir.

Denklemden, γ_{00} sınıfların fen başarı puanı ortalamasını, u_{0j} 2. düzey etkili ve r_{ij} birey etkili tek yönlü ANOVA modelini oluşturur. Bu bir tesadüfi etki modelidir çünkü grup etkileri tesadüfi olarak yorumlanmıştır ve modelin çıktısının varyansı aşağıdaki şekilde ifade edilmiştir:

$$\text{Var}(Y_{ij}) = \text{Var}(u_{0j} + r_{ij}) = \sigma^2 + \tau_{00}$$

Yukarıda açıklandığı gibi bu model Y 'deki varyansı açıklamaz, ancak bu varyansı alt seviye hataları varyansı " σ^2 " ve üst seviye hatalarının varyansı " τ_{00} " olarak ikiye böler (Raudenbush ve Bryk, 2002). Bu değerler kullanılarak da sınıflar arası ve sınıf içi korelasyon katsayısı " ρ " aşağıdaki formüllerde belirtildiği gibi hesaplanabilir:

$$\rho(\text{sınıflar arası}) = \tau_{00} / (\tau_{00} + \sigma^2)$$

$$\rho(\text{sınıf içi}) = \sigma^2 / (\sigma^2 + \tau_{00})$$

Seçkisiz Etkili Tek yönlü ANOVA modelinde gruplar içi ve gruplar arası korelasyon katsayısı (ρ) hesaplanarak sonuç üzerindeki varyansın ne kadarının birinci ve ikinci düzeyden kaynaklandığı belirlenir (De Leeuw ve Kreft, 1986).

Rastgele Etkiler Modeli (Ortalamaların Bağımlı Değişken Olduğu Model)

Araştırma sorusunun öğrenci değişkenlerine (1. düzey) cevap verecek olan Rastgele Etkiler modeli, diğer bir adı ile Seçkisiz Katsayılar Regrasyon modeli kurulmuştur. Bu model için HLM 7.00 programıyla sadece 1. düzey değişkenlerinin dâhil edildiği 2. model kurulmuştur. Seçkisiz Katsayılar

Regrasyon modellerinde alt modellerin hepsi sabit parametresi seçkisiz olarak değişen modeller varsayımı ile ele alınır. Modelde sabit ve eğim parametresini açıklayan 2. düzey bağımsız değişkenleri bulunmaz. Burada amaç araştırma desenine dâhil edilen kontrol değişkenlerinin öğrenci fen başarısı varyansını ne ölçüde etkilediğini kestirmektir.

$$\text{Düzyey-1: } FENBAS_{ij} = \beta_{0j} + \beta_{1j}*(CİNS_{ij}) + \beta_{2j}*(KİTAP_{ij}) + \beta_{3j}*(BİLG_{ij}) + \beta_{4j}*(İNER_{ij}) + r_{ij}$$

$$\text{Düzyey-2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

Birleştirilmiş Model:

$$FENBAS_{ij} = \gamma_{00} + \gamma_{10}*CİNS_{ij} + \gamma_{20}*KİTAP_{ij} + \gamma_{30}*BİLG_{ij} + \gamma_{40}*İNER_{ij} + u_{0j} + r_{ij}$$

Burada,

- β_{1j} : j. öğretmen sınıfındaki öğrencilerinin annesinin eğitim düzeyi değişkenin katsayısını,
- β_{2j} : j. öğretmen sınıfındaki öğrencilerinin babasının eğitim düzeyi değişkenin katsayısını,
- β_{3j} : j. öğretmen sınıfındaki öğrencilerin eğitim hedefi değişkenin katsayısını,
- β_{4j} : j. öğretmen sınıfındaki öğrencilerin okuldaki dersini ailesi ile paylaşım değişkenin katsayısını,
- γ_{00} : her bir sınıfın ortalama kestirilen başarı puanını,
- γ_{10} : j. sınıfta fen başarısına CİNS (Cinsiyet) etkisini,
- γ_{20} : j. sınıfta fen başarısına KİTAP (Kitap Sayısı) etkisini,
- γ_{30} : j. sınıfta fen başarısına BİLG (Evde Bilgisayar) etkisini,
- γ_{40} : j. sınıfta fen başarısına İNER (Evde İnternet) etkisini,
- u_{0j} : sınıf düzeyi değişkenlerin hata puanını göstermektedir.

Rastgele Etki Modelinin oluşturulmasında öğrenci düzeyindeki değişkenlerin modele tek tek alınarak çözümlenmesi önerilmektedir. Çünkü değişkenlerin modele tek tek alınması, birleştirilmiş modelin doğru oluşturulmasını sağlamaktadır (Raudenbush ve Bryk, 2002).

Koşullu Model (Ortalamaların Bağımlı Değişken Olduğu Model)

Araştırma sorusunun öğretmen değişkenlerine (2. düzey) cevap verecek olan Koşullu Model, diğer bir adı ile Ortalamaların Bağımlı Değişken Olduğu model kurulmuştur. Bu aşamada kurulan modellerin hepsi yalnızca 2. düzey yani öğretmen seviyesi değişkenlerini içermiştir. Bu modelde 2. düzey değişkenleri kullanılarak tahminler yapılır. Regresyon modeli, grup ortalamalarının 2. düzey değişkenleri tarafından açıklanması ile oluşur.

Koşullu model ‘Ortalamaların Bağımlı Değişken Olduğu’ modeli oluşturulurken araştırmadaki veri çözümlenme basamaklarının ve bulgularına ilişkin yorumların açık ve düzenli olması için öğretmen düzeyi değişkenleri üç kategori altında sırasıyla çözümlenmeye alınmıştır. Araştırma kapsamında bu modelin 1. düzeyi boş modelin ilk aşaması gibi kurulmuştur. İkinci düzeydeki öğretmen özellikleri ile ilgili değişkenler üç gruba ayrılmış ve üç farklı HLM modellemesinde ele alınmışlardır. Bunlardan bir tanesi aşağıdaki gibidir.

$$\text{Düzyey-1: } (FENBAS_{ij}) = \beta_{0j} + r_{ij}$$

$$\text{Düzyey-2: } \beta_{0j} = \gamma_{00} + \gamma_{01}*(BTBG05A_j) + \gamma_{02}*(BTBG05B_j) + \gamma_{03}*(BTBG05C_j) + u_{0j}$$

$$\text{Birleştirilmiş Model: } FENBAS_{ij} = \gamma_{00} + \gamma_{01}*BTBG05A_j + \gamma_{02}*BTBG05B_j + \gamma_{03}*BTBG05C_j + u_{0j} + r_{ij}$$

Burada,

- γ_{00} : her bir sınıfın ortalama kestirilen başarı puanını,

γ_{01} : sınıfın ortalama fen başarısına BTBG05A'ın (Matematik) etkisini,
 γ_{02} : sınıfın ortalama fen başarısına BTBG05B'in (Biyoloji) etkisini,
 γ_{03} : sınıfın ortalama fen başarısına BTBG05C'in (Fizik) etkisini,
 u_{0j} : sınıf düzeyi değişkenlerin hata puanını göstermektedir.

BULGULAR

Bu başlık altında yapılan analiz sonuçları “Fen Bilgisi Öğretmenlerinin Öğrencilerin Fen Başarısına Etkisi” ve “Matematik Öğretmenlerinin Öğrencilerin Matematik Başarısına Etkisi” olmak üzere iki başlık altında sunulmuştur.

Çalışmada incelenen fen ve matematik öğretmen özelliklerinin yüzde dağılımları ve oluşturulan koşullu modelin HLM analiz sonuçları “Ekler” kısmında yer almaktadır.

Fen Bilgisi Öğretmenlerinin Öğrencilerin Fen Başarısına Etkisi

Çalışmada ilk olarak kurulan HLM'nin boş model (etkisiz ANOVA) analizi sonrası ülkelerin öğrenci fen başarı puanlarını öğrenci düzeyi ve öğretmen düzeyi değişkenlerin açıklama oranlarına bakılmış bu bilgilere Tablo 2'de yer verilmiştir.

Tablo 2. Ülkelerin Fen Başarı Puanlarının Öğretmen ve Öğrenci Düzeyinde Açıklanma Oranları

| | Öğrenci Fen Başarı Puanı Ortalaması | Öğretmenler arası farklılıklar (%) | Öğretmen düzeyi açıklanan varyans (%) | Öğrenciler arası farklılıklar (%) | Öğrenci düzeyi açıklanan varyans (%) |
|-------------------|-------------------------------------|------------------------------------|---------------------------------------|-----------------------------------|--------------------------------------|
| Singapur | 590 | 42 | 10 | 58 | 9 |
| G. Kore | 560 | 6 | 26 | 94 | 14 |
| Japonya | 558 | 8 | 6 | 92 | 9 |
| Çin-Tayvan | 564 | 18 | 8 | 82 | 12 |
| Finlandiya | 552 | 12 | 6 | 88 | 11 |
| Slovenya | 543 | 10 | 4 | 90 | 14 |
| İngiltere | 533 | 55 | 13 | 65 | 12 |
| Türkiye | 483 | 29 | 11 | 71 | 7 |
| Malezya | 426 | 69 | 12 | 31 | 2 |
| Makedonya | 407 | 39 | 19 | 61 | 7 |

Araştırmaya dâhil edilen ülkelerin öğrenci ortalama fen başarı puanlarına göre, Singapur 590 puanla en yüksek fen başarı puanına sahipken, Makedonya 407 puanla en düşük fen başarı puanına sahiptir (Tablo 2).

Yapılan HLM analizinin birinci aşamasında oluşturulan boş modelde birinci düzey ve ikinci düzey değişkenlerin varyans değişkenleri arasında yapılan hesaplamalar ile Malezya'da öğrencilerin fen başarı puanlarının %69'unu öğretmenler arası farklılıklar, %31'ini ise öğrenciler arası farklılıklar ile açıklandığı bulunmuştur. G. Kore'de ise öğrencilerin fen başarı puanlarının %94'ünü öğrenciler arası farklılıklar, %6'sını ise öğretmenler arası farklılıklar ile açıklandığı bulunmuştur (Tablo 2).

İlk olarak kurulan boş modelden elde edilen varyans değerleri ve son kurulan koşullu modelden elde edilen varyans değerleri arasında yapılan hesaplamalar sonucunda analize dâhil edilen öğretmen değişkenlerinin öğrencilerin fen başarı puanlarının ne kadarını açıkladığı bulunmuştur. Bu hesaplama yöntemi ile araştırma kapsamında yer alan ülkelerin öğretmen düzeyi açıklanan varyans

değerlerine bakıldığında en yüksek açıklanma oranı %26 ile G. Kore, en az açıklanma oranı %6 ile Japonya ile Finlandiya'ya aittir.

Çalışmada ikinci olarak kurulan HLM'nin rastgele etkiler modeli ile elde edilen birinci düzey öğrenci özelliklerinin öğrencilerin fen başarı puanlarına aritmetik etkisi Tablo 3'de verilmiştir.

Tablo 3. Birinci Düzey Öğrenci Değişkenlerinin Öğrencilerin Fen Başarı Puanlarına Etkisi

| | Evde bulunan kitap Sayısı (25 kitap üzeri, %) | Evde bilgisayar olma durumu (%) | Evde internet olma durumu (%) |
|------------|--|------------------------------------|----------------------------------|
| Singapur | 15*** (59) | 17* (97) | 44*** (96) |
| G. Kore | 22*** (82) | -37* (99) | 51*** (97) |
| Japonya | 16*** (63) | 16** (90) | 9* (86) |
| Çin-Tayvan | 21*** (65) | 12 (97) | 8* (93) |
| Finlandiya | 18*** (76) | 11 (100) | -3 (99) |
| Slovenya | 24*** (62) | 46** (99) | 26** (98) |
| İngiltere | 17*** (62) | 14 (99) | 10 (98) |
| Türkiye | 20*** (44) | 4 (58) | 9* (45) |
| Malezya | 7*** (32) | -2 (62) | 5* (46) |
| Makedonya | 17*** (38) | 24** (90) | 21*** (81) |

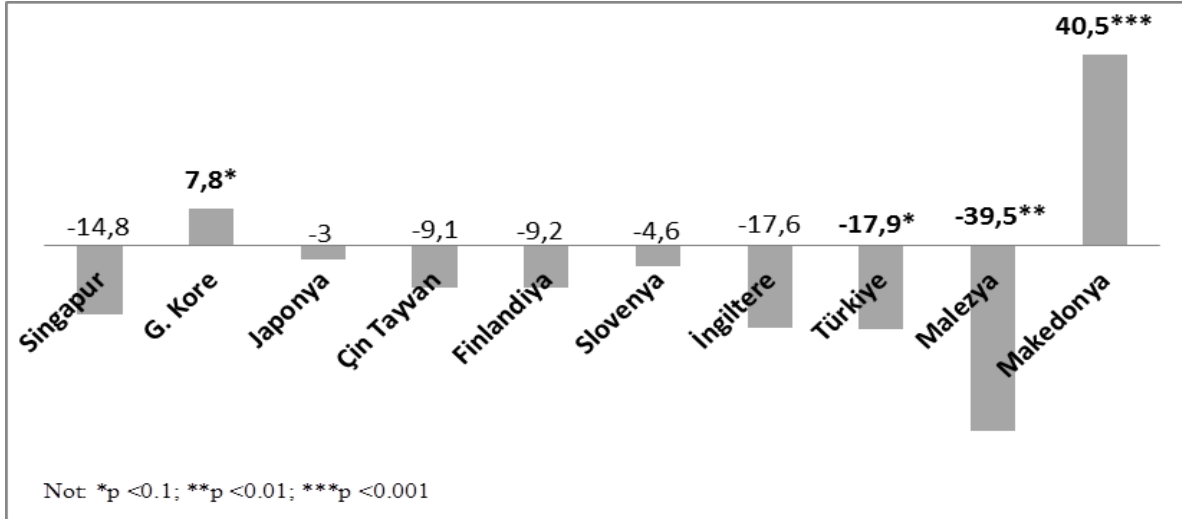
Not: *p <0.1; **p <0.01; ***p <0.001

Öğrencilerin Türkiye'de %44'ünün, Malezya'da %32'sinin ve Makedonya'da %38'inin evinde 25 kitaptan daha fazla kitap bulunmaktadır. Diğer ülkelerde ise bu oran %60'ın üzerindedir. Araştırmada yer alan ülkelerde öğrencilerin evinde bulunan kitap sayısı öğrencilerin fen başarı puanlarını istatistiksel olarak anlamlı (p <0.1) bir şekilde etkilemektedir. Öğrencilerin evinde bulunan kitap sayısındaki bir birimlik artış öğrenci fen başarı puanını Slovenya'da 24 puan artırırken G. Kore'de 22 puan ve Türkiye'de 20 puan artırmıştır (Tablo 3).

Türkiye ve Malezya dışındaki ülkelerde öğrencilerin %90'ının üzerinde öğrencilerin evinde bilgisayar bulunmaktadır. Öğrencilerin evinde bulunan bilgisayar durumuna göre öğrencilerin fen başarı puanına etkisi ülkeler arasında farklılık göstermektedir. Japonya, Slovenya ve Makedonya'da istatistiksel olarak anlamlı (p <0.01) ve pozitif bir etki meydana getirirken, G. Kore'de istatistiksel olarak anlamlı (p <0.001) fakat negatif bir etki göstermiştir.

Öğrencilerin Türkiye'de %45'inin ve Malezya'da %46'sının evinde internet erişimi bulunmaktadır. Diğer ülkelerde ise bu oran %80'in üzerindedir. Öğrencilerin evinde internet bulunma durumuna göre öğrencilerin fen başarı puanına etkisi ülkeler arasında farklılık göstermektedir. Singapur, G. Kore ve Makedonya'da öğrencilerin fen başarı puanına etkisi istatistiksel olarak anlamlı (p <0.1) bulunmuştur. Japonya, Slovenya ve Makedonya'da istatistiksel olarak anlamlı (p <0.01) ve pozitif bir etki meydana getirirken, G. Kore'de (p <0.001) istatistiksel olarak anlamlı fakat negatif bir etki göstermiştir (Tablo 3).

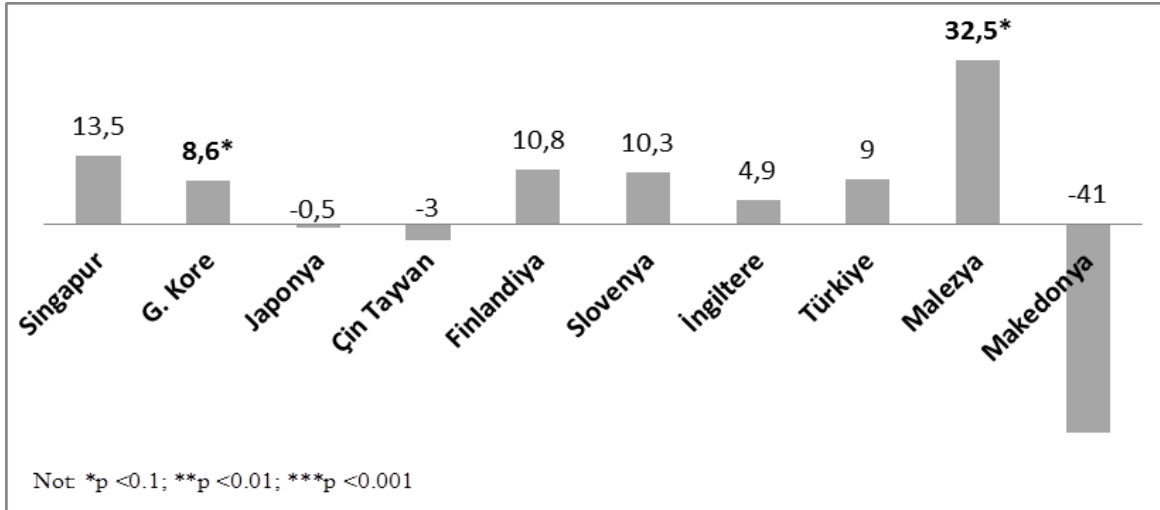
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden fen bilgisi öğretmenlerinin fen bilgisi/bilimleri alanında aldığı eğitimlerin öğrencilerin ortalama fen başarı puanlarına aritmetik etkisi Şekil 2'de yer almaktadır.



Şekil 2. Fen Bilgisi Öğretmenlerinin Yükseköğretiminde Fen Bilgisi/Bilimleri Öğretmenliği Alanında Aldığı Eğitimlerin Öğrencilerin Fen Başarısına Etkisi

Fen öğretmenlerinin yükseköğretimlerinde fen bilgisi/bilimleri öğretmenliği alanında almış oldukları eğitimin öğrencilerin ortalama fen başarı puanlarını G. Kore ve Makedonya dışındaki ülkelerde negatif olarak etkilediği görülmektedir (Şekil 2). Öğretmenlerin fen bilgisi/bilimleri öğretmenliği alanında aldığı eğitimin öğrencilerin fen başarılarına istatistiksel etkisi G. Kore, Makedonya, Türkiye ve Malezya'da anlamlı çıkmıştır. Öğretmenlerin fen bilgisi/bilimleri öğretmenliği alanında eğitim alması G. Kore'de öğrencilerin ortalama fen başarı puanını 7.8 puan artırırken, Türkiye'de ise 17.9 puan düşmesine sebep olmuştur.

Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden fen bilgisi öğretmenlerinin biyoloji alanında aldığı eğitimin öğrencilerin fen başarı puanlarına aritmetik etkisi Şekil 3'te yer almaktadır.

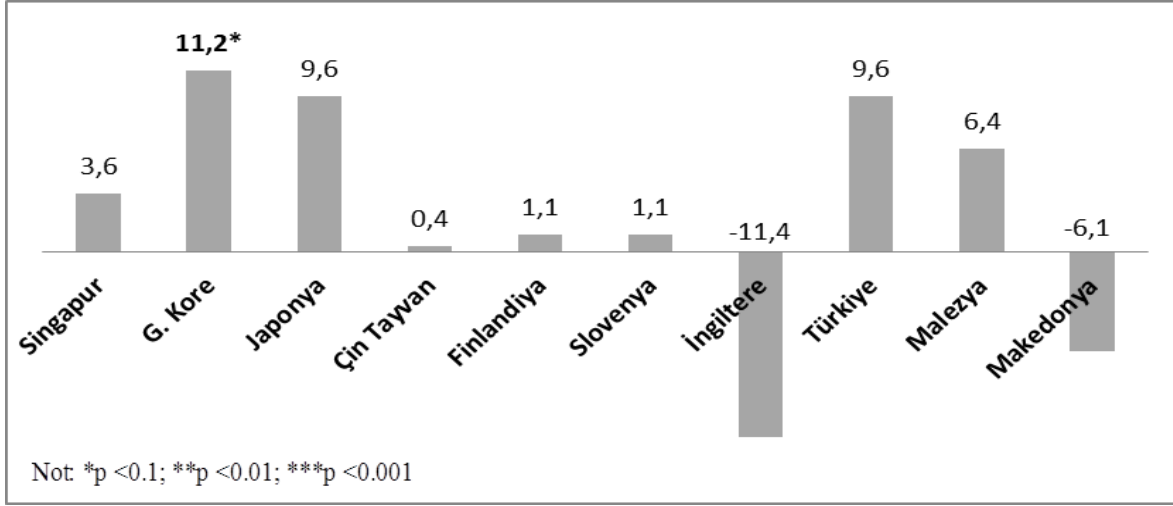


Şekil 3. Fen Bilgisi Öğretmenlerinin Yükseköğretiminde Biyoloji Alanında Aldığı Eğitimin Öğrencilerin Fen Başarısına Etkisi

Fen öğretmenlerinin yükseköğretimlerinde biyoloji alanında almış oldukları eğitimin öğrencilerin ortalama fen başarı puanlarını Makedonya dışındaki ülkelerde pozitif olarak etkilediği görülmektedir (Şekil 3). Öğretmenlerin biyoloji alanında aldığı eğitimin öğrencilerin fen başarılarına istatistiksel

etkisi G. Kore ve Malezya'da anlamlı çıkmıştır. Fen öğretmenlerinin biyoloji alanında eğitim alması Malezya'da öğrencilerin ortalama fen başarı puanını 32.5 puan artırırken, yalnızca Makedonya'da 41 puan düşmesine sebep olmuştur.

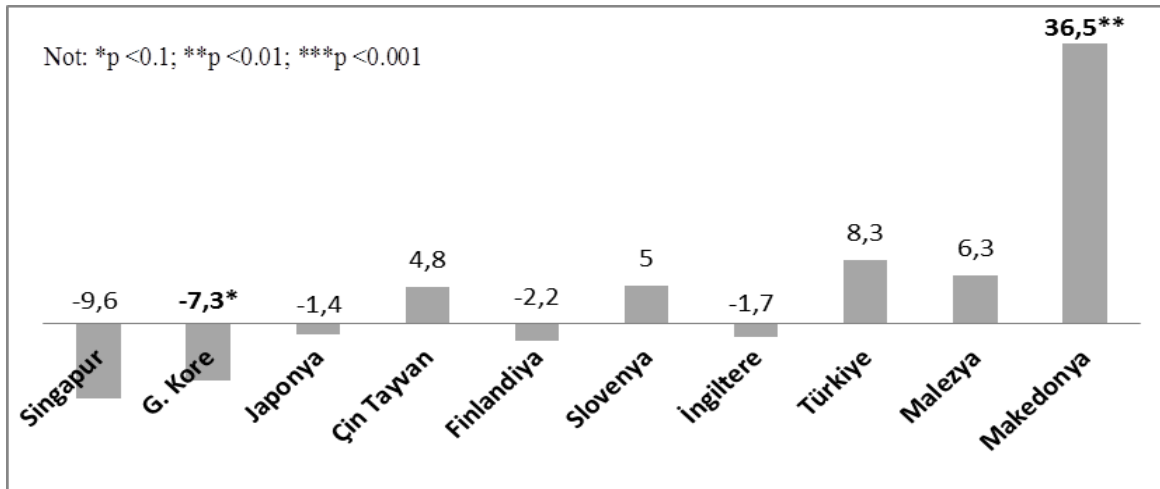
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden fen bilgisi öğretmenlerinin fizik alanında aldığı eğitimin öğrencilerin fen başarı puanlarına aritmetik etkisi Şekil 4'te yer almaktadır.



Şekil 4. Fen Bilgisi Öğretmenlerinin Yükseköğreniminde Fizik Alanında Aldığı Eğitimin Öğrencilerin Fen Başarısına Etkisi

Fen öğretmenlerinin yükseköğrenimlerinde fizik alanında almış oldukları eğitimin öğrencilerin fen başarı puanlarını İngiltere ve Makedonya dışındaki ülkelerde pozitif olarak etkilediği görülmektedir (Şekil 4). Öğretmenlerin fizik alanında aldığı eğitimin öğrencilerin fen başarılarına istatistiksel etkisi yalnızca G. Kore'de anlamlı çıkmıştır. Öğretmenlerin fizik alanında eğitim alması G. Kore'de öğrencilerin ortalama fen başarı puanını 11.2 puan artırırken, İngiltere'de ise 11.4 puan düşmesine sebep olmuştur.

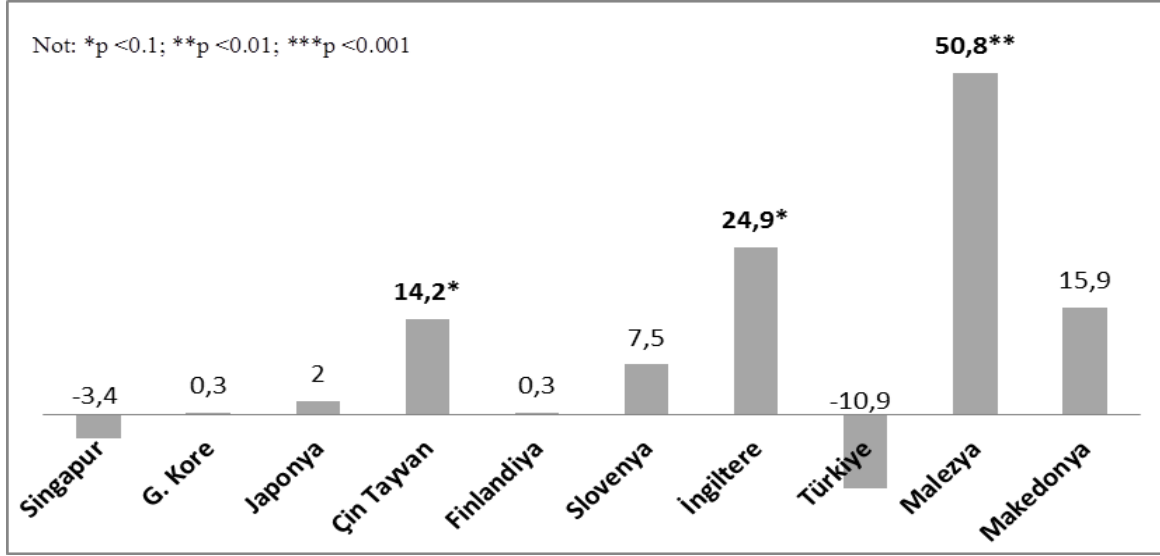
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden fen bilgisi öğretmenlerinin kimya alanında aldığı eğitimin öğrencilerin fen başarı puanlarına aritmetik etkisi Şekil 5'te yer almaktadır.



Şekil 5. Fen Bilgisi Öğretmenlerinin Yükseköğreniminde Kimya Alanında Aldığı Eğitimin Öğrencilerin Fen Başarısına Etkisi

Fen öğretmenlerinin yükseköğretimlerinde kimya alanında almış oldukları eğitimin öğrencilerin ortalama fen başarı puanlarına etkileri ülkeler arasında farklılık göstermektedir (Şekil 5). Öğretmenlerin kimya alanında aldığı eğitimin öğrencilerin fen başarılarına istatistiksel etkisi G. Kore ve Makedonya’da anlamlı çıkmıştır. Fen öğretmenlerinin kimya alanında eğitim alması Makedonya’da öğrencilerin ortalama fen başarı puanını 36.5 puan artırırken, Singapur’da ise 9.6 puan düşmesine sebep olmuştur.

Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden fen bilgisi öğretmenlerinin genel (eğitim) alanında aldığı eğitimin öğrencilerin fen başarı puanlarına aritmetik etkisi Şekil 6’da yer almaktadır.



Şekil 6. Fen Bilgisi Öğretmenlerinin Yükseköğretiminde Genel (Eğitim) Alanında Aldığı Eğitimin Öğrencilerin Fen Başarısına Etkisi

Fen öğretmenlerinin yükseköğretimlerinde genel (eğitim) alanında almış oldukları eğitimin öğrencilerin ortalama fen başarı puanlarını Türkiye ve Singapur dışındaki ülkelerde pozitif olarak etkilediği görülmektedir (Şekil 6). Öğretmenlerin genel (eğitim) alanında aldığı eğitimin öğrencilerin fen başarılarına istatistiksel etkisi yalnızca İngiltere’de anlamlı çıkmıştır. Öğretmenlerin genel (eğitim) alanında eğitim alması Malezya’da öğrencilerin ortalama fen başarı puanını 50.8 puan artırırken, Türkiye’de ise 10.9 puan düşmesine sebep olmuştur.

Matematik Öğretmenlerinin Öğrencilerin Matematik Başarısına Etkisi

Çalışmada ilk olarak kurulan HLM’nin boş model analizi sonrası ülkelerin öğrenci matematik başarı puanlarını öğrenci düzeyi ve öğretmen düzeyi değişkenlerin açıklama oranlarına bakılmış bu bilgilere Tablo 4’te yer verilmiştir.

Tablo 4. Ülkelerin Matematik Başarı Puanlarının Öğretmen ve Öğrenci Düzeyinde Açıklanma Oranları

| | Öğrenci Matematik Başarı Puanı Ortalaması | Öğretmenler arası farklılıklar (%) | Öğretmen düzeyi açıklanan varyans (%) | Öğrenciler arası farklılıklar (%) | Öğrenci düzeyi açıklanan varyans (%) |
|------------|---|------------------------------------|---------------------------------------|-----------------------------------|--------------------------------------|
| Singapur | 611 | 41 | 12 | 59 | 5 |
| G. Kore | 613 | 9 | 8 | 91 | 16 |
| Japonya | 570 | 13 | 3 | 87 | 8 |
| Çin-Tayvan | 609 | 23 | 7 | 77 | 9 |
| Finlandiya | 514 | 13 | 11 | 87 | 8 |
| Slovenya | 505 | 10 | 14 | 90 | 13 |
| İngiltere | 507 | 65 | 17 | 35 | 7 |
| Türkiye | 452 | 32 | 8 | 68 | 7 |
| Malezya | 440 | 68 | 5 | 32 | 3 |
| Makedonya | 426 | 34 | 14 | 66 | 6 |

Araştırmaya dâhil olan ülkelerin matematik başarı puanlarına göre, G. Kore 613 puanla en yüksek matematik başarı puanına sahipken, Makedonya 426 puanla en düşük matematik başarı puanına sahiptir (Tablo 4).

Yapılan HLM analizinin birinci aşamasında oluşturulan boş modelde birinci düzey ve ikinci düzey değişkenlerin varyans değerleri arasında yapılan hesaplamalar ile Malezya'da öğrencilerin matematik başarı puanlarının %68'ini öğretmenler arası farklılıklar, %32'sinin ise öğrenciler arası farklılıklar ile açıklandığı bulunmuştur. Yine G. Kore'de öğrencilerin fen başarı puanlarının %91'ini öğrenciler arası farklılıklar, %9'unu ise öğretmenler arası farklılıklar ile açıklandığı bulunmuştur (Tablo 4).

İlk olarak kurulan boş modelden elde edilen varyans değerleri ve son kurulan koşullu modelden elde edilen varyans değeri arasında yapılan hesaplamalar sonucunda analize dâhil edilen öğretmen değişkenleri ile öğrencilerin fen başarı puanlarının ne kadarının açıkladığı bulunmuştur. Bu hesaplama yöntemi ile araştırma kapsamında yer alan ülkelerin öğretmen düzeyi açıklanan varyans değerlerine bakıldığında ise en yüksek açıklanma oranı %17 ile İngiltere, en az açıklanma oranı %3 ile Japonya'ya aittir.

Bu hesaplama sonucunda, öğrenci düzeyi açıklanan varyans değerlerine bakıldığında, G. Kore ve Slovenya öğrencileri için çalışmaya dâhil edilen öğrenci faktörlerinin öğrenci matematik başarı puanı açıklanma oranının diğer ülkelere göre daha yüksek olduğu görülmektedir.

Çalışmada ikinci olarak kurulan HLM'nin rastgele etkiler modeli ile elde edilen birinci düzey öğrenci değişkenlerinin öğrencilerin matematik başarı puanlarına aritmetik etkisi Tablo 5'te verilmiştir.

Tablo 5. Birinci Düzey Öğrenci Değişkenlerinin Öğrencilerin Matematik Başarı Puanlarına Etkisi

| | Evde bulunan kitap Sayısı (25 kitap üzeri, %) | Evde bilgisayar olma durumu (%) | Evde internet olma durumu (%) |
|------------|---|---------------------------------|-------------------------------|
| Singapur | 9*** (59) | 21** (97) | 30*** (96) |
| G. Kore | 26*** (82) | -40** (99) | 75*** (97) |
| Japonya | 15*** (63) | 21*** (90) | 17** (86) |
| Çin-Tayvan | 21*** (65) | 23* (97) | 11* (93) |
| Finlandiya | 15*** (76) | 24 (100) | 17 (99) |
| Slovenya | 17* (62) | 34 (99) | 26* (98) |
| İngiltere | 11*** (62) | 25* (99) | 10 (98) |
| Türkiye | 23*** (44) | 6 (58) | 11** (45) |
| Malezya | 4*** (32) | 9** (62) | 12*** (46) |
| Makedonya | 13*** (38) | 28*** (90) | 26*** (81) |

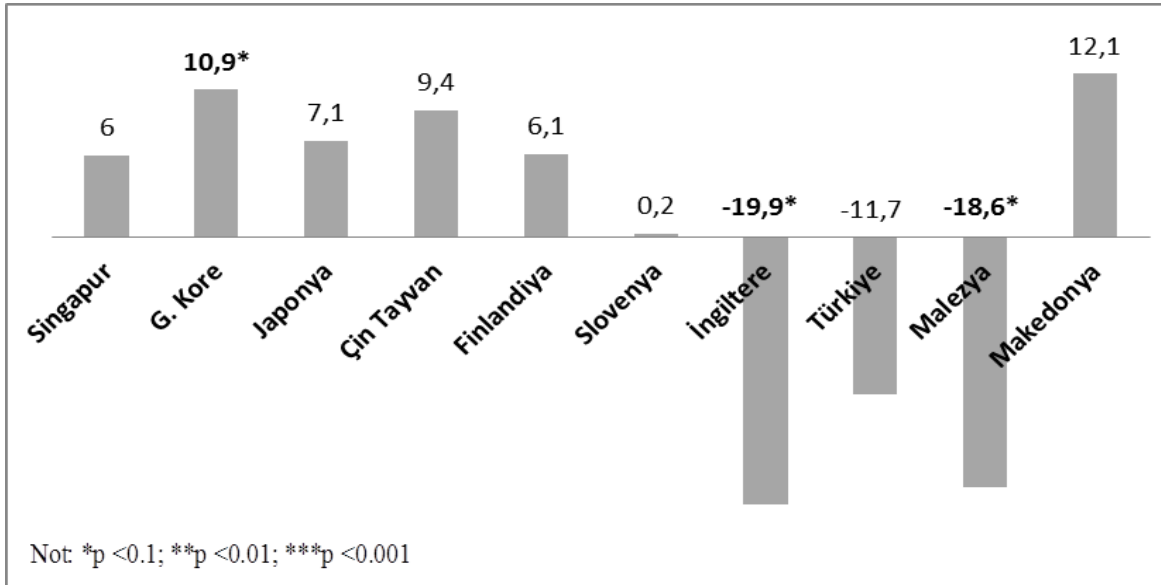
Not: *p < 0.1; **p < 0.01; ***p < 0.001

Öğrencilerin Türkiye’de %44’ünün, Malezya’da %32’sinin ve Makedonya’da %38’inin evinde 25 kitaptan daha fazla kitap bulunmaktadır. Diğer ülkelerde ise bu oran %60’ın üzerindedir. Araştırmada yer alan ülkelerde öğrencilerin evinde bulunan kitap sayısı öğrencilerin matematik başarı puanlarını istatistiksel olarak ($p < 0.1$) anlamlı bir şekilde etkilemektedir. Öğrencilerin evinde bulunan kitap sayısındaki bir birimlik artış öğrenci matematik başarı puanını G. Kore’de 26 puan arttırırken Türkiye’de 23 puan arttırmıştır (Tablo 5).

Türkiye ve Malezya dışındaki ülkelerde öğrencilerin %90’ının üzerinde öğrencilerin evinde bilgisayar bulunmaktadır. Öğrencilerin evinde bulunan bilgisayar durumuna göre öğrencilerin matematik başarı puanına etkisi ülkeler arasında farklılık göstermektedir. Japonya ve Makedonya’da istatistiksel olarak ($p < 0.1$) anlamlı ve pozitif bir etki meydana getirmiştir. Singapur ve Malezya’da istatistiksel olarak ($p < 0.01$) anlamlı ve pozitif bir etki meydana getirirken, G. Kore’de istatistiksel olarak anlamlı ($p < 0.01$) fakat negatif bir etki göstermiştir.

Öğrencilerin Türkiye’de %45’ünün ve Malezya’da %46’sının evinde internet erişimi bulunmaktadır. Diğer ülkelerde ise bu oran %80’in üzerindedir. Öğrencilerin evinde internet bulunma durumuna göre öğrencilerin matematik başarı puanına etkisi ülkeler arasında farklılık göstermektedir. Singapur, G. Kore, Malezya ve Makedonya’da öğrencilerin matematik başarı puanına etkisi istatistiksel olarak ($p < 0.1$) anlamlı ve pozitif etki meydana getirmiştir. Benzer şekilde Japonya ve Türkiye’de istatistiksel olarak ($p < 0.01$) anlamlı ve pozitif bir etki meydana getirmiştir (Tablo 5).

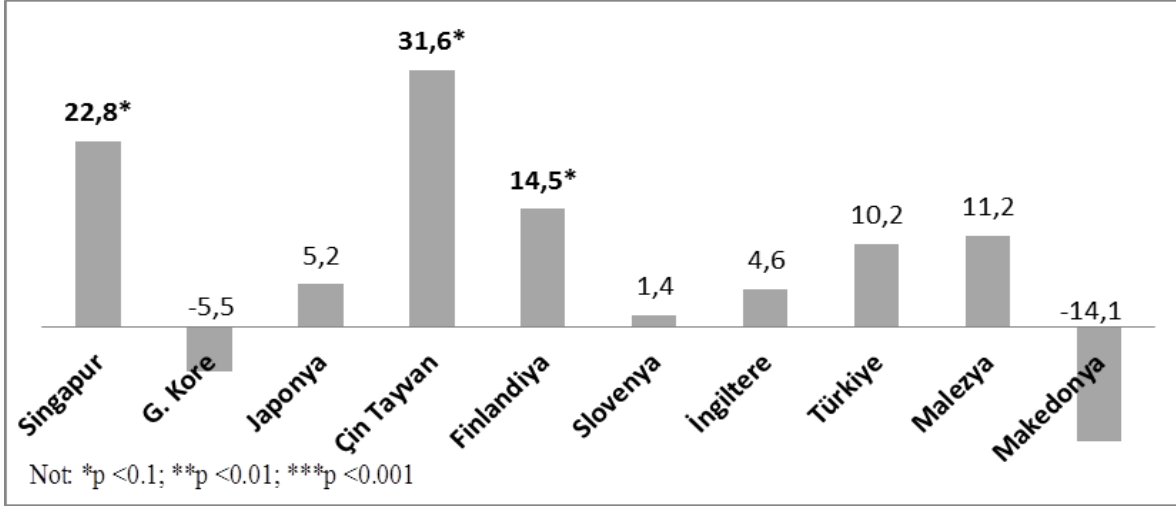
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden matematik öğretmenlerinin matematik öğretmenliği alanında aldığı eğitimin öğrencilerin ortalama matematik başarı puanlarına aritmetik etkisi Şekil 7’de yer almaktadır.



Şekil 7. Matematik Öğretmenlerinin Yükseköğreniminde Matematik Öğretmenliği Alanında Aldığı Eğitimin Öğrencilerin Matematik Başarısına Etkisi

Matematik öğretmenlerinin yükseköğrenimlerinde matematik öğretmenliği alanında almış oldukları eğitimin öğrencilerin matematik başarı puanlarını İngiltere, Türkiye ve Malezya dışındaki ülkelerde pozitif olarak etkilediği görülmektedir (Şekil 7). Öğretmenlerin matematik öğretmenliği alanında aldığı eğitimin öğrencilerin matematik başarılarına istatistiksel etkisi G. Kore, İngiltere ve Malezya’da anlamlı çıkmıştır. Öğretmenlerin matematik öğretmenliği alanında eğitim alması G. Kore’de öğrencilerin matematik başarı puanını 10.9 puan arttırırken, İngiltere’de ise 19.9 puan düşmesine sebep olmuştur.

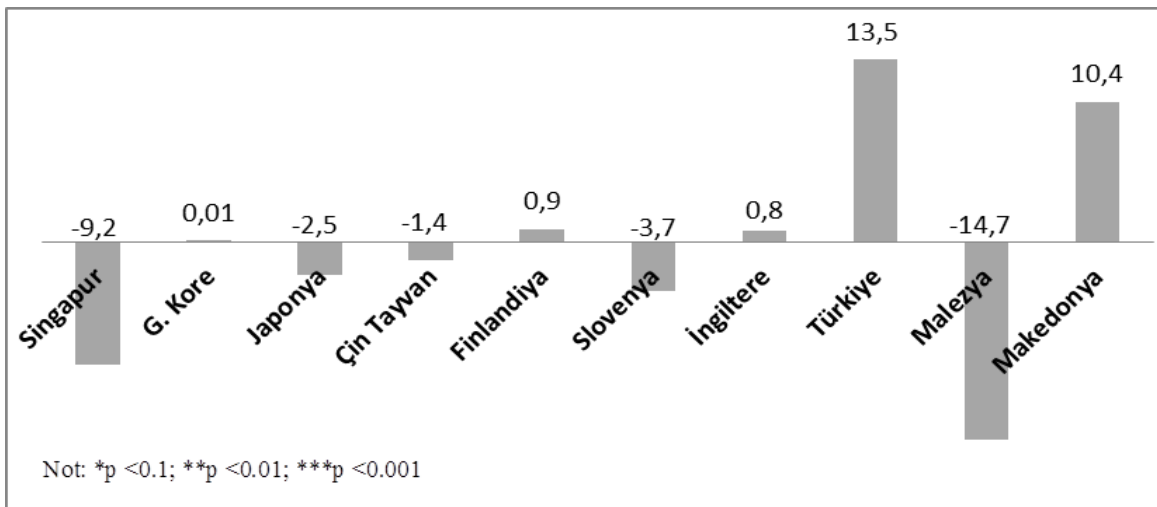
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden matematik öğretmenlerinin matematik alanında aldığı eğitimin öğrencilerin matematik başarı puanlarına aritmetik etkisi Şekil 8’de yer almaktadır.



Şekil 8. Matematik Öğretmenlerinin Yükseköğreniminde Matematik Alanında Aldığı Eğitimin Öğrencilerin Matematik Başarısına Etkisi

Matematik öğretmenlerinin yükseköğrenimlerinde matematik alanında almış oldukları eğitimin öğrencilerin matematik başarı puanlarını G. Kore ve Makedonya dışındaki ülkelerde pozitif olarak etkilediği görülmektedir (Şekil 8). Öğretmenlerin matematik alanında aldığı eğitimin öğrencilerin matematik başarılarına istatistiksel etkisi Singapur, Çin-Tayvan ve Finlandiya’da anlamlı çıkmıştır. Öğretmenlerin matematik alanında eğitim alması Çin-Tayvan’da öğrencilerin matematik başarı puanını 31.6 puan artırırken, Makedonya’da ise 14.1 puan düşmesine sebep olmuştur.

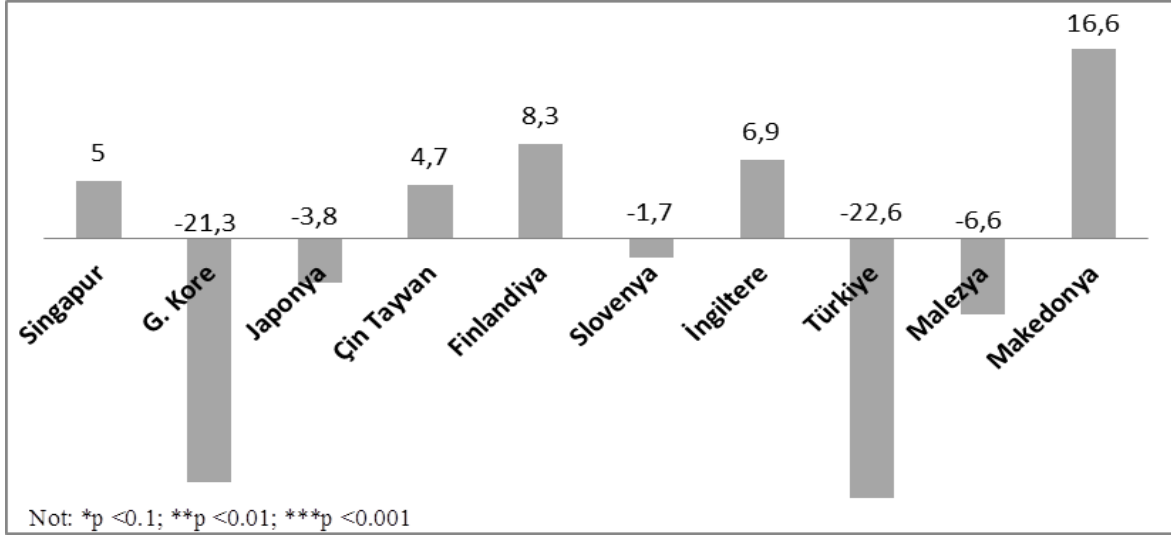
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden matematik öğretmenlerinin fizik alanında aldığı eğitimin öğrencilerin matematik başarı puanlarına aritmetik etkisi Şekil 9’da yer almaktadır.



Şekil 9. Matematik Öğretmenlerinin Yükseköğreniminde Fizik Alanında Aldığı Eğitimin Öğrencilerin Matematik Başarısına Etkisi

Matematik öğretmenlerinin yükseköğrenimlerinde fizik alanında almış oldukları eğitimin öğrencilerin ortalama matematik başarı puanlarını etkileri ülkeler arasında farklılık göstermektedir (Şekil 9). Matematik öğretmenlerinin fizik alanında aldığı eğitimin öğrencilerin matematik başarılarına istatistiksel etkisi çalışmaya dâhil olan ülkelerde anlamlı çıkmamıştır. Matematik öğretmenlerinin fizik alanında eğitim alması Türkiye’de öğrencilerin matematik başarı puanını 13.5 puan artırırken, Malezya’da ise 14.7 puan düşmesine sebep olmuştur.

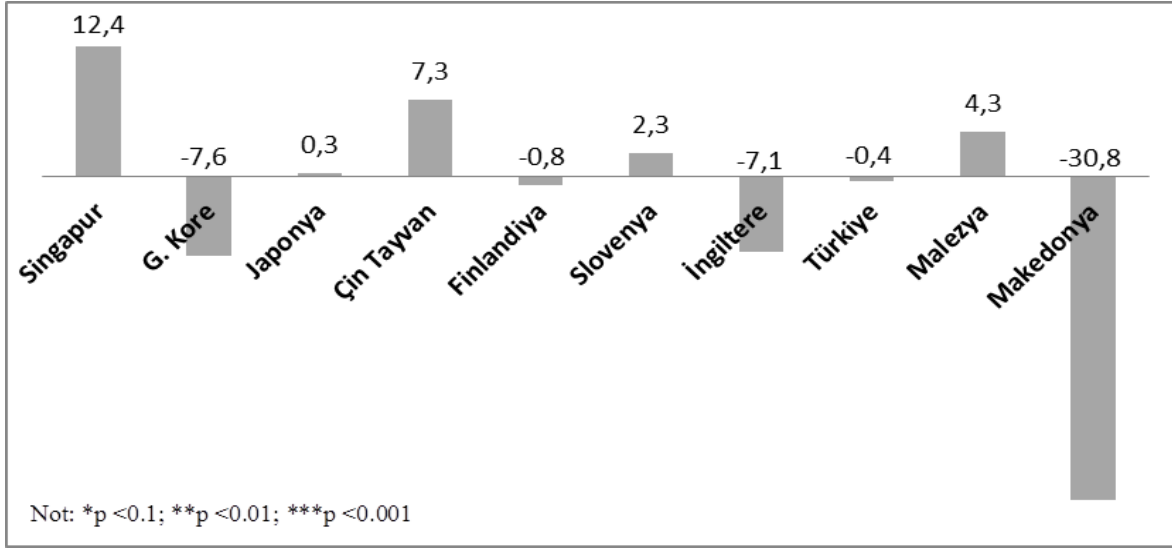
Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden matematik öğretmenlerinin fen bilgisi/bilimleri öğretmenliği alanında aldığı eğitimin öğrencilerin matematik başarı puanlarına aritmetik etkisi Şekil 10’da yer almaktadır.



Şekil 10. Matematik Öğretmenlerinin Yükseköğreniminde Fen Bilgisi/Bilimleri Öğretmenliği Alanında Aldığı Eğitimin Öğrencilerin Matematik Başarısına Etkisi

Matematik öğretmenlerinin yükseköğrenimlerinde fen bilgisi/bilimleri öğretmenliği alanında almış oldukları eğitimin öğrencilerin ortalama matematik başarı puanlarını etkileri ülkeler arasında farklılık göstermektedir (Şekil 10). Matematik öğretmenlerinin fen bilgisi/bilimleri öğretmenliği alanında aldığı eğitimin öğrencilerin matematik başarılarına istatistiksel etkisi çalışmaya dâhil olan ülkelerde anlamlı çıkmamıştır. Öğretmenlerin fen bilgisi/bilimleri öğretmenliği alanında eğitim alması Makedonya’da öğrencilerin matematik başarı puanını 16.6 puan artırırken, Türkiye’de ise 22.6 puan düşmesine sebep olmuştur.

Kurulan HLM analizinde ikinci düzey öğretmen değişkenlerinden matematik öğretmenlerinin genel (eğitim) alanında aldığı eğitimin öğrencilerin matematik başarı puanlarına aritmetik etkisi Şekil 11’de yer almaktadır.



Şekil 11. Matematik Öğretmenlerinin Yükseköğreniminde Genel (Eğitim) Alanında Aldığı Eğitimin Öğrencilerin Matematik Başarısına Etkisi

Matematik öğretmenlerinin yükseköğrenimlerinde genel (eğitim) alanında almış oldukları eğitimin öğrencilerin ortalama matematik başarı puanlarına etkileri ülkeler arasında farklılık göstermektedir (Şekil 11). Matematik öğretmenlerinin genel (eğitim) alanında aldığı eğitimin öğrencilerin ortalama matematik başarılarına istatistiksel etkisi çalışmaya dâhil olan ülkelerde anlamlı çıkmamıştır. Matematik öğretmenlerinin genel (eğitim) alanında eğitim alması Singapur'da öğrencilerin ortalama matematik başarı puanını 12.4 puan artırırken, Makedonya'da ise 30.8 puan düşmesine sebep olmuştur.

SONUÇLAR ve TARTIŞMA

Türkiye Sonuçları Üzerine

Yapılan analizler sonucunda fen bilgisi öğretmenlerinin yükseköğrenimde eğitim aldığı alan eğitimlerine göre en yüksek oran ortalama %71 ile fen bilgisi öğretmenliği alanı, en düşük oran ise %11 ile diğer alan eğitimleri olarak görülmektedir (Ek. 1). Yükseköğreniminde biyoloji, fizik ve kimya alanlarında eğitim alan fen bilgisi öğretmenlerinin öğrencilerinin ortalama fen başarı puanları, bu alanlarda eğitim almayan fen öğretmenlerine göre daha yüksek çıkmıştır. Bu sonuç fizik kimya ve biyoloji alanlarının fen bilgisi alanına kaynaklık etmesi olarak değerlendirilebilir. Yükseköğreniminde matematik, yer bilimleri, matematik öğretmenliği, fen bilgisi öğretmenliği, eğitim (genel) ve diğer olarak ifade edilen alanlarında eğitim alan fen bilgisi öğretmenlerinin öğrencilerinin ortalama fen bilgisi başarı puanları, bu alanlarda eğitim almayan öğretmenlere göre daha düşük çıkmıştır. Öğrencilerin ortalama fen başarı puanlarına istatistiksel olarak anlamlı etki eden yükseköğrenimde aldıkları alan eğitimleri matematik, yer bilimleri ve fen bilgisi öğretmenliği içerikli alanlar olarak görünmektedir. Öğrencilerin ortalama fen başarı puanını en fazla artıran yükseköğrenim alan eğitimleri fizik (9 puan), kimya (8 puan) ve biyoloji (8 puan) alanları olarak görülmektedir. Öğrencilerin ortalama fen başarı puanlarını en çok düşüren öğretmenlerin yükseköğrenimde aldıkları eğitim alanları ise yer bilimleri (27 puan), matematik (26 puan) ve fen bilgisi öğretmenliği (17 puan) alanlarıdır (Ek. 1).

Matematik öğretmenlerinin yükseköğrenimde eğitim aldıkları alan eğitimlerine göre en yüksek oran %78 ile matematik öğretmenliği alanı ve %72 ile matematik alanı, en düşük oran ise %11 ile diğer alan eğitimleri olarak görülmektedir (Ek. 2). Yükseköğreniminde matematik, yer bilimleri ve diğer

olarak ifade edilen alanlardan eğitim alan matematik öğretmenlerinin öğrencilerinin ortalama matematik başarı puanları, bu alanlarda eğitim almayan öğretmenlere göre daha yüksek çıkması beklenmeyen bir sonuçtur. Bu sonuç matematik öğretmeni yetiştiren kurumlara olan güveni azaltmaya sebep olabilir. Yükseköğretimde biyoloji, fizik, kimya, matematik öğretmenliği, fen bilgisi öğretmenliği ve eğitim (genel) alanlarında eğitim alan matematik öğretmenlerinin öğrencilerinin ortalama matematik başarı puanları, bu alanlarda eğitim almayan öğretmenlerin öğrencilerinin matematik başarı puanlarına göre daha düşük çıkması beklenen bir sonuç olarak değerlendirilebilir. Öğrencilerin matematik başarı puanlarına istatistiksel olarak anlamlı etki eden yükseköğretimde aldıkları alan eğitimleri biyoloji, kimya, yer bilimleri ve fen bilgisi öğretmenliği olarak görülmektedir. Öğrencilerin matematik başarı puanlarını en fazla artıran yükseköğretim alan eğitimi ise yer bilimleri (29 puan), diğer olarak ifade edilen alanlar (16 puan), fizik (13 puan) ve matematik (10 puan) alanları olarak çıkmıştır. Öğrencilerin ortalama matematik başarı puanını en çok düşüren yükseköğretim alan eğitimi ise biyoloji (33 puan), fen bilgisi öğretmenliği (22 puan), kimya (16 puan) ve matematik öğretmenliği (11 puan) alan eğitimleridir (Ek. 2).

Öğretmenlerin yükseköğretimde almış oldukları eğitim alanlarının öğrencilerin fen ve matematik başarılarına etkisi genel olarak değişkenlik göstermektedir. Öğretmenlerin yükseköğretimde öğretmenlik ve eğitim programları ile ilgili alanlarda almış oldukları eğitimler öğrencilerin başarı puanlarını negatif yönde etkilemesi incelenmesi gereken bir durumdur. Bunun aksine eğitim ve öğretmenlik alanı dışındaki aldıkları alan eğitimlerinin öğrenci başarısına olumlu etkisi çalışmada ele alınması, üzerinde çalışılması ve düşünülmesi gereken bir durumdur. Sonuç olarak, elde edilen veriler bize öğretmen yetiştirmede yükseköğretimde alınan alan eğitimi alanlarına değil, öğretmenlerin diğer niteliklerine veya alım süreçlerine; örneğin üniversitelerin öğretmenlik alanlarına seçilme süreçlerine odaklanılması gerektiği sonucunu ortaya koymaktadır.

Genel Araştırma Sonuçları Üzerine

“Herkes İçin Eğitim: Finlandiya” (UNESCO, 2000) adlı raporda Finlandiya’da üniversitelerin öğretmen yetiştiren programlarına talebin oldukça yüksek olduğu ve bu programlara başvuranların sadece %10’u ile %15’inin bu programlara kabul edildiği belirtilmektedir. Bu durum öğretmenlik mesleğinin ülkedeki statüsünü göstermesi açısından dikkate değer bulunmaktadır. Malaty (2006) yaptığı bir çalışmada, öğretmenliğe karşı isteklilik ve yakınlık derecesi yüksek öğretmen adaylarının seçimi ile bu kişilere mesleğe alındıktan sonra verilecek araştırma ve uygulama temelli mesleki eğitimin doğal bir sonucu olarak öğretmenlerin, öğrencilerin öğrenmesiyle daha çok ilgili ve kendilerinin meslekî gelişimine de daha istekli olduklarını vurgulamıştır. Bu isteklilik durumu ve öğretmenliğe yakınlık öğrencilerin akademik başarısını etkileyen öğretmen nitelikleri ve farklılıklar olarak görülmektedir.

TIMSS 2011 uygulaması verilerine göre iki düzeyli HLM analizine dâhil edilen öğrenci ve öğretmen değişkenlerine göre Türkiye’de öğretmenler arası farklılıklar öğrencilerin ortalama fen başarı puanının %29’unu açıklarken, öğrencilerin ortalama matematik başarı puanlarının ise %32’sini açıklamaktadır. Çalışmaya dâhil edilen ülkelere bakıldığında öğretmenler arası farklılıkların öğrencilerin fen ve matematik başarı puanını açıklama oranları 500 puan ve üzeri başarı gösteren ülkelerde Singapur ve İngiltere dışında daha düşük çıkmıştır (Tablo 2-4). Normalde başarılı olan ülkelerde öğretmenler arası farklılıkların öğrencilerin akademik başarısına etkisinin daha yüksek olması beklenirdi. Başarılı olan ülkelerde öğretmenler arası farklılıkların öğrenci başarısına etkisinin az olması ayrıntılı olarak incelenmeli ve araştırılmalıdır. Bunun nedeni öğretmenlik mesleğine geçişte diğer ülkelerdeki öğretmen seçme ve yetiştirme sistemindeki farklılıklar olabilir.

Ülkelerin genel olarak araştırma sonuçları incelendiğinde fen bilgisi öğretmenlerinin yükseköğretimde almış oldukları alan eğitimlerinde büyük farklılıklar görülmektedir. Bu durum matematik öğretmenleri için söylenemez. Matematik öğretmenlerinin yükseköğretime eğitimleri matematik ve matematik öğretmenliği alanında yoğunlaşmış durumdadır. Fen bilgisi öğretmenlerinin yükseköğretimde biyoloji, fizik ve eğitim alanlarında aldıkları eğitimler genel olarak (2-3 ülke dışında) öğrencilerin ortalama fen başarı puanını artırmıştır. Matematik öğretmenlerinin

yükseköğretimde matematik ve matematik öğretmenliği alanlarında almış oldukları eğitimler diğer alanlara göre (2-3 ülke dışında) öğrencilerin ortalama matematik başarı puanlarına olumlu yönde etkilemiştir. Öğrencilerin akademik başarısının yükselmesine neden olan öğretmenlerin yükseköğretimde almış oldukları alan eğitimlerinin içerikleri ve öğretmenlerin sahip oldukları nitelikler ayrıca araştırılması önem arz etmektedir.

Araştırmaya dâhil edilen ülkelerde araştırmaya katılan fen bilgisi öğretmenlerinin fen bilgisi öğretmenliği alanında eğitim alma durumları incelendiğinde Türkiye’de öğretmenlerin %72’si bu alanda eğitim alırken diğer ülkelerde bu oran İngiltere hariç %40’ın altındadır (Ek 1). Türkiye’de fen bilgisi öğretmenliği alanında eğitim alan fen bilgisi öğretmenleri öğrencilerin ortalama fen başarı puanını olumsuz etkilemektedir. G. Kore ve Makedonya dışındaki ülkelerde bu durum aynıdır. Türkiye’de üniversitelerin fen bilgisi öğretmenliği alanı ile ilgili bu yönde araştırmalara önem verilmeli ve öğrenci başarısının artırılmasına etki edecek çalışmalar yapılmalıdır.

Araştırmaya dâhil edilen ülkelerde araştırmaya katılan matematik öğretmenlerinin matematik öğretmenliği alanında eğitim alma durumları incelendiğinde Türkiye’de öğretmenlerin %79’u bu alanda eğitim alırken diğer ülkelerde bu oran %56’nın altındadır (Ek 2). Türkiye, İngiltere ve Malezya’da matematik öğretmenliği alanında eğitim alan matematik öğretmenleri öğrencilerin ortalama matematik başarı puanını olumsuz etkilemektedir. Türkiye’de üniversitelerin matematik öğretmenliği alanına yönelik araştırmalara önem verilmeli ve öğrenci başarısının artırılmasına etki edecek çalışmalar yapılmalıdır.

McKinsey ve Company’nin (2007), başarılı eğitim sistemlerine ilişkin yaptığı araştırma raporunda belirtildiği gibi, bir eğitim sisteminin kalitesi ancak öğretmen kalitesi kadardır. Bassett (2008) çalışmasında, eğitimin niteliğini arttıran temel faktörler arasında öğretmen seçiminin önemli olduğunu, öğretmen yetiştirme programları için en yetenekli öğretmen adaylarının seçilmesinin önemini vurgulamış ve uygulamada ise öğretmenlerden en iyiyi elde etme anlayışının öğretmen kalitesini artırdığını belirtmiştir. Yapmış olduğumuz bu çalışma sonuçlarında bize öğretmenin yükseköğretimde almış oldukları alan eğitimlerinin öğrencilerin akademik başarısına etkisinde değişkenlikler gösterdiğini ortaya koymuştur. Öğrencilerin akademik başarılarına etki eden en can alıcı değişkenlerin tespiti için daha boylamsal araştırmalara ihtiyaç duyulduğu ortaya çıkmaktadır.

KAYNAKÇA

- Akiba, M., LeTendre, G. K. & Scribner, J. P. (2007). Teacher Quality, Opportunity Gap and National Achievement in 46 Countries. *Educational Researcher*, 36(7), 369-387.
- Alexander, R. (2000). *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Basil Blackwell.
- Arnold, C. L. (1992). An Introduction to Hierarchical Linear Models. *Measurement and Evaluation in Counseling and Development*, 25 (2), 58-90.
- Bassett, P. F. (2008). What the Finns know shouldn’t surprise us (but it does). *International Educator*, 22 (4). 9-9.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Clotfelter, C., Ladd, H. & Vigdor, J. (2007). *How and Why Do Teacher Credentials Matter for Student Achievement?* (Working Paper No. 2). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.
- Coultas, J.C. & Lewin, K.M. (2002). Who Becomes a Teacher? The Characteristics of Student Teachers in Four Countries. *International Journal of Educational Development*, 22, 243–260.
- Croninger, R. G., Rice, J. K., Rathbun, A. & Nishio, M. (2007). Teacher Qualifications and Early Learning: Effects of Certification, Degree and Experience on First-Grade Student Achievement. *Economics of Education Review*, 26, 312-324.
- Cruikshank, D. (2001). Good teachers, plural. *Educational Leadership*, 58:(5), 26-30.
- De Leeuw, J. and Kreft, I. (1986). Random Coefficient Models for Multilevel Analysis. *Journal of Educational Statistics*, 11, 57-85.

- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Available: <http://epaa.asu.edu/epaa/v8n1>
- Darling-Hammond, L., Berry, B. & Thoreson, A. (2001). Does Teacher Evaluation Matter? Evaluating the Evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57-77.
- Darling-Hammond, L., Holtzman, D., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? *Education Policy Analysis Archives*, 13(42). Available: <http://epaa.asu.edu/epaa/v13n42>
- Decker, P., Mayer, D. & Glazerman, S. (2004). The effects of Teach for America on Students: Findings from a National Evaluation. Princeton, NJ: Mathematical Policy Research.
- Eberts, R. W. & Stone, J. A. (1984). *Unions and Public Schools*. Lexington, MA: D.C. Heath and Company
- Goldhaber, D. & Anthony, E. (2004). Can Teacher Quality Be Effectively Assessed? The Urban Institute.
- Goldhaber, D. & Brewer, D. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.
- Goldhaber, D. & Brewer, D. J. (1998). When Should We Reward Degrees for Teachers? *Phi Delta Kappan*, 80(2), 134-138.
- Goldhaber, D. & Brewer, D. J. (1998). When Should We Reward Degrees for Teachers? *Phi Delta Kappan*, 80(2), 134-138.
- Goldhaber, D. (2002). The Mystery of Good Teaching. *Education Next*, 2(1), 50-55.
- Greenwald, Rob, Larry Hedges, and Richard Laine. (1996). The Effect of School Resources on Student Achievement. *Review of Educational Research* 66:361-96.
- IEA Data Processing Center, 2013.: <http://www.iea-dpc.de/>
- Kane, T., Rockoff, J. & Staiger, D. (2006). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City (Working Paper No. 12155). Cambridge, MA: National Bureau of Economic Research.
- Malaty, G. (2007). What are the reasons behind the success of Finland in PISA? *International Journal for Mathematics Teaching and Learning*, No. 28.06, 8 p.
- McKinsey & Company. (2007). How the world's best performing school systems come out on top. http://www.mckinsey.com/client/service/socialsector/resources/pdf/Worlds_School_systems_final.pdf.
- Murnane, R. J. & Phillips, B. (1981). Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance. *Economics of Education Review*, 1(4), 453-465.
- Nye, Barbara, Spyros Konstantopoulos and Larry Hedges. (2004). How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis* 26: 237- 57.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage
- Raudenbush, S.W., Bryk, A.S. & Congdon, R. (2004). *Hierarchical Linear and Nonlinear Modeling: HLM for Windows (Version 6.00) [Computer Software]*. Lincolnwood, IL: Scientific Software International.
- Rivkin, Steven, Eric Hanushek, and John Kain, (2005). Teachers, Schools, and Academic Achievement. *Econometrica* 73: 417-58.
- Rowan, B. Correnti, R. & Miller, R. J. (2002). What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Board*, 104, 1525-1567.
- Rowan, B. Correnti, R. & Miller, R. J. (2002). What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Board*, 104, 1525-1567.
- Scott-Kassner, C. (1999). Developing Teachers for Early Childhood Programs. *Music Educators Journal*, 86(1), 19-25.
- Tondeur, J., Valcke, M. & Van Braak, J. (2008). A Multidimensional Approach to Determinants of Computer Use in Primary Education: Teacher And School Characteristics. *Journal of Computer Assisted Learning*, 24(6), 494-506.
- UNESCO (2000). Education for All (EFA) 2000 assesment country reports. Finland. http://www.unesco.org/education/wef/countryreports/finland/rapport_1.html
- Wayne, A. J. & Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research*, 73(1), 89-122.
- Zuzovsky, R. (2009). Teachers' Qualifications And Their Impact On Student Achievement: Findings From TIMSS 2003 Data For Israel. *Issues And Methodologies in Large-Scale Assessments*, 37-62.

EXTENDED ABSTRACT

Introduction

In this study, "International Mathematics and Science Trends Survey" (TIMSS 2011) joined the practice in Singapore, South Korea, Japan, China-Taiwan, Finland, Slovenia, England, Turkey, Malaysia, and using data Macedonia "Eighth grade students in science and mathematics achievements of teachers in undergraduate education (Tertiary) shows how a change according to the courses they have taken? "has to answer the question. As a result, the countries studied in this research in science and mathematics teachers in undergraduate education (higher education) which areas the courses the students of science and impacts statistically significant academic achievement in mathematics and has been studied to determine which help boost academic achievement. In this study, Trends in International Mathematics and Science Study (TIMSS 2011) joined the practice was carried out using data from 10 countries together with Turkey. Countries of the TIMSS application; Planning the basic principles of math and science courses and to obtain data relating to the implementation, participate with the aim to make comparisons at the international level. In addition, mathematics and science teachers, class, and students of the school environment to determine the impact of science and mathematics achievement at the national level and to make comparisons at the international level, countries see the main policy differences in education, identifying shortcomings and measures needed to be corrected was conducted to improve the quality of education.

Method

The data sources used to perform this analysis were derived from TIMSS 2011 application. For correlation and regression analysis of the data was performed with SPSS and SPSS HLM analysis based employee programs. In this context, students of mathematics and science teachers have received my higher education courses in their area of influence on the content of science and mathematics achievement Hierarchical Linear Modeling (HLM) was tried to be analyzed by the method. The scope of research, "I have taken my area of science and mathematics teachers in higher classes leads to a change in how students' achievement in science and mathematics?" Has to answer the question. This model of care is a descriptive study of this work. Descriptive analysis is a type of qualitative data analysis including summary and interpretation of the data according to predetermined themes obtained with different data collection techniques.

Survey questions covered by the TIMSS 2011 mathematics and science teacher questionnaire application; "Do what your Yükseköğrenimi received training in the majors or areas?" The question was directed at teachers. Separately created a two-level science and mathematics teachers to work in hierarchical linear modeling (HLM), the following steps were followed. First, empty model where the teacher and student level has been established. This model, including student and teacher level 1 level 2 level contained two sub-models. This model variance values obtained the first level and second level of how much of the variance in student achievement between students from differences have tried to determine how much of which arise from differences between teachers.

Results and Discussion

Graduated in biology, physics and chemistry courses in academic science achievement scores of students of science teachers were higher than the teachers who take courses in these areas. Graduated in mathematics, earth sciences and other areas referred to as academic courses in mathematics achievement scores of students math teacher, was higher compared to teachers who take courses in these areas. They have received their education in higher education of science and mathematics teachers on students' overall success varies. Teachers of higher education teaching and they have taken courses in fields related to their training programs negatively affect student achievement scores. In contrast, education and student success courses taken outside the teaching field are

outstanding positive effect. As a result, the obtained data reveals us to the conclusion that teacher training courses are taken not on Higher Education teachers should focus on the selection process. Mathematics in higher education, earth sciences and average math scores of students for success in other areas of mathematics teachers expressed as the area is higher than the teachers who take courses in these areas. Biology in higher education, physics, chemistry, math teacher, science teacher, and education (general) average mathematics achievement scores of students in their intended course area math teacher space, students of teachers who take courses in these areas has increased significantly lower than the mathematics achievement scores. Of course they take me to higher mathematics achievement scores were statistically significantly affecting biology, chemistry, earth sciences and science seem to be a teacher.

EK 1.

Fen Bilgisi Öğretmenlerinin Yükseköğreniminde Almış Oldukları Ders Alanlarının Yüzde Dağılımları

| | Matematik | | Biyoloji | | Fizik | | Kimya | | Yer Bilimleri | | Matematik Öğretmenliği | | Fen Bilgisi Öğretmenliği | | Eğitim (Genel) | | Diğer |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|------------------------|--------------|--------------------------|--------------|----------------|--------------|-------|
| | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | |
| Singapur | 58-42 | 40-60 | 48-52 | 58-42 | 48-52 | 3-97 | 22-78 | 36-64 | 25-75 | 18-82 | | | | | | | |
| G. Kore | 2-98 | 30-70 | 24-76 | 28-72 | 24-76 | 13-87 | 0-100 | 30-70 | 4-96 | 4-96 | | | | | | | |
| Japonya | 3-97 | 33-67 | 34-66 | 40-60 | 34-66 | 14-86 | 0-100 | 31-69 | 17-83 | 15-85 | | | | | | | |
| Çin Tayvan | 22-78 | 22-78 | 58-42 | 65-35 | 58-42 | 26-74 | 6-94 | 38-62 | 34-66 | 15-85 | | | | | | | |
| Finlandiya | 38-62 | 32-68 | 28-72 | 29-71 | 28-72 | 26-74 | 6-94 | 9-91 | 22-78 | 10-90 | | | | | | | |
| Slovenya | 12-88 | 43-57 | 23-77 | 37-63 | 23-77 | 24-76 | 7-93 | 20-80 | 2-98 | 13-87 | | | | | | | |
| İngiltere | 16-84 | 60-40 | 34-66 | 55-45 | 34-66 | 15-85 | 4-96 | 50-50 | 17-83 | 16-84 | | | | | | | |
| Türkiye | 40-60 | 44-56 | 51-49 | 51-49 | 51-49 | 15-85 | 9-91 | 72-28 | 49-51 | 12-88 | | | | | | | |
| Malezya | 41-59 | 41-59 | 30-70 | 44-56 | 30-70 | 8-92 | 14-86 | 42-58 | 19-81 | 24-76 | | | | | | | |
| Makedonya | 12-88 | 33-67 | 30-70 | 40-60 | 30-70 | 24-76 | 5-95 | 12-88 | 5-95 | 3-97 | | | | | | | |

Fen Bilgisi Öğretmenlerinin Yükseköğreniminde Almış Oldukları Ders Alanlarının Öğrenci Fen Başarılarına Etkisi (HLM Analizi)

| | Matematik | | Biyoloji | | Fizik | | Kimya | | Yer Bilimleri | | Matematik Öğretmenliği | | Fen Bilgisi Öğretmenliği | | Eğitim (Genel) | | Diğer |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|------------------------|--------------|--------------------------|--------------|----------------|--------------|-------|
| | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | Evet - Hayır | |
| Singapur | -7,3 | 13,5 | 3,6 | -9,6 | 3,6 | -61,7* | -1,2 | -14,8 | -3,4 | 26,5* | | | | | | | |
| G. Kore | -19,8* | 8,6* | 11,2** | -7,3* | 11,2** | -18*** | | 7,8* | 0,3 | -11 | | | | | | | |
| Japonya | -5 | -0,5 | 9,6 | -1,4 | 9,6 | -2 | | -3 | 2 | -5,4 | | | | | | | |
| Çin Tayvan | 6,1 | -3 | 0,4 | 4,8 | 0,4 | -4,4 | 14,1 | -9,1 | 14,2* | -22,5** | | | | | | | |
| Finlandiya | 3,4 | 10,8 | 1,1 | -2,2 | 1,1 | 5,6 | -4 | -9,2 | 0,3 | -8,9 | | | | | | | |
| Slovenya | -8,2 | 10,3 | 1,1 | 5 | 1,1 | -1,2 | -3,5 | -4,6 | 7,5 | -4 | | | | | | | |
| İngiltere | -15,3 | 4,9 | -11,4 | -1,7 | -11,4 | -14,6 | -68,4*** | -17,6 | 24,9* | -36,4* | | | | | | | |
| Türkiye | -26,6* | 9 | 9,6 | 8,3 | 9,6 | -27,7 | -1,7 | -17,9* | -10,9 | -10 | | | | | | | |
| Malezya | 5 | 32,5* | 6,4 | 6,3 | 6,4 | 18 | 4,3 | -39,5** | 50,8** | -20,2 | | | | | | | |
| Makedonya | 25* | -41 | -6,1 | 36,5** | -6,1 | -38 | 27,8* | 40,5*** | 15,9 | -49,2* | | | | | | | |

Not: *p < 0.05; **p < 0.01; ***p < 0.001.

EK 2.

Matematik Öğretmenlerinin Yükseköğrenimde Almış Oldukları Ders Alanlarının Yüzde Dağılımları

| | Matematik | | Biyoloji | | Fizik | | Kimya | | Yer Bilimleri | | Matematik Öğretmenliği | | Fen Bilgisi Öğretmenliği | | Eğitim (Genel) | | Diğer | |
|-------------------|-----------|-------|----------|-------|-------|-------|-------|-------|---------------|-------|------------------------|-------|--------------------------|-------|----------------|-------|-------|-------|
| | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır | Evet | Hayır |
| Singapur | 76 | 24 | 7 | 93 | 41 | 59 | 27 | 73 | 3 | 93 | 38 | 62 | 17 | 83 | 24 | 76 | 45 | 55 |
| G. Kore | 51 | 49 | 1 | 99 | 1 | 99 | 0 | 100 | 0 | 100 | 54 | 46 | 1 | 99 | 7 | 93 | 6 | 94 |
| Japonya | 80 | 20 | 1 | 99 | 7 | 93 | 2 | 98 | 0 | 100 | 53 | 47 | 3 | 97 | 32 | 68 | 26 | 74 |
| Çin Tayvan | 89 | 11 | 5 | 95 | 14 | 86 | 10 | 90 | 3 | 97 | 56 | 44 | 19 | 81 | 53 | 47 | 26 | 74 |
| Finlandiya | 66 | 34 | 1 | 99 | 35 | 65 | 29 | 71 | 1 | 99 | 7 | 93 | 5 | 95 | 23 | 77 | 19 | 81 |
| Slovenya | 81 | 19 | 2 | 98 | 49 | 51 | 2 | 98 | 1 | 99 | 50 | 50 | 12 | 88 | 4 | 96 | 14 | 86 |
| İngiltere | 74 | 26 | 5 | 95 | 20 | 80 | 8 | 92 | 4 | 96 | 46 | 54 | 3 | 97 | 19 | 81 | 41 | 59 |
| Türkiye | 69 | 31 | 30 | 70 | 41 | 59 | 32 | 68 | 2 | 98 | 79 | 21 | 32 | 68 | 54 | 46 | 12 | 88 |
| Malezya | 84 | 16 | 29 | 71 | 31 | 69 | 29 | 71 | 4 | 96 | 41 | 59 | 9 | 91 | 21 | 79 | 36 | 64 |
| Makedonya | 7 | 93 | 4 | 96 | 38 | 62 | 2 | 98 | 4 | 96 | 25 | 75 | 10 | 90 | 6 | 94 | 3 | 97 |

Matematik Öğretmenlerinin Yükseköğrenimde Almış Oldukları Ders Alanlarının Öğrenci Fen Başarılarına Etkisi (HLM Analizi)

| | Matematik | | Biyoloji | | Fizik | | Kimya | | Yer Bilimleri | | Matematik Öğretmenliği | | Fen Bilgisi Öğretmenliği | | Eğitim (Genel) | | Diğer | |
|-------------------|-----------|----------|----------|---------|---------------|------------------------|--------------------------|----------------|---------------|--|------------------------|--|--------------------------|--|----------------|--|-------|--|
| | Matematik | Biyoloji | Fizik | Kimya | Yer Bilimleri | Matematik Öğretmenliği | Fen Bilgisi Öğretmenliği | Eğitim (Genel) | Diğer | | | | | | | | | |
| Singapur | 22,8* | -28* | -9,2 | 22,6** | 0,1 | 6 | 5 | 12,4 | 1 | | | | | | | | | |
| G. Kore | -5,5 | | 0,01 | | | 10,9* | -21,3 | -7,6 | -10,2 | | | | | | | | | |
| Japonya | 5,2 | -19,2 | -2,5 | -16,4 | | 7,1 | -3,8 | 0,3 | -2,5 | | | | | | | | | |
| Çin Tayvan | 31,6* | 11,6 | -1,4 | 11,9 | -30,3* | 9,4 | 4,7 | 7,3 | -10,1 | | | | | | | | | |
| Finlandiya | 14,5** | -31,7* | 0,9 | 0,6 | -27,2 | 6,1 | 8,3 | -0,8 | -4,8 | | | | | | | | | |
| Slovenya | 1,4 | -0,7 | -3,7 | 3,3 | | 0,2 | -1,7 | 2,3 | 0,2 | | | | | | | | | |
| İngiltere | 4,6 | 78,7** | 0,8 | 84,7*** | 81,1** | -19,9* | 6,9 | -7,1 | -2,5 | | | | | | | | | |
| Türkiye | 10,2 | -33,6* | 13,5 | -16* | 29* | -11,7 | -22,6 | -0,4 | 16,3 | | | | | | | | | |
| Malezya | 11,2 | 27,8* | -14,7 | 15,5 | 22,6 | -18,6* | -6,6 | 4,3 | 7,7 | | | | | | | | | |
| Makedonya | -14,1 | -41* | 10,4 | 86,2* | -15,4 | 12,1 | 16,6 | -30,8 | -86,5*** | | | | | | | | | |

Not: *p<0,1; **p<0,01; ***p<0,001.

The Development of IRT Based Attitude Scale towards Educational Measurement Course*

Ölçme ve Değerlendirme Dersine Yönelik Tutum Ölçeğinin Madde Tepki Kuramına Dayalı Olarak Geliştirilmesi*

Nükhet DEMİRTAŞLI ** Seher YALÇIN *** Cansu AYAN ****

Abstract

In this study, the Scale of Attitude towards Educational Measurement and Evaluation (SAEM) developed by Demirtaşlı (2002) is reconstructed based on polytomous Item Response Theory (IRT) models and its psychometric features are identified. In this context, the best polythomous IRT model was investigated which is fitted SAEM data. IRT models gives invariant person and item parameters, when data-model fit. A version of SAEM has 41 Likert type items with four points was administered to 519 teacher candidates attending teacher education programs at several universities in Turkey. The data were analyzed according to polythomous IRT models: Samejima's graded response model (S-GRM), the partial credit model (PCM) and a nominal response model (NRM). The results of the analysis showed that a new version of SAEM, which is based on S-GRM, consists of 33 items, has lower chi-square value than the other models and the classic internal reliability was found to be 0.97. The findings of the study indicate that the validity and reliability features of the scale are fairly good.

Key Words: Attitude toward educational measurement and evaluation, polytomous item response model, attitude scale.

Öz

Bu araştırmada, ölçme ve değerlendirme dersine yönelik tutumu ölçmek üzere geliştirilen Likert tipi, Ölçme ve Değerlendirme Dersine Yönelik Tutum (ÖDET) ölçeğinin (Demirtaşlı, 2002) madde tepki kuramına dayalı (MTK) çok kategorili puanlanan modeller çerçevesinde yeniden ölçeklenerek psikometrik özelliklerinin karşılaştırılması amaçlanmıştır. MTK'ya dayalı modeller verilerle uyum gösterdiğinde, değişmez birey ve madde parametreleri kestirilebilir amaca uygun ölçek geliştirmede daha güvenilir ve geçerli sonuçlara ulaşılabilir. Ölçeğin ilk versiyonu, dörtlü Likert tipi, dereceli toplamalı tepki vermeye uygun 41 maddeden oluşmaktadır. Ölçeğin bu formu, Türkiye'deki farklı illerden üç devlet üniversitesinin öğretmen yetiştiren fakülte ve programlarına devam eden 519 üniversite öğrencisine uygulanmıştır. Maddeler çok kategorili puanlanan maddeler için geliştirilen MTK modellerinden Samejima (S) kademeli tepki modeli (S-Graded Response model), kısmi puan modeli (partial credit model) ve sınıflandırılmalı tepki modeline (nominal response model) göre ölçeklenmiştir. Ki-kare veri-model uyum testi sonucunda, S-kademeli tepki modeline göre ölçeklenen ölçeğin 33 maddelik yeni versiyonunun veriyle daha uyumlu olduğu görülmüştür. İç tutarlık anlamındaki klasik güvenilirlik katsayısının da 0.97 olduğu bulunmuştur. S-kademeli tepki modeline göre psikometrik özellikleri MTK'ya göre kestirilen ÖDET'in geçerli ve güvenilir bir ölçme aracı olduğu görülmüştür.

Anahtar Kelimeler: Ölçme ve değerlendirmeye yönelik tutum, çok kategorili MTK, tutum ölçeği.

* Paper presented at The 4th Congress on Measurement and Evaluation in Education and Psychology on June 9th-13th 2014, Ankara

** Prof. Dr., Ankara University, Faculty of Educational Sciences, Ankara-Turkey, e-mail: rnukhet@yahoo.com

*** Res. Asst., PhD., Ankara University, Faculty of Educational Sciences, Ankara-Turkey, e-mail: yalcins@ankara.edu.tr

**** Res. Asst., Ankara University, Faculty of Educational Sciences, Ankara-Turkey, e-mail: cnsayan@gmail.com

INTRODUCTION

Educational measurement and evaluation is a compulsory course in undergraduate and teacher certification programs. In spite of, teachers spend as much as a third of their professional time in assessment related activities and many of these activities require skills in testing and measurement (Wise, Lukin and Roos, 1991), some pre-service and in-service teachers have concerns and negative attitude about succeeding in these math-based subjects (Brady & Bowd, 2005; Gresham, 2010; Jaggernouth, 2010; Kottke, 2000). As an affective trait, attitude is a tendency to respond in direction of approaching or avoiding to an object, person, institution, or event (Ajzen, 2005). This tendency can have an indirect positive or negative impact on learning behavior (Perkins, Adams, Pollock, Finkelstein & Wieman, 2005; Reed, Drijvers & Kirschner, 2010; Shih & Gamon, 2001). Several studies have investigated the attitudes of pre-service and in-service teachers towards the measurement and evaluation course and their self-efficacy in this course (Aktaş & Alıcı, 2012; Kılınç, 2011; Kilmen & Demirtaşlı, 2009; Ozan & Köse, 2013; Özbaşı & Demirtaşlı, 2013; Ulutaş, 2003). Recognizing the attitudes of pre- and in-service teachers towards the measurement and evaluation course can be used to create a more positive learning environment in education and training programs. Searching and analyzing the negative attitudes of student teachers in relation to a course using a valid and reliable attitude scale helps to identify the pedagogic action to be taken to change teacher candidates's attitudes from negative to positive. This situation contributes to the establish a positive learning climate.

In education and psychology, test construction is based on primarily two test theories; the classical test theory (CTT) and the item response theory (IRT). The theoretical foundations of IRT dates back to the 1950s however, since IRT-based estimations require complex mathematical and statistical processes, the remarkable progress in this area was observed after the 1980s with the significant innovations in computer and software technology. When studied on a dataset that meets its basic assumptions, IRT can overcome the limitations of CTT and provides several advantages for the scaling process. In scale-development studies based on IRT, when the basic assumptions of IRT are fulfilled and the data fit the model, invariant person and item parameters can be estimated (De Ayala, 2009; Hambleton et al., 1991). Therefore, IRT based tests are not necessarily to establish conventional test norms for items measure in the same way at subsamples from the same population (Embretson & Reise, 2000, p. 25; Hambleton et al., 1999). An IRT-based scale can be used as a valid and reliable instrument to estimate the traits of subsamples. With this advantage, IRT can also be used to solve other measurement problems such as those related to the test equating, computer based adaptive testing, detecting of biased items.

In this context, the purpose of IRT based SAEM is to benefit from IRT's advantages such invariant item and theta parameters when model-data fit. By means of invariance, no further norm studies in interpretation of SAEM scores, comparison of groups. Besides, since IRT models give individual error estimations in item and person level, IRT based SAEM will be able to measure attitudes towards educational measurement course more reliably. In addition to this advantage, it can be detect possible item bias for several participants' background variables like type of under graduate program (Social sciences, Science), level of attitudes towards numerical content courses. Finally, when SAEM developed based on IRT, paralell forms of SAEM can be construct more easily and reliably.

Purpose

In this study, IRT was used to reconstruct a Likert-type CTT-based scale (SAEM) developed by Demirtaşlı (2002) to measure the attitude towards educational measurement and evaluation. In this context, the best polythomous IRT model that fits attitude data was investigated. For this purpose, the psychometric characteristics of the SAEM were tested under Samejima's Graded Response Model (GRM), Partial Credit Model (PCM) and Nominal Response Model (NRM) (Embretson & Reise, 2000)

METHOD

Research Model

This study is a descriptive research since the aim was to identify the psychometric features of the SAEM based on IRT (Glass & Hopkins, 1984; Kaptan, 1995).

Study Group

This scale was administered to 519 pre-service teachers enrolled in teacher college in three public universities in three different provinces of Turkey. All the participants had already taken the measurement and evaluation course in teacher college programs. Of the participants, 67% were female and 29% were male. Participation in the study was voluntary.

Data Collection

The scale reconstructed in this study was developed to measure the attitudes towards the measurement and evaluation course, which is compulsory in teacher education and teacher certification programs. This scale is a four-point graded Likert scale consisting of 41 items, and was found to measure valid and reliable with three factors. The results of Cronbach's alfa correlation coefficients showed that the reliability for SAEM's each factor were ranged from .82 and .92 (Demirtaşlı, 2002). The following four categories was used to respond to all items in the scale; 1=strongly disagree, 2=disagree, 3=agree and 4=strongly agree. The items were scored reverse that are express negative attitudes: 4 for "strongly disagree" and 1 for "strongly agree". The minimum and maximum scores of the scale are 41 and 164, respectively. A higher score means that the participant has a more positive attitude towards the measurement and evaluation course, and a lower score indicates a negative attitude.

Data Analysis

Data analysis was performed in three stages. First, the participants' responses to the items were scored. Then, the data were analyzed in terms of basic IRT assumptions namely unidimensionality and local independence. When the data fit the IRT-based models, invariant person and item parameters can be estimated (Embretson & Reise, 2000; Hambleton et al., 1991). This feature of IRT helps to construct tests for the expected features, and also, equate test forms and develop computerized adaptive testing.

The scale dimensionality was detected by a Principal Component Analysis (PCA). The data were analyzed by the SPSS 15.0. The statistical convenience of the items to the PCA was determined using their Kaiser-Mayer-Olkin (KMO) value and the results obtained from Bartlett's test. The KMO value was found to be 0.97, and according to the result of Bartlett's test, the chi-square statistic was significant ($\chi^2(820) = 13163.31$; $p < 0.05$). These findings indicate that the items of the scale fit the PCA. In the first analysis, 41 items were loaded under five components. In initial analysis, the five-component structure was observed that accounted for 60% of the total variance and ten items had loadings more than one factor (factor loading > 0.40). The scree plot (Figure 1) of the data shows a rapid decrease in the eigenvalue from the first to the second factor. Based on this result, it was concluded that SAEM had a dominant one factor.

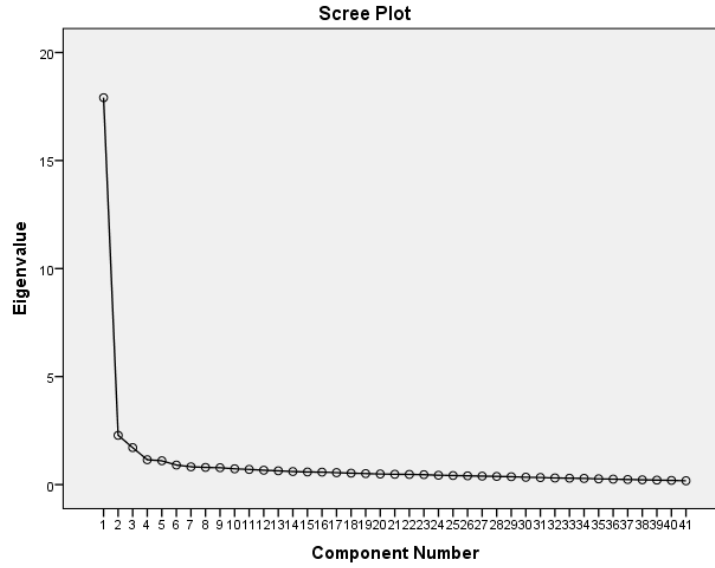


Figure 1. Scree Plot of the SAEM Factor Structure

After that, factor structure of the scale was re-analyzed by restricting it to a single factor with varimax rotation. The results of the PCA restricted to a single factor showed that 41 items explained 44% of the total variance and factor loadings were varied from 0.35 to 0.77. Based on these results, it can be concluded that SAEM has a dominantly unidimensional structure and thus met the unidimensionality assumption of IRT. Another assumption of IRT is local independence, which means that at a given trait level, a test taker's response to an item is independent from the other items. In other words, a response to any of the items in the scale (e.g. endorsing or rejecting a certain attitude) is not dependent on the response to another item. This is observed when the unidimensionality assumption is met. In a test identified as unidimensional, the covariance between the items is zero for subjects at the same latent trait. This indicates that once the unidimensional assumption is met, the local independence assumption is also met (Hambleton & Swaminathan, 1985). As a result, the 41-item scale used in the current study was considered to have fulfilled the assumption of local independence.

In the second phase of the data analysis, items were detected in terms of bias. The probable source of bias for this data is gender. In testing procedure, the individual differences resulted from the measured trait, rather than the gender of the participants with the same latent trait. To this end, the items in the scale were analyzed to determine whether they displayed differential item functioning (DIF) in terms of gender. To detect the DIF of polytomous items, the Polytomous Simultaneous Item Bias Test (PSIBTEST) and IRT Likelihood Ratio Test (IRT-LRT) were used. In the PSIBTEST method, DIF is determined through a regression-based correction that can determine Type I error (Clauser & Mazor, 1998).

IRT-LRT is based on a comparison of observed and theoretical models (Thissen, Steinberg & Wainer, 1993) which requires restricted and extended models. In the restricted model, which assumes that none of the items has DIF, the probability of the parameters of all items being equal is calculated. In the extended model, the likelihood of item parameters, for which DIF is detected, being different in the focal and reference groups with other parameters being equal is found. The G^2 value is calculated by subtracting the two $-2\log$ likelihood values obtained from the likelihood ratio of the restricted and extended models (Thissen, 2001). The calculated G^2 value is then compared to the chi-square value with the degrees of freedom. The degrees of freedom is the number of parameters in the model, and thus in the current study, it was four ($df =$ three threshold parameters and one discrimination parameter). If the G^2 value is less than 9.49, it is interpreted that a negligible DIF level is present; if it is higher, then there is a medium or high level of DIF against the focal

group of the relevant item (Greer, 2004). The IRT-LRT analysis method uses anchor items to equate the groups. For the selection of anchor items, the following criteria are used; having a high level of discrimination, having a high range of difficulty level, displaying no DIF according to other DIF detection methods, producing a small error variance in the PCA and having high factor loadings (Yıldırım, 2006). In this study, the criteria for the selection of anchor items were that they represented both way of the attitude, display no DIF according to the result of PSIBTEST and have high factor loadings in PCA. As a result, items 19, 27, 29 and 34 were selected as anchor. DIF analyses were performed DIFPACK 1.7 and IRTLRDIF 2.0b packages. Table 1 presents the results of the analyses performed using two DIF detection methods.

Table 1. Results of the DIF Analyses Under Two Different DIF Methods

| Items | PSIBTEST | IRT-LRT |
|-------|--|----------------------------|
| | B or C Level DIF | B or C Level DIF |
| | 3, 8, 10, 13 , 17 , 18, 19, 25, 26, 31, 37 | 13 , 17 , 41 |

As shown in Table 1, two items were found to display DIF in both methods (items 13 and 17). Item 13 was, “I would like to take other measurement courses” and item 17 was, “I wish I could take more measurement and evaluation courses”. Both items were in favor of the male participants. In other words, when male and female students with the same level of attitude were compared, the probability of male students moving from “agree” to “strongly agree” was found to be significantly higher. Following the analysis performed, these items were excluded from the scale.

In the third stage of data analysis, the remaining 39 items were analyzed according to Samejima’s GRM, PCM and NRM using the MULTILOG 7.03 package. Samejima’s GRM is used to measure items that have ordered categorical responses such as Likert type scale items, and it is an extension of the two-parameter logistic (2PL) model. In GRM, the threshold values of response categories should be ordered, which is not required by PCM or generalized PCM (Embretson & Reise, 2000). PCM was originally developed for items that require responses in multiple steps. It is also used for the analysis of responses to items in scales that measure traits, in which two or more categorical responses are possible (such as attitude and personality traits). NRM is used to measure responses of similar format items but it does not require item choices to be ordered or identified numerically. The purpose of this model is to plot options characteristic curves based on the frequency of the selected choices in multiple-choice items. This model can also be applied to attitude and personality scales. All three models are used in items that are scored using grades and they have different advantages and disadvantages. For example, Samejima’s GRM does not require the items to have the same number of categories. Therefore, it is appropriate for scales consisting of items with different response formats. Furthermore, this model is an extension of the 2PL model and allows the discrimination index to be different among items. PCM, on the other hand, is an extension of the Rasch Model, and as a result, raw score is sufficient statistics to estimate the ability level of an individual. However, in the PCM model, the slopes of all items in this model are considered to be equal. In other words, the model assumes that the discrimination index among items is equal, which is not that easy to realize in practice (Baker, Rounds & Zevon, 2000; Embretson & Reise, 2000).

RESULTS

Thirty-nine items of the scale were scaled using the three models, and Table 2 presents the maximum item information obtained from each model.

Table 2. Amount of Information Obtained From Items Using Different Polythomous IRT Models

| Items | GRM | PCM | NRM | Items | GRM | PCM | NRM |
|--------------------|-----------------|----------------|---------------|--------------------|--------------|--------------|--------------|
| M1 | 0.982 | 0.876 | 0.963 | M21 | 1.441 | 1.992 | 1.557 |
| M2 | 1.052 | 1.135 | 1.357 | M22 | 0.982 | 1.658 | 1.028 |
| M3 | 0.577 | 0.774 | 0.596 | M23 | 1.254 | 1.731 | 1.171 |
| M4 | 0.825 | 1.109 | 0.927 | M24 | 1.250 | 1.814 | 1.246 |
| M5 | 1.561 | 2.075 | 1.540 | M25 | 2.135 | 3.568 | 2.386 |
| M6 | 0.388 | 0.394 | 0.684 | M26 | 1.470 | 1.942 | 1.306 |
| M7 | 0.351 | 0.469 | 0.300 | M27 | 2.749 | 4.418 | 3.161 |
| M8 | 0.491 | 0.688 | 0.477 | M28 | 0.880 | 1.404 | 1.112 |
| M9 | 0.784 | 0.849 | 0.836 | M29 | 0.138 | 0.147 | 0.134 |
| M10 | 0.231 | 0.246 | 0.222 | M30 | 1.424 | 1.686 | 1.578 |
| M11 | 0.533 | 0.630 | 0.551 | M31 | 1.684 | 3.321 | 2.201 |
| M12 | 1.886 | 3.222 | 2.022 | M32 | 2.891 | 7.068 | 3.011 |
| M13 | 0.792 | 0.910 | 0.939 | M33 | 0.825 | 0.923 | 0.961 |
| M14 | 1.451 | 1.916 | 1.373 | M34 | 1.996 | 4.063 | 2.439 |
| M15 | 1.754 | 2.779 | 1.825 | M35 | 1.669 | 2.497 | 1.921 |
| M16 | 0.445 | 0.483 | 0.631 | M36 | 1.822 | 2.976 | 2.076 |
| M17 | 1.773 | 2.198 | 1.727 | M37 | 1.760 | 2.469 | 1.979 |
| M18 | 1.868 | 3.008 | 2.115 | M38 | 1.250 | 2.104 | 1.196 |
| M19 | 1.804 | 3.617 | 1.739 | M39 | 0.441 | 0.487 | 0.520 |
| M20 | 1.931 | 3.075 | 3.104 | | | | |
| Total | 49.84 | 67.56 | 45.97 | Marginal | 0.973 | 0.974 | 0.970 |
| Information | (-1.40)* | (-1.20) | (0.60) | Reliability | | | |

*The values in parentheses indicate the level of trait (attitude) with the highest amount of information.

As shown in Table 2, according to GRM, PCM and NRM, item information ranges from 0.14 to 2.89, from 0.15 to 7.07 and from 0.13 to 3.10, respectively. The total test information values obtained from the three models are presented in Table 2. The highest test information (67.56) provided from PCM at -1.20 theta (attitude) level. The highest test informations obtained from GRM and NRM respectively. Although the reliability coefficient of all three models was close to each other, the highest reliability coefficient value, 97.4, was obtained from PCM.

The model-data fit level was determined by comparing $-2 \log$ likelihood values from polythomous model pairs. First, PCM and GRM were compared in terms of differences in $-2 \log \chi^2$ values, chi-square statistics and degrees of freedom (Df). Df is computed by multiplying the number of items with the number of parameters calculated for the estimation model. The number of parameters varies according to the model used for estimation; however, the number of “step difficulty/threshold/intercept” parameters substituting item difficulty equals the “number of categories-1” (Embretson & Reise, 2000). In PCM, for each item with four categories, three step difficulty parameters and two item slope parameters were calculated, and thus the degrees of freedom is 195 (39*5). In GRM, for each item with four categories, three threshold parameters and one item slope parameter were estimated, resulting in a degrees of freedom of 156 (39*4). According to this, $\chi^2(195, 156) = 26115.8 - 25886.5 = 229.3$ and the approximate table value is $\chi^2(39; 0.05) = 55.75$. Since the calculated value is higher than the table value, the difference between the $-2 \log \chi^2$ values is significant. Therefore, it can be concluded that GRM is more appropriate for this type of data. In the second stage, the difference in $-2 \log \chi^2$ values obtained from GRM and NRM was determined and compared with the Chi-Square statistic using the 0.05 significance level and related degrees of freedom. In NRM, for each item with four categories, three intercept and three item slope parameters are computed, which results in degrees of freedom being 234 (39*6). $\chi^2(156, 234) = 25886.5 - 25832.4 = 54.1$ and the approximate table value is $\chi^2(78; 0.05) = 101.88$. Since the calculated value is lower than the table value, the difference between the $-2 \log \chi^2$ values is not significant. This indicates that there is no difference between the GRM and NRM models. Furthermore, in GRM, the reliability and maximum information values were found to be 0.973 and 49.84, respectively while in NRM, these were 0.970 and 45.97, respectively. Although no significant difference was observed between GRM and NRM in terms of data fit, estimations were performed using GRM since the highest maximum information and reliability was achieved with this model. Through the estimations on the Multilog program, the slope, threshold parameters and threshold

information functions of all the items were obtained and analyzed. According to GRM, six of the 39 items (6, 7, 10, 18, 31 and 41) had a low level of item information (under .45), and therefore, they were excluded from the scale. Furthermore, differences in the observed and expected values were analyzed for each category and the items were found to be fit to the data. Table 3 presents the slope and threshold parameters of items estimated by the GRM model (33 items).

Table 3. Estimated Item Parameters According to GRM

| Items | a | b1 | b2 | b3 | Items | a | b1 | b2 | b3 |
|-------|------|-------|-------|------|-------|------|-------|-------|------|
| M1 | 1.84 | -2.03 | -1.45 | 0.18 | M18 | 1.89 | -2.23 | -0.94 | 1.23 |
| M2 | 1.91 | -1.91 | -1.39 | 0.44 | M19 | 2.10 | -1.67 | -0.59 | 1.22 |
| M3 | 1.37 | -1.98 | -0.78 | 1.48 | M20 | 2.06 | -1.73 | -0.82 | 1.08 |
| M4 | 1.74 | -1.89 | -0.14 | 1.95 | M21 | 2.72 | -1.69 | -0.93 | 0.84 |
| M5 | 2.36 | -1.54 | -0.50 | 1.26 | M22 | 2.32 | -1.57 | -0.39 | 1.40 |
| M6 | 1.28 | -2.83 | -1.17 | 1.10 | M23 | 3.11 | -1.47 | -0.67 | 0.99 |
| M7 | 1.65 | -2.36 | -1.29 | 0.84 | M24 | 1.83 | -1.81 | -0.33 | 1.67 |
| M8 | 1.38 | -1.62 | 0.04 | 2.62 | M25 | 2.31 | -1.42 | -0.01 | 1.67 |
| M9 | 2.52 | -1.92 | -1.12 | 0.83 | M26 | 2.55 | -1.70 | -0.47 | 1.35 |
| M10 | 1.72 | -1.97 | -0.30 | 1.75 | M27 | 3.27 | -1.66 | -1.13 | 0.83 |
| M11 | 2.22 | -1.60 | -0.78 | 0.87 | M28 | 1.77 | -1.71 | 0.17 | 1.84 |
| M12 | 2.61 | -2.20 | -1.15 | 0.73 | M29 | 2.76 | -1.98 | -1.12 | 0.81 |
| M13 | 2.59 | -1.30 | -0.07 | 1.67 | M30 | 2.48 | -1.63 | -0.66 | 0.97 |
| M14 | 2.59 | -1.73 | -1.08 | 0.67 | M31 | 2.52 | -1.82 | -1.06 | 0.81 |
| M15 | 2.53 | -2.17 | -1.37 | 0.62 | M32 | 2.66 | -2.24 | -0.98 | 0.72 |
| M16 | 2.72 | -1.77 | -0.78 | 0.96 | M33 | 2.20 | -2.45 | -1.22 | 0.92 |
| M17 | 2.34 | -1.62 | -0.29 | 1.57 | | | | | |

As shown in Table 3, parameter ‘a’ is between 1.28 and 3.27. DeMars (2010) stated that the discrimination level of polytomous items is interpreted in the same way as dichotomous items. The discrimination level of items is classified as; very low (0.01-0.34), low (0.35-0.64), medium (0.65-1.34), high (1.35-1.69) and very high (>1.70) (Baker, 2001). On this basis, 33 items in the current scale had a high or very high level of discrimination, with item 27 having the highest and item 6 having the lowest level.

The threshold values of categories vary from -2.83 to 2.62. Most of the first two threshold parameter values were found to be negative, which indicates that the responses to the first three categories had been endorsed by participants with much lower attitude levels ($\theta < 0$). The category threshold values in Table 3 show that the first threshold parameter is around “-2”, the second was around “-1” and the third was around “0”. This indicates that the scale better differentiates people with a lower attitude. Furthermore, the threshold parameter increasing in parallel to the attitude level suggests that the categories performs hierarchically as expected. Considering all these results, it is concluded that the discrimination characteristic and threshold values of the items in the scale are sufficiently high. Figure 2 presents the standard errors related to total information and measurement error obtained from the 33-item GRM-based SAEM.

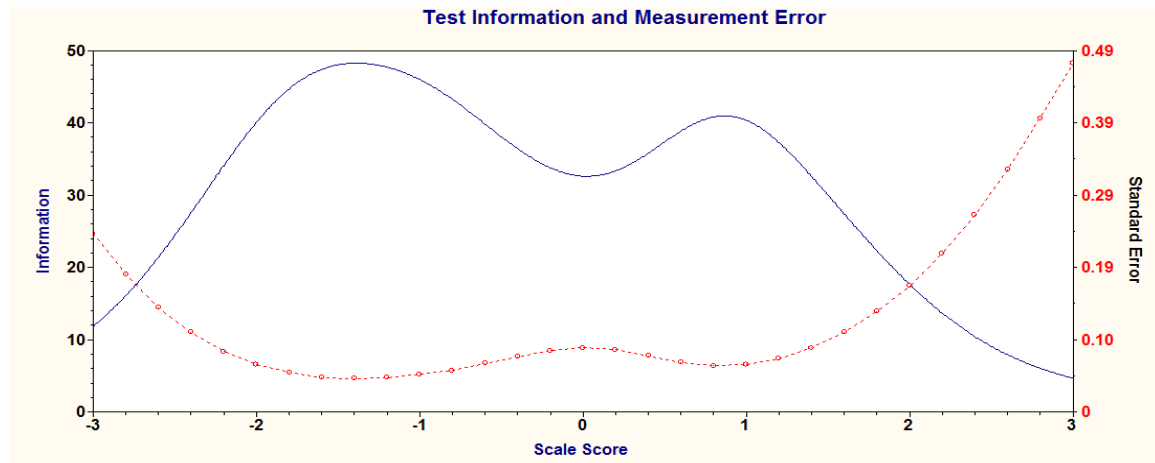


Figure 2. Standard Errors and the Total Test Information Obtained from the GRM-based SAEM

Figure 2 shows that in the last version of the 33-item GRM-based SAEM, the information obtained from the scale is considerably high and falls within a wide range of attitude levels ($-2 \leq \Theta \leq +2$). The scale only provided less information for individuals with an attitude level at the lowest or highest end. The maximum information obtained from the GRM-based scale was found to be 48.23, which was achieved at the attitude level of -1.40.

CONCLUSION and DISCUSSION

In related literature, there are many examples of cognitive test construction studies under IRT models. On the other hand, scales on affective traits have relatively limited that developed by IRT based models. In this study, As an attitude scale SAEM was rescaled based on polyhomous IRT models. Attitude is remarkable affective trait which is effective on our behaviours (Ajzen, 2005). Searching and displaying the negative attitudes of pre-service teachers in relation to a course using a valid and reliable attitude scale helps to identify the pedagogic efforts and activities to be taken to change negative attitudes in direction of positive attitudes. This contributes to the establishment of a positive learning environment. Differently from the CTT based scales, tests and scales based on IRT models separately estimates the probability of individuals at different trait (attitude) levels endorsing each category in each item, which provides more valid and reliable results in terms of the measures of individual differences. Thus, the functionality of both items and response categories can be estimated independently from the other items in the scale and the participant samples. In other words, from the total information and item information values obtained with this approach, it is possible to identify items and response categories that reveal the individual differences at different attitude levels (Le, 2013; Matteucci & Stracqualursi, 2006). In the current study, the model-data fit was higher in GRM than PCM. This may have resulted from the GRM criterion that the threshold values of response categories should be ordered. Thus, the statements “strongly disagree”, “disagree”, “agree” and “strongly agree” can be considered to indicate ordered responses. This study shows that the IRT based SAEM with 33-item is a valid and reliable instrument to determine the attitudes of pre-service teachers towards the measurement and evaluation course.

Future studies will be able to carry out with this instrument. This IRT based version of the scale is a valid and reliable scale that can be used in studies that compare the attitudes of teacher candidates from teachers and non-teachers college programs. And also, this scale can be used to investigate the effectiveness of a variety of actions undertaken during the teaching of the course to change any negative attitudes held by the pre-service teachers. Furthermore, this scale can be used to investigate the relationship between pre-service teachers' achievement in the measurement and evaluation course and their attitude towards it.

REFERENCES

- Aktaş, M., & Alıcı, D. (2012). Development of likert type attitude scale towards measurement and evaluation in education course. *Journal of Qafqaz University*, 33, 66-73.
- Ajzen, I. (2005). *Attitudes, personality and behavior*. Milton-Keynes, England: Open University Press/McGraw- Hill.
- Baker, F. B. (2001). *The basis of item response theory*. USA: ERIC Clearing house on Assessment and Evaluation.
- Baker, J. G., Rounds, J. B. & Zevon, M. A. (2000). A comparison of graded response and rasch partial credit models with subjective well-being. *Journal of Educational and Behavioral Statistics*, 25, 253-70.
- Brady, P. & Bowd, A. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among pre-service education students. *Teachers and Teaching: Theory and Practice*, 11(1), 37-46.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedure to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.
- Demirtaşlı, Ç. N. (2002). Developing a scale for attitudes toward the measurement and evaluation course. *Education and Science*, 27, 125, 44-48.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. LEA publishers: NJ.
- Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Englewood Cliffs, NJ: Prentice Hall.
- Greer T. G. (2004). *Detection of differential item functioning (DIF) on the SATV: A comparison of four methods: Mantel-Haenszel, logistic regression, simultaneous item bias and likelihood ratio test* (Doctoral dissertation, University of Houston, Texas). Retrieved from <http://search.proquest.com/pqdtglobal/docview/305196900/EE5E5B40D7A34E5DPQ/1?accountid=8319>
- Gresham, G. (2010). A study exploring exceptional education pre-service teachers' mathematics anxiety. *IUMPST: The Journal*, 4, 1-14.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Jaggernauth, J. S. (2010). *Mathematics anxiety and the primary school teacher: An exploratory study of the relationship between mathematics anxiety, mathematics teacher efficacy, and mathematics avoidance*. EDRS6900: Project Report. The University of the West Indies.
- Kaptan, S. (1995). *Bilimsel araştırma ve istatistik teknikleri* (10. Baskı). Ankara: Rehber Yayınevi.
- Kılınç, M. (2011). A perceptual scale for measurement and evaluation of prospective teachers self-efficacy in education. *Journal of Kırşehir Education Faculty*, 12, 4, 81-93.
- Kilmen, S., & Çıkrıkçı-Demirtaşlı, N. (2009). The Perceptions of primary school teachers about their application levels of measurement and evaluation principles. *Ankara University Journal of Faculty of Educational Sciences*, 42(2), 027-054.
- Kottke, J. L. (2000). Mathematical proficiency, statistics knowledge, attitudes toward statistics, and measurement course performance. *The College Student Journal*, 34, 334-347.
- Le, D. T. (2013). *Applying item response theory modeling in educational research* (Doctoral dissertation, Iowa State University). Retrieved from <http://search.proquest.com/pqdtglobal/docview/1450045295/DD5A94235C534A49PQ/1?accountid=8319>.
- Matteucci, M., & Stracqualursi, L. (2006). Student assessment via graded response model. *STATISTICA*, 4, 435-447.
- Ozan, C., & Köse, E. (2013). Adaptation of attitudes toward educational measurement inventory (ATEMI) to Turkish. *E-International Journal of Educational Research*, 4 (2), 29-47.
- Özbaşı, D., & Demirtaşlı-Çıkrıkçı N. (2013). Primary school teachers' perceptions of their competencies regarding measurement and evaluation in terms of some variables. *Ankara University Journal of Faculty of Educational Sciences*, 46(2), 25-46.
- Perkins, K. K., Adams, W. K., Pollock, S. J., Finkelstein, N. D., & Wieman, C. E. (2005). Correlating student beliefs with student learning using the Colorado learning attitudes about science survey. *AIP Conference Proceedings*, 790(1), 61-64.
- Reed, H. C., Drijvers, P., & Kirschner, P. A. (2010). Effects of attitudes and behaviours on learning mathematics with computer tools. *Computers & Education*, 55(1), 1-15.

- Shih, C. C., & Gamon, J. (2001). Web-based learning: Relationships among student motivation, attitudes, learning styles, and achievement. *Journal of Agricultural Education*, 42(4), 12-20.
- Thissen, D. (2001). IRTL RDIF v.2.0b: *Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland and H. Wainer (Eds.). *Differential item functioning*. Lawrence Erlbaum.
- Ulutaş, S. (2003). *Investigating the competency of teachers in high schools in measurement and evaluation and level of application principles of measurement and evaluation* (Master thesis, University of Ankara). Retrieved from <https://tez.yok.gov.tr/>.
- Wise, S.L., Lukin, L.E. & Roos, L.L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, 42(1), 37-42.
- Yıldırım, H. H. (2006). *The differential item functioning (DIF) analysis of mathematics items in the international assessment programs* (Doctoral dissertation, METU). Retrieved from <https://tez.yok.gov.tr/>.

GENİŞ ÖZET

Giriş

Ölçme ve değerlendirme dersi öğretmen yetiştiren lisans programlarında ve öğretmenlik sertifikası programlarında okutulan zorunlu bir derstir. Bu dersin, genel öğretmenlik yeterliklerini geliştirmek üzere verilen temel dersler içinde matematiksel ve istatistiksel işlemlerle ilgili konuları da içermesi nedeniyle “sayısal” bir ders olarak algılanması, öğretmen adaylarının bu derste başarılı olma konusunda kaygılı olmaları, bu derse ilişkin öğretmenlerin ve öğretmen adaylarının tutumlarının bilinmesinin gerekliliği göz önünde bulundurulduğunda geçerli ve güvenilir bir ölçme aracına ihtiyaç vardır. Literatürdeki ölçme araçlarının Klasik Test Kuramına (KTK) dayalı olarak geliştirilmesi, farklı gruplarda tekrar geçerlilik ve güvenilirlik kanıtları toplanmasını gerektirmektedir. Madde Tepki Kuramı’nda (MTK) ise varsayımlar sağlandığında, bir ölçme aracı aynı evrenden geldiği bilinen alt örneklemelerde de ölçme amacına hizmet etmektedir. Bu nedenle bu çalışmada, ölçme aracının yapısına uygun MTK modeli ile ölçmeleme yapılmasına gereklilik görülmüştür.

Bu çalışmada, ölçme ve değerlendirme dersine yönelik tutumu (ÖDET) ölçmek üzere Demirtaşlı (2002) tarafından geliştirilen Likert tipi tutum ölçeğinin madde tepki kuramına dayalı çok kategorili puanlanan modellerden Samejima (S) Kademeli Tepki Modeli-KTM (Graded Response Model-S-GRM), Kısmi Puan Modeli-KPM (Partial Credit Model) ve Sınıflamalı Tepki Modeli’ne-STM (Nominal Response Model) (Embretson ve Reise, 2000) göre yeniden ölçeklenip psikometrik özelliklerinin karşılaştırılarak en uygun MTK modelinin saptanması ve bu modele göre ölçeğin psikometrik niteliklerinin betimlenmesi amaçlanmıştır.

Yöntem

Betimsel türdeki bu araştırmanın katılımcıları, Türkiye’nin üç ilindeki üç devlet üniversitesinin öğretmen yetiştiren fakülte ve programlarına devam eden, ölçme ve değerlendirme dersi almış, toplam 519 öğretmen adayından oluşmaktadır.

Araştırma kapsamında kullanılan ölçeğin dördümlü Likert tipi dereceli toplamalı tepki vermeye uygun 41 maddeden oluşan ilk versiyonunda (Demirtaşlı, 2002), her bir madde dört kategoriden birinde tepkide bulunmaya uygundur (1=hiç katılmıyorum, 2=katılmıyorum, 3=katılıyorum ve 4=tamamen katılıyorum).

Bu çalışmada, verilerin analizi üç aşamada gerçekleştirilmiştir. İlk aşamada veriler, MTK’nın temel varsayımları olan tek boyutluluk ve yerel bağımsızlık bakımından incelenmiştir. Ölçeğin başat bir boyutu ölçüp ölçmediği Temel Bileşenler Analizi (TBA) ile incelenmiştir. Veriler SPSS 15.0

programı ile analiz edilmiştir. Bulgular, maddelerin TBA için uygun olduğunu göstermektedir. Analiz sonucu, ÖDET'in başat bir faktörü yokladığı sonucuna ulaşılmıştır. Bu durum, tek boyutluluk varsayımının karşılanmasıyla yerel bağımsızlık varsayımının da karşılandığını göstermektedir (Hambleton ve Swaminathan, 1985).

Verilerin analizinin ikinci aşamasında, ölçekteki maddelerin kız ve erkek öğretmen adayları arasında ölçülen özellik dışında cinsiyete göre madde işlev farklılığı (MİF) (Differential Item functioning) gösterip göstermediği incelenmiştir. Bu amaçla çoklu puanlanan maddeler için MİF'i belirlemeyi sağlayan PSIBTEST (Polytomous Simultaneous Item Bias Test) ve Madde Tepki Kuramı Olabilirlik Oran Testi (MTK-OO) (Item Response Theory Likelihood Ratio Test) yöntemleri kullanılmıştır. MİF analizlerinde DIFPACK 1.7 ve IRTLRDIF 2.0 paketleri kullanılmıştır. İki yöntemle de ortak olarak erkekler lehine MİF gösteren iki madde olduğu görülmüş ve uzman görüşleri sonucu çıkarılmasının uygun olduğuna karar verilmiştir.

Verilerin analizinin üçüncü aşamasında, kalan 39 madde MULTILOG 7.03 paket programında Samejima KTM, KPM ve STM modeline göre analiz edilmiştir. Model tercihi test bilgi fonksiyonu değeri ve güvenilirlik değerine göre yapılmıştır.

Sonuç ve Tartışma

Bu araştırmada öğretmen yetiştirme programında zorunlu ders kategorisinde olan ölçme ve değerlendirme dersine yönelik tutumları ölçmek üzere geliştirilmiş olan bir aracın (ÖDET) MTK'ya göre yeniden ölçeklenerek, geliştirilmesi amaçlanmıştır. Çok kategorili veriler için geliştirilen MTK modellerine göre madde ve test parametreleri kestirilen ÖDET'in en yüksek uyum gösterdiği MTK modelini saptamak üzere, modellerden kestirilen -2 loglikelihood değerleri ikili olarak karşılaştırılmıştır. İlk olarak KPM ve KTM'ye ait -2 log χ^2 değerlerinin farkı, Kay-Kare istatistiği ve serbestlik derecesi değerlendirilmiş ve veriler için Kademeli Tepki Modeli-KTM'nin daha uygun olduğu görülmüştür. İkinci aşamada, KTM ve STM'ye ait kestirime ilişkin -2 log χ^2 değerlerinin farkı alınarak ilgili serbestlik derecesi ile tabloda verilen Kay-Kare istatistiği ile karşılaştırılmıştır. Modeller arasında bir farklılık olmadığı tespit edilmiştir. Analizler sonucu, KTM ve STM'nin veriye uyumu açısından manidar farklılık olmamakla beraber KTM'nin maksimum bilgi ve güvenilirliğinin daha yüksek olması sebebiyle, kestirimler bu modele göre yapılmıştır. KTM'ye göre yapılan incelemeler sonucu, 39 maddeden altısının (6, 7, 10, 18, 31, 41 nolu maddeler) madde bilgi düzeylerinin düşük olduğu (.45'in altında) görülmüş ve ölçekten çıkarılmıştır. Ayrıca her bir kategori için gözlenen ve beklenen oranlar arasındaki farklar da incelenmiş, maddelerin veriye uyumlu olduğu görülmüştür.

KTM ile elde edilen sonuçların kısmi puanlama modelinden daha fazla uyum göstermesi, Kademeli Tepki Modeli'nde tepki kategori eşik değerlerinin sıralı olması koşulunun bulunmasından kaynaklı olabilir. Bu durum "hiç katılmıyorum, katılmıyorum, katılıyorum ve tamamen katılıyorum" ifadelerinin sıralı tepkiler ifade ettiğinin de bir göstergesi olarak yorumlanabilir. KTM'ye göre ölçekleme sonucu, ölçekten elde edilen maksimum bilgi 48.23 ve bu bilgi miktarını veren tutum düzeyi -1.40'dır.

Yapılan çalışma, öğretmen adaylarının ölçme ve değerlendirme dersine yönelik tutumunu belirlemek üzere MTK'ya dayalı olarak geliştirilen 33 maddelik ÖDET'in oldukça geçerli ve güvenilir bir ölçme aracı olduğunu göstermektedir.

MTK'ya dayalı bir model olan KTM ile ölçeklenen ölçeğin model veri uyumunun sağlanması, ölçeğin farklı gruplarda uygulansa da değişmez parametre kestirimleri elde edilmesini sağlar. Bu araç, yeni haliyle farklı öğretmenlik programlarındaki aday öğretmenlerin ölçme ve değerlendirme dersi öncesi ve sonrası tutumlarının karşılaştırmasını konu alan araştırmalarda, olumsuz tutuma sahip olduğu saptanan gruplarda dersin öğretimi sürecinde yürütülecek çeşitli manipülasyonların etkisinin değerlendirileceği çalışmalarda geçerli ve güvenilir bir araç olarak kullanılabilir. Bunun yanında,

ölçme ve değerlendirme derslerindeki başarı ve bu derse yönelik tutum arasındaki ilişkinin inceleneceği arařtırmalarda da kullanılabilir.

Ölçeklerin Yapı Geçerliğini Belirlemede Bir Yöntem Olarak Yapay Sinir Ağı Modelinin İncelenmesi: Kendini Düzenleyen Haritalama

Investigation On Artificial Neural Network Model As A Method In Determining Construct Validity Of Scales: Self-Organizing Mapping

Esin TEZBAŞARAN *

Öz

Bu araştırmada, yapay sinir ağı modellerinden biri olan kendini düzenleyen haritalama yönteminde kullanılan nöron sayısının, ölçeklerin yapılarını ortaya çıkarmadaki etkililiği incelenmektedir. Bu yöntemde haritalama yaparken kullanılan nöron sayısı değiştiğinde, ölçek yapısı farklılaşabilmektedir. Bu çalışmada ölçek yapısını ortaya çıkarmada en uygun nöron sayısının tespiti için kullanılan yöntemler doğrulayıcı faktör analizi ve farklı gruplar yöntemidir. Araştırma bulguları, nöron sayısı arttıkça ölçek yapısının tek faktöre indirildiği aynı zamanda, tek faktörde ortaya çıkan bu ölçeğin uyum ve hata indeksleri açısından diğerlerine göre daha uyumlu bir model meydana getirdiğini ortaya çıkarmıştır. Bunun yanı sıra farklı gruplar yönteminden elde edilen kanıt da bu durumu desteklemektedir. Sonuç olarak, yapı geçerliği için kullanılan kendini düzenleyen haritalama yönteminde, ölçeğin ilişkili maddeleri tek nöronda toplanıncaya kadar nöron sayısının artırılması önerilmektedir. Buna ek olarak, ortaya çıkan yapının bağlamsal olarak ölçülen tutum değişkeni açısından analiz edilmesi de önerilmektedir.

Anahtar Kelimeler: Yapı geçerliği, kendini düzenleyen haritalama

Abstract

This research investigates the efficiency of number of neurons used in self-organizing mapping, one of the artificial neural network models, on detecting construct of scales. In this method, the construct of a scale can differ as the number of neurons used for mapping changes. In this study, the methods for determining the optimum number of neurons to detect the construct of a scale are confirmatory factor analysis and distinct groups method. The research findings reveal that increasing neuron numbers causes reduction through one factor of the scale construct and, at the same time, the fit and error indexes of this scale constructed in one factor have more fitted model than those of others. In addition to this, the evidence from distinct groups method supports this finding. As a result, it is recommended that the number of neurons should increase till the related items of a scale are gathered in a neuron in the method of self-organizing mapping used for construct validity. In addition to this rule, it is recommended to analyze the revealed construct with respect to related attitude variable contextually.

Keywords: Construct validity, self-organizing mapping

GİRİŞ

Yapı geçerliği psikolojik değişkenleri ölçme amacıyla geliştirilen araçlar için geçerlik kanıtı olabilecek önemli bir psikometrik özelliktir. Yapı geçerliği, psikolojik bir yapıyı ya da özelliği, ölçme aracının ne derecede ortaya koyabildiği ile ilgilidir (Anastasi, 1988). Bir ölçme aracının maddeleri, yapının değerlendirilmesine dönük bir araçlar takımındır. Çünkü, değişkene ait yapı doğrudan gözlenemez (De Vellis, 2003). O halde soyut olan psikolojik yapı ile gözlenebilir, somut davranışların ilişkisini ortaya koyma durumu, yapı geçerliğini belirlemede temel anahtardır (Murphy ve Davidshofer, 2001).

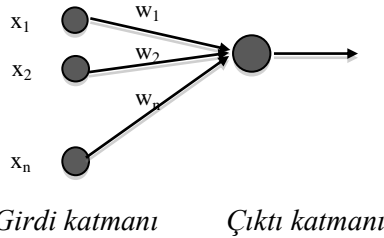
* Dr., İstanbul Üniversitesi, Hasan Ali Yücel Eğitim Fakültesi, İstanbul-Türkiye, tezbasaranesin@gmail.com

Bir ölçme aracının yapı geçerliği, farklı yöntemlerle kanıtlar elde edilerek ve bu kanıtlar değerlendirilerek saptanır. Bu yöntemlerden birisi, özellikle Likert tipi ölçek geliştirme uygulamalarında, temel bileşenler analizi tekniği kullanarak, ölçme aracının boyutlarını ortaya koyan faktör analizi yöntemidir.

Temel bileşenler analizinin genel amacı, veri azaltma ve veriyi yorumlamadır (Johnson ve Wichern, 1982). Bu analiz doğrusal bir analiz tekniği olması nedeni ile doğrusal olmayan ilişkileri ortaya çıkarmada yetersizdir. Ayrıca bu analiz kullanılabilmesi için değişkenlerin sürekli olması ve normallik sayıtlarının karşılanması gerekir. 1943 yılından itibaren yapay sinir ağı modellerinin gelişmesi ile bu modellerin veri analizlerinde kullanım alanları da genişlemiş ve birçok parametrik yöntemler için gerekli olan sayıtlara gerek duymayan yapay sinir ağı modelleri kullanımı yaygınlaşmıştır. Artık birçok istatistiksel yöntemler, yapay sinir ağları üzerinden modellenenmektedir (Cheng ve Titterington, 1994). Literatürde, Kohonen ağı olarak da bilinen Kendini Düzenleyen Haritalama (KDH) modelinin ise teori ve uygulamada, faktör analizi yerine kullanılabildiğini gösteren çalışmalar mevcuttur. Özellikle Kohonen (2013), KDH'nın istatistiksel yöntemlerden açılımlayıcı veri analizi yerine kullanılabileceğini belirtir.

Kohonen (1982), kendini düzenleyen bir haritalama yapan ağın temel prensibinin, gözlenen olayların özelliklerine göre geometrik düzlemde gerçeğe uygun haritalarını otomatik olarak düzenlemesi olarak tanımlar. KDH, dış dünyadan gelen sinyallere göre girdi özelliklerinin haritasını çıkaran bir ağ modelidir. Bu özelliği nedeni ile KDH, çok boyutlu bir yapıya sahip olan veri setinin boyut sayısını indirgeyerek, iki boyutlu bir haritada görselleştirir. Bu ağ, çok boyutlu girdi sinyallerini işleyerek, iki boyutlu düzlemde soyut bilgiyi ortaya koyabilen bir yapay sinir ağı yapısıdır (Kohonen,1990).

KDH tek katmanlı mimari yapıda, ileri beslemeli, denetimsiz öğrenme yöntemine sahip bir ağ modelidir (Şekil 1). KDH için kullanılan öğrenme yöntemi, denetimsiz öğrenme yöntemidir; beklenen bir çıktı değeri yoktur. O yüzden ölçüt, girdilerin benzerlik ya da farklılıkları üzerinden belirlenir ve çıktı katmanında ağın tespit ettiği benzerliklere göre kümeler oluşur. Bunun için her bir nöron, girdi vektörü ile ağırlık vektörü arasında oklit uzaklığını hesaplar.



Şekil 1. Tek Katmanlı İleri Beslemeli Yapay Sinir Ağı Mimarisi

KDH, yarışmacı öğrenme kuralı kullanır. Yarışmacı öğrenme kuralı kullanan ağ yapısı ile, girdilerin özelliklerini göz önüne alarak kümelemesi, KDN'nın temel işlevidir. Buna göre algoritmada eğitim girdi vektörlerinin ($x_i(t)$), c kümelerine sınıflandırılması için ağırlık güncellemesini ($w_c(t+1)$) koşullu olarak formüllendirilir. $\alpha(t)$, her adımda monoton olarak azalan skaler bir değerdir. Genellikle $\alpha(t)$ için 0,01 ya da 0,02 başlangıç değeridir.

$$w_c(t+1) = w_c(t) + \alpha(t)[x_i(t) - w_c(t)] \quad \text{Eğer } x \text{ doğru sınıflandırılmış ise,}$$

$$w_c(t+1) = w_c(t) - \alpha(t)[x_i(t) - w_c(t)] \quad \text{Eğer } x \text{ yanlış sınıflandırılmış ise,}$$

$$w_i(t+1) = w_i(t) \quad \text{Eğer } i \neq c$$

KDH veri hazırlama, modelleme ve kümeleme amacıyla kullanılabilmesinin yanında, bilginin keşfedilmesi sürecinde veriyi anlama açısından da oldukça uygun bir yöntemdir (Vesanto, Himberg, Alhoniemi ve Parhankangas, 2000). KDH ile veriler (girdiler) üzerinden soyut bilginin ortaya çıkarılması ve yapı geçerliği arasında önemli bir örtüşme olduğu düşüncesiyle, denemelik ölçeklerin yapı geçerliği için KDH kullanımının incelenmesi önemlidir.

Araştırmanın Amacı

KDH modeli için bir takım desen parametreleri vardır: Nöron serisinin boyutluluğu, her bir boyuttaki nöron sayısı, nöronların birbiri ile komşuluk ilişkisi/şekli, komşuluğun yakınlaşma planı ve ağıın öğrenme hızı (Jain, Mao ve Mohiuddin, 1996). KDH sonuçlarını etkileyen parametreler içinde, modelde önceden belirlenen nöron sayısı, çıktı haritasıyla ortaya çıkan bilgiyi değiştirebilmektedir. Bu değişkenlik, ölçeklerin yapısı hakkında elde edilen bilginin de değişmesi anlamına gelmektedir. Bu düşünceden yola çıkarak bu çalışmada, çıktı haritasında bulunan nöronların sayılarındaki artış ya da azalışın, bir ölçeğin yapı geçerliği çalışmasına nasıl yansıtacağına ve yapı geçerliği çalışmasında en uygun nöron sayısının ne olacağına yönelik KDH modelinin incelenmesi amaçlanmıştır.

Bu amaçla tasarlanan çalışmada, yapay sinir ağı modellerinden biri olan KDH modelinde farklı nöron sayıları üzerinden yapılan haritalamada, Tezbaşaran ve Kelecioğlu (2004) tarafından hazırlanan denemelik Sigaraya İlişkin Tutum Ölçeğinin yapı geçerliği saptamada en uygun nöron sayısını tespit edebilmek için aşağıdaki sorulara cevap aranmaktadır:

1. Farklı nöron sayıları üzerinden uygulanan KDH yöntemi sonuçlarına göre ölçek yapısı nasıldır?
2. KDH yönteminde değişen nöron sayıları sonucunda ortaya çıkan ölçek yapılarının doğrulayıcı faktör analizi sonuçlarına göre uyum ve hata indeksleri nedir?
3. KDH yönteminde değişen nöron sayıları sonucunda ortaya çıkan ölçek yapılarının farklı gruplardan elde edilen kanıtlara göre yapı geçerliği nasıldır?

Sınırlılıklar

Bu araştırmanın sonuçları yalnızca katılımcılardan elde edilen verilerle sınırlıdır, genelleme yapılamaz. Ayrıca, cinsiyet ve sigara kullanımı açısından araştırma grubunun frekans dağılımları arasındaki açıklık araştırma sonuçları açısından önemli bir sınırlılıktır. Bunun yanı sıra, KDH’da kullanılan ağıın başlangıç ağırlık değerleri ve girdi sırası her seferinde rasgele atandığından dolayı, KDH analizinin verdiği sonuçlar tekrarda aynı olmayabilir.

YÖNTEM

Araştırmanın Türü

Bu araştırma, KDH modeli kullanan yapay sinir ağı modelinin ölçeklerin yapı geçerliğine kanıt oluşturmak üzere en uygun nöron sayısının tespitine yönelik olduğundan, temel çalışmadır. Temel araştırmalar, var olan bilgiye yenilerini katan araştırmalardır (Karasar, 1998). Bu araştırmalar tarafından üretilen bilgiler pratikte hemen uygulanabilir bilgiler değildir ancak, uygulamaya dönük çalışmalar için gerekli bilgi kaynağını ya da altyapısını oluştururlar (Elmes, Kantowitz ve Roediger III, 1992; Bogden ve Biklen, 1998).

Çalışma Grubu

Bu çalışmaya üniversitede okuyan ya da mezun olmuş 18-56 yaş grubundaki 540 birey, gönüllü olarak katılımcı olmuşlardır. Sigaraya İlişkin Tutum Ölçeğinin uygulandığı grubun %83,5’i kadın,

%16,5 'i erkektir. Kadınların %15'i, erkeklerin ise %75'i sigara kullandığını belirtmişlerdir. Bireylerin %59,6'sı üniversite öğrencisi, %40,4'ü ise üniversite mezunu olup; katılımcılar 48 farklı meslek gruplarına ait bölümlere dağılım göstermektedir.

Veri Toplama Aracı

Veri toplama aracı iki bölümden oluşmaktadır. Birincisi demografik bilgilerin alındığı ve sigaraya ilişkin sorulardan oluşan anket ve Tezbaşaran ve Kelecioğlu (2004) tarafından hazırlanan Sigaraya İlişkin Tutum Ölçeğidir. 41 maddeden oluşan denemelik ölçek 5'li Likert tipinde bir ölçek olup, 19 maddesi olumlu, 22 maddesi olumsuz ifadelerden oluşmaktadır. Orjinal çalışmada, ölçeğin güvenilirliğinin 1322 kişi üzerinden yapılan uygulama ile 0,92 olduğu ve açımlayıcı faktör analizi sonucu ölçekte iki bileşen gözlemlendiği belirtilmektedir. Birinci bileşenin açıkladığı varyans %25,41, ikinci bileşenin açıkladığı varyans ise %7,25'dir (Tezbaşaran ve Kelecioğlu 2004).

Verilerin Analizi

Öncelikle Sigaraya İlişkin Tutum Ölçeğindeki sigaraya ilişkin olumsuz maddeler tersten puanlanarak, yüksek puanların olumlu tutum, düşük puanların olumsuz tutumu ifade edecek şekilde düzenlenmesi yapılmıştır. Araştırmanın birinci alt problemini cevaplamak üzere, ölçek üzerinden elde edilen veriler girdi değerleri olarak, yapay sinir ağı modeli olan KDH yazılımının bulunduğu Matlab-SOM toolbox programında farklı nöron sayıları üzerinden uygulanmış; ölçeğin yapısı faktörler ve bu faktörleri oluşturan maddeler üzerinden belirlenmiştir. Sonuçlar vuruş (hit) histogramı üzerinden görselleştirilmiştir. Bu çizim verinin haritadaki dağılımını görmek açısından iyi bir araçtır (Westerlund, 2005). Aynı zamanda yazılıma, nöronlarda toplanan maddelerin hangileri olduğunu tanımlayacak bir kod eklenmiştir. Yazılımda kullanılan iterasyon sayısı 500 olarak belirlenmiştir. Ağın çıktısı sırasıyla 2x2, 3x3, 4x4, 5x5, 6x6, 7x7 nöron haritası üzerinden elde edilmiştir. KDH iki boyut üzerinden haritalandırma yapar ve bu araştırma için her bir boyuttaki nöron sayısı n olmak üzere, $n \times n$ boyutunda, altıgen ızgara yapısında görüntüleme kullanılmıştır.

Araştırmanın ikinci alt problemi için 6 farklı haritalamadan çıkan ölçek yapıları doğrulayıcı faktör analizine tabi tutulmuştur. DFA, denence olarak ortaya konan yapının veri seti ile iyi bir uyum gösterip göstermediğini belirler (Holtzman ve Vezzu, 2011). Bu amaçla, Lisrel 9.2 programı ile yapılan doğrulayıcı faktör analizi sonuçlarında ortaya çıkan her bir yapının uyum ve hata indeksleri belirlenmiştir.

Ayrıca üçüncü alt problemde nöron sayılarına göre farklılaşan ölçek yapılarının yapı geçerliği açısından kuvvetli olanı tespit için, diğer bir yapı geçerliği belirleme yöntemi olan farklı gruplar yöntemi kullanılmıştır. Eğer bir ölçme aracı istenen yapıyı geçerli bir biçimde ölçüyor ise, bu yapıya ilişkin farklı olduğu bilinen grupları, ölçüm sonucu üzerinden ayırabilmelidir. (Cohen, Montague, Nathanson ve Swerdlık, 1988). Bu amaçla farklı gruplar sigara içenler ve içmeyenler olarak belirlenmiştir. Çünkü Feldman (1997), davranışlar ve tutumlar arasındaki bağı kuvveti değişmekle birlikte, genel olarak insanların tutum ve davranışları arasında bir tutarlılık sağlamak için çaba gösterdiklerini belirtir. Buna göre, 6 farklı ölçek yapısına ait puanlar üzerinden, iki grup arasındaki puan ortalamaları farkının manidarlığı, bağımsız gruplar için t testi ile SPSS 21.0 programı kullanılarak test edilmiştir. Ancak araştırma grubunun sigara kullananların sayısı (90) ile kullanmayanların sayısı (450) arasında yüzde olarak büyük açıklık vardır (yaklaşık %66,7). Bu nedenle analiz, sigara kullanmayanlar arasından rasgele yöntemle seçilen 90 kişilik grup belirlenerek yapılmıştır.

BULGULAR

Bulgular, çalışmada belirlenen problemlerin sırasına göre birer başlık altında verilmiştir.

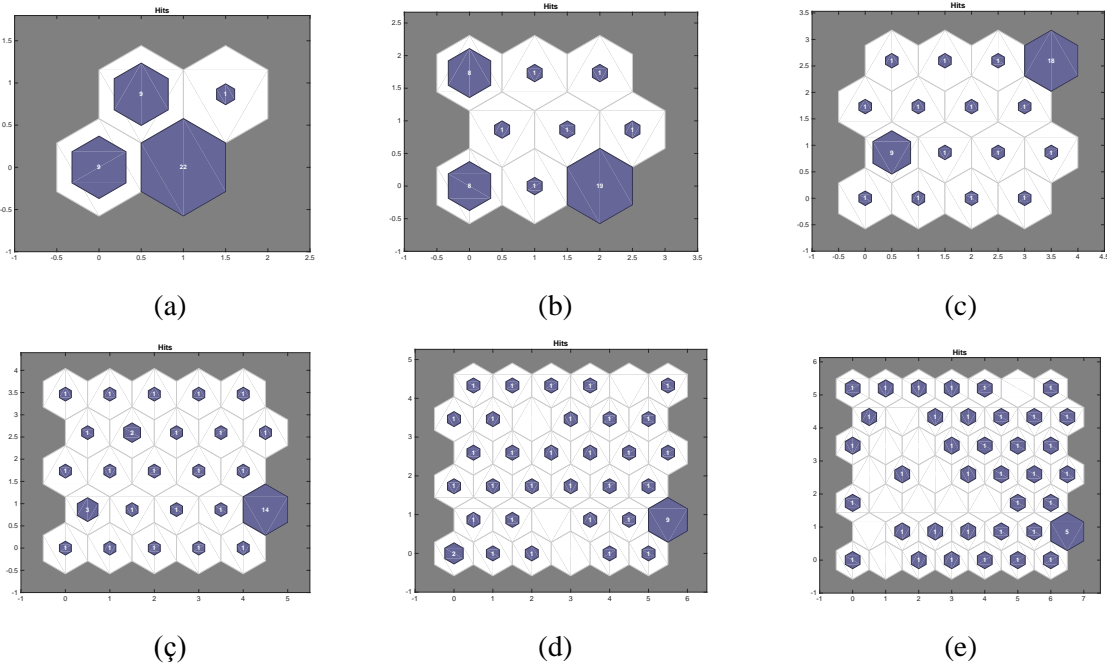
Farklı Nöron Sayılarında Haritalanan Ölçek Yapıları

Başlangıç KDH için 2x2 üzerinden 4 adet nöron kullanılmıştır. Bunun nedeni tek nöron üzerinde ölçeğin tüm maddelerinin görüntülenecek olmasıdır. İki boyutlu haritalamada, Sigaraya İlişkin Tutum Ölçeğinin maddeleri 4 nörona dağılmıştır. Ölçekteki 2. madde diğerlerinden ayrılarak dördüncü nöronda olmak üzere, ilk nöronda 9 madde, ikincisinde 22 ve üçüncüsünde 9 madde bulunmaktadır (Şekil 2-a). Buna göre 2. maddenin bulunduğu nöron dışta tutulmak kaydı ile ölçek yapısının üç faktörlü olduğu söylenebilir (Tablo 1).

İki boyutlu haritalamada nöron sayısı 3x3 olmak üzere 9 olarak belirlendiğinde ortaya yine üç faktörlü bir yapı ortaya çıkmıştır. Nöronların birincisinde 8, üçüncüsünde 19 ve yedincisinde ise 8 madde toplanmıştır. Kalan nöronlar diğer 6 maddeyi paylaşmıştır (Şekil 2-b).

Nöron sayısı 4x4 üzerinden 16'ya çıkarıldığında beşinci nöronda 8, on altıncı nöronda 18 madde görülmektedir. Böylelikle faktör sayısının ikiye düştüğü gözlenmiştir (Şekil2-c). Nöron sayısı 5x5 üzerinden 25 olduğunda, maddeler farklılıklarına ve benzerliklerine göre daha da ayrılarak altıncı nöronda 3, onuncu nöronda 14 madde, on yedinci nöronda 2 madde ile faktör yapısı tekrar üç olmaktadır (Şekil2-ç). Ancak bir faktörde 2 maddenin bulunması istenmeyen bir durumdur. Kline (2005), özellikle örnek sayısı küçük ise faktör başına en az üç değişkenin bulunmasını önerir.

Haritadaki toplam nöron sayısı 36'ya çıktığında, 5x5'lik haritadaki maddelerin de birbirinden ayrıştığı ancak, 3 maddelik faktöründe ayrılarak ikiye düştüğü ve birinci nöronda iki maddenin kaldığı görülmektedir (Tablo 1). Birincil faktör olarak 9 maddelik bir yapı ortaya çıkmıştır (Şekil 2-d). Maddelerin ayrışmasını boyuttaki nöron sayısını bir artırarak 7x7'lik harita üzerinden incelendiğinde (Şekil 2-e), 5 maddelik tek boyutlu bir yapının ortaya çıktığı görülmektedir. Bunun üzerine nöron artışı durdurulmuştur. Çünkü nöron sayısı artırıldığında maddelerin her biri bir nöronda olmak üzere dağılmıştır.



Şekil 2. KDH Sonucu Maddelerin Nöronlara Dağılımını Gösteren Vuruş Histogramı: a) 2x2 b) 3x3 c) 4x4 ç) 5x5 d) 6x6 e) 7x7

Tablo 1. Farklı Nöron Sayılarında KDH Sonuçlarına Göre Maddelerin Faktörlere Dağılımı

| Nöron Sayısı | Faktörler* | | |
|--------------|-----------------------------------|--|-----------------------------------|
| | 1. Faktör | 2. Faktör | 3. Faktör |
| 2x2 | 6, 13, 18, 19, 27, 31, 32, 38, 41 | 1, 3, 5, 7, 8, 9, 10, 12, 15, 16, 21, 23, 24, 25, 28, 29, 30, 33, 34, 37, 39, 40 | 4, 11, 14, 17, 20, 22, 26, 35, 36 |
| 3x3 | 6, 13, 18, 19, 27, 31, 32, 41 | 1, 5, 7, 8, 9, 10, 12, 15, 21, 23, 24, 25, 28, 29, 30, 33, 34, 37, 39 | 4, 11, 14, 17, 20, 22, 35, 36 |
| 4x4 | 4, 13, 14, 17, 20, 26, 36, 38 | 1, 2, 3, 5, 8, 15, 21, 23, 24, 28, 29, 30, 32, 33, 34, 37, 39, 40 | |
| 5x5 | 14, 35, 36 | 1, 5, 9, 10, 15, 21, 23, 24, 28, 30, 32, 33, 37, 39 | 11, 17 |
| 6x6 | 35, 36 | 1, 3, 5, 8, 21, 23, 24, 34, 39 | |
| 7x7 | 1, 5, 7, 21, 23** | | |

*Faktör numaraları nöron numaralarının sıralamasına göre verilmiştir.

** Önceki Nöron sayılarına göre KDH sonuçlarında 2. Faktör içinde yer alan maddelerdir.

Ölçek Yapılarının Uyum ve Hata İndeksleri

Farklı nöron sayıları üzerinden yapılan KDH analizinde, Sigaraya İlişkin Tutum Ölçeğinin yapısında, farklı madde ve faktör sayısı olduğu gözlenmiştir. Bu durumda hangi yapının kabul edilebilir olduğunu tespit etmek için doğrulayıcı faktör analizi yapılarak elde edilen uyum ve hata indekslerinden yararlanılmıştır.

İlk olarak X_M^2 uyum istatistiğinin serbestlik derecesine bölümü üzerinden elde edilen uyum indeksine göre hepsi kabul edilebilir uyuma sahip olmakla birlikte 7x7'lik KDH sonucunda oluşan 5 maddeden oluşan tek faktörlü yapı en iyi uyumu vermektedir. Ancak Ullman (2001), bu oranın model uyumu için kaba bir gösterge olduğunu belirtir.

İkinci olarak RMSEA hata indeksi ele alındığında, Hesaplanan değer 0,05'den küçük olması beklenir ancak, 0,05-0,08 arasındaki değerle de kabul edilebilir bir uyumu sergiler (Kline, 2005). Buna göre hepsi kabul edilebilir sınırlar içinde olmakla birlikte 7x7'lik KDH sonucu en düşük hata indeksine sahiptir.

GFI ve AGFI değerlerinin kabul edilebilir sınırları 0,90 ve yukarıdır (Kline, 2005). Buna göre 6x6 ve 7x7'lik KDH sonuçlarında oluşan yapılar kabul edilebilir düzeydedir. NFI, NNFI ve CFI uyum indeksleri de 0,90 ve yukarıdır değerleri iyi bir uyum indeks değeri olarak kabul edilir (Kline, 2005). Buna göre ölçek yapılarının hepsi kabul edilebilir düzeydir. Ancak KDH analizinde, nöron sayısı arttıkça, ortaya çıkan daha az sayıda madde ve faktör sayısı üzerinden oluşan yapıların uyum indekslerinin arttığı gözlenmektedir (Tablo 2).

Tablo 2. Farklı Nöron Sayılarında KDH Sonuçlarına Göre SİTÖ yapısının DFA Uyum ve Hata İndeksleri

| Uyum ve Hata İndeksleri | 2x2 | 3x3 | 4x4 | 5x5 | 6x6 | 7x7 |
|-------------------------|-------------------|-------------------|------------------|------------------|-----------------|--------------|
| X_M^2 / sd | 1998,79/732= 2,73 | 1432,75/524= 2,73 | 806,96/298= 2,71 | 438,09/149= 2,94 | 102,03/43= 2,37 | 9,62/5= 1,92 |
| RMSEA | 0,065 | 0,063 | 0,061 | 0,063 | 0,050 | 0,042 |
| GFI | 0,82 | 0,85 | 0,89 | 0,92 | 0,97 | 0,99 |
| AGFI | 0,79 | 0,83 | 0,87 | 0,89 | 0,95 | 0,98 |
| NFI | 0,92 | 0,92 | 0,92 | 0,93 | 0,94 | 0,96 |
| NNFI | 0,94 | 0,94 | 0,95 | 0,95 | 0,96 | 0,96 |
| CFI | 0,95 | 0,95 | 0,95 | 0,95 | 0,97 | 0,98 |

Ölçek Yapılarının Farklı Gruplar Üzerinden İncelenmesi

Nöron sayılarına göre KDH'dan elde edilen farklı ölçek yapıları, sigara içenler ve içmeyenler olarak belirlenen iki farklı grubun, altı farklı ölçek puanlarının ortalamaları bağımsız gruplar için t testi ile karşılaştırılmıştır. Yapılan karşılaştırmada farklı nöron sayılarındaki KDH sonuçlarına göre ölçek yapılarının tümünde, sigara kullanan ve kullanmayanların ölçek puanlarının ortalamaları arasındaki farklılığın istatistiksel olarak anlamlı olduğu görülmektedir (Tablo 3). Yalnızca 6x6 ve 7x7'lik nöron haritasına göre çıkan ölçek yapılarından elde edilen grup varyansları homojenliği, Levene testi sonucu reddedilmiştir ($p < 0,05$). Nöron sayılarındaki değişim nedeni ile, KDH analizi sonucu ortaya çıkan 6 farklı ölçek yapısında, sigara kullananların sigaraya ilişkin tutum puanları ortalamaları yüksek ve kullanmayanların ise düşük olduğu görülmektedir.

Tablo 3. Farklı Ölçek Yapıları Açısından Sigara Kullanım Durumlarına Göre Ölçek Puanları İstatistikleri

| Ölçek Yapıları | Sigara Kullanımı | N | \bar{X} | Ss | Sd | Ortalama Fark | t |
|----------------|------------------|----|-----------|-------|--------|---------------|---------|
| 2x2 | Evet | 90 | 104,60 | 18,00 | 178 | 28,51 | 11,391* |
| | Hayır | 90 | 76,01 | 15,59 | | | |
| 3x3 | Evet | 90 | 90,72 | 15,63 | 178 | 24,32 | 11,244* |
| | Hayır | 90 | 66,40 | 13,30 | | | |
| 4x4 | Evet | 90 | 65,09 | 12,34 | 178 | 17,68 | 10,327* |
| | Hayır | 90 | 47,41 | 10,55 | | | |
| 5x5 | Evet | 90 | 47,52 | 9,06 | 178 | 13,16 | 10,364* |
| | Hayır | 90 | 34,37 | 7,93 | | | |
| 6x6 | Evet | 90 | 25,18 | 5,47 | 168,98 | 6,79 | 9,237* |
| | Hayır | 90 | 18,39 | 4,32 | | | |
| 7x7 | Evet | 90 | 8,58 | 2,50 | 165,52 | 1,78 | 5,387* |
| | Hayır | 90 | 6,80 | 1,86 | | | |

* $p < 0,05$

SONUÇLAR ve TARTIŞMA

Yapay sinir ağı üzerinden KDH analizi sonuçları incelendiğinde, nöron sayısı arttıkça faktörlerde yer alan madde sayısı azalmakta, aynı zamanda son olarak faktör sayısı 1'e düşmektedir. Eğer 2 maddeden bir boyut oluşturan faktörler dışta tutularak KDH sonuçları incelenirse 5x5'lik KDH sonucunun 2 faktörlü, 6x6'lik KDH sonucunun ise tek faktörlü olduğu söylenebilir. Bunun yanı sıra 7x7'lik haritada 5 maddeli tek faktörlü bir yapı kendiliğinden oluşmuştur.

Madde sayıları ve boyutları gittikçe azalan ölçek yapıları doğrulayıcı faktör analizine tabi tutularak uyum ve hata indeksleri açısından karşılaştırılarak incelenmiştir. Buna göre NFI, NNFI ve CFI uyum indeksleri açısından yapılar kabul edilebilir düzeyde saptanmış, 5x5'den başlamak üzere, nöron sayısı 6x6 ve 7x7 üzerinden haritalandığında, NNFI ve CFI indeks değerleri keskin olmayan, küçük artışlar göstermiştir. GFI ve AGFI indeks değerleri ise nöron sayısı arttıkça model uyumlarında gittikçe yükselme gözlenmiştir. Ancak RMSEA hata indeksi ile X_M^2 / sd uyum oranı 4x4 ile iki faktörde görülen yapıdan 5x5 ile tekrar üç faktöre çıkması değerlerinde olumsuz açıdan gerileme gözlenmiş 6x6 ile faktör yapısının azalması ile tekrar yapının uyum oranı artmış ve nihayetinde 7x7 üzerinden yapılan haritalamada 5 maddeli tek faktörlü yapıda en iyi değerler gözlenmiştir. Tezbaşaran (2016) tarafından yapılan çalışmada, nöron sayısındaki artış sonucunda ölçek yapısına ait uyum ve hata indekslerinin daha iyi değerler aldığı belirtilmektedir.

Nöron sayısı arttıkça madde ve faktör sayısı azalan ölçek yapılarının farklı gruplar yöntemi ile yapı geçerlikleri incelenmiştir. Sigara içen ve içmeyen olarak tanımlanan farklı grupların 6 farklı ölçek puanlarına göre, hepsinde sigara kullananların sigaraya ilişkin tutum puanları ortalamaları yüksek, kullanmayanların ise düşüktür. Ayrıca gruplar arası ortalama farkının istatistiksel olarak anlamlı

olması, ortaya çıkan ölçeklerin başka bir açıdan yapı geçerliği kanıtını sağladıkları anlamına gelmektedir. Özellikle madde sayısı 5 olan tek boyutlu yapıda da bu durumun saptanmış olması, madde sayısının azalmasına rağmen istenen yapıyı ölçmedeki yeterliğini ortaya koyması açısından önemlidir.

Bu araştırmada, ölçülen yapının ölçek maddelerinin kapsamı ve toplanan verilerin farklılaşması nedeni ile yapı geçerliği yöntemi olarak kullanılacak KDH modeli için, başlangıçta sabit bir nöron sayısı belirlemenin mümkün olmadığı ortaya çıkmıştır. Bununla birlikte araştırma sonucu olarak, kaç adet nöron kullanılacağına ilişkin ölçüt miktar yerine, bir kural önerilmektedir: Ölçeğin maddeleri tek nöronda toplanıncaya, diğer bir deyişle, yapı tek faktöre indirgeninceye kadar nöron sayısı artırılabilir. Ancak bu kuralı uygularken ortaya çıkan boyutların adlandırılması ve denemelik ölçek oluşturulurken seçilen kuramsal tutum tanımı ve boyutları üzerinden karşılaştırılarak desteklenmesi de önerilmektedir.

KDH denetimsiz öğrenme yöntemine sahiptir. Bir başka deyişle çıktının doğru olup olmadığını belirleyecek bir iç denetimi yoktur. Ölçme araçlarının yapı geçerliğinde sonuçların yapı ile bağlamsal ilgileri oldukça önemlidir. Bu nedenle önceden yapısal boyutları kuramsal ve işevuruk olarak daha belirgin bir ölçme aracı örneğin bir bilişsel testin KDH sonuçları, dış denetim yolu ile incelenmesi ileriki araştırmalar için önerilebilir.

KDH yönteminin denemelik ölçekten nihai ölçeğe madde seçmek için bazı avantajları olduğu söylenebilir. KDH aynı zamanda çok boyutlu bir yapıya sahip olan veri setinin boyut sayısını indirgeyerek, genellikle iki boyutlu bir haritada görselleştirir (Kohonen,1990). Öncelikle bu tekniğin en önemli avantajlarından biridir. Çünkü yapılar çok boyutlu ise maddeleri boyutları üzerinden görselleştirmek mümkün olmamaktadır. Ayrıca iki boyutlu çıktı haritasında doğrusal ilişkilerin yanı sıra, doğrusal olmayan ilişkileri üzerinden de maddelerin boyutlandırılmasına imkan vermektedir. KDH modeli kullanımının, parametrik istatistiksel yöntemlerin aksine, değişkenler üzerinde her hangi bir sayılısı yoktur. Ayrıca tutumların tek boyutluluk sayılısını karşılayabilecek şekilde maddelerin tek boyuta indirgenebilmesini sağlar. KDH yöntemi bir araştırmacının temel aldığı kurama göre geliştirdiği test ya da ölçek maddelerinin, ölçülen yapıyı ortaya koyabilmesinde, önemli bir kanıt elde etme yolu olarak düşünülebilir.

Bir başka açıdan incelendiğinde, KDH'nın yapı geçerliğinde yöntem olarak kullanılmasının bazı dezavantajlar bulunmaktadır. KDH için gerekli başlangıç parametrelerinin belirlenmesinde güçlükler vardır ve her farklı parametre değerleri farklı sonuçlar verecektir (Taşkın ve Emel, 2010). Ayrıca bu yöntemin bir diğer sınırlılığı ise, aynı veriler üzerinden ağ her çalıştırıldığında farklı çıktı haritalarının gözlenebilme olasılığıdır. Çünkü KDH modelinde kullanılan yarışmacı öğrenme kuralı gereği, McClelland (2015), başlangıçta ağırlık değerleri ve girdilerin sırası rasgele belirlendiğini ve bunun da, tekrarda, farklı gruplandırmalara neden olabileceğini belirtir. KDH çıktı haritası için doğru sayıda satır ve sütun nöron sayısını belirlemek zordur çünkü, bu karar bir çok deneme-yanılma sürecini gerektirmektedir (Matignon, 2007; Callica, 2011).

Fritzke (1994), nöron sayısını artırma işlemi yapan yeni bir kendini düzenleyen ağ temelli genişleyen hücre yapısı (GHY) algoritması önermiştir. Bu artırma, istenen performans ölçütü sağlanıncaya kadar ya da istenen ağ büyüklüğüne ulaşıncaya kadar, işleme devam edebileceğini belirtir (Fritzke 1995). Aynı zamanda Kohonen (2013), Fritzke'nin çalışmasına atıf yaparak önerilen yeni KDH versiyonunda ızgara yapısı ve nöron sayısının hareketli olarak belirlenebileceğini belirtir. Yine Alahakoon, Halgamuge ve Srinivasan (2000) tarafından önerilen ve dinamik yapıda bir KDH olan Genişleyen kendini düzenleyen haritalama (GKDH) modelinin bilgiyi keşfetmede büyük avantajlar sağladığını belirtilmektedir. Bu modellerin ölçeklerin yapı geçerliğinde kullanılabilirliğine dair çalışmalar yapılabilir. Ayrıca KDH modeli aynı veriler üzerinde tekrarlandığında, farklı haritalamanın ortaya çıkması olasılığı yüzünden çıkan yapıların kararlılık düzeyinin incelenmesi, ayrı bir çalışma konusu olabileceği düşünülmektedir.

KDH modelinde nöron sayısı dışında, nöron serisinin boyutluluğu, nöronların birbiri ile komşuluk ilişkisi/şekli, komşuluğun daralma planı ve ağı öğrenme hızı gibi parametreler mevcuttur. Bu parametrelerin her birinin ve kombinasyonlarının, yapı geçerliği çalışması üzerindeki etkilerinin

incelenmesi, ileriki çalışmalar için önerilebilir. Bu çalışmalarda, ölçülen değişkeni etkileyebileceği düşünülen değişkenler açısından, araştırma grubunun frekans dağılım dengesi göz önünde tutulmalıdır.

KAYNAKÇA

- Alahakoon, D., Halgamuge, S.K. & Srinivasan, B. (2000). Dynamic self-organizing maps with controlled Growth for knowledge discovery. *IEEE Transactions on Neural Networks*. 11(3), 601-614.
- Anastasi, A. (1988). *Psychological Testing*. New York: Macmillan Publishing Company.
- Bogdan, R. C. & Biklen, S. K. (1998). *Qualitative research for education*. Boston: Allyn and Bacon.
- Callica, R. (2011). *Customer segmentation and clustering using SAS enterprise miner*. USA: North Carolina Institute.
- Cheng, B. & Titterington, D.M. (1994). Neural Networks: A review from a statistical perspective. *Statistical Science*. 9(1), 2-54.
- Cohen, R.J., Montague, P., Nathanson L.S. & Swerdlik, M.E. (1988). *Psychological testing*. Mountain View: Mayfield Publishing Company.
- DeVellis, R.F. (2003). *Scale Development Theory and Applications*. Thousand Oaks: SAGE Publications.
- Elmes, D. G., Kantowitz, B. H. & Roediger III, H. L. (1992). *Research methods in psychology*. St. Paul: West Publishing Company.
- Feldman, R.S. (1997). *Essentials of understanding psychology*. New York: The McGraw-Hill Companies, Inc.
- Fritzke, B. (1994). Growing cell structure: A self organizing network for supervised and un-supervised learning. *Neural Networks*. 7.1441-1460.
- Fritzke, B. (1995). Growing grid - a self-organizing network with constant neighborhood range and adaptation strength. *Neural Processing Letters*. 2(5), 9-13.
- Holtzman, S. & Vezzu, S. (2011). *Confirmatory factor analysis and structural equation modeling of noncognitive assessments using PROC CALIS*. Retrieved from www.lexjansen.com/nesug11/sa/sa07.pdf
- Jain, A.K., Mao, J. & Mohiuddin, K.M. (1996). *Artificial neural network*. Retrieved from http://www.cogsci.ucsd.edu/~ajyu/Teaching/Cogs202_sp12/Readings/jain_ann96.pdf
- Johnson, R.A. & Wichern, D.W. (1982). *Applied multivariate statistical analysis*. London: Prentice-Hall International, Inc.
- Karasar, N. (1998). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Kline, R.B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of The IEEE*. 78(9), 1464-1480.
- Kohonen, T. (2013). Essentials of self-organizing map. *Neural Networks*. 37, 52-65.
- Matignon, R. (2007). *Data mining using SAS enterprise miner*. New Jersey: John Wiley & Sons.
- McClelland, J.L. (2015). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Retrieved from web.stanford.edu/group/pdplab/pdphandbook/handbookch7.html
- Murphy, K.R. ve Davidshofer, C.O. (2001). *Psychological Testing Principles and Applications*. New Jersey: Prentice Hall Inc.
- Taşkın, Ç. ve Emel, G.G. (2010). Veri madenciliğinde kümeleme yaklaşımları ve kohonen ağları ile perakendecilik sektöründe bir uygulama. *Süleyman Demirel Üniversitesi İktisari ve İdari Bilimler Fakültesi Dergisi*. 15(3), 395-409.
- Tezbaşaran, A. A. ve Kelecioğlu, H. (2004). Madde-ölçek korelasyonlarına, alt-üst grup ortalamalarına ve aşamalı tepki modeline göre geliştirilen sigaraya ilişkin tutum ölçeğinin madde ve ölçek özelliklerinin incelenmesi. *XIII. Ulusal Eğitim Bilimleri Kurultayı*, Malatya, Türkiye
- Tezbaşaran, E. (2016). *Temel bileşenler analizi ve yapay sinir ağı modellerinin ölçek geliştirme sürecinde kullanılabilirliğinin incelenmesi*. (Yayınlanmamış Doktora tezi, Mersin Üniversitesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Mersin).
- Ullman, J. B. (2001). Structural equation modeling. In B.G. Tabachnick and L.S. Fidell (Eds.), *Using Multivariate Statistics* (p. 653-771). Boston: Allyn & Bacon.
- Vesanto, J., Himberg, J., Alhoniemi, E. & Panhankangas, J. (2000, February). *Self-organizing map in matlab: the SOM toolbox*. Updated version of the paper published in Proceedings of the Matlab-DSP 1999 Conference. Retrieved from <http://lib.tkk.fi/Diss/2002/isbn951226093X/article2.pdf>
- Westerlund, M. L. (2005). *Classification with Kohonen self-organizing maps*. Retrieved from www.hi.is/~benedikt/Courses/Mia_report2.pdf

EXTENDED ABSTRACT

Introduction

Construct validity is an important psychometric property in terms of an evidence for the validity of a scale developing to measure psychometric variables. Construct validity of a scale is determined by evidences getting from various methods. One of these methods is factor analysis that detects the factor construct of a scale.

The aim of the research is to find out /to determine what the number of neuron (grid size) should be used in self-organizing mapping when it is used for construct validity of a Likert type scale instead of the factor analysis done with the technique of the principal component analysis used in the construct validity studies in practice. Self-organizing mapping which is an artificial neural network model developed by Kohonen is used for classification and data reduction. Network outputs can be visualized in a two dimensional map. This network detects the input features and gives a map of characteristics of the inputs on a plane.

There are three research questions to investigate the self-organizing mapping as a method for construct validity of Attitudes Towards Smoking Scale:

1. What are the constructs of the scale obtained from self-organizing mapping using different grid size?
2. What are the fit end error indexes of the different constructs of the scale obtained from self-organizing mapping using different grid size according to confirmatory factor analysis?
3. What is the construct validity of the different constructs of the scale obtained from self-organizing mapping using different grid size according to evidence from distinct groups method?

Method

540 individuals aged between 18-56 voluntarily participated in this research as study group. Nearly 60% of the participants were university students and 40% of them was university graduate. Attitudes Towards Smoking Scale preparing by Tezbaşaran and Kelecioğlu (2004) was used to investigate the construct validity. The 41-item scale is a Likert type scale consisting of 19 positive and 22 negative attitude expressions.

As a first step, factor constructs of the scale were analyzed with self-organizing mapping using 2x2, 3x3, 4x4, 5x5, 6x6, 7x7 grid sizes respectively. Each output of network is visualized on two dimensional hexagonal grid. Then, the scale constructs obtained from 6 different grid size were analyzed with confirmatory factor analysis. Results of the analysis that give fit end error indexes of different construct of the scale obtained from different grid size were compared. On the other hand, independent samples t-test was used to explore the construct validities of 6 different scale obtained in the way of distinct groups method. Distinct groups were defined as smokers and non-smokers. It was analyzed repeatedly for each scale construct to find whether there is a statistically significant difference between their means of scale scores.

Results and Discussion

Since all items are in one neuron, at the beginning 2x2 grid size self-organizing mapping having outputs on the 4-neuron map was used to get the scale construct. In two-dimensional mapping, the items of Attitudes Towards Smoking Scale fall apart in 4 neurons. As one item of the scale is in a different neuron, it was excluded and 3 factors with 40 items were observed. In two dimensional mapping, a construct with 3 factors and 37 items was emerged on the condition that the number of neuron was 3x3. When the number of neuron is raised to 16 on 4x4, the total item number lowers to 26 and factor number to 2. When the number of neuron is 25 on 5x5, the factor construct is again 3 and the number of items lowers to 19 since the items are separating more in terms of their

similarities and differences. When the total number of neuron in the map raises 36 on 6x6, it was observed that the number of factor lowers to 2 and of items to 11 as the items are separating much more from each other. It was seen that a construct with 5 items and 1 dimension emerges when the separation of the items is investigated on a map with 7x7. At this point, the neuron rise was stopped as each of the items fall apart in one neuron whenever the number is raised.

The constructs of the scale whose item and factor numbers were continuously decreasing, were investigated by using confirmatory factor analysis and comparing in terms of the fit and error indexes. As a consequence, by lowering the item and factor numbers with increasing the grid size through 7x7, the fitting of the construct raised and finally on the mapping done on 7x7, the construct with 5 items in one factor had the best values.

On the other hand, the attitude the mean scores of smokers are high whereas the mean scores of non-smokers are low, and the difference between them is statistically meaningful. This is also an evidence that the scales have the construct validity. It is important to see the efficiency for measuring the construct in spite of lowering the number of items, particularly in one dimension with 5 items.

Although differentiation of scale items of different psychological construct measured and the differentiation of the gathered data hinder giving an exact neuron number of self-organizing mapping to be used as a construct validity method, a rule can be given as a result of this research: Related to making decision on the number of neuron to be used, it can be increased until the related items of a scale is observed in one neuron, in other words, until a construct with one factor is observed. In addition to this rule, it is recommended to analyze the revealed construct with respect to related attitude variable contextually.

Akıllı Telefon Bağımlılığı Ölçeği'nin (ATBÖ) Türk Lise Öğrencileri İçin Uyarlama Çalışması

The Study Of Adapting Smartphone Addiction Scale (SAS) For Turkish High School Students

Mehmet ŞATA *
Zafer Ertürk ***

İlyas ÇELİK **
Umut Erkin TAŞ ****

Öz

Bu araştırmada, Kwon ve arkadaşları (2013) tarafından geliştirilen Akıllı Telefon Bağımlılığı Ölçeği (ATBÖ)'ni Türk lise öğrencileri için uyarlamak ve psikometrik özelliklerini incelemek amaçlanmıştır. Araştırma örneklemini, Ankara, Elazığ ve Kayseri illerinde yer alan devlet liselerinde öğrenim gören 175 erkek ve 281 kız öğrenciden oluşan 456 katılımcıdan oluşmaktadır. Orjinali İngilizce olan Akıllı Telefon Bağımlılığı Ölçeği'nin Türkçe formunu elde edebilmek için niteliksel yöntemlerden yararlanılmıştır. Uzman görüşlerine dayanılarak elde edilen ölçek formunun geçerlik ve güvenilirliğine ilişkin kanıtlar elde etmek amacıyla doğrulayıcı faktör analizi yapılmıştır. Doğrulayıcı faktör analizi sonuçlarına göre uyum indeksi değerleri RMSEA=.058, GFI=.84, AGFI=.82, NNFI=.96, CFI=.96 olarak bulunmuştur. Ayrıca iç tutarlılık anlamında güvenilirlik düzeylerine ilişkin kanıtlar elde etmek üzere, ilgili ölçümlere ilişkin Cronbach α değerleri ve ölçeğin geneli için tabakalı Cronbach α değeri hesaplanmıştır. Cronbach α değerleri, altı alt ölçeğin ve ölçeğin bütünü için kabul edilebilir düzeyde güvenilir ölçümler sağlayabildiğine işaret etmektedir. Sonuç olarak ATBÖ'nün Türkiye'deki lise öğrencilerinin akıllı telefon bağımlılıklarını ölçmede kullanılabilecek geçerli ve güvenilir ölçümler sağladığı söylenebilir.

Anahtar Kelimeler: Akıllı telefon bağımlılığı, ölçek uyarlama, doğrulayıcı faktör analizi.

Abstract

The aim of this study was to adapt Smartphone Addiction Scale, developed by Kwon et al. (2013), for Turkish high school students and to specify its psychometric characteristics. The sampling of study is comprised of 456 students, 175 boys and 281 girls, studying in state high schools in Ankara, Elazığ and Kayseri. Qualitative methods were used to get the Turkish version of Smartphone Addiction Scale, which is originally English. Confirmatory factor analysis were carried out in order to get evidences about validity and reliability of scale which was formed based on experts' views. According to results of confirmatory factor analysis, fit index values were found as follows: RMSEA=.058, GFI=.84, AGFI=.82, NNFI=.96, CFI=.96. Moreover, Cronbach α values were calculated concerning to relevant measurements to obtain evidences about reliability levels for internal consistency. Cronbach α values indicated that six sub-factors and the scale as whole generate reliable measurements in acceptable level. As a result, it is concluded that SAS can generate reliable and valid measurements to be used for assessing smartphone addictions of high school students in Turkey.

Keywords: Smartphone addiction, scale adapting, confirmatory factor analysis.

GİRİŞ

Bağımlılık, bir madde ya da davranışı kullanmayı bırakamama veya kontrol edememe şeklinde tanımlanabilmektedir (Egger ve Rauterberg, 1996). Bağımlılık, alkol ve uyuşturucu gibi belli bir

* Arş. Gör., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme, Ankara-Türkiye, e-posta: mehmetwsata@gmail.com

** Yüksek Lisans Öğrencisi, Kara Harp Okulu, Savunma Bilimleri Enstitüsü, Ankara-Türkiye, e-posta: celikilyas80@gmail.com

*** Arş. Gör., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme, Ankara-Türkiye, e-posta: zerturk35@gmail.com

**** Uzman Yardımcısı, Milli Eğitim Bakanlığı, Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, Ankara-Türkiye, e-posta: umut_erkintas@hotmail.com

maddeye biyolojik olarak bağlanmayla ve sinir sisteminin bir işlevi olarak ortaya çıkmaktadır. Bununla beraber, bazı araştırmacılar bağımlılıkların biyolojik olabileceği gibi davranışsal da olabileceğini ileri sürmektedirler (Comings, 1995; Stein ve ark., 1994). Özellikle teknolojinin gelişmesi ile teknolojik aletler toplumda çığır açan bir değişime yol açmış ve davranışsal bağımlılıkları da arttırmıştır.

Toplumda büyük bir değişim yaratan bu teknolojik aletlerden biri de akıllı telefonlardır. Uluslararası İstatistik Kurumu (International Data Corporation, IDC)'nin yaptığı bir araştırmaya göre, 2010 yılında dünya çapında 305 milyon adet akıllı telefon satışı gerçekleşmiştir. 2011 yılında ise bu rakam %62'lik bir artış göstererek 494 milyon sayıya ulaşmıştır. Aynı araştırmaya göre, 2012 yılında toplam akıllı telefon satışı 660 milyon adet olarak belirlenirken, 2015 yılı itibariyle akıllı telefon satışının dünya çapında toplam bir milyara ulaşılması öngörülmektedir. Akıllı telefon satışları yıllar geçtikçe artmakla birlikte, bu rakamlar akıllı telefonların günlük yaşamda ve iş yaşamındaki öneminin artacağına da bir göstergesidir. Başlangıçta sadece görüşme ya da kısa mesaj fonksiyonlarını kullandığımız cep telefonları, internet erişimi ve daha birçok özelliğin de eklenmesiyle günümüzde vazgeçilmez bir nesne haline almıştır. İnsanlar günlük yaşamında iletişimin yanı sıra internet erişimi, e-posta veya anlık mesaj yollama, fotoğraf çekme ve sosyal medyada paylaşma vb. işlerini akıllı telefonlar sayesinde yapmaktadırlar. Ayrıca video sohbet gerçekleştirme ve navigasyon sayesinde gidilecek yeri kolayca bulabilme akıllı telefonların hayatımızı kolaylaştıran özellikleri arasındadır.

Diğer yandan, akıllı telefonların çok fazla kullanılmasından dolayı oluşan olumsuz etkileri de görmek mümkündür. Özellikle akıllı telefon aracılığıyla sözde sosyalleşirken, gerçek yaşamda ise daha çok bireyselleşmeye gidilmektedir. Karaca (2007), ergen bireyin gerçek hayatta ortak ilgilere sahip bireylerle bir araya gelme imkânlarının azalması sebebiyle kendisiyle aynı fikir, duygu ve düşünceleri paylaştığı bireylerle bağlantı kurma ihtiyacını sanal sohbet ve sanal topluluklara katılarak sağlamaya çalıştığını belirtmektedir. Günüş (2009), bu sosyalleşme unsurlarının internet bağımlılığı için bir risk faktörü olduğuna dikkat çeker. İnternet bağımlısı bireylerin, bağımlı olmayan bireylere göre daha çok çevrimiçi oyunları tercih ettiklerini belirtir.

Akıllı telefon ve benzeri cihazların kullanılmasının yaygınlaşması toplumdaki şiddeti artırdığı gibi, aile içi ve iş yaşamındaki iletişimi de olumsuz etkilemektedir. Shinse ve Baek (2013) tarafından yapılan "ergenlerin akıllı telefon bağımlılıklarının saldırganlık davranışı üzerindeki etkileri" adlı çalışmada katılımcılardan topladıkları veri seti analiz edilmiştir. Analiz sonuçlarına göre; ergenlerdeki saldırganlık davranışı üzerinde akıllı telefon bağımlılığının önemli etkisi olduğu görülmüştür. Aşırı akıllı telefon kullanımının ergenlerin, insanlara ve eşyalara karşı saldırganlık davranışı göstermesine neden olduğu ortaya çıkmıştır.

Bunlara ilave olarak, aşırı akıllı telefon kullanımı sebebiyle öğrenciler de farklı sorunlarla karşı karşıya kalabilmektedirler. Jino (2014) tarafından Güney Kore'de yapılan araştırmada, ilköğretimde öğrenim gören öğrencilerin "Dikkat Bozukluğu ve Hiperaktivite" belirtileri üzerinde akıllı telefon bağımlılığının etkileri incelenmiştir. Bu çalışmaya, 4. ve 6. sınıftaki ilköğretim öğrencilerinden toplam 991 öğrenci katılmıştır. Araştırma sonuçlarına göre "Dikkat Bozukluğu ve Hiperaktivite" belirtisi bulunan öğrencilerin yaklaşık olarak %8'inin akıllı telefon bağımlılığı riski taşıdığı görülmüştür. Altıncı sınıfta düşük akademik başarıya sahip olan öğrencilerin akıllı telefon bağımlılığının daha yüksek olduğu görülmüştür. Akıllı telefon bağımlılık düzeyinin öğrencilerin dikkatleri ve hiperaktiflikleri üzerinde negatif etkisi olduğu sonucu ortaya çıkmıştır.

Akıllı telefonun günlük yaşamda kullanımının büyük bir artış göstermesi, akıllı telefona olan bağımlılığı artırmış ve bu durum bu alanda yapılan araştırmaların da artmasına yol açmıştır. Lane ve Manner (2011) ise kişilik özellikleri ve akıllı telefon sahipliği ve kullanımı arasındaki ilişkiyi tespit etmeye yönelik çalışmalarında, dışadönük bireylerin akıllı telefona sahip olmaya ve kullanmaya daha yatkın oldukları sonucuna varmıştır. Park ve Chen (2007) yaptıkları araştırmada doktor ve hemşirelerin akıllı telefon kullanımını etkileyen motivasyonel faktörleri incelemiştir. Bu araştırmanın bulgularına göre, kullanışlılık ve akıllı telefon kullanımına yönelik tutumlar, akıllı telefon kullanım niyetini pozitif olarak etkilemektedir. Park ve Lee (2012) akıllı telefon kullanımı ve

psikolojik bağlamda iyi olma ile ilgili yaptıkları çalışmada akıllı telefon kullanımının yalnızlık, depresyon ve özgüveni arttırabileceğini belirtmişlerdir. Ayrıca Kim, Lee, Lee, Nam ve Chung (2014), telefon bağımlılığın yeni bir bağımlılık türü olarak internet bağımlılığı ile birlikte daha fazla ilgi odağı olduğunu belirtmişlerdir.

Araştırmalarda, ergen ve gençlerde mobil telefona sahip olma oranının % 76'ya ulaştığı ve bu oranın %40'ının ise ikinci bir mobil telefona sahip olduğu sonucuna ulaşılmıştır. Yüksek düzeyde mobil telefon kullanan kişilerin düşük öz saygı düzeyine sahip oldukları ve bu kişilerin öz saygılarını arttırabilmek için mobil telefonunu sıklıkla kullandıkları görülmüştür (Phillips, Ogeil ve Blaszczyński, 2011).

Park, Hyun ve Ha (2014) tarafından Güney Kore'de yapılan ve ergenlerin internet ve cep telefonu bağımlılıklarının değişiminin incelendiği çalışmada; çevresel faktörlerin (aile ve arkadaş etkileri) ve kişisel faktörlerin (teknolojik araçların kullanımına ayrılan zaman) analizi yapılmıştır. İnternet ve cep telefonu bağımlılıkları arasındaki farklılıkları incelemek amacıyla toplam 1420 öğrenci analize dahil edilmiştir. Araştırma sonuçlarına göre, internet ve cep telefonu bağımlılıklarına sebep olan faktörler arasında yakın ilişkiler bulunmuştur.

Gençler ergenlik döneminde suskunluk, içe kapanıklık, çevreden kopma, can sıkıntısı, ailede çatışma, arkadaşlarıyla sorun yaşaması, çevresinde beğenilmeme korkusu, kendi veya çevresi ile ilgili geleceği hakkında endişeye kapılması ve kötümserlik gibi psikolojik ve sosyolojik durumlarını yaşayabilmektedir (Saygılı, 2002). Birey, bu dönemde karşılaşacağı bu sorunları untabileceği ve bu problemlerden uzaklaşabileceği bir ortamın ihtiyacını hissetmektedir. Bundan dolayı ergen birey mevcut ihtiyaçlarını giderebilecek, zevk ve eğlence unsurlarını içeren, daha iyi anlaşılabilirdiği ve kabul gördüğü, duygu ve düşüncelerine önem verildiği bir ortama yönelebilecektir. Ergen birey internet ile tanıştıktan sonra kendi sosyal çevresinden bulamadığı desteği bu ortamdan sağladıkça, internete daha çok bağlanabilmekte zamanının büyük bir kısmını bu ortamda geçirebilmektedir (Günüç, 2009).

Genç bireyler arasında akıllı telefon kullanılarak internete erişim sağlaması popüler bir etkinlik olarak görülmektedir. Bu etkinliğin popüleritesi ise aşırı internet kullanımının sebebi olabilmektedir (Lin ve Tsai, 2002). İnternet kullanımının bu derece yaygın olarak tercih edilmesi ve kullanılmasının önemli sebeplerinden biri, akıllı telefon yardımıyla kolayca ve günün her anında internete ulaşılabilirliğidir. Çünkü insanlar boş zamanlarını değerlendirebilmek için, genellikle kolay erişebilecekleri ve uygulayabilecekleri etkinlikleri tercih etmektedirler (Bayraktutan, 2005).

Toplumda artan teknoloji bağımlılığını tespit etmek için alan yazında az sayıda ölçek bulunmaktadır. Bianchi ve Philips (2005) tarafından geliştirilen ve güvenilirlik çalışması yapılan Problemlü Mobil Telefon Kullanım Ölçeği, 27 maddeden oluşmaktadır ve problemlü telefon kullanımıyla haftalık telefon için harcanan zaman arasında güçlü bir ilişki olduğu göstermektedir. Kim ve arkadaşları (2014) tarafından geliştirilen ve geçerliliği test edilen Akıllı Telefon Bağımlılığına Yatkınlık Ölçeği (Smartphone Addiction Proness Scale-SAPS), 15 maddeli, 4 faktörlü yapıdan oluşmaktadır ve 4'lü likert ölçeği kullanılmıştır (1:hiçbir zaman ve 4: her zaman). Akıllı telefon bağımlılığını değerlendirmek için Kwon ve arkadaşları (2013) 33 madde ve 6'lı likert ölçeğinden oluşan Akıllı Telefon Bağımlılık Ölçeği (Smartphone Addiction Scale-SAS)'ni geliştirmişlerdir. 18-53 yaş aralığındaki 197 kişinin katıldığı bu çalışmanın iç güvenilirlik hesaplanmasında geçerli ve güvenilir olduğu bulunmuştur. Akıllı telefon bağımlılığını belirlemek için geliştirilen ilk ölçek olan bu ölçekte altı faktörlü yapının olduğu bulunmuştur (günlük yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal ilişkiler, aşırı kullanım ve dayanma). Kwon ve ark. (2013) geliştirmiş oldukları SAS ölçeğini gözden geçirmek ve ergenler için geçerliliğini test etmek için bu ölçeğin kısa versiyonunu oluşturmuşlardır: Akıllı Telefon Bağımlılık Ölçeği –Kısa Versiyonu (SAS-Short Version). Ölçeğin kısa versiyonu, 10 madde ve 6'lı likert ölçeğinden oluşmaktadır ve yaş ortalaması 14,5 olan 540 katılımcı ile yürütülmüştür. Ayrıca bu çalışmaya katılan erkek ve kadın katılımcı sayıları arasında fark olduğundan cinsiyete yönelik karşılaştırma yapılamamıştır.

Akıllı telefonların sayısı ülkemizde de hızla artmaktadır. Özellikle lise çağındaki gençler akıllı telefon gibi ileri teknoloji ürünlerle tanışan bir nesil olduğundan diğer yaş gruplarına göre daha fazla

etkilenmektedirler. Alan yazına göre 12-18 yaş dönemi internet bağımlılığında çok kritik bir dönem olduğu belirtilmiştir (Tsai ve Lin, 2003). Bundan dolayı akıllı telefonların özellikle gençler arasında daha yaygın olduğunu görmek mümkündür. Uyarlama çalışmasını yaptığımız Akıllı Telefon Bağımlılık Ölçeği (Kwon ve ark., 2013), ergen yaşlardaki bireylerin gerek madde bağımlılığı gerekse davranışsal bağımlılığa karşı daha duyarlı olmasından dolayı ülkemizde lise çağındaki öğrencilere uygulanmıştır. Çalışmadan elde edilecek sonuçlar, akıllı telefon bağımlılığını engellemek diğer değişkenlerle olası etkilerini belirlemek amacıyla yapılacak araştırmalarda kullanılabileceğinden önem arz etmektedir.

Araştırmanın Amacı

Bu çalışma bireylerin akıllı telefon bağımlılıklarını ölçen “Akıllı Telefon Bağımlılığı Ölçeği”nin Türkiye’deki lise öğrencilerine kültürel olarak uyarlamak ve yapı geçerliği için kanıtlar aramayı amaçlamaktadır.

YÖNTEM

Araştırmanın Modeli

“Akıllı Telefon Bağımlılığı Ölçeği”nin Türk lise öğrencileri için uyarlanmasını ve psikometrik özelliklerinin incelenmesini amaçlayan bu araştırma, temelde bir tarama çalışmasıdır. Tarama modeli var olan durumu olduğu gibi betimlemeyi amaçlayan bir araştırma yaklaşımıdır. Bu çalışmada da lise öğrencilerinden elde edilen verilerin orijinal “Akıllı Telefon Bağımlılığı Ölçeği”ne uyumu için kanıtların ortaya konması amaçlanmıştır.

Evren ve Örneklem

Bu araştırmada, “Akıllı Telefon Bağımlılığı Ölçeği”nin Türkiye’deki lise öğrencileri için kültürel uyarlanmasının yapılması amaçlanmıştır. Araştırma, seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme yöntemi ile seçilen Ankara ilinden iki, Kayseri ilinden iki ve Elazığ ilinden bir lise olmak üzere toplam beş lisede gerçekleştirilmiştir. Örneklem ilgili illerde seçilme nedeni araştırmacıların ulaşılabilirliği olmuştur. Araştırmanın örnekleme, araştırmaya gönüllük esasına göre katılan lise 1-4. sınıfta öğrenim gören 715 öğrenciden oluşmaktadır. Öncelikli olarak öğrencilere akıllı telefon kullanma durumlarına yönelik sorular sorulmuş, akıllı telefon kullanmadığı ve anketi tamamlamadığı tespit edilen katılımcılar araştırma için uygun olmadığı değerlendirilerek çıkarılmış ve nihai olarak 175 erkek ve 281 kız öğrenciden oluşan 456 katılımcı araştırmaya dâhil edilmiştir. Araştırmaya katılan öğrencilerin ortalama yaşı Ort.=16,56’dır. Örneklemde yer alan bireylerin cinsiyet, sosyo-ekonomik düzeyi ve kendi beyanlarına göre bağımlılık durumları Tablo 1’de sunulmuştur.

Tablo 1. Araştırmaya Katılan Öğrencilerle İlgili Cinsiyet, Sosyo-ekonomik Düzeyi ve Kendi Beyanlarına Göre Bağımlılık Durumları

| Değişken | n = 456 | Frekans | Yüzde(%) |
|----------------------|---------------|---------|----------|
| Cinsiyet | Kız | 221 | 62 |
| | Erkek | 175 | 38 |
| Sosyo-ekonomik statü | Alt | 148 | 32 |
| | Orta | 241 | 53 |
| | Üst | 67 | 15 |
| Öz değerlendirme | Bağımlı | 98 | 22 |
| | Bağımlı değil | 289 | 63 |
| | Bilmiyor | 69 | 15 |

Veri Toplama Araçları

Akıllı Telefon Bağımlılığı Ölçeği (ATBÖ), altı boyutta (günelik yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal yönelimli ilişki, aşırı kullanım ve dayanma) akıllı telefona olan bağımlılığı ölçen kendini değerlendirme türünde bir ölçme aracıdır. Ölçeğin orijinal formu, bireyin akıllı telefona bağımlılığını yukarıda belirtilen boyutlarıyla ölçen ve 6'lı Likert türü derecelemeyi (1:kesinlikle katılmıyorum, 6:kesinlikle katılıyorum) kullanan 33 maddeden oluşmaktadır. Ölçek; günelik yaşam bozuklukları (5 madde), olumlu beklenti (8 madde), yoksunluk hissi (6 madde), sanal yönelimli ilişki (7 madde), aşırı kullanım (4 madde) ve dayanma (3 madde)'e ilişkin ölçümler sağlamaktadır. Bu ölçekten elde edilecek puan dağılımı 33 ila 198 puan arasındadır. Yüksek puanlar ciddi telefon bağımlılığına işaret etmektedir.

Kwon ve arkadaşları (2013) tarafından yapılan araştırmada yaşları 18 ila 53 arasında değişen (= 26,06) 64 erkek ve 133 kadın katılımcıdan oluşan toplam 197 kişiden elde edilen veriler üzerinde açımlayıcı faktör analizleri uygulanmıştır. Analizler sonucunda, ölçeğin altı faktörlü bir yapı sergilediği ve bu faktörlerin toplam varyansın yaklaşık %61'ini açıklayabildiği bulunmuştur. Ölçeğin bütünü için Cronbach α değeri $\alpha=0,967$ ve her bir alt faktör için sırasıyla $\alpha=0,858$, $\alpha=0,913$, $\alpha=0,876$, $\alpha=0,904$, $\alpha=0,825$, $\alpha=0,865$ olarak hesaplanmıştır. Bu araştırmalardan elde edilen bulgular, ölçeğin bireylerin akıllı telefona bağımlılık düzeylerine ilişkin geçerli ve güvenilir ölçümler sağlayabildiğine yönelik kanıtlar olarak değerlendirilmiştir.

İşlem

Bu ölçeğin Türk lise öğrencileri için uyarlanması sürecinde “Uluslararası Test Komisyonu (ITC) Rehber Kitabı: Bir Psikolojik Ölçme Aracını Uyarlama Süreci”nde yer alan ölçek uyarlama yönergesi izlenmiştir. Buna göre öncelikle bu ölçeğin ülkemiz için gerekliliğine karar verilmiş ve daha sonra da orijinal ölçeği geliştirenlerden uyarlama için gerekli izinler alınmıştır. Akıllı Telefon Bağımlılığı Ölçeği'nin uyarlama sürecinde öncelikle orijinali İngilizce olan ölçek maddeleri, tepki seçenekleri ve yönergesi dört araştırmacı tarafından Türkçe'ye çevrilmiştir. Çeviri formunun gözden geçirilmesinin ardından orijinal formdaki maddelerin çeviri formundaki maddeler ile eşdeğer olup olmadığının sınanması gerekmektedir. Bu amaç doğrultusunda niteliksel yöntemlerden tek yönde çeviri ve çevirinin başka çevirmenlerce değerlendirilmesi yöntemi kullanılmıştır. Bu doğrultuda ölçeğin yönergesi, maddeleri ve tepki seçeneklerinin hem orijinalinin hem de çevirilerinin yer aldığı bir uzman değerlendirme formu hazırlanmıştır. Bu form, alan ve dil uzmanlarından oluşan sekiz kişilik bir uzman grubu (iki dil uzmanı, bir psikolojik-danışmanlık ve rehberlik uzmanı ve beş ölçme ve değerlendirme uzmanı) tarafından ayrı ayrı değerlendirilmiştir. Değerlendirme sonucunda alan ve dil uzmanları tarafından her bir maddeye ait üzerinde en çok uzlaşılan çeviri ölçeğe alınmıştır. Yapılan işlemler sonucunda ölçeğin çeviri formu oluşturulmuştur. Ardından ölçeğin çeviri formu örnekleme de yer alan kişilere uygulanmıştır.

Veri Analizi

Bu araştırmada, Akıllı Telefon Bağımlılığı Ölçeği'ni (ATBÖ) Türk lise öğrencileri için uyarlamak ve psikometrik özelliklerini incelemek amaçlanmıştır. Bu amaç doğrultusunda doğrulayıcı faktör analizi yapılmıştır. Bu çalışmada ilgili ölçümlerin yapı geçerliğine ilişkin kanıt olarak, “ölçme modelindeki maddelerin ilgili yapının iyi birer temsilcileri olmaları” ve “ölçme modellerinin veri setine kabul edilebilir düzeyde uyum göstermesi” kriterleri ele alınmıştır. Test edilen modelin uyum düzeyinin değerlendirilmesi için çok sayıda uyum indeksi kullanılmaktadır. Bu çalışmada yapılan DFA için Ki-kare değeri, İyilik Uyum İndeksi (Goodness of Fit Index, GFI), Düzeltilmiş İyilik Uyum İndeksi (Adjusted Goodness of Fit Index, AGFI), Karşılaştırmalı Uyum İndeksi (Comparative Fit Index, CFI), Normlaştırılmamış Uyum İndeksi (Non-Normed Fit Index, NNFI) ve Yaklaşık Hataların Ortalama Karekökü (Root Mean Square Error of Approximation, RMSEA) uyum indeksleri dikkate alınmıştır.

Bu uyum indekslerinde genelde GFI, AGFI, CFI ve NNFI $>.90$ ve RMSEA $<.06$ ölçüt olarak alınmaktadır (Hu ve Bentler, 1999). Ki-kare (χ^2) değerinin serbestlik derecesine bölünmesiyle elde edilen değerin (χ^2/sd değer) iki ve altında olması uyumun iyi, beş veya altında bir değer olması ise kabul edilebilir bir uyum iyiliğinin olduğunu gösterir (Şimşek, 2007). Schermelleh, Moosbrugger ve Müller'e (2003) göre ise χ^2/sd değerinin 3'ten küçük olması veri-model uyumu için kabul edilebilir seviyelerdir.

Uyum indekslerinden GFI ve AGFI genelde $>.90$ kriterini sağlaması gerekmektedir. Ancak bu değerler örneklem büyüklüğünden etkilenmektedirler. Örneklem büyüklüğü, kestirilmesi gereken parametre sayısının en az 10 katı olduğunda bu değerleri yorumlamak daha anlamlı olmaktadır. Ancak bu çalışmada kestirilen parametre sayısı 87 (madde yükleri, madde hataları, faktör sayısı ve faktörler arası korelasyon) ve bunun için gerekli örneklem büyüklüğü olan 870'e ulaşamadığından GFI ve AGFI değerleri düşük çıkmıştır. Örneklem büyüklüğünün yeterli olmadığı ve verilerin çok değişkenli normal dağılım sayılıtısını sağlamadığı durumlarda GFI ve AGFI değerleri yerine NNFI ve CFI değerleri esas alınır.

Ölçek maddelerine ilişkin hesaplanan faktör yük değerleri ise, o göstergelerin ilgili yapıyı temsil edebilme gücüdür. Bu doğrultuda araştırmalarda, faktör yük değeri 0.30'dan düşük maddeler ölçekten çıkartılmalıdır (Kline, 2000, s.6). Doğrulayıcı faktör analizi sonucunda faktör yük değeri 0.30'dan düşük olan herhangi bir madde olmadığından ölçme modelinden madde çıkarılmamıştır. İlgili ölçümlerin güvenilirlik düzeylerine ilişkin kanıtlar sağlamak üzere, bu ölçümlere ilişkin Cronbach α değerleri SPSS 21 paket programı kullanılarak hesaplanmıştır.

BULGULAR ve YORUMLAR

Akıllı Telefon Bağımlılığı Ölçeği (ATBÖ)'nin faktör yapısını incelemek amacıyla doğrulayıcı faktör analizi yapılmıştır. Ölçümlerin yapı geçerliğini incelemek için yapılan doğrulayıcı faktör analizinde ilk olarak örneklemin yeterli büyüklükte olup olmadığını test etmek için Barlett Sphericity testi yapılmıştır. Verilerin faktör analizine uygunluğu için KMO değerinin .70'dan yüksek ve Barlett testinin anlamlı çıkması gerekmektedir. (Büyüköztürk, 2013). Yapılan analiz sonucu Kaiser-Mayer-Olkin (KMO) değeri 0,909 ve Barlett küresellik testi; χ^2 değeri ise 6221.88 ($p < .001$) olarak anlamlı düzeyde olduğu bulunmuş ve örneklemin AFA için uygun olduğu görülmüştür. Veri setinin çok değişkenli normal dağılım sergileyip sergilemediğini test etmek ve uç değerleri tespit etmek için Cook's ve Leverage değerlerine bakılmıştır. Veri setinin çok değişkenli normal dağılım sergilemediği ve uç değer olmadığı görülmüştür. Değişkenler arasında doğrusal ilişki olup olmadığını belirlemek için ZPRED-ZRESID saçılma diyagramına bakılmış ve doğrusal ilişkilerin bulunduğu gözlenmiştir. Madde puanları arası ilişkilere yönelik hesaplanan korelasyon katsayıları incelenmiş ve 0,80'den büyük değer olmadığı, dolayısıyla çoklu bağlantı sorunu bulunmadığı tespit edilmiştir. Tüm bu sayılıtların incelenmesinden sonra doğrulayıcı faktör analizine başlanmıştır.

Yapı geçerliği bağlamında geçerlik düzeylerine ilişkin kanıtlar elde etmek üzere, ölçeğin orijinaline ilişkin tanımlanan altı faktörlü birinci düzey ölçme modeli (Model1) test edilmiştir. Ayrıca orijinal ölçekte tüm faktörlerin "akıllı telefon bağımlılığı" yapısının alt ölçekleri oldukları ifade edilmektedir. Bu sebeple altı faktörlü ikinci düzey ölçme modeli (Model2) test edilmiştir. Analiz sonucunda modifikasyon indisleri incelendiğinde; madde 6 ile 7, madde 21 ile 23, madde 24 ile 25 arasındaki hata kovaryansları eşleştirildiğinde Ki-kare değerlerinde anlamlı bir azalma olacağı görülmüştür. Söz konusu maddeler arasındaki hata kovaryansları eşleştirilerek ikinci düzey DFA modeli (Model3) test edilmiştir. Model1'de akıllı telefon bağımlılığının bileşenleri olan "gündelik yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal yönelimli ilişki, aşırı kullanım ve dayanma" faktörleri (gizil değişkenler) olarak, ilgili maddeler ise göstergeler olarak tanımlanmıştır. Model2 ve Model3'de ise "akıllı telefon bağımlılığı" ikinci düzey gizil değişken ve bileşenleri olan "gündelik yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal yönelimli ilişki, aşırı kullanım ve dayanma" ise birinci düzey gizil değişkenler (faktörler) olarak, ilgili maddeler ise göstergeler olarak tanımlanmıştır. Doğrulayıcı faktör analizi sonucunda faktör yük değeri .30'dan düşük veya

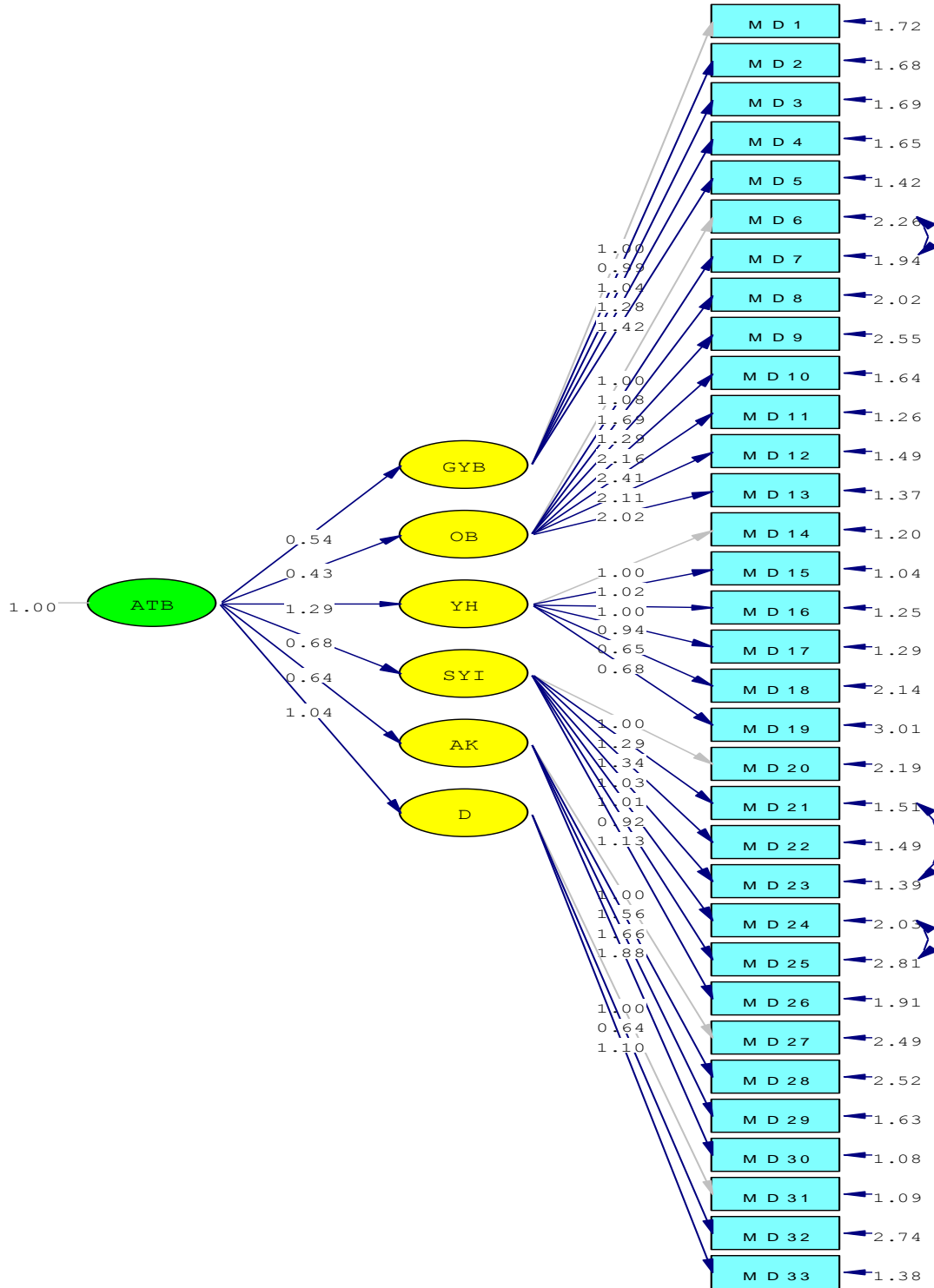
hata değeri .90'dan yüksek olan herhangi bir madde bulunmadığından ölçme modellerinden madde çıkarılmamıştır. Bu modellere ilişkin hesaplanan uyum indeksleri Tablo 2'de sunulmuştur.

Tablo 2. Akıllı Telefon Bağımlılığı Ölçek Verilerinin Model-Veri Uyum Değerleri

| Model | S-B2 /df | p değeri | GFI | AGFI | CFI | NNFI | RMSEA |
|--------|-----------------------|----------|------|------|------|------|-------|
| Model1 | 1350,59/480 (2,81) | 0,00000 | 0,83 | 0,80 | 0,96 | 0,95 | 0,063 |
| Model2 | 1464,05/489 (2,99) | 0,00000 | 0,81 | 0,79 | 0,95 | 0,95 | 0,066 |
| Model3 | 1220,11/486 (2,51) | 0,00000 | 0,84 | 0,82 | 0,96 | 0,96 | 0,058 |

Tablo 2'de görüldüğü gibi, Model1 ve Model3'e ilişkin hesaplanan uyum indekslerinden GFI ve AGFI değerlerinin, ölçüt değerden biraz düşük oldukları görülmektedir. CFI ve NNFI ve RMSEA değerleri ise bu araştırmada dikkate alınan uyum ölçütlerini ($GFI \geq 0.90$ ve $RMSEA \leq 0.08$) karşılamaktadır.

Model1 ve Model2'nin ilgili veri setlerine uyum düzeyleri karşılaştırıldığında ise; Model3'e ilişkin hesaplanan AGFI, GFI, NNFI ve RMSEA değerlerinin, Model1'e göre biraz daha iyi uyuma işaret ettiği, Model1 ve Model3'e ilişkin hesaplanan CFI değerlerinin ise aynı olduğu görülmektedir. Bunun yanı sıra örneklem grubu için; Model1'e dayalı olarak göstergelere ilişkin hesaplanan faktör yük değerlerinin ($\lambda=0.32-0.89$), Model3'e dayalı olarak hesaplanan değerler ile ($\lambda=0.33-0.89$) aynı ve Model1'e dayalı olarak hesaplanan hata değerlerinin de ($\epsilon=0.33-0.82$) Model3'e dayalı olarak hesaplanan hata değerlerinden ($\epsilon=0.33-0.89$) daha düşük oldukları görülmektedir. Tüm bu bulgular, Model3'ün, Model1'e göre, ilgili veriye daha iyi uyum sergilediğini ve Model3'te yer alan göstergelerin ilgili yapıları daha iyi temsil ettiklerini göstermektedir. Bu bulgular ölçeğin 33 maddelik formununun, Türk lise öğrencileri için, ölçmeyi amaçladığı yapıyı "gündelik yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal yönelimli ilişki, aşırı kullanım ve dayanma" yönleriyle (bileşenleriyle) geçerli bir şekilde ölçebildiğine işaret etmektedir. Ayrıca Model3'ün test edilmesi sonucunda hesaplanan birinci düzey gizil değişkenler ile ikinci düzey gizil değişkenler arasındaki korelasyonlara bakıldığında, yüksek korelasyon katsayıları birinci düzey gizil değişkenler ile ikinci düzey gizil değişkenin birbiriyle ilişkili olduğunu göstermektedir. Dolayısıyla bu bulgular da, ilgili ölçümlerin yapı geçerliğine ilişkin ek bir kanıt olarak değerlendirilmiştir. Bu doğrultuda Model3, ölçeğin Türkçe Formu'nun faktör yapısına ilişkin geçerli model olarak kabul edilmiştir. Model3'e ilişkin şekilsel gösterim ve parametre kestirimleri Şekil 1'de sunulmuştur.



Şekil 1. ATBÖ Türkçe Formu'nun Faktör Yapısına İlişkin Tanımlanan İkinci Düzey Modifikasyonlu Ölçme Modeli (Model3) (ATB: Akıllı Telefon Bağımlılığı, GYB : Gündelik Yaşam Bozuklukları, OB: Olumlu Beklenti, YH : Yoksunluk Hissi, SYİ: Sanal Yönelimli İlişki, AK: Aşırı Kullanım, D: Dayanma

Şekil 1’de sunulan faktör yük değerleri ve hata değerleri, söz konusu maddelerin ilgili faktörlerin iyi/uygun birer temsilcileri olduklarına ve kabul edilebilir düzeyde hata içerdiklerine işaret etmektedir.

Ölçeğin 33 maddelik formundan elde edilen ölçümlerin iç tutarlılık anlamında güvenilirlik düzeylerine ilişkin kanıtlar elde etmek üzere, ilgili ölçümlere ilişkin Cronbach α değerleri ve ölçeğin geneli için tabakalı Cronbach α değeri hesaplanmıştır. Hesaplanan bu Cronbach α değerleri ile ölçeğin orijinal formundan elde edilen ölçümlere ilişkin değerleri Tablo 3’de sunulmuştur.

Tablo 3. ATBÖ Türkçe Formundan ve Orijinal Formu’ndan Elde Edilen Ölçümlere İlişkin Güvenirlik Değerleri

| Alt Ölçekler | Tabakalı Cronbach α değeri | | Cronbach α değerleri | |
|--|-----------------------------------|--------|-----------------------------|----------------|
| | Model3 | Model3 | Model3 | Orijinal Ölçek |
| Gündelik yaşam bozuklukları | -- | 0,769 | 0,769 | 0,858 |
| Olumlu beklenti | -- | 0,788 | 0,788 | 0,913 |
| Yoksunluk hissi | -- | 0,841 | 0,841 | 0,876 |
| Sanal yönelimli ilişki | -- | 0,777 | 0,777 | 0,904 |
| Aşırı kullanım | -- | 0,721 | 0,721 | 0,825 |
| Dayanma | -- | 0,716 | 0,716 | 0,865 |
| Tüm Ölçek (Akıllı Telefon Bağımlılığı) | 0,935 | -- | -- | 0,967 |

Tablo 3’de verilen tabakalı Cronbach α ve Cronbach α değerleri incelendiğinde, eldeki verilerin ölçeğin tamamı için iyi düzeyde, alt faktörlerde ise kabul edilebilir düzeyde güvenilir ölçümler sağlayabildiği anlaşılmaktadır.

TARTIŞMA ve SONUÇ

Akıllı telefon kullanıcılarının akıllı telefona olan bağımlılık düzeyine ilişkin geçerli ve güvenilir ölçümlerin elde edilmesi, akıllı telefon kullanımı sonucu ortaya çıkan bu bağımlılığı önlemek amacıyla toplumsal programlar geliştirilmesi açısından oldukça önem arz etmektedir. Akıllı telefon üretim sanayisine sahip olmayan bir ülke olmamıza rağmen, akıllı telefonun toplumun hemen her kesiminde oldukça yaygın kullanıma sahip olduğu görülmektedir. Araştırmaya göre akıllı telefon kullanma amacı daha çok eğlence ve oyun amaçlı kullanıldığı görülmektedir. Bu bağlamda akıllı telefon bağımlılığı iyi bir şekilde tanımlanıp toplumsal bazda bu bağımlılığa karşı politikalar geliştirmek önemlidir. Bu çalışmada, gelecek araştırmalara ön ayak olması ve hızla yayılan bu bağımlılığa karşı yeni program ve önerilerin hazırlanması amacı doğrultusunda “Akıllı Telefon Bağımlılığı Ölçeği’nin” Türk lise öğrencileri için uyarlama çalışması yapılmıştır.

Bu çalışmada DFA kullanılarak, ölçeğin Türk kültüründeki faktör yapısı ile orijinalinin faktör yapısı karşılaştırılmış ve ilgili ölçümlerin yapı geçerliğine ilişkin kanıtlar elde edilmiştir. Ölçeğin orijinaline ilişkin AFA sonuçlarına göre, ölçek altı faktörlü bir yapı sergilemektedir (Kwon ve ark., 2013). Bu çalışmada, verilerin orijinal ölçekteki faktörlere olan uyumunu test etmek amacıyla doğrulayıcı faktör analizi (DFA) yapılmıştır. Bunun için orijinal ölçeği esas alan bir ölçme modeli tanımlanıp test edilmiştir.

Bu çalışmada, Türk lise öğrencileri için uyarlaması yapılan ölçek için hesaplanan güvenilirlik katsayıları, uyum indeksleri, göstergelere ilişkin faktör yük ve hata değerleri, ölçeğin orijinalinde olduğu gibi, “gündelik yaşam bozuklukları, olumlu beklenti, yoksunluk hissi, sanal yönelimli ilişki, aşırı kullanım ve dayanma” alt boyutlarından meydana gelen altı faktörlü modelin, veri setine iyi uyum sergilediğini göstermektedir.

Ölçeğin 33 maddelik Türkçe Formu’na ilişkin hesaplanan tabakalı Cronbach α değerleri (tüm ölçek tabakalı $\alpha = .935$; faktörler $\alpha = 0.716-0.841$), ölçeğin akıllı telefona bağımlılık ile bunların bileşenlerine yönelik güvenilir ölçümler sağlayabildiğine işaret etmektedir. Ayrıca bu çalışmada

bulunan Cronbach α değerleri orijinal ölçekteki Cronbach α değerleriyle karşılaştırılmıştır. Buna göre ölçeğin Türkçe Formu'nun orijinal ölçeğe göre değerlerinin biraz düşük olduğu fakat yine de güvenilir ölçümler sağlayabilecek değerlere sahip olduğu görülmüştür. Bu durumun temel nedeni orijinal ölçek için verinin toplandığı örneklem ile Türk lise öğrencilerine uyarlaması yapılan ölçeğe ilişkin verilerin toplandığı örneklemin niteliklerin (orjinal ölçekte yaş aralığı 18-53 arası iken Türk lise öğrencileri için uyarlaması yaş aralığı 15-20) ve büyüklüklerinin (orjinal ölçekte 197 iken Türkçe uyarlaması 456) farklı olmasından kaynaklandığı düşünülmektedir.

ATBÖ'nün alt ölçekleri şu hususları içermektedir: “Gündelik yaşam bozuklukları” alt ölçeği, planlanan işleri yetiştirememeye, sınıftayken veya çalışırken konsantre olmakta zorlanma, görme bulanıklığı ya da baş dönmesi yaşama, boyunda veya el bileklerinde ağrı hissetme ve uyku bozukluğunu içermektedir. “Olumlu beklenti” alt ölçeği, akıllı telefonla keyifli hissetme ve onu kullanarak stres atma olarak tanımlanmaktadır. “Yoksunluk hissi” alt ölçeği, akıllı telefon olmadığına sabırsız, huysuz ve tahammülsüz olmayı, akıllı telefonunu kullanmasa bile birinin sürekli aklında telefonunun olmasını, akıllı telefonunu kullanmayı asla bırakmamayı ve akıllı telefon kullanırken rahatsız edildiğinde öfkelenmeyi kapsamaktadır. “Sanal yönelimli ilişki” alt ölçeği, bir kimsenin akıllı telefonla edindiği arkadaşları ile olan ilişkilerinin gerçek yaşamdaki arkadaşları ile olan ilişkilerinden daha samimi hissetmesi ile ilgili soruları ve sürekli kişinin akıllı telefonunu kontrol etmesini içermektedir. “Aşırı kullanım” alt ölçeği, birinin akıllı telefonunu kontrol edilemez bir şekilde kullanmasını, başka insanlardan yardım istemektense akıllı telefonda araştırmayı tercih etmeyi, akıllı telefonu kullanmayı bıraktıktan hemen sonra bile tekrar kullanma isteği duymayı ifade etmektedir. “Dayanma” alt ölçeği ise, bir kimsenin sürekli olarak akıllı telefonunu kullanımını kontrol etmeyi denemesi ancak her seferinde başarısız olması olarak belirlenmektedir.

Bu çalışma bazı sınırlılıklara sahiptir. Birincisi örneklem küçüktür ve olasılığa dayalı bir örnekleme yöntemi ile seçilmemiştir. İkincisi ise cinsiyette kız öğrencilerin sayısı anlamlı bir şekilde erkeklerden fazladır. Bu yüzden cinsiyet farkından kaynaklanan hatalar olabilir. Çünkü cinsiyet değişkeni açısından erkeklerin, kızlara göre daha çok bağımlı ve risk altında olduğu görülmüştür (Ögel ve Cömert, 2009). Bu yüzden cinsiyeti dikkate alan iyi tasarlanmış bir araştırma yapılmalıdır. Üçüncüsü, bu alandaki literatür henüz çok yeterli değildir ve bu yüzden bu çalışmanın kuramsal temeli oldukça zayıftır. Dördüncüsü, bu çalışmada cep telefonunun ve internetin özellikleri göz ardı edildiğinden, daha sonraki araştırmalarda internet bağımlılığı ve cep telefonu bağımlılığı ile bir karşılaştırma yapılarak akıllı telefon bağımlılığı doğrulanmalıdır.

Bu sınırlılıklara rağmen, bu çalışmanın sonuçlarından elde edilen çıkarımlar akıllı telefon bağımlı bağımlı davranışların anlaşılmasını kolaylaştırmakta ve akıllı telefon bağımlılığı için bir tanılama kılavuzu sağlamaktadır. Araştırma bulguları ve bu tartışmalar ışığında, ölçeğin 33 maddelik Türkçe Formu'nun, lise öğrencilerinin akıllı telefona bağımlılık düzeyini geçerli ve güvenilir bir şekilde ölçebileceği sonucuna ulaşılmıştır. Bu araştırmanın sonuçlarına dayalı olarak ölçeğin, akıllı telefona olan bağımlılığının bileşenlerini değerlendirmek için bağımlılıkla ilgilenen araştırmacılar tarafından genel bir uygulama şeklinde kullanılabilir yararlı bir ölçme aracı olduğu söylenebilir. Araştırmacılar bu ölçeği kullanarak akıllı telefon bağımlılığın çeşitli yönlerine ilişkin bilgiler elde edebilirler. Akıllı telefon bağımlılığın bu yönlerini ölçmek, araştırmacılara bu bağımlılığın bireyin akademik başarısı, kişiler arası iletişim gibi değişkenler üzerine etkilerini araştırma imkanı sağlar.

Ölçeğin lise öğrencileri için geçerli ve güvenilir ölçümler sağladığına ilişkin kanıtlar elde edilmesi ise, öğrencilerin akıllı telefon bağımlılığın bu yönleri açısından gelişimsel olarak izlenmesine ve bağımlılık düzeylerindeki değişimlerin belirlenmesine imkan sağlayacaktır.

Daha sonraki çalışmalarda, akıllı telefon bağımlılığı tanılama sistemin yapısal modelini kanıtlamak için akıllı telefon bağımlılığı ile kaygı, depresyon, yalnızlık, düşük özsaygı, dürtü ve sosyal çevreye uyumsuzluk arasındaki ilişkiler araştırılmalıdır. Ayrıca, ATBÖ'nün güvenilirlik ve uygunluğunu artırmak için kısa bir formunun geliştirilmesine ya da uyarlanan SAS ölçeğinin kısa formunun (SAS-Short Version) uyarlanmasına ihtiyaç vardır.

Kwon ve arkadaşlarına (2013) göre, SAS akıllı telefon bağımlılığının kliniksel tanımına hizmet eden ilk ölçektir. Bu bağlamda bu çalışma ile uyarlaması yapılan ATBÖ'de, Türk lise öğrencileri için akıllı telefon bağımlılığının kliniksel tanımına hizmet eden ilk ölçektir.

KAYNAKÇA

- Bayraktutan, F. (2005). *Aile içi ilişkiler açısından internet kullanımı*. Yüksek Lisans Tezi, Sosyal Bilimler Enstitüsü, İstanbul Üniversitesi, İstanbul.
- Bianchi, A., & Phillips, J. G. (2005). Psychological predictors of problem mobile phone use. *Journal of Cyberpsychology & Behavior*, 8 (1), 39-51.
- Büyüköztürk, Ş. (2013). *Sosyal Bilimler İçin Veri Analizi El Kitabı*. Ankara: Pegem Akademi.
- Comings, D. E. (1995). *Tourette's syndrome: A Behavioral Spectrum Disorder*. *Behavioral Neurology of Movement Disorders*. New York: Raven.
- Egger, O., & Rauterberg, M. (1996). *Internet behaviour and addiction*. Yüksek lisans tezi, Swiss Federal Institute of Technology Üniversitesi, Zurih, Switzerland.
- Günüç, S. (2009). *İnternet bağımlılık ölçeğinin geliştirilmesi ve bazı demografik değişkenler ile internet bağımlılığı arasındaki ilişkilerin incelenmesi*. Yüksek Lisans Tezi, Sosyal Bilimler Enstitüsü, Yüzüncü Yıl Üniversitesi, Van.
- Hu, L.T., & Bentler, P.M.(1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6 (1), 1-55.
- Jinoh, C. (2014). Analysis of influence of smart phone addiction practice on ADHD symptoms of elementary school students. *Korean Review of Crisis and Emergency Management*, 10(5), 159-178.
- Karaca, M. (2007). *Sosyolojik bir olgu olarak internet gençliği: Elazığ örneği*. Doktora tezi, Sosyal Bilimler Enstitüsü, Fırat Üniversitesi, Elazığ.
- Kim, D., Lee, Y., Lee, J., Nam, J.K. & Chung, Y. (2014). Development of Korean smartphone addiction proneness scale for youth. *PLOS ONE*, 9(5), 1-7.
- Kline, P. (2000). *An Easy Guide to Factor Analysis*. London and New York: Routledge.
- Kwon, M., Lee, J.Y., Won, W.Y., Park, J.W. & Min, J.A. (2013). Development and validation of a smartphone addiction scale (SAS). *PLOS ONE*, 8(2), 1-7.
- Lane, W., & Manner, C. (2011). The impact of personality traits on smartphone ownership and use. *International Journal of Business and Social Science*, 17(2), 22-28.
- Lin, S. S. J., Tsai, C. C. (2002). Sensation seeking and internet dependence of taiwanese high school adolescents. *Computers in Human Behavior*, 18(4), 411-426.
- Ögel, K., & Cömert, İ.T. (2009). İnternet ve bilgisayar bağımlılığının yaygınlığı ve farklı etkenlerle ilişkisi. *Türkiye Klinikleri J Foren Med*, 6(1), 9-16.
- Park, J.C., Hyun, S.J. & Ha, H. (2014). Comparison between the internet and mobile phone addiction of adolescent with structural equation model and dominance analysis. *The Journal of Korean Association of Computer Education*, 16(1), 11-22.
- Park, N., & Lee, H. (2012). Social implications of smartphone use: Korean college students' smartphone use and psychological well-being. *Cyberpsychology, Behavior, and Social Networking*, 15(9), 491-497.
- Park, P., & Chen, J.V. (2007). Acceptance and adoption of the innovative use of smartphone. *Industrial Management & Data Systems*, 9(107), 1349-1365.
- Phillips, J. G., Ogeil, R. P. & Blaszczynski, A. (2011). Electronic interests and behaviors associated with gambling problems. *Int J Ment Health Addiction*, 1(2), 33-42.
- Saygılı, S. (2002). *Ergenlik sorunları*. İstanbul: Elit Yayınları.
- Schermelleh E. K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Shinse, O. & Baek, S. (2013). The Influence of Adolescents Smart Phone Addiction on Aggression. *Korean Review of Crisis and Emergency Management*, 9(11), 345-362.
- Stein, D.J., Hollander, E., Simeon, D. Cohen, L., Islam, M.N. & Aronowitz, B. (1994). Neurological soft signs in female trichotillomania patients, obsessive-compulsive disorder patients, and health control subject.. *J Neuropsychiatry Clin Neurosci*, 6, 184-187.
- Şimşek, Ö.F. (2007). *Yapısal Eşitlik Modellemesine Giriş: Temel İlkeler ve Lisrel Uygulamaları*. Ankara: Ekinoks Yayınları.
- Tsai, C. C., & Lin, S. S. J. (2003). Internet addiction of adolescents in Taiwan: An Interview study. *CyberPsychology & Behavior*, 6(6), 649-652.

EXTENDED ABSTRACT

Addiction is described as not be able to stop or control any substance or behavior (Egger and Rauterberg, 1996). Addiction emerges with biological affiliation to a specific substance such as alcohol and drug and as a function of nervous system. In addition, some researchers suggest that addiction is because of not only biological but also behavioral (Comings, 1995; Stein et. al., 1994). Technological devices causes a pioneer change in society especially with development of technology and also increases behavioral addictions.

One of technological devices causing this change in society is smartphone. According to a research made by International Data Corporation, IDC, there were 305 millions smartphone sales in 2010 around the world. By increasing about 62, this quantity of sales reached 494 millions in 2011. Accessing internet by using smartphones among teenagers is seen as a popular activity. The popularity of this activity can be the cause of excessive internet use (Lin and Tsai, 2002). One of the significant reasons why internet is commonly preferred and used is the availability of internet easily and at any moment in day with the help of smartphone. Because people usually prefer activities that are easy accessible and practicable to make use of their spare time (Bayraktutan, 2005).

The number of smartphones are rapidly increasing in our country. Teenagers especially at high school ages are affected more than other age groups as they are the generation that are introduced advanced technological devices like smartphone. In literature, it is indicated as 12-18 age group is a critical period for internet addiction (Tsai and Lin, 2003). As a result, it is possible to see that smartphones are especially common among teenagers. Smartphone Addiction Scale (Kwon et al., 2013), is administered to high school students in our country as individuals in adolescents are more sensitive to both drug addiction and behavioral addiction. Results of the study become more of an issue as they may be used in other studies which are aimed to prevent smartphone addiction and specify possible effects with other variables. It is aimed in this study to culturally adapt Smartphone Addiction Scale, which assesses individuals' smartphone addiction, for high school students in Turkey and to find proof for construct validity.

Method

Aimed to adapt Smartphone Addiction Scale for Turkish high school students and analyze psychometric characteristic, this study is a survey research. A survey research is an research approach that aspires to describe a substantial case as is.

Population and Sample

This study was conducted in five high schools; two in Ankara, two in Kayseri and one in Elazığ, that were chosen with convenience sampling method, one of purposive sampling methods. Sampling of the study is comprised of 715 students in grade 9-12 on a volunteer basis. At first, students were asked if they used smartphone or not, then students that did not use smartphone or complete the questionnaire were excluded from study; consequently, 456 participants, 175 boys and 281 girls, were included in this study. Mean age of participants in the study was 16,56.

Procedure

During the process of adapting this scale for Turkish high school students, test adaptation instructions in "International Test Commission Guidelines: Adaptation Process of a Psychological Measurement Scale" were followed. First, necessity of this scale for our country was decided and then necessary permissions were obtained from scale developers. During the process of adaptation Smartphone Addiction Scale, original English items, follow up options and instruction were translated to Turkish by four researchers. It was necessary to test if the original items were equipollence with translated items after reviewing translation form. For this purpose, methods of one

way translation and evaluation of the translation by another translators which are forms of qualitative processes were used. Accordingly, an expert review form which included original and translation forms of instruction, items and follow up options was prepared. This form was separately evaluated by eight field experts and linguists (two linguists, a psychological counseling and guidance expert and five assessment and evaluation experts. After evaluation, the most agreed translation of each item by field experts and linguists was included in scale. As a result, the translated form of the scale was constituted. Then, this translated scale was implemented to individuals in sampling.

Results

First, exploratory factor analysis was carried out to analyze factor structure of Smartphone Addiction Scale (SAS). In exploratory factor analysis aimed for analyzing construct validity of measurements, Bartlett sphericity test was done to see if sampling size was enough or not. Values of KMO should be above .70 and Bartlett test should be meaningful for values' consistency in factor analysis (Büyüköztürk, 2013). It was found from the analysis that Kaiser-Mayer-Olkin (KMO) was 0.909 and χ^2 value of Bartlett Sphericity test was 6221.88 ($p < .001$) and these values were significant and it was considered as sampling was appropriate for EFA. Cook's and Leverage values were examined to test if data set displayed multivariate normal distribution and to identify extreme values. ZPRED-ZRESIL scatter plot was examined to identify if there was linear relationship among variables and there was. Calculated correlation coefficient was analyzed for relation among item scores and it was detected that there was no value more than .80, hence there was no multicollinearity. After all premises were fulfilled, factor analysis was done.

Scale's factor structure in Turkish culture was compared with original factor structure and proofs related to construct validity of related measurements were found by using EFA and CFA in this study. According to EFA result for original scale there was a structure with six factors (Kwon et al., 2013). As factor structure obtained with exploratory factor analysis was not corresponded with six factors structure in original scale, confirmatory factor analysis was used to test the compatibility of data with factors in original scale in this study. For this purpose, an assessment model based on original scale was designated and tested.

In this study; reliability coefficients, fit indices, factor load and error values related to indicators calculated for adapting scale for Turkish high school students point that six factors model arising from sub-dimensions "daily-life disturbance, positive anticipation, withdrawal, cyberspace-oriented relationships, overuse and tolerance" like in original scale showed a good compatibility with data set.

Discussion

Obtaining valid and reliable measurements related to smartphone users' addiction level has great importance in terms of developing social programs to prevent this addiction as a result of use of smartphone. Although it is our country that does not have a smartphone manufacturing industry, it's seen that the smartphone has a quite widespread use in almost all segments of the society. According to the survey, the smartphone seems to be used mostly for fun and games. In this context, it is important to define the smartphone addiction well and to develop policies against this addiction in social base. In this study, "Smart Phone Addiction Scale" was adapted for Turkish high school students for the purpose of setting the ball rolling for the future researches and preparing a new program and recommending against the spread of this addiction.

This study has some limitations. First, sample is small and a probability based sampling methods was not selected. Second, the number of female students in gender was significantly higher than males. So, there may be errors due to the difference of gender. Because, it was found that in terms of gender, males are much more dependent and at higher risk than females (Ögel and Cömert, 2009). For this reason, a well-designed research taking gender into account should be made. Third, literature in this area is not yet far enough and so the theoretical basis of the study is very weak. Last, smartphone addiction should be confirmed by making a comparison with the mobile phone

addiction and internet addiction in later researches as the mobile phone and internet facilities were not investigated in this study.

Though these limitations, the inferences obtained from the results of this study make easier to understand dependent behaviors related with smartphone and provide an identification guide for smartphone addiction. In the light of research findings and these discussions, it has been found that Turkish form of the 33-point scale can measure high school students' smartphone addiction in a valid and reliable way. Based on the results of this study, it can be said this scale is a useful assessment tool that can be used as a general practice by researchers dealing with addiction to evaluate the components of addiction on smartphones. Researchers can obtain information on various aspects of the smartphone addiction using this scale. Measuring these aspects of the smartphone addiction gives opportunity to the researchers to search the effects of the addiction on variables such as the academic achievement of the individual and interpersonal communication.

Obtaining evidence that the scale provides reliable and valid measurement for high school students gives chance to monitoring the students' smartphone addiction in terms of these developmental aspects and identifying the changes in the level of addiction.

In later researches, the relationship between smartphone addiction and depression, loneliness, low self-esteem, impulse (motivation) and non-compliance to the social environment should be investigated to prove the structural model of the smartphone addiction identification system. Also there is a need to develop a short form of SAS to improve the reliability and conformity or an adaptation of the scale of the short form of the SAS (SAS-Short Version).

Sato Test Kuramı Yöntemleriyle Farklı Örneklemelerden Elde Edilen Madde Parametrelerinin Tutarlılığının İncelenmesi

Examination of the Consistency of the Sato Test Theory Item Parameters Obtained From Different Samples

Sait ÇÜM *

Selahattin GELBAL**

Ching-Pin TSAI***

Öz

Bu araştırmada, aynı evrenden tesadüfi olarak seçilen örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUİ) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır. Bu amaç doğrultusunda, 20 madde ve bu maddelere dayalı 1000 adet yanıt örüntüsü simülasyon yoluyla elde edilmiş, yanıt örüntüleri içerisinde seçkisiz olarak seçilen ve 30 yanıt örüntüsü içeren 10 farklı veri seti (örneklem) oluşturulmuştur. Örneklemelerin her birinden elde edilen MİSGİD değerlerinin farklı örneklemeler arası tutarlılığının belirlenebilmesi amacıyla Spearman Sıra Farkları Korelasyon Katsayıları hesaplanmış ve buna ek olarak değer ortalamaları arasındaki farkın anlamlılığı Kruskal Wallis-H Testi ile incelenmiştir. Ayrıca örneklemelerden elde edilen MUİ ile yapılan sınıflandırmaların ne oranda uyum gösterdiği yüzde oranlar yardımıyla incelenmiştir. Elde edilen bulgular, Sato Test Kuramında maddelerin kalitesinin belirlenebilmesi için üretilen MİSGİD parametrelerinin ve MUİ parametresine dayalı yapılan sınıflandırmaların tutarlı sonuçlar verdiğini göstermiştir.

Anahtar Kelimeler: Sato test kuramı, öğrenci-problem çizelgesi, madde analizi.

Abstract

On this research, it has been aimed to the examination of consistency of Localized Grey Relational Grade-Problem (LGRG-P) values and accommodation item classifications which are made via Problem Caution Index (PCI) between different samples that are randomly chosen from the same universe. In accordance with this purpose, 20 items and 1000 response patterns based on these items have been obtained by simulation, 10 different dataset (sample) which are randomly chosen among response patterns and involve 30 response patterns have been constructed. In order to examination of consistency of obtained LGRG-P values among different samples, Spearman Rank-Order Correlation Coefficients have been calculated and additionally significance of difference among value averages has been examined by Kruskal Wallis-H Test. Moreover, with the help of percentages at what rate of item classifications made by PCI obtained from samples show accommodation. Findings which are obtained have showed that Sato Theory Parameters which are produced in order to determine the quality of items have given consistent results.

Keywords: Sato test theory, student-problem chart, item analysis.

GİRİŞ

Modern toplumda eğitimde ve psikolojide kullanılan testler bireylerin hayatında çok önemli bir rol oynamaktadır. Test puanları, bireylerin eğitim görecekları okulları büyük ölçüde belirlemede dolayısıyla uzmanlaşacakları mesleklerin belirlenmesi sürecinde kritik bir unsur olarak ortaya çıkmaktadır. Başka bir bakış açısıyla test puanları, kişinin hayatı boyunca kendisini nasıl göreceği ve başkalarının onu nasıl göreceğini belirleyen bir etmen haline gelmiştir. Bu anlamda test puanlarının nasıl elde edileceği ve nasıl yorumlanacağı konuları üzerinde dikkatle durulmalıdır. Eğitimde ve

*Doktora Öğrencisi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, e-posta:saitcum@hotmail.com

** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta:sgelbal@gmail.com

*** Dr., National Taichung University, Eğitimde Ölçme ve İstatistik Enstitüsü, Taichung-Tayvan, e-posta: tsai.chenbin@msa.hinet.net

psikolojide test geliştirme ve ölçme süreçleri, oldukça karmaşık ve geniş kapsamlı bilgiler bütünü üzerine inşa edilmiş bir bilimdalı olan psikometri çerçevesinde ele alınır. Psikometristler, test maddelerinin nasıl geliştirileceği, nasıl seçileceği, verilen cevapların sayılara nasıl dönüştürüleceği, bunların nasıl yorumlanacağı ve elde edilen verilerin kalitesinin nasıl değerlendirileceği gibi sorulara matematiksel ve istatistiksel teorilere dayalı olarak yanıt ararlar. Söz konusu teorilere “test kuramı” ya da “psikometrik kuram” adı verilmektedir (Suen, 1990).

Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) geçmişten günümüze üzerinde en fazla çalışılan başlıca iki test kuramı olarak öne çıkmaktadır. Alanyazında, Klasik Test Kuramı ve Madde Tepki Kuramının bazı üstün ve zayıf yönlerinin tartışıldığı görülmektedir. Hambleton, Swaminathan ve Rogers (1991), Klasik Test Kuramının sınırlılıklarını şu şekilde açıklamaktadırlar: Madde ve test istatistikleri gruba bağlı olarak hesaplanır. Bu sınırlılık madde ve test özelliklerinin birbirinden ayıramamasıyla ilişkilidir. Eğer bir test zor ise kişi düşük yeteneğe, eğer kolay ise yüksek yetenek düzeyine sahipmiş gibi görünebilir. Katılımcıların yeteneği ise test maddelerinin güçlük düzeyini belirlemektedir çünkü KTK’da maddelerin ayırt ediciliği ve güçlüğü belli bir grup için geçerlidir. Bu nedenle madde parametreleri uygulamanın yapıldığı gruba bağlı olarak değişecektir. Bu durum paralel formlar oluşturma ve farklı testlere katılan bireyleri karşılaştırma konularında sorunların yaşanmasına neden olur. Ayrıca KTK’ya göre doğru yanıtlanan toplam soru sayısından elde edilen puan bireyin yetenek düzeyi olarak kabul edilir. Bu tür bir hesaplamada farklı sorulara doğru yanıt vermiş olmalarına rağmen eşit sayıda doğru yanıt bulunan bireyler aynı yetenek düzeyinde kabul edilmektedir. Örneğin, 20 maddelik bir matematik testini alan ve farklı yedi maddeye doğru yanıt veren iki öğrenci olduğu varsayılırsa bu öğrencilerin ayırt edebilmesi sorunu ortaya çıkacaktır. Doğru yanıt verilen maddelerin güçlüklerinin ve ayırt ediciliklerinin önemli derecede birbirinden farklı olması durumunda iki bireyin aynı yetenek düzeyinde kabul edilmesinin ölçme sonuçlarının hatalı yorumlanmasına yol açacağı öngörülebilir. Böyle bir durumda, test sonuçlarını eğitimin iyileştirilmesi için bir dönüt olarak kullanacak olan öğretmenin yedi kolay soruya doğru yanıt veren öğrenci ile yedi zor soruya doğru yanıt veren öğrenci için uygulayacağı yöntemleri seçerken test puanlarını göz önünde bulundurması yetersiz olmanın ötesinde yanıltıcı da olacaktır. Özellikle biçimlendirici değerlendirme (formative assesment) durumlarında KTK’ya dayalı ölçme süreçlerinden istenilen düzeyde etkili sonuçlar elde edilemeyeceği görüşü öne sürülebilir. MTK modelleri farklı yanıt örüntüsüne sahip eşit sayıda doğru yanıt bulunan bireylerin yetenek düzeyi bakımından birbirlerinden ayırt edilebilmesi konusunda etkili çözümler sunmaktadır fakat MTK’ya dayalı parametrelerin daha hatasız kestirilebilmesi için uygulamaların geniş örneklem üzerinde yapılması gerekmektedir (Hulin, Lissak ve Drasgow, 1982; Ree ve Jensen, 1983; Hambleton ve van der Linden, 1997; de Ayala, 2009). Okullarda yapılan ölçmeler ise küçük katılımcı grupları üzerinde yürütülmekte bu durum başlı başına MTK’ya dayalı test uygulamalarını olanaksız kılmaktadır. Üstelik MTK’ya dayalı test geliştirmek ve sonuçları değerlendirmek alan uzmanlığı gerektiren bir süreçtir. Söz konusu nedenlerden dolayı öğretmen yapımı testlerin geliştirilmesi ve test sonuçlarının değerlendirilmesi sürecinde KTK’ya dayalı tekniklerin kullanılmasının daha uygun olduğu görüşü öne sürülebilir. Buna karşın KTK’nın sınırlılıkları düşünüldüğünde ölçme sonuçlarının geçerliği, güvenilirliği ve karşılaştırılabilirliği bakımından önemli sorunların ortaya çıkması küçük örneklemelerde daha etkili ölçmelerin yapılabilmesini sağlayacak farklı yöntem ve tekniklerin geliştirilmesi ihtiyacını doğurmaktadır.

Takahiro Sato 1969’da diğer test kuramlarının sınırlılıklarından doğan ihtiyaçtan hareketle katılımcıların cevap örüntüleri üzerinden performanslarını ayırt etmeye yönelik küçük örneklem üzerinde de etkili sonuçlar üretebilecek bir sistem üzerinde çalışmalarını yürütmüştür. Sato’nun geliştirmiş olduğu Öğrenci-Madde Çizelgesi Analizi (Student-Problem Chart Analysis) ile hem maddelerin kalitesi hem de öğrencilerin bilme ve öğrenme durumları belirlenebilmektedir. Bu anlamda Sato Test Kuramı modern test kuramları arasında yerini almaktadır (Switzer ve Connell, 1990; Wu, 1999; Sheu, Nguyen, Nguyen, Pham, Tsai ve Nagai, 2014).

Öğrenci-Madde Çizelgesi Analizi ile *Öğrenci Uyarı İndeksi* (Student Caution Index) ve *Madde Uyarı İndeksi* (Problem Caution Index) olmak üzere iki tür indeks üretilmektedir. Öğrenci Uyarı İndeksi (ÖUI) aracılığıyla öğrenciler sınıflandırılarak her bir öğrencinin testle ölçülen konu

alanlarını öğrenme ve maddeleri dikkatli bir şekilde yanıtlama durumlarına ilişkin dönüt alınmaktadır. Benzer şekilde Madde Uyarı İndeksi (MUI) aracılığıyla maddeler sınıflandırılarak her bir maddenin işlerliği hakkında dönüt alınmaktadır. Bu indeksler aşağıdaki formüllerle hesaplanır (Sheu, Nguyen, Ngyuyen, Pham, Tsai ve Nagai, 2014; Lin ve Yih, 2015).

$$\text{ÖÜİ} = 1 - \frac{\sum_{m=1}^M (Y_{nm})(Y_m) - (Y_n) \cdot (u')}{\sum_{m=1}^{Y_n} Y_m - (Y_n) \cdot (u')}$$

$$\text{MUI} = 1 - \frac{\sum_{n=1}^N (Y_{nm})(Y_n) - (Y_m) \cdot (\mu')}{\sum_{n=1}^{Y_m} Y_n - (Y_m) \cdot (\mu')}$$

m: Maddeler ($m=1,2,3,\dots,M$)

n: Öğrenciler ($n=1,2,3,\dots,N$)

Y_{nm}: *n*. öğrencinin *m*. maddeden aldığı puan (doğru:1, yanlış: 0)

Y_m: Öğrencilerin *m*. maddeden aldıkları puanların toplamı

Y_n: *n*. öğrencinin toplam test puanı

$$u' : \frac{\sum_{m=1}^M Y_m}{M}, \quad \mu' : \frac{\sum_{n=1}^N Y_n}{N},$$

Madde Uyarı İndeksi (MUI), maddeleri dört farklı sınıfa atamak için kullanılır. Her bir madde için hesaplanan MUI değerinin 0,50'den daha büyük bir değer alması ilgili maddede bazı sorunların olabileceğine ilişkin bilgi verir. Elde edilen MUI değeri maddenin doğru yanıtlanma oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Düzgün çalışan bir madde.

A': Kısmi olarak sorunlu bir madde. Tekrar gözden geçirilebilir.

B: Zor bir madde.

B': Sorunlu bir madde. Ters işliyor olabilir. Testten atılmalı ya da yeniden yazılmalı.

Öğrenci Uyarı İndeksi (ÖÜİ) ise, öğrencileri altı farklı sınıfa atamak için kullanılır. Her bir öğrenci için hesaplanan ÖÜİ değerinin 0,50'den daha büyük bir değer alması ilgili öğrencinin öğrenme eksikliklerinin olabileceğine ya da dikkatsizlikle kaçırıldığı soruların olabileceğine ilişkin bilgi verir. Elde edilen ÖÜİ değeri öğrencinin maddeleri doğru yanıtlama oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Etkili öğrenme gerçekleşmiş.

A': Öğrenme var fakat çok dikkatsiz.

B: Genel olarak iyi ancak biraz daha çalışmaya ihtiyacı var.

B': Biraz dikkatsiz ve daha fazla çalışmaya ihtiyacı var.

C: Öğrenme düzeyi yetersiz.

C': Öğrenme gerçekleşmemiş.

Söz konusu ÖÜİ ve MUI indekslerinin öğrencilerin yetenek düzeyleri ve maddelerin kalite düzeyleri hakkında yeterli bilgi vermediği iddia edilebilir. Bunun nedeni, nitel veri düzeyindeki çıktılarının, bireyleri ve maddeleri karşılaştırabilme, yapılan değerlendirme sonucu çeşitli kararlar alabilme anlamında sınırlı kalmasıdır. Öğrenci-Madde Çizelgesi'nin bu sınırlılığını ortadan kaldırmak

amacıyla Nagai (2010) bulanık mantığa (fuzzy logic) dayalı Gri İlişki Analizi'nin Sato Test Kuramı'na dahil edilmesini önererek Gri Öğrenci-Madde Çizelgesi Analizi'ni (Grey Student-Problem Chart Analysis) geliştirmiştir. Bu analize dayalı olarak öğrencilerin test performansları üzerinden üretilen *Öğrenciler için Sınırlandırılmış Gri İlişki Derecesi –ÖİSGİD-* (Localized Grey Relational Grade-Student) ve maddelere verilen yanıtlar üzerinden üretilen *Maddeler için Sınırlandırılmış Gri İlişki Derecesi –MİSGİD-* (Localized Grey Relational Grade-Problem) elde edilmektedir (Sheu, Nguyen, Ngyuyen, Pham, Tsai ve Nagai, 2014). Gri ilişki dereceleri hesaplanırken kullanılan vektörler ve formüller aşağıda verilmiştir (Sheu, Nguyen, Tsai, Pham, Ngyuen ve Nagai, 2014).

Gri ilişki analizinde karşılaştırılan vektörler,

$$x_0 = (x_0(1), x_0(2), \dots, x_0(k), \dots, x_0(n)) : \text{Referans vektör}$$

$$x_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n)) : \text{İncelenen vektör}$$

Gri öğrenci-madde çizelgesi analizinde incelenen vektörler,

$$S_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n)) : i. \text{ öğrencinin tüm maddelere verdiği yanıtlar vektörü, } k = 1, 2, 3, \dots, n$$

$$P_k = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_m(k)) : k. \text{ maddeye tüm öğrencilerin verdiği yanıtlar vektörü, } i = 1, 2, 3, \dots, m$$

$$\text{ÖİSGİD} = \Gamma_{(x_0, x_i)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{0i}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$$\text{MİSGİD} = \Gamma_{(x_j, x_0)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{k0}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$$\bar{\Delta}_{0i} = \left(\sum_{k=1}^n (x_0(k) - x_i(k))^p \right)^{\frac{1}{p}} : \text{Referans vektör ve incelenen vektör (öğrenciler için) arasındaki Minkowski uzaklık ölçüsü}$$

$$\bar{\Delta}_{k0} = \left(\sum_{i=1}^m (x_0(k) - x_i(k))^p \right)^{\frac{1}{p}} : \text{Referans vektör ve incelenen vektör (maddeler için) arasındaki Minkowski uzaklık ölçüsü}$$

$$\bar{\Delta}_{max} = \bar{\Delta}_{0i} \text{ ve } \bar{\Delta}_{k0} \text{ için maksimum değer}$$

$$\bar{\Delta}_{min} = \bar{\Delta}_{0i} \text{ ve } \bar{\Delta}_{k0} \text{ için minimum değer}$$

Gri İlişki Analizi'nin dahil edilmesiyle birlikte Sato Test Kuramı öğrenciler ve öğretim sürecinin etkililiği ile ilgili dönüt alabilmek için kullanılabilmesinin (biçimlendirici değerlendirme) yanı sıra düzey belirlemeye ve buna bağlı olarak öğrencilerin eğitim durumları ile ilgili karar almaya (özetleyici değerlendirme) yönelik olarak kullanılabilir.

Sato Test Kuramı üzerinde Türkiye'de çok az sayıda çalışma yapıldığı ve kurama ilişkin uygulamaların genellikle Uzak Doğu Ülkelerinde yaygınlaştığı belirtilebilir. Günümüzde eğitim, psikoloji, sağlık ve endüstri gibi pek çok farklı alanda ölçme ve değerlendirme çalışmalarına verilen

önem yaygınlaşmakta ve ölçme faaliyetlerinin önemli bir bölümü küçük katılımcı grupları üzerinde gerçekleştirilmektedir. Bu bağlamda değerlendirildiğinde Sato Kuramcılarının küçük örneklem üzerinde etkili sonuçların elde edilmesine olanak sağlayan yöntemler sunduklarına dair iddialarının dikkate alınmaya değer olduğu sonucuna varılabilir. Kurama ilişkin yöntemlerin geçerliği ve güvenilirliğine ilişkin çalışmaların sayısının artmasının kuramın işlerliğinin ve kullanılabilirliğinin ortaya koyulması bakımından büyük önem taşıdığı açıktır.

Araştırmanın Amacı

Bu araştırmanın amacı, aynı evrenden tesadüfi olarak seçilen örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUI) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesidir. Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- 1) Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Dereceleri arasında ne düzeyde bir ilişki bulunmaktadır?
- 2) Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin ortalamaları arasındaki fark istatistiksel olarak anlamlı mıdır?
- 3) Örneklemelerden elde edilen Madde Uyarı İndeksleri ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?

YÖNTEM

Bu çalışmada, simülatif olarak oluşturulmuş 20 maddelik bir testi yanıtlayan 1000 kişi içerisinde seçilen 30'ar kişilik 10 farklı örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUI) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır. Araştırma, bilgi üretmeye yönelik kuramsal bir çalışma özelliği taşıması bakımından temel araştırma niteliğindedir.

Verilerin Oluşturulması

Araştırmada, WinGen3 programı ile yetenek düzeyleri -3 ve +3 aralığında değişen 1000 kişilik bir katılımcı grubu ile a parametresi 0 ve 2 aralığında; b parametresi -3 ve +3 aralığında; c parametresi 0 ve 0,20 aralığında değişen 20 maddeden oluşan, 1-0 şeklinde puanlanan bir test oluşturularak her bir katılımcı için yanıt örüntüsü elde edilmiştir. WinGen programı genellikle Madde Tepki Kuramı kapsamında yapılan çalışmalar için simülasyon veri üretmek amacıyla kullanılmaktadır. Bu nedenle maddeler MTK parametrelerine dayalı olarak üretilmiş olmasına karşın bu çalışma için parametrelerin çok fazla öneminin olmadığı belirtilebilir. Araştırmada WinGen programı gerçeğe daha uygun yanıt örüntüleri elde etmek amacıyla tercih edilmiştir.

Yanıt örüntüleri elde edildikten sonra bunların içerisinde seçkisiz olarak seçilen ve 30'ar kişinin yanıt örüntülerinden oluşan 10 farklı veri seti (örneklem) oluşturulmuştur. Söz konusu örneklemeler ve maddelerin daha iyi betimlenebilmesi amacıyla her bir örnekleme seçilen yanıt örüntüleri üzerinden Klasik Test Kuramı yaklaşımıyla madde güçlük indeksleri (p, doğru yanıt puan toplamı/30) ve madde ayırt edicilik indeksleri (r, madde-toplam test puanları arası Çift Serili Korelasyon Katsayısı) hesaplanmıştır. Elde edilen bulgular Tablo 1 ve Tablo 2 de sunulmuştur.

Tablo 1. Farklı örneklerden elde edilen madde güçlük indeksleri

| Maddeler | Örn1 | Örn2 | Örn3 | Örn4 | Örn5 | Örn6 | Örn7 | Örn8 | Örn9 | Örn10 |
|----------|------|------|------|------|------|------|------|------|------|-------|
| | p | p | p | p | p | p | p | p | p | p |
| 1 | 0,13 | 0,13 | 0,10 | 0,13 | 0,07 | 0,23 | 0,07 | 0,27 | 0,17 | 0,20 |
| 2 | 0,50 | 0,67 | 0,50 | 0,43 | 0,50 | 0,57 | 0,47 | 0,53 | 0,40 | 0,50 |
| 3 | 0,30 | 0,40 | 0,37 | 0,40 | 0,40 | 0,57 | 0,17 | 0,37 | 0,30 | 0,57 |
| 4 | 0,70 | 0,53 | 0,67 | 0,80 | 0,60 | 0,77 | 0,60 | 0,67 | 0,70 | 0,67 |
| 5 | 0,40 | 0,23 | 0,30 | 0,43 | 0,27 | 0,43 | 0,43 | 0,40 | 0,33 | 0,30 |
| 6 | 0,43 | 0,47 | 0,40 | 0,50 | 0,47 | 0,60 | 0,40 | 0,50 | 0,60 | 0,57 |
| 7 | 0,43 | 0,37 | 0,33 | 0,63 | 0,47 | 0,47 | 0,40 | 0,43 | 0,40 | 0,47 |
| 8 | 0,83 | 0,87 | 0,83 | 0,90 | 0,77 | 0,90 | 0,97 | 0,77 | 0,97 | 0,87 |
| 9 | 0,40 | 0,33 | 0,37 | 0,57 | 0,40 | 0,57 | 0,43 | 0,60 | 0,43 | 0,50 |
| 10 | 0,30 | 0,43 | 0,30 | 0,43 | 0,37 | 0,60 | 0,30 | 0,43 | 0,50 | 0,57 |
| 11 | 0,90 | 0,87 | 0,80 | 0,87 | 0,80 | 0,87 | 0,77 | 0,77 | 0,87 | 0,80 |
| 12 | 0,33 | 0,23 | 0,30 | 0,53 | 0,37 | 0,53 | 0,33 | 0,50 | 0,37 | 0,40 |
| 13 | 0,27 | 0,23 | 0,23 | 0,30 | 0,23 | 0,40 | 0,27 | 0,40 | 0,23 | 0,23 |
| 14 | 0,87 | 0,87 | 0,87 | 0,90 | 0,97 | 0,90 | 0,90 | 0,93 | 0,87 | 0,90 |
| 15 | 0,23 | 0,27 | 0,20 | 0,37 | 0,23 | 0,30 | 0,37 | 0,30 | 0,30 | 0,33 |
| 16 | 0,73 | 0,53 | 0,57 | 0,77 | 0,57 | 0,70 | 0,57 | 0,67 | 0,73 | 0,70 |
| 17 | 0,40 | 0,50 | 0,43 | 0,50 | 0,43 | 0,63 | 0,33 | 0,67 | 0,47 | 0,50 |
| 18 | 0,33 | 0,17 | 0,20 | 0,17 | 0,17 | 0,23 | 0,27 | 0,17 | 0,23 | 0,13 |
| 19 | 0,50 | 0,70 | 0,57 | 0,60 | 0,67 | 0,77 | 0,57 | 0,37 | 0,63 | 0,63 |
| 20 | 0,47 | 0,50 | 0,53 | 0,47 | 0,40 | 0,63 | 0,50 | 0,47 | 0,53 | 0,50 |

Tablo 1 incelendiğinde, 1, 13, 15 ve 18. maddelerin zor; 8,11 ve 14. maddelerin ise kolay maddeler olduğu görüşü öne sürülebilir.

Tablo 2. Farklı örneklerden elde edilen madde ayırt edicilik indeksleri

| Maddeler | Örn1 | Örn2 | Örn3 | Örn4 | Örn5 | Örn6 | Örn7 | Örn8 | Örn9 | Örn10 |
|----------|------|------|------|------|------|------|------|------|------|-------|
| | r | r | r | r | r | r | r | r | r | r |
| 1 | 0,21 | 0,26 | 0,07 | 0,25 | 0,24 | 0,02 | 0,06 | 0,30 | 0,13 | 0,05 |
| 2 | 0,33 | 0,18 | 0,18 | 0,43 | 0,21 | 0,44 | 0,24 | 0,52 | 0,33 | 0,68 |
| 3 | 0,28 | 0,50 | 0,66 | 0,65 | 0,40 | 0,01 | 0,56 | 0,53 | 0,30 | 0,37 |
| 4 | 0,56 | 0,68 | 0,56 | 0,46 | 0,70 | 0,52 | 0,79 | 0,69 | 0,55 | 0,67 |
| 5 | 0,56 | 0,65 | 0,48 | 0,51 | 0,34 | 0,60 | 0,64 | 0,51 | 0,43 | 0,57 |
| 6 | 0,53 | 0,58 | 0,70 | 0,59 | 0,68 | 0,06 | 0,62 | 0,71 | 0,65 | 0,86 |
| 7 | 0,53 | 0,79 | 0,77 | 0,76 | 0,72 | 0,64 | 0,64 | 0,64 | 0,77 | 0,72 |
| 8 | 0,64 | 0,50 | 0,50 | 0,38 | 0,58 | 0,29 | 0,21 | 0,64 | 0,28 | 0,52 |
| 9 | 0,42 | 0,28 | 0,43 | 0,51 | 0,62 | 0,01 | 0,68 | 0,52 | 0,40 | 0,68 |
| 10 | 0,41 | 0,41 | 0,38 | 0,58 | 0,74 | 0,17 | 0,57 | 0,31 | 0,53 | 0,37 |
| 11 | 0,41 | 0,35 | 0,50 | 0,37 | 0,34 | 0,08 | 0,58 | 0,32 | 0,31 | 0,55 |
| 12 | 0,36 | 0,28 | 0,28 | 0,41 | 0,45 | 0,40 | 0,44 | 0,53 | 0,22 | 0,27 |
| 13 | 0,43 | 0,26 | 0,60 | 0,50 | 0,16 | 0,60 | 0,21 | 0,62 | 0,46 | 0,41 |
| 14 | 0,62 | 0,33 | 0,41 | 0,43 | 0,13 | 0,40 | 0,41 | 0,21 | 0,31 | 0,47 |
| 15 | 0,53 | 0,57 | 0,38 | 0,49 | 0,40 | 0,73 | 0,33 | 0,57 | 0,46 | 0,09 |
| 16 | 0,51 | 0,75 | 0,62 | 0,71 | 0,79 | 0,16 | 0,78 | 0,74 | 0,41 | 0,67 |
| 17 | 0,31 | 0,52 | 0,31 | 0,40 | 0,46 | 0,10 | 0,31 | 0,30 | 0,33 | 0,23 |
| 18 | 0,53 | 0,40 | 0,35 | 0,33 | 0,24 | 0,66 | 0,04 | 0,34 | 0,31 | 0,35 |
| 19 | 0,21 | 0,39 | 0,28 | 0,05 | 0,26 | 0,48 | 0,04 | 0,48 | 0,04 | 0,06 |
| 20 | 0,62 | 0,26 | 0,43 | 0,24 | 0,36 | 0,03 | 0,03 | 0,17 | 0,26 | 0,10 |

Tablo 2 incelendiğinde, Çift Serili Korelasyon Katsayısı yöntemi ile hesaplanan madde ayırt edicilik indekslerinin farklı örneklem arasında büyük farklılıklar gösterdiği görülmektedir. Örneğin, 3. maddenin ayırt edicilik indeksi Örneklem 3 verileri ile hesaplandığında 0,66 değerini alırken Örneklem 6 verileri ile hesaplandığında 0,01 değerini almaktadır. Benzer şekilde 20. maddenin ayırt edicilik indeksi Örneklem 1’de 0,62 olarak hesaplanmış fakat Örneklem 6 ve Örneklem 7 de 0,03 olarak belirlenmiştir. Tablodan hareketle birçok madde için benzer durumlardan söz edilebilir. Klasik Test Kuramı yöntemleri küçük örneklemelerde madde parametreleri hesaplamak için sıklıkla kullanılmaktadır fakat karşılaşılan bu tip durumlar KTK’nın bir dezavantajı olarak ifade edilebilir.

Verilerin Analizi

Çalışmada, Maddeler için Sınırlandırılmış Gri İlişki Dereceleri (MİSGİD) seçkisiz olarak belirlenen 10 farklı örneklemin her birinden, Sheu, Pham, Nguyen ve Nguyen (2013) tarafından Öğrenci-Madde Çizelgesi Analizi ve Gri Öğrenci-Madde Çizelgesi Analizi için geliştirilen Matlab Paketi ile elde edilmiştir. Paket, Matlab R2013b versiyonu üzerinde çalıştırılmıştır. Elde edilen MİSGİD değerlerinin farklı örneklemler arası tutarlılığının belirlenebilmesi amacıyla Spearman Sıra Farkları Korelasyon Katsayıları hesaplanmış ve buna ek olarak değer ortalamaları arasındaki farkın anlamlılığı Kruskal Wallis-H Testi ile incelenmiştir. Söz konusu analizler için SPSS V22 programı kullanılmıştır. Son olarak örneklemelerden elde edilen Madde Uyarı İndeksleri ile yapılan sınıflandırmaların ne oranda uyum gösterdiği yüzde oranlar yardımıyla incelenmiştir.

BULGULAR

Öncelikle, “Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Dereceleri (MİSGİD) arasında ne düzeyde bir ilişki bulunmaktadır?” sorusuna yanıt bulmak amacıyla, 10 farklı veri setinden (örneklem) elde edilen MİSGİD değerleri hesaplanmış ve elde edilen bulgular Tablo 3’te sunulmuştur.

Tablo 3. Farklı Örneklemelerden Elde Edilen MİSGİD Değerleri

| Maddeler | Örn1 | Örn2 | Örn3 | Örn4 | Örn5 | Örn6 | Örn7 | Örn8 | Örn9 | Örn10 |
|----------|------|------|------|------|------|------|------|------|------|-------|
| 1 | 0,07 | 0,07 | 0,05 | 0,07 | 0,12 | 0,03 | 0,03 | 0,14 | 0,09 | 0,11 |
| 2 | 0,42 | 0,29 | 0,29 | 0,25 | 0,34 | 0,29 | 0,27 | 0,32 | 0,23 | 0,32 |
| 3 | 0,23 | 0,16 | 0,20 | 0,23 | 0,34 | 0,23 | 0,09 | 0,20 | 0,16 | 0,16 |
| 4 | 0,32 | 0,45 | 0,42 | 0,55 | 0,52 | 0,37 | 0,37 | 0,42 | 0,45 | 0,42 |
| 5 | 0,12 | 0,23 | 0,16 | 0,25 | 0,25 | 0,14 | 0,25 | 0,23 | 0,18 | 0,16 |
| 6 | 0,27 | 0,25 | 0,23 | 0,29 | 0,37 | 0,27 | 0,23 | 0,29 | 0,37 | 0,34 |
| 7 | 0,20 | 0,25 | 0,18 | 0,39 | 0,27 | 0,27 | 0,23 | 0,25 | 0,23 | 0,27 |
| 8 | 0,63 | 0,59 | 0,59 | 0,68 | 1,00 | 0,52 | 0,82 | 0,52 | 0,82 | 0,63 |
| 9 | 0,18 | 0,23 | 0,20 | 0,34 | 0,34 | 0,23 | 0,25 | 0,37 | 0,25 | 0,29 |
| 10 | 0,25 | 0,16 | 0,16 | 0,25 | 0,37 | 0,20 | 0,16 | 0,25 | 0,29 | 0,34 |
| 11 | 0,63 | 0,68 | 0,55 | 0,63 | 0,63 | 0,55 | 0,52 | 0,52 | 0,63 | 0,55 |
| 12 | 0,12 | 0,18 | 0,16 | 0,32 | 0,32 | 0,20 | 0,18 | 0,29 | 0,20 | 0,23 |
| 13 | 0,12 | 0,14 | 0,12 | 0,16 | 0,23 | 0,12 | 0,14 | 0,23 | 0,12 | 0,12 |
| 14 | 0,63 | 0,63 | 0,63 | 0,68 | 0,68 | 0,82 | 0,68 | 0,74 | 0,63 | 0,68 |
| 15 | 0,14 | 0,12 | 0,11 | 0,20 | 0,16 | 0,12 | 0,20 | 0,16 | 0,16 | 0,18 |
| 16 | 0,32 | 0,48 | 0,34 | 0,52 | 0,45 | 0,34 | 0,34 | 0,42 | 0,48 | 0,45 |
| 17 | 0,29 | 0,23 | 0,25 | 0,29 | 0,39 | 0,25 | 0,18 | 0,42 | 0,27 | 0,29 |
| 18 | 0,09 | 0,18 | 0,11 | 0,09 | 0,12 | 0,09 | 0,14 | 0,09 | 0,12 | 0,07 |
| 19 | 0,45 | 0,29 | 0,34 | 0,37 | 0,52 | 0,42 | 0,34 | 0,20 | 0,39 | 0,39 |
| 20 | 0,29 | 0,27 | 0,32 | 0,27 | 0,39 | 0,23 | 0,29 | 0,27 | 0,32 | 0,29 |

Tablo 3 incelendiğinde, 10 farklı örneklem üzerinden yapılan hesaplamalarda 8. madde haricinde diğer tüm maddelerin 0,75'in altında MİSGİD değeri elde ettiği görülmektedir. MİSGİD değerinin 0-1 aralığında değiştiği göz önünde bulundurulduğunda madde kalitelerinin yüksek olmadığı belirtilebilir. Daha sonraki aşamada, elde edilen MİSGİD değerleri arasındaki ilişkiler Spearman Sıra Farkları Korelasyon Katsayısı tekniği ile belirlenmiş ve elde edilen bulgular Tablo 4'te sunulmuştur.

Tablo 4. MİSGİD Değerleri Arasındaki Korelasyonlara İlişkin Bulgular

| | Örn1 | Örn2 | Örn3 | Örn4 | Örn5 | Örn6 | Örn7 | Örn8 | Örn9 | Örn10 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Örn1 | | | | | | | | | | |
| Örn2 | ,874* | | | | | | | | | |
| Örn3 | ,950* | ,933* | | | | | | | | |
| Örn4 | ,770* | ,866* | ,857* | | | | | | | |
| Örn5 | ,934* | ,841* | ,954* | ,844* | | | | | | |
| Örn6 | ,943* | ,926* | ,956* | ,893* | ,904* | | | | | |
| Örn7 | ,829* | ,929* | ,879* | ,852* | ,809* | ,852* | | | | |
| Örn8 | ,741* | ,780* | ,813* | ,832* | ,793* | ,781* | ,733* | | | |
| Örn9 | ,899* | ,886* | ,919* | ,895* | ,959* | ,899* | ,873* | ,814* | | |
| Örn10 | ,922* | ,858* | ,900* | ,867* | ,935* | ,912* | ,853* | ,819* | ,972* | |

*Korelasyon ,01 düzeyinde anlamlıdır.

Tablo 4 incelendiğinde, farklı örneklemelerden 20 madde için elde edilen MİSGİD değerlerinin birbirleri ile yüksek düzeyde pozitif korelasyon gösterdiği ve söz konusu korelasyon değerlerinin ,01 düzeyinde anlamlı olduğu görülmektedir. Bu bulgudan hareketle, Sato Test Kuramı kapsamında hesaplanan ve maddelerin kalitesi hakkında bilgi veren Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin, aynı maddeleri yanıtlayan katılımcılar arasından seçilen farklı örneklemelerde tutarlı sonuçlar verdiği iddia edilebilir. Bu iddiayı sağlamlaştırmak amacıyla, “Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin ortalamaları arasındaki fark istatistiksel olarak anlamlı mıdır?” sorusuna yanıt aranmış ve farklı örneklemelerden elde edilen MİSGİD değerlerinin ortalamaları arasındaki farkın anlamlılığı nonparametrik bir test olan Kruskal Wallis-H testi ile incelenmiştir. Elde edilen bulgular Tablo 5'te sunulmuştur.

Tablo 5. MİSGİD Değerlerinin Ortalamaları Arasındaki Farkın Anlamlılığına İlişkin Bulgular

| | MİSGİD |
|---------|--------|
| Ki-Kare | 8,929 |
| sd | 9 |
| p | ,444 |

Tablo 5 incelendiğinde, MİSGİD değerlerinin ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmadığı ($p=0,444$) görülmektedir. Bu bulgunun, farklı örneklemelerden elde edilen MİSGİD parametrelerinin tutarlı sonuçlar verdiği ilişkin ikincil bir kanıt oluşturduğu ifade edilebilir.

Son olarak, “Örneklemelerden elde edilen Madde Uyarı İndeksleri (MUI) ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?” sorusuna yanıt aranmış, yapılan sınıflandırmalar ve bunlara ilişkin uyum yüzdeleri Tablo 6'da sunulmuştur.

Tablo 6. Madde Uyarı İndeksleri ile Yapılan Sınıflandırmaların Uyumuna İlişkin Bulgular

| Maddeler | Örn1 | Örn2 | Örn3 | Örn4 | Örn5 | Örn6 | Örn7 | Örn8 | Örn9 | Örn10 | Uyum |
|----------|------|------|------|------|------|------|------|------|------|-------|------|
| 1 | B | B | B' | B | B' | B | B' | B' | B' | B' | %60 |
| 2 | A' | A' | A' | B' | A' | A' | B' | A' | B' | A' | %70 |
| 3 | B | B' | B | B | A' | B' | B | B | B' | B | %60 |
| 4 | A | A | A | A | A | A | A | A | A | A | %100 |
| 5 | B | B | B | B | B | B | B | B | B | B | %100 |
| 6 | B | B | B | A | A | B | B | B | A | A | %60 |
| 7 | B | B | B | A | B | B | B | B | B | B | %90 |
| 8 | A | A | A | A | A | A | A | A | A | A | %100 |
| 9 | B' | B | B | A | A' | B | B | A | B | A | %50 |
| 10 | B | B | B | B | A' | B | B | B' | A | A' | %60 |
| 11 | A | A | A | A | A | A | A | A' | A | A | %90 |
| 12 | B' | B' | B' | A | A | B' | B | A | B' | B' | %60 |
| 13 | B' | B | B | B | B | B' | B' | B | B' | B | %60 |
| 14 | A | A | A | A | A | A | A | A | A | A | %100 |
| 15 | B | B | B | B | B | B | B' | B | B | B' | %80 |
| 16 | A | A | A | A | A | A | A | A | A | A | %100 |
| 17 | A | B' | B' | A | A' | B | B' | A' | B' | A' | %40 |
| 18 | B | B | B | B | B | B' | B' | B' | B' | B | %60 |
| 19 | A' | A' | A' | A' | A' | A' | A' | B' | A' | A' | %90 |
| 20 | A' | B | A | B' | A' | B' | A' | B' | A' | A' | %50 |

Tablo 6'da yer alan sınıflandırmaların farklı örneklemeler üzerinden yapılan hesaplamalar sonucu değişkenlik gösterebildiği belirlenmiştir. Buna karşın 4, 5, 8, 14 ve 16. maddelerin 10 farklı örnekleme de aynı şekilde (%100 uyum) sınıflandırıldığı, 2, 7, 11, 15 ve 19. maddelerin de sınıflandırma uyumlarının yüksek olduğu belirlenmiştir (uyum yüzdeleri için modu en yüksek olan sınıf baz alınmıştır). Sınıflandırılma bakımından en fazla değişkenlik gösteren maddenin 17. madde olduğu (%40 uyum) ifade edilebilir. Söz konusu madde haricindeki tüm maddelerin %50 ve üzeri uyum oranında sınıflandırıldığı bulgusu göz önünde bulundurulduğunda Madde Uyarı İndekslerine dayalı olarak yapılan sınıflandırmaların tutarlı olduğu görüşü öne sürülebilir.

SONUÇLAR ve TARTIŞMA

Bu araştırmada, simülatif olarak oluşturulmuş 20 maddelik bir testi yanıtlayan 1000 kişi içerisinde seçilen 30'ar kişilik on farklı örneklemden elde edilen Maddeler için Sınırlanmış Gri İlişki Derecelerinin tutarlılığının ve Madde Uyarı İndeksleri aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır.

Farklı örneklemelerden 20 madde için elde edilen MİSGİD değerlerinin birbirleri ile yüksek düzeyde pozitif korelasyon gösterdiği ve söz konusu korelasyon değerlerinin ,01 düzeyinde anlamlı olduğu belirlenmiştir. Buna ek olarak Kruskal Wallis-H testi ile yapılan incelemelerde MİSGİD değerlerinin ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmadığı ($p= 0,444$) görülmüştür. Elde edilen bulgulardan hareketle, Sato Test Kuramında maddelerin kalitesinin belirlenebilmesi için üretilen MİSGİD parametrelerinin oldukça küçük örneklemeler (30 kişi) üzerinden hesaplanmasına karşın tutarlı ve yakın değerler aldığı sonucuna ulaşılmıştır.

Madde Uyarı İndeksleri ile yapılan sınıflandırmaların örneklemeler arası değişkenlik gösterdiği belirlenmiş buna karşın beşi %100 uyum göstermekle birlikte toplamda 10 maddenin yüksek düzeyde uyum oranıyla sınıflandırıldığı ve bir madde haricinde %50'nin altında uyum oranıyla sınıflandırılan madde olmadığı görülmüştür. Elde edilen bulgular, Sato Test Kuramı kapsamında yapılan test analizlerinin küçük örneklemelerde bile iyi sonuçlar verdiği iddialarına kanıt

oluşturabilecek niteliktedir. Sheu ve diğerleri (2014), Gri Öğrenci-Problem Çizelgesi Analizinin yalnızca küçük örneklemelerde değil büyük örneklemelerde de kullanılabileceği sonucuna ulaşmışlardır. Bu çalışma ise, küçük örneklemeler üzerinde Sato Test Kuramına dayalı üretilen madde parametrelerinin tutarlılığının incelenmesi süreciyle sınırlandırılmıştır. Alanyazında benzer çalışmalara ulaşılamaması bu konu üzerinde yapılabilecek tartışmaları sınırlandırmaktadır. Sato Test Kuramıyla ilgili çalışmaların son yıllarda yalnızca Uzak Doğu Ülkelerinde bir artış gösterdiği bunların da kuramı tanıtmaya (Lin ve Chen, 2006; Wang, Sheu ve Nagai, 2011; Wang ve Chen 2013) ya da kurama ilişkin geliştirilen bir yazılımı tanıtmaya (Wu, 1999; Sheu, Pham, Nguyen ve Nguyen, 2013; Sheu, Nguyen, Nguyen, Pham, Tsai ve Nagai, 2014; Pham, Sheu ve Nagai, 2015) amacıyla yapıldığı belirlenmiştir. Bu anlamda, kuram çerçevesinde üretilen parametrelerin çeşitli koşullar altında incelenmesi ve bu alanda yapılan çalışmaların sayısının artırılması gerekliliği ortaya çıkmaktadır. Benzer çalışmaların sayısının artması ve kuramın güvenilir ve geçerli sonuçlar üretebildiğine dair bulguların fazlalaşması, küçük örneklemelerde Klasik Test Kuramına alternatif yöntemlerin kullanımının yaygınlaşması bakımından oldukça önemlidir. Söz konusu çabalar, daha etkili ölçme araçları geliştirme, bireylerin bir takım özellikleriyle ilgili daha doğru bilgiler elde edebilme ve bireyler hakkında yapılan değerlendirmelerin isabetliliği bakımından alandaki ilerlemeye katkı sağlayacaktır. Tüm bunların yanısıra, kuramın uygulanabilirliği, iç geçerliliğin sağlanmasıyla ilişkili olduğu kadar dış geçerliliğin sağlanmasını da zorunlu kılmaktadır. Söz konusu nedenle, kurama ilişkin parametrelerin hem kendi içerisinde tutarlılığının belirlenebilmesi hem de diğer test kuramlarına dayalı yöntemlerle üretilen parametrelerle karşılaştırılarak sonuçların incelenmesi, daha sonra yapılacak çalışmalarda ele alınması bakımından önerilebilir.

KAYNAKÇA

- de Ayala, R. J. (2009). The theory and practice of item response theory. New York: The Guilford Press.
- Hambleton, R. K., Swaminathan, H., ve Rogers, H. (1991). Fundamentals of item response theory. Newbury Park CA: Sage.
- Hambleton, R. K., ve van der Linden, W. J. (1997). Handbook of modern item response theory. New York:Springer.
- Hulin, C. L., Lissak, R. I., ve Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. Applied Psychological Measurement, 6, 249-260.
- Lin, Y.H., ve Chen, S.M. (2006). The investigation of S-P chart analysis on the test evaluations of equality axiom concepts for sixth graders. Proceedings of the 2nd International Conference on Educational Technologies, Romania, Bucharest.
- Lin, Y.H., ve Yih, J.M. (2015). Application of IIRS in mathematics instruction to promote pupils decimal concept. The International Conference on Language, Education and Psychology, Taiwan.
- Pham, D.H., Sheu, T.W., ve Nagai, M. (2015). PCSP 1.0 software for partial credit S-P chart analysis. International Journal of Hybrid Information Technology. 8(6), 309-322.
- Ree, M. J., ve Jensen, H. E. (1983). Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss & R. D. Bock (Eds.), New horizons in testing: Latent trait test theory and computerized adaptive testing (pp. 135-146). New York: Academic Press.
- Suen, H.K. (1990). Principles of test theories. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sheu, T. W., Nguyen, P. T., Nguyen, P. H., Pham, D. H., Tsai, P. C., ve Nagai, M. (2014). A MATLAB toolbox for misconceptions analysis based on S-P chart, grey relational analysis and ROC. Transactions on Machine Learning and Artificial Intelligence, 2, 72-85.
- Sheu, T.W., Nguyen, P.T., Tsai, C.P., Pham, D.H., Nguyen, P.H. ve Nagai, M. (2014). Using grey student-problem chart in the evaluation of tests with large data sets. Education Practise and Innovation, 1(2), 2372-3106.
- Sheu, T.W., Pham, D.H., Nguyen, P.T., ve Nguyen, P.H. (2013). Amatlab toolbox for student-problem chrt and grey student-problem chart and its application. International Journal of Kansei, 4(2), 75-86.
- Switzer, D. M. ve Connell, M. L. (1990). Practical applications of student response analysis, Educational Measurement: Issue and Practice, 9, 15-17.
- Wang, B.T., Sheu, T.W., ve Nagai, M. (2011). Evaluating the english-learning of engineering students using the grey S-P chart: a facebook case study in Taiwan. Global Journal of Engineering Education, 13(2), 51-56.

Wang, C.H. ve Chen, C.P. (2013). Employing online S-P diagnostic table for qualitative comments on test results. *The Electronic Journal of e-Learning*, 11(3), 263-271.

Wu, H. (1999). Software Based on S-P Chart Analysis and Its Applications. *Proceedings of the National Science Council*, 8, 102-107.

EXTENDED ABSTRACT

Introduction

Psychometrists search response for questions such as how test items will be improved, how they will be chosen, how given answers will be turn into numbers, how they will be commented and how quality of obtained data will be assessed, as depending on mathematical and statistical theories. Aforementioned theories are called "test theory" or "psychometric theory" (Suen, 1990).

Classical Test Theory (CTT) and Item Response Theory (IRT) have become prominent as main two test theories that they have been studied mostly from past up to the day. On the literature, it has been seen that some of strengthening and weaknesses of Classical Test Theory and Item Reaction Theory have been discussed. Applications require to be made on large samples in the order that parameters based on IRT are to be estimated more faultlessly (Hulin, Lissak and Drasgow, 1982; Ree and Jensen, 1983; Hambleton and van der Linden, 1997; de Ayala, 2009). Measurements made at schools are conducted on small participant groups, this case precludes test applications based on IRT at there. Furthermore, developing test based on IRT and assessing results are a process which requires domain expertise. Due to aforementioned reasons, a view has been suggested about that using techniques based on CTT is more appropriate for development of teacher - made tests and assessment of test results, but still limitations of CTT are considered, that important problems occur in terms of measurement results validity, reliability and comparability. Because of this there is a need to develop different method and techniques which will provide to make more efficient measurements on small samples. When it is assessed with this sense, it can be concluded that suggestions on which Sato Theorists present methods providing opportunity to obtain effective results on small samples are things to consider.

The goal of this research is to the examination of consistency of Localized Grey Relational Grade-Problem (LGRG-P) values and accommodation item classifications which are made via Problem Caution Index (PCI) between different samples that are randomly chosen from the same universe. In accordance with this purpose, the following questions searched for answers:

- 1) What a level of the relation is there among Localized Grey Relational Grade Problem obtained from Samples?
- 2) Is the difference among averages of Localized Grey Relational Grade Problem obtained from samples significant as statistic?
- 3) At what rate classifications made by Problem Caution Indexes obtained from samples show accommodation?

Method

The research has a basic research quality in terms of having feature to be a theoretic study related to produce information. On this research, 1000 response patterns for each participants that their ability levels change at the range -3 and +3 with consists of 20 items (dichotomously) which change at the range of -3 and +3 for "b" parameters; at the range of 0 and 0,20 for "c" parameters; at the range of 0 and 2 for "a" parameter on WinGen3 program. After response patterns have been obtained, 10 different dataset (sample) which are randomly chosen among them and consist of response patterns of 30 persons.

On the study, Used Matlab Pack which is developed for Analysis of Student-Problem Chart and Analysis of Grey Student-Problem Chart by Sheu, Pham, Nguyen and Nguyen (2013) to obtain that

Localized Grey Relational Grades. The pack has been run on Matlab R2013b version. In order to examination of consistency of obtained LGRG-P values among different samples, Spearman Rank-Order Correlation Coefficients have been calculated and additionally significance of difference among value averages has been examined by Kruskal Wallis-H Test via SPSSV22.

Results and Discussion

It has been determined that LGRG-P values obtained for 20 items from different samples show highly positive correlation with each other and aforementioned correlation values are significant at 0.01 level. Additionally, on the examinations which are made by Kruskal Wallis-H Test, it has been seen that the difference among averages of LGRG-P values are not statistically significant ($P=0,444$). From this point of obtained findings, although LGRG-P parameters produced to determine quality of items on Sato Test Theory are calculated through quite small (30 persons) samples, it has been concluded that consistent and close values are gotten.

It has been determined that Classifications made by Problem Caution Index have showed change among samples but it has been seen that five of them have showed 100% accommodation and total 10 items have been classified with high level accommodation ratio and except from one item, there is not any items classified with accommodation ratio under 50%.

Findings which are obtained have quality to provide evidence against suggestions that test analyses made within the concept of Sato Test Theory have given well-results even on small samples. This study has been limited with examination process of item parameters consistency produced only as depending on Sato Test Theory. With this sense, a necessity reveals about that parameters produced for individuals responding test within the frame of the theory are to be examined under various conditions. It can be emphasized on a view that many studies which are conducted under different conditions are required in the order that consistency of parameters related to the theory in itself and also accuracy of results is to be discussed as comparing with parameters produced by methods based on other test theories.

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 Differential 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 separate 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 applied 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

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 özelliklerle 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 kanılarıyla yardımcıyla 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 güvenilirliğini 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 deseni (Kelecioğlu, 1994), Rasch modeli kullanılarak ortak test deseni (Şahhüseyinoğlu, 2005) ve dikey eşitleme ile (Çetin, 2009) çalışıldığı; KTK'ya dayalı daha çok tek grup deseni (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 performansını 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 örneklem araştırma 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 yıllarında SBS'de çıkan 100 madde olmak üzere toplam 275 maddenin cinsiyete dayalı DMF analizi yapılmıştır. Madde Tepki Kuramı'na dayalı yöntemlerden elde edilen 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 örneklemelerden; $N=750$ ve $N=1000$ büyüklüğünde alt gruplar oluşturulmuştur. Kullanılan iki yöntemde 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 $(\Delta R^2) \geq 0.010$ olan maddeler ile MH yöntemine göre $-1 \leq \Delta MHI < -1.5$ 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 $-1.5 \leq \Delta MHI$ 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 öğretmenin 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 Madde | Ortak Maddeler | | DMF'siz Madde | 15 | 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 Madde | Ortak Maddeler | | DMF'siz Madde | 13 | 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 kullanılarak 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 parametrelili 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 | | B FORMU | |
|-------------------------------------|----------------|---------------|----------------|---------------|
| | Ortak Maddeler | | 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 r _{yx} (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 $\alpha= 0.01$ 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 örneklemeden 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 veriye uyumlu olduğuna işaret etmektedir (Carmines ve McIver, 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 parametrelilik 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 varsayımlarını 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 | | |
|------------------------------------|--|--------------------|-------------------|
| | 1PLM | 2PLM | 3PLM |
| $n_A=691, n_B=659, N=17$ | | | |
| 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 çıkması oldukça olası bir durumdur (Kang ve Cohen, 2007). Araştırmada maddelerin tek tek model uyumu incelendiğinde, iki parametrelili lojistik modellerle 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) |
| $\mu(b)$ Toplam | 0.152 | -0.053 | 0.084 | -0.118 |
| $\mu(b)$ Ortak Madde | 0.164 | -0.705 | 0.108 | -0.756 |
| $\mu(a)$ Toplam | 1.011 | 1.148 | 0.975 | 0.998 |
| $\mu(a)$ Ortak Madde | 0.529 | 1.132 | 0.839 | 0.905 |

* $\alpha=0.01$ 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 güvenilirlik 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 güvenilirlik katsayıları en yüksek 2PLM'de elde edilirken, 1PLM ve 3 PLM'de Lord'un güvenilirlik katsayıları KR-20 güvenilirlik katsayılarından düşük bulunmuştur. Özdemir'in (2004) araştırmasında benzer şekilde 2PLM'de Lord'un güvenilirlik katsayısı yardımıyla elde edilen güvenirlığın, KTK'ya dayalı KR-20 güvenirligiden yüksek sonuçlar verdiği bulunmuştur.

Eşitleme Hatasının Hesaplanması

Bu araştırmada A ve B formlarının ortak maddeleri DMF'li ve DMF'siz olduğu durumlarda yapılan eşitlemelere karşın 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 kareköküne dayanmaktadır (Paek ve Young, 2005; Harris ve Crouse, 1993):

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

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 formula ortak maddeler içeren A formunu almayan bireylerin gerçek yeteneklerinden (θ) yola çıkarak, almadıkları A formundaki kestirilen yetenekleri (θ^*) aşağıdaki eşitlik (2) ile hesaplanmıştır:

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

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 |

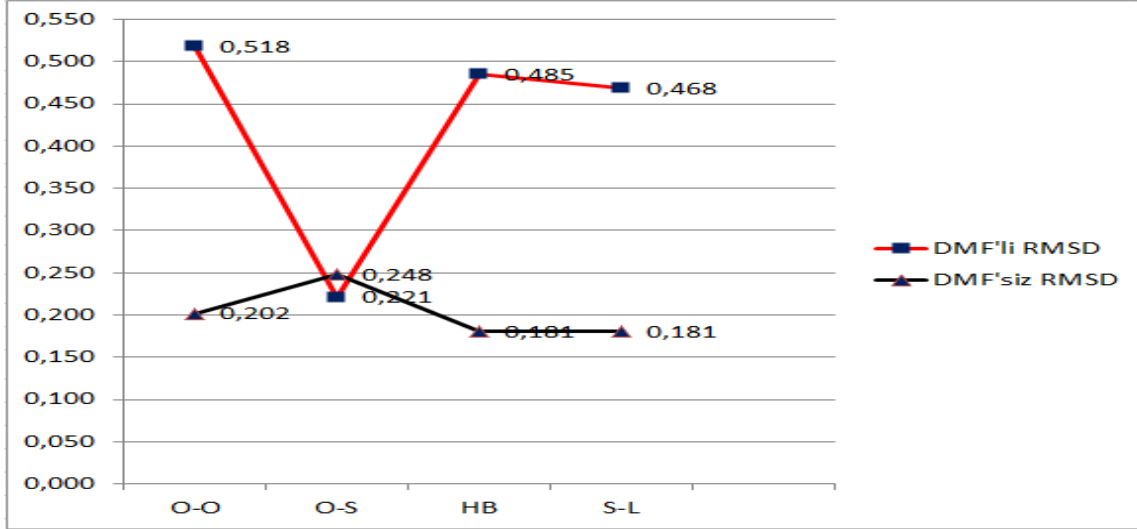
Tablo 9'da ortak maddelerin DMF'siz olduğu durumlarda HB ve S-L yöntemlerine ait eşitleme hataları birbirine eşit elde edilirken diğer hata değerlerinin farklılaştığı göze çarpmaktadır. Eşitleme hatalarını hesaplamada kullanılan A ve B eşitleme katsayıları Tablo 10'da verilmiştir.

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 arasındaki 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) ayrıricılık 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:

$$\begin{array}{l} A = \mu(a_B) / \mu(a_A) \text{ ve,} \\ \mu(a_B) = 0.839 \\ \mu(a_A) = 0.529 \end{array} \left. \vphantom{\begin{array}{l} A = \mu(a_B) / \mu(a_A) \text{ ve,} \\ \mu(a_B) = 0.839 \\ \mu(a_A) = 0.529 \end{array}} \right\} \mathbf{A = 1.586}$$
$$\begin{array}{l} B = \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \mu(b_A) = 0.164 \\ \mu(b_B) = 0.108 \end{array} \left. \vphantom{\begin{array}{l} B = \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \mu(b_A) = 0.164 \\ \mu(b_B) = 0.108 \end{array}} \right\} \begin{array}{l} \mathbf{B = 0.164 - 1.586(0.108)} \\ \mathbf{B = -0.007 \approx -0.01} \end{array}$$

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

$$\begin{array}{l} A = \mu(a_B) / \mu(a_A) \text{ ve,} \\ \mu(a_B) = 0.905 \\ \mu(a_A) = 1.132 \end{array} \left. \vphantom{\begin{array}{l} A = \mu(a_B) / \mu(a_A) \text{ ve,} \\ \mu(a_B) = 0.905 \\ \mu(a_A) = 1.132 \end{array}} \right\} \mathbf{A = 0.80}$$
$$\begin{array}{l} B = \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \mu(b_A) = -0.705 \\ \mu(b_B) = -0.756 \end{array} \left. \vphantom{\begin{array}{l} B = \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \mu(b_A) = -0.705 \\ \mu(b_B) = -0.756 \end{array}} \right\} \begin{array}{l} \mathbf{B = -0.705 - 0.80(-0.756)} \\ \mathbf{B = -0.10} \end{array}$$

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.θ+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:

$$\begin{aligned} A &= \sigma(b_A) / \sigma(b_B) \text{ ve} & B &= \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \left. \begin{aligned} \sigma(b_A) &= 1.758 \\ \sigma(b_B) &= 1.406 \end{aligned} \right\} & \mathbf{A} &= \mathbf{1.250} & \left. \begin{aligned} \mu(b_A) &= 0.164 \\ \mu(b_B) &= 0.108 \end{aligned} \right\} & B &= 0.164 - 1.250 \cdot (0.108) \\ & & & & & \mathbf{B} &= \mathbf{0.029 = 0.03} \end{aligned}$$

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:

$$\begin{aligned} A &= \sigma(b_A) / \sigma(b_B) \text{ ve} & B &= \mu(b_A) - A \cdot \mu(b_B) \text{ ve,} \\ \left. \begin{aligned} \sigma(b_A) &= 0.580 \\ \sigma(b_B) &= 0.761 \end{aligned} \right\} & \mathbf{A} &= \mathbf{0.760} & \left. \begin{aligned} \mu(b_A) &= -0.705 \\ \mu(b_B) &= -0.756 \end{aligned} \right\} & B &= (-0.705) - 0.760(-0.756) \\ & & & & & \mathbf{B} &= \mathbf{-0.13} \end{aligned}$$

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 Heabara 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 konusuken 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. Araştırmanın 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 örnekleme 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 bulunmaktadır. 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 çalışmada 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 çalışmadan farklı olarak uzun test (40 maddeli) ve büyük örneklem (5000) kullanılmıştır. Yine iki çalışmada DMF'li ortak madde oranlarında farklılıklar bulunmaktadır. Bu çalışmada 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 çalışmada MPK'lı madde oranı arttıkça ortak maddelerin güçlük parametre varyansları artmış ve A katsayısı azalmıştır. Böylece bu çalışmadan farklı olarak O-S yönteminde eşitleme hatasının büyümüştür. Bu durumun nedeni DMF-MPK farklılığı başta olmak üzere iki çalışmadaki 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 çalışmalarında O-S yönteminin zayıf performans göstermesine neden olarak, bir ortak maddenin güçlüğünde yaşanan ciddi değişimin temel formun güçlük parametreleri varyansını büyük çapta arttırmasına bağlamışlardır. Çalışmada 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 çalışmada S-L yöntemi için elde edilen bulguyu destekler niteliktedir. Chen (2013) simülatif veri ile yürüttüğü çalışmasında S-L yöntemiyle eşitlemede ortak maddelerin a ve b parametrelerinde ayrı ayrı ele alınan MPK'nın yetenek kestirimlerine etkisini incelemiştir. Çalışmada 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. Çalışma 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 çalışmanın 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 çalışmanın koşulları kısmen benzer seçildiğinde sonuçlarda tutarlılık gözlemlendiği 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). *Değişen madde fonksiyonunun test eşitlemeye etkisi*. Yayınlanmamış 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*. Yayınlanmamış 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. *Journal of Educational Measurement*, 28(2), 147-162.
- Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi*. Yayınlanmamış Yüksek Lisans Tezi, Mersin: Mersin Üniversitesi, Sosyal Bilimler Enstitüsü.
- Büyüköztürk, Ş. (2005). Sosyal bilimler için veri analizi el kitabı. 5. Baskı Ankara: Pegem 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). *Değişen madde fonksiyonlarının sibtest, Mantel Haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. Yayınlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü.
- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması*. Yayınlanmamış 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 procedure? *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-Haenzsel 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*. Yayınlanmamış 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 comparison 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). *Akademik 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*. Yayınlanmamış 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*. Yayınlanmamış 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 differential item functioning: Logistic regression modelling as a unitary framework for binary and likert-type item scores*. Ottawa: 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 separate 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.

TIMSS 2011 Sekizinci Sınıf Öğrencilerinin Matematik Başarısını Etkileyen Değişkenlerinin Bölgelere Göre İncelenmesi

The Investigation of the Variables That Affecting Eight Grade Students' TIMSS 2011 Math Achievement According To Regions

Rukiye ÖLÇÜOĞLU *

Sevda ÇETİN **

Öz

Bu çalışmada, Türkiye'deki Uluslararası Matematik ve Fen Eğilimleri Araştırması (Third International Mathematics and Science Study) TIMSS 2011 uygulamasına katılan sekizinci sınıf öğrencilerinin matematik başarısını etkileyen bazı değişkenler modellenmekte ve modelin coğrafi bölgelere göre ölçme değişmezliği incelenerek, değişkenlerin bölgeler bakımından farklılık gösterip göstermediği tespit edilmeye çalışılmaktadır. Bu araştırmanın örneklemini Türkiye'deki 239 okuldan seçilen 14 yaş grubundaki 6928 öğrenciden oluşmaktadır.

İlk olarak, açılımlayıcı faktör analizi (AFA) ile öğrenci anketinde yer alan ve matematik başarısını etkilediği düşünülen maddeler incelenmiş, 12 madde matematik ile ilgili duyuşsal özellikler, dokuz madde ev ortamı ve beş maddenin de okul ortamını olmak üzere üç boyutta toplandığı görülmüştür. Daha sonra, öğrencilerin matematik başarıları ile modele alınan değişkenler arasındaki ilişkileri test etmek için, yapısal eşitlik modellemesi ile bir model kurulmuştur. Bulgular sonucunda, matematik başarısını belirlemede en yüksek payın pozitif yönde duyuşsal özellikler değişkenine ait olduğu bulunurken, ev ortamı değişkeni pozitif yönde en fazla etki eden ikinci değişken, okul ortamı ise negatif yönde istatistiksel olarak manidar olmayan bir ilişki göstermiştir.

Araştırmanın ikinci aşamasında, modelin bütün bölgelerde geçerli olduğu saptanmıştır. Daha sonra, değişkenlerin coğrafi bölgeler arasında karşılaştırılabilmesi için, modelin Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA) ile ölçme değişmezliği incelenmiştir. Bölgeler alt gruplarında sadece yapısal ve metrik değişmezlik koşullarının yerine getirildiği ve yalnızca bu aşamalarda yapılan karşılaştırmalarda gruplar arası farklılıkların ölçme aracından kaynaklanamayacağı neticesine ulaşılmıştır. Sonuç olarak, ölçek değişmezlik koşulu sağlanamadığından, tam eşdeğerlik saptanamamıştır.

Bu sonuçlar ışığında, değişmezlik ile ilişkili yapılacak karşılaştırmaların yorumlarında farklılıkların bölgelerden mi yoksa ölçme aracından mı kaynaklandığı belirlenemediğinden, bölgeler arası matematik başarı puan karşılaştırmalarının yorumlanmasında dikkatli olunmalıdır.

Anahtar Kelimeler: TIMSS, matematik başarısı, yapısal eşitlik modellemesi ve ölçme değişmezliği

Abstract

In this study, some of the variables affecting the mathematics achievement of Third International Mathematics and Science Study (TIMSS) 2011 eighth-grade in Turkey sample were modeled and whether the variables showed differences among the regions according to the measurement invariance examination. The sample of this research consists of 6928 students in the age group of 14 selected from 239 schools in Turkey.

In the first step, items in the student survey were examined by explanatory factor analysis (EFA) and found out that 12 items were loaded into mathematics related affective domain, five items into home and three items into school environment factors. Then, the model was established by structural equation modeling (SEM). According to the results, mathematics related affective behaviors shows positively the highest effect. The second variable also positively was the home environment. Lastly, a negative non-statistically significant relationship was determined between the school environment and mathematics achievement.

In the second step, the applicability of the model in all regions was found. Then, the invariance of model was examined by means of Multi-Grup Confirmatory Factor Analysis (MG-CFA) to compare the variables between

*Arş. Gör. Rukiye Ölçüoğlu, Yıldırım Beyazıt Üniversitesi, Tıp Fakültesi, Ankara-Türkiye,
e-posta:rukiyeyuce@gmail.com

**Yrd. Doç. Dr. Sevda Çetin, Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye,
e-posta:tsevda@hacettepe.edu.tr

geographical regions. Only the configural and weak invariance conditions in the sub-groups are met by the model. Moreover, because strong invariance condition cannot be met, measurement invariance cannot be exactly determined.

In terms of these results, when we make comparison among the regions, we should be cautious about interpretations of math achievement scores because the differences can be result of the regions' differences as well as the assessment and evaluation instrument.

Keywords: TIMSS, math success, the structural equation modeling and the measurement invariance.

GİRİŞ

Avrupa ve Orta Asya bölgelerinde 29 ülkeyi içeren araştırmaya göre; eğitim, Türk vatandaşlarının endişelendiği en önemli konudur. Bu bölgelerde devlet yatırımlarında öncelikli alanın eğitim olması gerektiğini düşünen ülkelerde, en yüksek oran Tacikistan'dan sonra Türkiye'ye aittir (Çelebi, Güner, Kaya ve Korumaz, 2014). Böyle önemli bir alanda sorunların olması kaçınılmazdır. Geçmişten günümüze bu alanda yapılan çalışmalarda en sık rastlanılan eğitim ve fırsat eşitliği kavramlarının ilişkisi birçok etmene bağlı karmaşık sorunlardan biridir. Örneğin, Türkiye'de eğitim kalitesi açısından Dünya Bankasının 2011 verilerine göre bölgeler arası fark kapanmaya başlamasına rağmen, bazı bölgelerde okula kayıt olma oranı bölgeye ve cinsiyete göre halen değişiklik göstermektedir (Ersoy, 2007).

TIMSS eğitim alanında tüm dünyada altmıştan fazla ülkenin katıldığı geniş çaplı ve ortak teşebbüslü uluslararası bir çalışmadır (Mullis, Martin, Ruddock, O'Sullivan ve Preuschoff, 2009). Bu çalışma, eğitim sistemindeki performansların daha iyi anlaşılması ve bu konudaki uzmanlara yön vermesi amacıyla yapılmakta olup, matematik ve fen alanında iki ayrı eğitim seviyesindeki (4. ve 8. sınıf) öğrencilerin bilgilerini karşılaştırmaktadır. Bu karşılaştırmalar yapılırken bir takım veriler de toplanır. Bunlar; eğitim sistemi, idareciler, öğretmen, okul bilgileri ve öğrenci anketleri gibi verilerdir. Böylece, matematik ile fen başarısında eğitim sisteminin nasıl işlediği konusunda bilgiler sağlanırken, eğitimde yapılacak reform ve gelişmeler için de değerli birer kaynak oluşturulmaktadır (Mullis, Martin, Foy ve Arora, 2012). Mathforum (2006)'da TIMSS gibi uluslararası çalışmaların politikacılar tarafından bu denli takip edilmesinin, matematik ve fen okuryazarlığının bahsedilen ekonomik üretkenliği etkilemesinden kaynaklandığı söylenmiştir.

Feniger, Livneh ve Yogev (2012) ise bu tip uluslararası sınavlarla ülkelerin, diğer ülkeler arasındaki kendi eğitim başarı sıralamalarını tespit edebildikleri gibi bu buldukları sıralama ile gelecekte küresel ekonomide ne kadar yarışabilir bir nesle sahip olacaklarını görebildiklerinden bahsederler. Bu yüzden, önemli bir proje olan uluslararası sınavlar, sevilsen de sevilmesen de sadece akademik bir araştırma olarak görülemez. Dört yıllık dönemler halinde gerçekleştirilerek, bir önceki veya sonraki değerlendirmesi ile bağlantısı kurulabilen TIMSS çalışmaları, eğitimciler ve eğitim politikalarını belirleyenlere düzenli ve güncel bilgi akışı sağlamaktadır. İlk uygulamada dördüncü sınıfta eğitim gören öğrenciler, bir sonraki uygulamada sekizinci sınıfta bu uygulamaya katılmaktadır. Böylece, aynı öğrenci grubu birden fazla değerlendirmeye tabi tutulup, boylamsal bir çalışma gerçekleştirilerek başarı değişimleri izlenebilmektedir (Mullis ve ark., 2009).

Oral ve McGivney'in (2013) TIMSS 2011 analiz raporunda, matematikte dördüncü ve sekizinci sınıf öğrencilerinin dağılım yüzdeleri incelendiğinde, Türkiye'de dikkat çeken en önemli faktör ileri düzey ve düşük düzey altı yeterlik varyans yüzdelerindeki dağılımın çarpıklığıdır. Tüm ülkelerin ortalaması incelendiğinde; ülkemizde düşük düzey altındaki öğrenci sayısı oldukça fazlayken, ileri düzey yeterlikte ortalama ya da daha üstünde performanslar da gözlenmektedir. Yani, Türkiye'deki öğrencilerin eğilimi çok iyi veya çok kötü performansa doğrudur. Özellikle 8. sınıf örneğininin % 40'ı, bu düzeylerde performans göstermiş olan öğrencilerden oluşmuştur. Bu iki farklı grup öğrenci profili özellikle Türkiye'nin coğrafi konum, nüfus büyüklüğü ve ekonomik kalkınması açısından büyük farklılığa sahip bölgelerinde kendini direk belli etmektedir (Marmara ve Güneydoğu bölgelerinin karşılaştırılması gibi). Yeterlik dağılımının bu denli ayrı uç noktalarda toplanması Oral ve McGivney'e (2013) göre, eşitsizliğin eğitim sistemimizde dikkat çekici bir unsur olduğuna işaret etmektedir. Ayrıca, bu kadar öğrencinin düşük düzeyin altında kalmasını, ülkemizdeki öğrenme

süreçleri ile eğitim kalitesinin olması gerekenin altında seyretmesinden kaynaklandığını da vurgulamaktadırlar.

Sonuç olarak; Türkiye gibi eğitim sisteminde sorunları olan ve bu konuda çözüm arayışında olan ülkeler için, TIMSS gibi uluslararası sınavların sonuçlarının takibi ve ortaya çıkan veri kaynaklarından yararlanılması ne olursa olsun gelişmekte olan ülkeler için önemlidir. Sonuçlarının eğitim politikalarının şekillenmesinde yapıcı role sahip olduğu unutulmadan objektif bir biçimde değerlendirilmelidir.

Araştırmanın Amacı

Uluslararası sınavların sonuçlarına bakarak, öğrenciler arasındaki başarı farklarının kapatılması için eğitimde fırsat eşitliğinin sağlanmasına yönelik çalışmaların yapılabilmesi, bu ölçmeler ancak ve ancak değişmez olduklarında mümkündür. Bu yüzden, TIMSS vb. öğrenci başarılarında karşılaştırma yapmayı amaçlayan uluslararası sınavlar farklı gruplarda aynı özelliği ölçebilmelidir. Yani, yapılan gruplar arası karşılaştırmalar ölçme araçlarından bağımsız aynı sonucu verebilmelidir. Farklılığın ölçme aracından mı yoksa bölgelerdeki herhangi bir eşitsizlikten mi kaynaklandığının gösterilebilmesi, bu alanda yapılacak araştırmaların yapılmasının anlamlı olabilmesinde büyük önem taşımaktadır.

Bu yüzden, çalışmanın amacı Türkiye'deki sekizinci sınıf öğrencilerinin matematik başarısını etkileyen bazı değişkenleri TIMSS 2011 verilerini kullanarak modellemek ve bunun bölgeler bakımından farklılık gösterip göstermediğini tespit etmektir. Böylece, eğitimde fırsat eşitliği ve matematiğin çağımızdaki önemi düşünüldüğünde; eğitim sistemimizin kalitesinin ve başarısının yükseltilmesi yönünde ihtiyaç duyulan yönlerin belirlenerek, ülke eğitimi için destekleyici bir çalışma olacağı düşünülmektedir.

Bu amaçla çalışmada aşağıdaki sorulara yanıt aranmıştır:

1. Matematik ile ilgili duyuşsal özellikler, ev ortamı ve okul ortamı değişkenlerinden oluşturulan model, sekizinci sınıf matematik başarısını yordamakta mıdır?
2. Sekizinci sınıf matematik başarısı modeli bölgelere göre ölçme değişmezliğini sağlamakta mıdır?
3. TIMSS 2011 sonuçlarına göre, Türkiye'deki sekizinci sınıf öğrencilerin matematik başarısını yordama modeline alınan hangi değişkenler bölgeler arasında farklılık göstermektedir?

YÖNTEM

Araştırmanın Türü

Çalışmanın ilk aşamasında, oluşturulan modelin sekizinci sınıf matematik başarısını yordayıp yordamadığına bakılmıştır. Bu amaçla önce açılımlı faktör analizi ile TIMSS 2011 çalışmasından elde edilen anket verileri incelenmiş, sonra da yapısal eşitlik modellemesiyle öğrencilerin matematik başarılarına ilişkin bir model kurulmuştur. Bu yüzden, çalışma özü itibarıyla ilişkisel bir araştırmadır.

Çalışmanın ikinci aşamasında ise oluşturulan matematik başarısı modelinin ölçme değişmezliği ile bölgelere göre aynı olup olmadığına bakılmıştır. Böylece, TIMSS 2011 sonuçlarına göre, Türkiye'deki sekizinci sınıf öğrencilerin matematik başarısını yordama modeline alınan hangi değişkenlerin bölgeler arasında farklılık göstermekte olduğunun incelenmesi anlamlı olmuştur. Araştırmanın ikinci aşaması bu yönüyle TIMSS 2011 çalışmasının geçerlilik düzeyini saptamaya yönelik olduğundan, betimsel bir araştırma olarak adlandırılabilir.

Örneklem

Türkiye genelinde TIMSS örneklem planı doğrultusunda tüm ilköğretim okullarının listesi hazırlanarak, tesadüfi (seçkisiz) yöntemle uygulamanın yapılacağı Türkiye'nin yedi bölgesinden sekizinci sınıf düzeyinde toplam 239 okul ve 6928 öğrenci katılmıştır. TIMSS prensiplerine göre, örneklem seçiminde göz önünde bulundurulmuş diğer bir önemli nokta ise, sekizinci sınıf öğrencilerinin uygulamanın yapıldığı tarih itibarı ile 13.5 yaş üstünde olmalarıdır. Türkiye için yaş ortalaması 14 olarak belirtilmiştir (Mullis ve ark, 2012; Martin ve Mullis, 2012). Yukarıdaki bilgiler doğrultusunda, bu araştırmanın evreni Türkiye'deki 14 yaş grubu öğrencilerden, örneklemini ise 239 okuldaki seçilen 14 yaş grubundaki 6928 öğrenciden oluşmaktadır. Ancak, sayıtların kontrolü sonrasında, analizler Tablo 1'de bölgelere göre yüzde dağılımları verilen 5868 kişi üzerinden gerçekleştirilmiştir.

Tablo 1. Coğrafi Bölgelere Göre Öğrencilerin Frekans Dağılımı

| Kod | Bölge | Frekans | Yüzde |
|-----|-------------------|---------|-------|
| 8 | Ege | 682 | 11.6 |
| 9 | Karadeniz | 663 | 11.3 |
| 10 | İç Anadolu | 727 | 12.4 |
| 11 | Doğu Anadolu | 643 | 11.0 |
| 12 | Marmara | 1730 | 29.5 |
| 13 | Akdeniz | 713 | 12.2 |
| 14 | Güneydoğu Anadolu | 710 | 12.1 |
| | TOPLAM | 5868 | 100.0 |

Veri Toplama Araçları

Bu çalışmada kullanılan veriler, öğrencilerin TIMSS 2011 uygulamasından aldıkları matematik başarı testi puanları ile öğrenci anketlerinde yer alan bazı maddelere verdikleri cevaplardan oluşmaktadır. Tablo 2'de görülmektedir. Veriler TIMSS çalışmasının uluslararası internet sitesinden indirilmiştir (TIMSS, 2011).

İşlem

TIMSS temelde ülke genelinin başarı düzeyini belirlemeye çalıştığı için, aslında öğrenciler için tek tek başarı puanı hesaplamaz. Yapılan şey, testlerde benzer performans gösteren öğrencilerin tümü için başarı dağılımlarını hesaplamaktır. Bahsedilen bu dağılımlardan, ilgili öğrenciler için rastgele beşer değer seçilmektedir. Bu çalışmada yer alan bulgular, öğrencilerin performanslarını gösteren bu beş adet olası değer (plausible value) puanlarının hepsi kullanılarak oluşturulmuştur.

TIMSS 2011 anket başlıkları altında modele alınacak değişkenleri ve bu değişkenlerin boyutlarını belirlemek amacıyla, ankette yer alan matematik öğrenme algısı ile ilgili olabilecek tüm maddelere açımlayıcı faktör analizi (AFA) yapılmıştır.

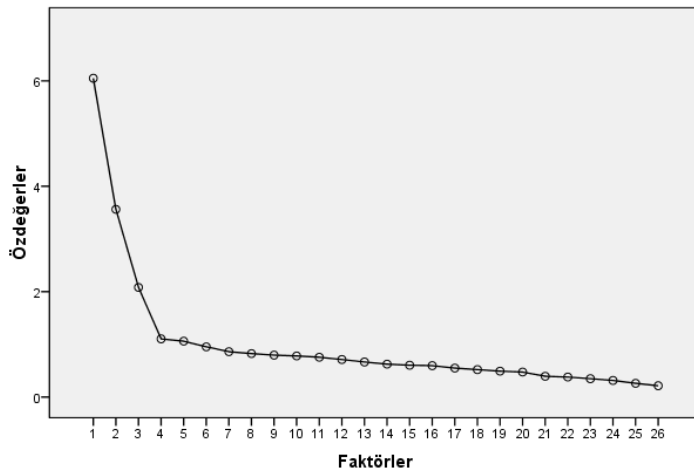
Analize başlamadan önce; olumsuz maddelere dönüştürme işlemi yapılmış, çok fazla eksik veri bulduran maddeler analize alınamamıştır. Örneklem büyüklüğüne göre tolere edilebilecek eksik veri miktarı ile ilgili bir kılavuz yoktur. Ancak Tabachnick ve Fidell'e (2007, s. 63) göre büyük örneklemdeki eksik veri, toplam veri sayısından %5 veya daha az ise eksik veri analizi kullanılabilir.

Tablo 2. Modele alınan maddeler

| Kod | Maddeler |
|---------|---|
| BSBG04 | Evdeki kitap sayısı |
| BSBG05A | Evdeki bilgisayar |
| BSBG05B | Çalışma masası |
| BSBG05D | Kendisine ait oda |
| BSBG05E | Evde internet bağlantısı |
| BSBG05F | Ülkeye özgü varlık göstergesi 1 |
| BSBG05H | Ülkeye özgü varlık göstergesi 2 |
| BSBG05J | Ülkeye özgü varlık göstergesi 3 |
| BSBG13A | Okulda alay edilme |
| BSBG13C | Okulda hakkında yalan söylenmesi |
| BSBG13D | Okulda hırsızlık durumuna maruz kalma |
| BSBG13E | Okulda diğer öğrenciler tarafından saldırıya uğrama |
| BSBM14A | Matematik öğreniminden zevk alıyorum |
| BSBM14B | Keşke matematik çalışmak zorunda olmasam |
| BSBM14C | Matematik sıkıcıdır |
| BSBM14D | Matematikte birçok ilgi çekici şey öğreniyorum |
| BSBM14E | Matematiği seviyorum |
| BSBM16A | Genellikle matematikte iyiyimdir |
| BSBM16C | Matematik benim güçlü yönlerimden biri değildir |
| BSBM16D | Matematikteki şeyleri çabucak öğrenirim |
| BSBM16F | Zor matematik problemleri çözmede iyiyimdir |
| BSBM16G | Öğretmenim matematik derslerindeki zor maddelerde bile iyi olduğumu düşünür |
| BSBM16H | Öğretmenim matematikte iyi olduğumu söyler |
| BSBM16N | Matematiği içeren bir iş isteyebilirim. |

AFA sonucunda, verilerin faktör analizi yapımındaki uygunluğuna bakmak için Kaiser-Meyer-Olkin katsayısı (0,895) ve Barlett küresellik testi ($\chi^2 = 54233,237$; $p=0,000 < 0,05$) incelenmiştir. Bu iki sayıların karşılanmasıyla, AFA sonuçlarının değerlendirilmesine devam edilmiş ve değişkenlerin kaç boyutta toplandığına karar vermek için yamaç-birikinti grafiği incelenmiştir (Şekil 1).

Çizgi Grafiği



Şekil 1. Faktör Analizi Yamaç-Birikinti Grafiği

Tablo 3'te görüldüğü gibi, tüm maddelerin faktör yük değerlerinin kabul düzeyini ($>.32$) karşıladığı saptanmıştır. Yapılan analiz sonucunda, teorik anlamda maddelerin tanımlanabilecek yapılar altında toplandığı görülmektedir. Her bir faktöre uygun isimlendirme yapılmıştır. Birinci boyuttaki 12 madde matematik ile ilgili duyuşsal özellikleri, ikinci boyuttaki 9 madde ev ortamını ve üçüncü boyuttaki 5 madde de okul ortamını kapsamaktadır.

Verilerin Analizi

Birinci aşamada analizler için gerekli olan sayıtların kontrolü yapıldıktan sonra öğrencilerin matematik başarılarını etkileyen faktörlere ilişkin model oluşturulmuş, ikinci aşamada ise bölgelere göre modelin ölçme değişmezliği incelenmiştir.

Alan yazında, ölçme değişmezliğine bakmak için kullanılan yaklaşımlardan Madde Tepki Kuramı'na dayalı Madde ve Test İşlev Farklılıklarını inceleyen DIF ve DTF gibi modeller ile Yapısal Eşitlik Modellemesi (YEM)'nin sıklıkla göze çarptığı görülür (Somer ve ark., 2009). Yapısal eşitlik modeli ilişkisel bir araştırma yöntemidir; bu yüzden eksik veya uç veriler ya da verilerin normal dağılmaması gibi durumlar değişkenler arasındaki varyans-kovaryansı etkilediğinden YEM analizini de etkilemektedir (Schumacker ve Lomax, 2004). Verilerin analiz için hazırlanması ve bu açılarından taranması iki nedenle önem taşımaktadır. İlki, YEM için yaygın olarak kullanılan tahmin yöntemlerinin, verilerin dağılım özellikleri ile ilgili belirli bazı varsayımlara ihtiyaç duymalarıdır. Diğeri ise, YEM için kullanılan bilgisayar programlarının verilerle ilgili problemler olduğunda mantıksal çözümler üretememeleri veya çökmeleridir (Kline, 2015). Bu yüzden, araştırmanın modeli kurulmadan önce eksik veri analizi, uç değer analizi, normallik, çoklu bağlantı sayıtları kontrol edilmiştir.

Tablo 3. Maddelere Ait Faktör Yükleri

| Maddeler | Boyutlar | | |
|----------|----------|--------|-------|
| | 1 | 2 | 3 |
| BSBM14E | 0.814 | | |
| BSBM16A | 0.799 | | |
| BSBM16D | 0.772 | | |
| BSBM14A | 0.770 | | |
| BSBM16H | 0.739 | | |
| BSBM16F | 0.732 | | |
| BSBM14C | 0.698 | | |
| BSBM16C | 0.696 | | |
| BSBM16N | 0.635 | | |
| BSBM14B | 0.584 | | |
| BSBM16G | 0.573 | | |
| BSBM14D | 0.552 | | |
| BSBG05A | | 0.734 | |
| BSBG05E | | 0.703 | |
| BSBG05F | | 0.673 | |
| BSBG05B | | 0.613 | |
| BSBG05H | | 0.609 | |
| BSBG05D | | 0.588 | |
| BSBG04 | | -0.502 | |
| BSBG05J | | 0.500 | |
| BSBG03 | | 0.439 | |
| BSBG13C | | | 0.696 |
| BSBG13E | | | 0.689 |
| BSBG13A | | | 0.608 |
| BSBG13F | | | 0.558 |
| BSBG13D | | | 0.541 |

Yapısal eşitlik modellemesi; ölçüm modelini (yani, göstergeler ve gizil değişkenler arasındaki ilişkileri) test etmede kullanılan ve gizil değişkenler arasındaki ilişkilerin yapısal modelini inceleyen, genel ve kapsamlı bir analiz topluluğudur (Harrington, 2009). Araştırmacıdan araştırmacıya nispeten değişiklik gösterse de, YEM uygulaması beş basamaktan oluşmaktadır (Bollen ve Long, 1993; In'nami, ve Koizumi, 2013; Schumacker ve Lomax, 2004; Tabachnick ve Fidell, 2007).

- 1) Model betimleme
- 2) Model tanımlama
- 3) Parametre hesaplaması
- 4) Uyumu test etme
- 5) Yeniden betimleme

AFA sonucunda belirlenen faktörler ile model 3 boyutlu olarak kurulmuştur. 12 madde matematik ile ilgili duyuşsal özellikleri, 8 madde ev ortamını ve 4 madde de okul ortamını kapsamaktadır. Tablo 4'te bu üç gizil değişken ile onlarla ilişkili olduğu düşünölen gözlenen değişkenlere yer verilmiştir.

Tablo 4. Gizil ve Gözlenen Değişkenler

| Gizil Değişkenler | Gözlenen Değişkenler |
|---------------------|--|
| DUYUŞSAL ÖZELLİKLER | BSBM14A, BSBM14B, BSBM14C, BSBM14D, BSBM14E, BSBM16A, BSBM16C, BSBM16D, BSBM16F, BSBM16G, BSBM16H, BSBM16N |
| EV ORTAMI | BSBG04, BSBG05A, BSBG05B, BSBG05D, BSBG05E, BSBG05F, BSBG05H, BSBG05J |
| OKUL ORTAMI | BSBG13A, BSBG13C, BSBG13D, BSBG13E |

Bu bağlamda, modele alınan her bir gizil değişken en az dört gözlenen değişkenle ölçülecek şekilde model tanımlaması yapılmıştır. Boyutlarda en fazla faktör yüküne sahip olan değişkenler referans değişken olarak karar verilmiş ve faktörlerdeki diğer değişkenlerin serbestçe değişmesi sağlanmıştır. Referans değişkenler; duyuşsal özellikler boyutu için ise BSBM14E, ev ortamı boyutu için BSBG05A ve okul ortamı boyutu için BSBG13C olarak seçilmiştir.

Verilerin sayıltıları sağlamlasından dolayı, araştırmada kestirim yöntemi olarak En Çok Olabilirlik Yöntemi (Maximum Likelihood Estimation Method) kullanılmıştır. Sıralanan bu uyum indekslerine ilişkin iyi ve kabul edilebilir uyum ölçütleri, Tablo 5'te sunulmuştur.

Tablo 5. Model Uyum Ölçütleri

| Uyum İndeksi | İyi Uyum | Kabul Edilebilir Uyum |
|----------------|--|------------------------------|
| χ^2 | manidar olmayan bir p değeri ile serbestlik derecesi ile ilgili düşük χ^2 | |
| $\chi^2/ s.d.$ | $0 \leq \chi^2/ s.d. \leq 2$ | $2 \leq \chi^2/ s.d. \leq 5$ |
| RMSEA | $0 \leq RMSEA \leq 0.05$ | $0.05 \leq RMSEA \leq 1.00$ |
| NFI | $0.95 \leq NFI \leq 1.00$ | $0.90 \leq NFI \leq 0.95$ |
| NNFI | $0.95 \leq NNFI \leq 1.00$ | $0.90 \leq NNFI \leq 0.95$ |
| CFI | $0.95 \leq CFI \leq 1.00$ | $0.90 \leq CFI \leq 0.95$ |
| GFI | $0.95 \leq GFI \leq 1.00$ | $0.90 \leq GFI \leq 0.95$ |
| AGFI | $0.90 \leq AGFI \leq 1.00$ | $0.85 \leq AGFI \leq 0.90$ |
| SRMR | $0 \leq SRMR \leq 0.05$ | $0.05 \leq SRMR \leq 0.10$ |

Kaynak: Hooper, D., Coughlan, J. & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.; Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th Edition). USA: Pearson Education, Inc.; Kline, R. B. (2005). *Principles and Practice of Structural Equation Modelling*. (2nd Edition). New York: Guilford Publications, Inc.

Araştırmanın ikinci aşamasında ise, YEM çalışmasında kurulan modelin coğrafi bölgelere göre ölçme değişmezliğini sağlayıp sağlamadığı, çoklu grup doğrulayıcı faktör analizi (MG-CFA) ile kontrol edilmiştir. Drasgow'a (1984) göre bir grup bireyi özelliklerinin derecesi bakımından karşılaştırmak için ya da özellik seviye puanlarının gruplar arası farklılık ilişkileri gösterip göstermediğini araştırmak için, incelenmekte olan sayısal değerlerin aynı ölçüm göstergesinde olduğu varsayılmalıdır. Yani, testin gruplar arası "ölçme değişmezliği"ne sahip olması gerekmektedir. Bir ölçeğin faktör yükleri, madde kesişimi ve artık varyansları gibi birçok parametrenin aynı olması, ölçme modelinin gruplar arasında aynı olmasını sağlamaktadır. Yapılan karşılaştırmalarda tüm parametreler aynı olduğu takdirde modelin gruplar için ölçme değişmezliğini sağladığı söylenebilir (Başusta, 2010; Gregorich, 2006).

Bu yüzden, ölçme değişmezliğinin iç içe geçmiş dört hiyerarşik aşaması bulunmaktadır. Araştırma modelinin coğrafi bölgelere göre ölçme değişmezliği, aşağıda yer alan bu dört aşamada test edilmiştir (Meredith, 1993; Somer ve ark., 2009; Wu, Li ve Zumbo, 2007).

- 1) Yapısal (biçimsel) değişmezlik
- 2) Metrik (zayıf / faktöryel) değişmezlik
- 3) Ölçek (güçlü / skalar) değişmezlik
- 4) Katı (tam) değişmezlik

Değişmezlik aşamalarında en çok katılımcıya sahip olan Marmara bölgesi referans grup olarak seçilmiştir.

BULGULAR

"Matematik ile ilgili duyuşsal özellikler, ev ortamı ve okul ortamı değişkenlerinden oluşturulan model, sekizinci sınıf matematik başarısını yordamakta mıdır?" sorusunu cevaplayabilmek için gözlenen değişkenlerle gizil yapılar arasındaki ilişkiyi gösteren ölçme modelinin testinde, modeldeki tüm maddeler için manidar t-değerleri elde edilmiştir. Ardından hata varyansları incelendiğinde, standartlaştırılmış yol diyagramına ait hiçbir hata varyansının 0.90'nın üstünde olmadığı saptanmıştır.

Tablo 6. Ölçme Modelinin Uyum İndeksleri

| | χ^2 | s.d. | RMSEA | SRMR | NFI | NNFI | CFI | GFI | AGFI |
|--------------|----------|------|-------|-------|------|------|------|------|------|
| Ölçme Modeli | 12254.41 | 371 | 0,074 | 0.058 | 0,95 | 0,95 | 0,96 | 0,87 | 0,85 |

Model uyumunun değerlendirilmesinde kullanılan indekslere bakıldığında ise, Tablo 6'daki kimi değerlerin (RMSEA, SRMR,AGFI) kabul edilebilir ölçüde kiminin de (NFI, NNFI, CFI) iyi uyum aralığında yer aldığı görülmektedir (Bkz. Tablo 5). Yalnız GFI değeri 0.87 olmasına rağmen, referanslarda alt değer olarak önerilen 0.90 değerinin çok az altında seyri ve diğer indekslerin iyi uyuma sahip olması nedeniyle bu değer kabul edilmiştir. Bu sonuçlar doğrultusunda, modele alınan değişkenlerin iyi uyum gösterdiği söylenebilir.

Duyuşsal boyutun en çok 0.81 ile BSBM14E (matematiği seviyorum) gözlenen değişkeni tarafından açıklandığı görülmektedir. Ev ortamı gizil değişkeninde en yüksek faktör yüküne sahip değişkenin ise BSBG05A (evdeki bilgisayar) olduğu saptanmıştır. Aynı şekilde, okul ortamı ise en çok BSBG13E (okulda diğer öğrenciler tarafından saldırıya uğrama) tarafından açıklanmaktadır.

Bir sonraki aşamada, gizil değişkenler arasındaki ilişkileri özetleyen yapısal modelin testine geçilmiştir. Tablo 7'de, matematik başarısını yordayan modelin uyum değerlendirilmesinde kullanılan indekslere yer verilmiştir. Modele ilişkin uyum ölçütlerine göre, model kapsamında elde edilen verinin kovaryans yapısının ana kovaryans yapısı ile uyumlu olduğu görülmektedir.

Tablo 7. Yapısal Modelin Uyum İndeksleri

| | χ^2 | s.d. | RMSEA | SRMR | NFI | NNFI | CFI | GFI | AGFI |
|---------------|----------|------|-------|-------|------|------|------|------|------|
| Yapısal Model | 12254.41 | 371 | 0.074 | 0.058 | 0.95 | 0.95 | 0.96 | 0.87 | 0.85 |

Yapısal modelinin testinde de yine t-değerlerinin manidarlığı incelenmiştir. Modeldeki okul değişkeni hariç tüm değişkenler için manidar t-değerleri elde edilmiştir. Bu durum okul gizil değişkeninin matematik başarısını yordamadığını göstermektedir.

Tablo 8. Duyuşsal, Ev ve Okul Değişkenleri için Yapısal İlişkiler

| Yapısal ilişkiler | Standartlaştırılmış Yükler | t-değeri |
|-------------------|----------------------------|----------|
| Duyuşsal → Başarı | 0.43 | 35.90* |
| Ev → Başarı | 0.39 | 31.18* |
| Okul → Başarı | -0.03 | -1.92 |

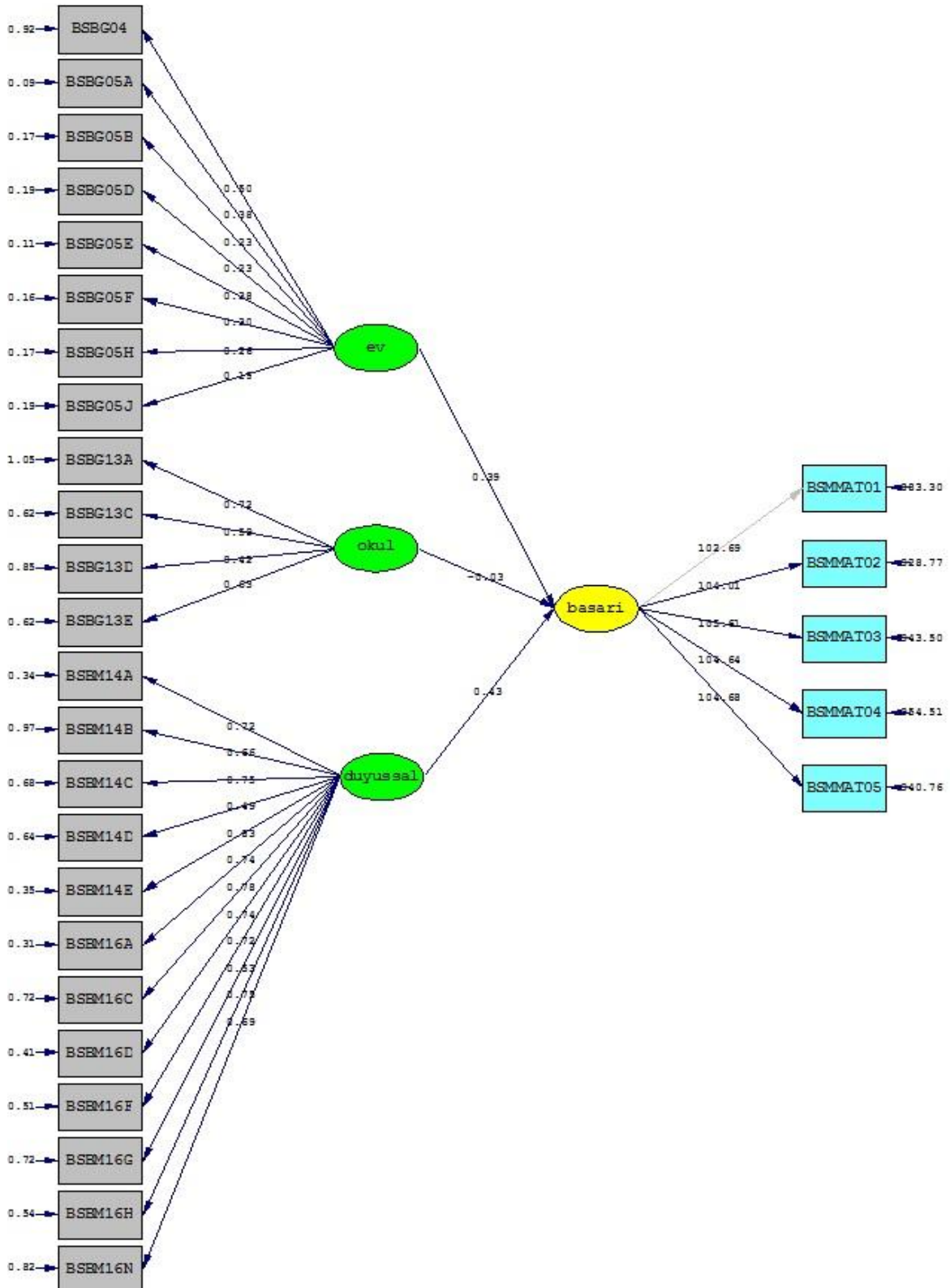
Tablo 8’de görüldüğü gibi, duyuşsal özelliklerle başarı arasında 0.43’lük pozitif yönde istatistiksel olarak manidar bir ilişki bulunmuştur. Ev ortamı değişkeni ile matematik başarısı arasında 0.39’luk ve pozitif yönde istatistiksel olarak anlamlı bir ilişki bulunurken, okul ortamı ve matematik başarısı arasında ise -0.026’lık negatif yönde istatistiksel olarak zayıf ve manidar bir ilişki mevcuttur.

Matematik başarısına ait regresyon denklemi aşağıdaki gibidir:

$$\text{BAŞARI} = 0.39 \cdot \text{ev} - 0.026 \cdot \text{okul} + 0.43 \cdot \text{duyuşsal}; \text{Hata varyansı} = 0.64; R^2 = 0.36$$

Regresyon denklemi incelendiğinde, belirtme katsayısı olan R^2 ’nin 0.36 olduğu saptanmıştır. Bu değer ile TIMSS 2011 Türkiye sekizinci sınıf öğrenci anketinden bu araştırma için belirlenen; matematik duyuşsal özellikleri, ev ve okul ortamı değişkenlerinin, öğrenci matematik başarılarının %36’sını açıkladığı belirlenmiştir. Matematik başarısını belirlemede en yüksek pay pozitif yönde duyuşsal özelliklere aitken, onu yine pozitif yönde ev ortamı takip etmektedir. Okul ortamı ile matematik başarısı arasında ise negatif yönde istatistiksel olarak anlamlı olmayan bir ilişki mevcuttur.

Yapısal modelin analizinden elde edilen yol diyagramı ise, Şekil 2’de (standartlaştırılmamış katsayılar) görülmektedir.



Şekil 2. Yapısal Modele Ait Yol Diyagramı (standartlaştırılmamış katsayılar)

Tablo 9. Modelin Bölgelere Göre Uyum İndeksleri

| | χ^2 | s.d. | RMSEA | SRMR | NFI | NNFI | CFI | GFI | AGFI |
|-------------------|----------|------|-------|-------|------|------|------|------|------|
| Ülke Genel | 12254.41 | 371 | 0.074 | 0.058 | 0.95 | 0.95 | 0.96 | 0.87 | 0.85 |
| Ege | 1787.86 | 371 | 0.075 | 0.066 | 0.94 | 0.95 | 0.96 | 0.85 | 0.82 |
| Karadeniz | 1674.23 | 371 | 0.073 | 0.067 | 0.95 | 0.96 | 0.96 | 0.85 | 0.83 |
| İç Anadolu | 1800.25 | 371 | 0.073 | 0.060 | 0.95 | 0.95 | 0.96 | 0.85 | 0.83 |
| Doğu Anadolu | 1577.47 | 371 | 0.071 | 0.063 | 0.93 | 0.94 | 0.95 | 0.86 | 0.83 |
| Marmara | 4028.19 | 371 | 0.076 | 0.061 | 0.95 | 0.95 | 0.95 | 0.86 | 0.84 |
| Akdeniz | 1786.73 | 371 | 0.073 | 0.061 | 0.94 | 0.95 | 0.95 | 0.85 | 0.83 |
| Güneydoğu Anadolu | 1772.08 | 371 | 0.073 | 0.060 | 0.93 | 0.94 | 0.94 | 0.85 | 0.83 |

Matematik başarısı modelinin 7 coğrafi bölgeye göre uyum indeksleri Tablo 9’da özetlenmiştir. Bu bağlamda, tüm bölgelerdeki GFI ve AGFI değerlerinin kabul edilebilir sınırların çok az altında seyrettiği görülmüştür. Ancak, AGFI’nın son yıllarda alanyazında daha az görülmesi (Kline, 2015) ile GFI’nın örneklem genişliğine duyarlı olduğundan büyük verilerde daha küçük değerler vermesinin normal olması (Özer ve Anıl, 2011) ve diğer uyum indekslerinin iyi uyuma işaret etmelerinden dolayı, modelin tüm bölgelerde iyi uyum gösterdiği söylenebilir.

“Sekizinci sınıf matematik başarısı modeli bölgelere göre ölçme değişmezliğini sağlamakta mıdır?” sorusuna yanıt aramak için çoklu grup doğrulayıcı faktör analizi ile yapılan ölçme değişmezliği testinin aşamalarına ait uyum indeksleri incelenmiş ve elde edilen bulgular Tablo 10’da verilmiştir.

Tablo 10. Ölçme Değişmezliği Aşamalarının Uyum İndeksleri

| | χ^2 | RMSEA | CFI | Δ CFI | NFI | NNFI | Δ NNFI |
|---------------------|-------------------------|-------|------|--------------|------|------|---------------|
| Yapısal değişmezlik | 12517.01 (s.d. 1743) | 0.086 | 0.92 | - | 0.91 | 0.91 | - |
| Metrik değişmezlik | 13087.03 (s.d. 1887) | 0.084 | 0.92 | 0 | 0.90 | 0.92 | -0.01 |
| Ölçek değişmezlik | 20554.22 (s.d. 2052) | 0.10 | 0.85 | 0.07 | 0.83 | 0.86 | 0.06 |
| Katı değişmezlik | 16444.48 (s.d. 2214) | 0.088 | 0.60 | 0.25 | 0.57 | 0.65 | 0.21 |

Bu bağlamda, ilk aşama olan yapısal değişmezlik aşamasının uyum indeksleri model uyumunun değerlendirilmesinde kabul edilebilir aralıktadır (RMSEA < 1.00 ve CFI, NFI, NNFI > 0.90). Yani, matematik başarısı modelinin yapısının her coğrafik grupta da aynı olduğu söylenebilir. Böylece, yapısal değişmezlik sağlandığından bir sonraki aşama olan metrik değişmezliğe bakılabilir.

Alt gruplar arasında faktör yüklerine eşdeğer olma kısıtlaması getirilen metrik değişmezlikte, yine uyum indekslerine göre modelin veri ile kabul edilebilir uyum gösterdiği saptanmıştır.

Bu farklar incelendiğinde kabul edilir aralıkta oldukları ($-0.01 \leq \Delta$ CFI ≤ 0.01 , $-0.01 \leq \Delta$ NNFI ≤ 0.01) görülmektedir. Yani, modele alınmış değişkenlerin faktör yükleri alt gruplarda değişmemektedir. Böylece, farklı coğrafi bölgelerde ölçülen özellikler ile ölçeğin boyutları arasındaki ilişkilerin benzer olduğu yorumu yapılabilir. Bu durumda, bir sonraki aşama olan ölçek değişmezlik testine geçilebilir.

Tablo 10 incelendiğinde, ölçek değişmezlik aşamasında uyum indekslerinin kabul edilebilir sınırların biraz altında seyrettiği görülebilir. Ek olarak, CFI ve NNFI değerleri arasındaki farklar incelendiğinde ise, modelin ölçek değişmezliği sağlamadığı belirlenmiştir. Bu bağlamda, maddelere ilişkin faktör yükleri coğrafi bölgeler arasında değişmez iken, faktörler arası korelasyonların değişmezlik sergileyemediklerini söylemek yanlış olmaz. Bu durum, gizil değişken ortalamalarının

bölgeler arasında farklılaştığı ve coğrafi bölgelere ilişkin ortalamaların karşılaştırılmasının anlamlı olmayacağı şeklinde düşünülebilir.

Ölçme değişmezliğinde bir değişmezlik koşulunun sağlanamadığı durumda diğer aşamaya geçilememektedir. Yani, ölçek değişmezliği sağlanamadığından, katı değişmezliğin incelenmesi doğru olmayacaktır. Tablo 10'dan elde edilecek veriler, katı değişmezlik değerlerinin zaten bu aşamayı da sağlayamadığının bir göstergesi olarak düşünülebilir.

“TIMSS 2011 sonuçlarına göre, Türkiye'deki sekizinci sınıf öğrencilerin matematik başarısını yordama modeline alınan hangi değişkenler bölgeler arasında farklılık göstermektedir?” sorusuna cevap bulabilmek için değişkenlerin bölgelere ait regresyon katsayıları incelenmiş ve Tablo 11'de verilmiştir.

Tablo 11. Coğrafi Bölgelere Göre Faktörlere Ait Regresyon Katsayıları

| FAKTÖR | BÖLGELER | | | | | | | ÜLKE GENELİ |
|----------------|----------|-----------|------------|--------------|---------|---------|-------------------|-------------|
| | EGE | KARADENİZ | İÇ ANADOLU | DOĞU ANADOLU | MARMARA | AKDENİZ | GÜNEYDOĞU ANADOLU | |
| DUYUŞSAL | 0.44 | 0.46 | 0.46 | 0.43 | 0.44 | 0.39 | 0.43 | 0.43 |
| EV | 0.36 | 0.40 | 0.41 | 0.34 | 0.37 | 0.38 | 0.31 | 0.39 |
| OKUL | -0.029 | -0.039 | 0.082* | -0.002 | -0.030 | 0.006 | -0.110* | -0.03 |
| R ² | 0.35 | 0.44 | 0.43 | 0.30 | 0.36 | 0.31 | 0.32 | 0.36 |

Tablodaki çoğu bölge, okul gizil değişkenin ülke genelindeki gibi matematik başarısını yordamadığını göstermektedir. Ancak, İç Anadolu ve Güneydoğu Anadolu bölgesi, okul değişkeni bakımından manidar t-değerine sahip olup matematik başarısını yordamaktadır.

Coğrafi bölgelere ait regresyon katsayıları incelendiğinde, hem okul ortamı değişkeninin ülke genelinde modeli yordamasından hem de duyuşsal özelliklerdeki farklılıklar daha öğrenci temelli olduğu için bu durumun katsayılar da çok farklılık göstermemesinden dolayı, matematik başarısını yordama modeline alınan değişkenlerden hangilerinin bölgeler arasında farklılık gösterdiğinin tespitinde ev ortamına ait değişkenler incelenmiştir.

Ev ortamı gizil değişkenine ait bazı gözlenen değişkenlerin bölgelere göre frekans dağılımları incelendiğinde; evdeki kitap sayısı, evde bilgisayarın ve internet bağlantısının olup olmama durumu ile kendisine ait bir oda ve çalışma masasına sahip olma durumlarının, Güneydoğu ve Doğu Anadolu bölgelerinde azaldığı görülmektedir. Ülke geneline bakıldığında, bu oranlar diğer bölgelerde fazlasıyla yüksektir. Evdeki kitap sayısı, bilgisayar ve internet bağlantısına sahip olma en yüksek Marmara bölgesinde iken, çalışma masası ve kendisine ait bir oda bulunması ise Ege bölgesinde yüksek oranda bulunmuştur.

SONUÇLAR ve TARTIŞMA

Bu araştırmada, Türkiye'den TIMSS 2011 çalışmasına katılan sekizinci sınıf öğrencilerinin görüşlerine dayalı anket sonuçları ile üç boyutlu ve 24 maddeden oluşan ve öğrencilerin matematik başarısını yordayan bir model elde edilerek, ölçme değişmezliğine bakılmış ve bu değişkenlerin coğrafi bölgelere göre nasıl farklılıklar gösterdiği incelenmiştir. Yıldırım ve arkadaşlarının (2013) da bahsettiği gibi, TIMSS vb. uluslararası çalışmalar ülkeler arasındaki karşılaştırmaların yanı sıra, ülke içindeki çeşitli farklılıkların ortaya çıkarılmasına da hizmet etmektedir. Ülke verisinin coğrafi bölgeler gibi alt gruplara ayrılarak analiz edilmesi ve bu yüzden bazen yorumlanması zor sonuçların

ortaya çıkmasına sebep olsa da, bütün olarak incelenen ülke verisinde gözden kaçabilecek bulguları ortaya çıkarabilmektedir.

Bu çalışmanın sonuçları, alt problemlerin sırasına göre aşağıda maddeler halinde yer almaktadır.

1. Matematik başarısını yordamak için kurulan model iki aşamalı olarak test edilmiş ve elde edilen bulgular sonucunda modelin verilerle iyi uyum gösterdiği bulunmuştur. Modele alınan değişkenlerden, Türkiye'deki sekizinci sınıf öğrencilerinin matematik başarısına en büyük etkisi olan değişkenin matematikle ilgili duyuşsal özellikler olduğu ve ev ortamı değişkeni sonucunun da bu değişkeni yakın bir şekilde takip ettiği görülmüştür. Modele alınan üçüncü değişken olan okul ortamı ile matematik başarısı arasında ise, negatif yönde istatistiksel olarak anlamlı olmayan bir ilişki bulunmuştur.

Bu sonuçlar, matematik başarısını etkileyen değişkenler üzerine yapılan bazı çalışmalarla paralellik gösterirken, bazıları ile farklılaşabilmektedir. Değişkenlerden ilk olarak matematik ile ilgili duyuşsal özellikler incelendiğinde; yapılan birçok araştırma göstermektedir ki; öğrenme ile ilgili en kritik öneme sahip yapı matematiğe karşı tutumdur. Birçok ülkede matematiğe karşı olumlu tutum gösteren öğrencilerin başarılarının da yüksek olduğu saptanmıştır (Else-Quest ve ark., 2010; House, 2006; Singh ve ark., 2002; Winheller ve ark., 2013). Ancak, matematiğe karşı tutum çok boyutlu bir yapıdır ve Vandecandelaere ve arkadaşları (2012) tarafından açıkça üç farklı bölüme ayrılmıştır. Son otuz yılda yapılan araştırmalar, matematiğe karşı tutumun tüm bu yönlerinin matematik başarısında dikkat çeken belirgin yordayıcılar olarak ortaya çıktıklarını göstermektedir. (Chiu ve Klassen, 2010; Marsh ve Hau, 2004; Singh ve ark., 2002).

Bu durum MEB'in TIMSS 2011 raporunda öğrenci anketlerinin değerlendirildiği kısımda; matematik öğrenmeyi seven, matematiğe değer veren ve matematikte kendine güvenen öğrenciler alt başlıkları ile de desteklenmektedir (Büyüköztürk, Çakan, Tan ve Atar, 2014). Bu çalışmada, TIMSS 2011 öğrenci anketlerinde matematik hakkında yer alan tüm alt başlıklar tek bir boyutta toplandığı için, matematik ile ilgili duyuşsal özellikler değişkeni olarak adlandırılmıştır.

Ev ortamı değişkeni ise MEB'in (Büyüköztürk ve ark., 2014) TIMSS 2011 raporunda da bahsedildiği gibi, hem Türkiye genelinde hem de TIMSS genelinde evdeki kaynaklar ve matematik başarısı arasında önemli bir ilişki gözlenmesi ile paralellik göstermektedir. Yapılan TIMSS uygulamalarının sonuçlarına baktığımızda, hem dördüncü hem de sekizinci sınıflarda akademik başarı ile destekleyici ev ortamı arasında güçlü ve pozitif bir ilişkinin varlığı gözlenmektedir (Bayar ve Bayar, 2013). Brese ve Mirazchiyski (2010) araştırmalarını TIMSS 2007 ve PISA 2006 verilerinde yaparak, birçok ev ortamı ile ilişkili maddenin matematik başarısı ile güçlü ve ortalama ilişkide olduklarını göstermişlerdir.

Okul ortamı ise Akyüz'ün (2014) çalışmasında da bahsettiği gibi güvensiz olduğunda, öğrencilerin matematik başarısı üstünde negatif etkiye sahiptir. Bu değişkeni açıklayan dört gözlenen değişken de incelendiğinde, akran zorbalığı ile ilişkili anlam olarak olumsuz içerikte maddeler olduğu anlaşılmaktadır. Bu yüzden, elde edilen bulgunun matematik başarısı üstünde negatif etkiye sahip olması doğal bir sonuç olarak ortaya çıkmaktadır. Ayrıca, MEB TIMSS 2011 raporunda, Türkiye'de nerdeyse hiç zorbalık yaşanmayan okullarla, sıklıkla zorbalığa uğrayan okullarda öğrenim gören öğrencilerin başarı ortalamaları karşılaştırılarak; okullardaki zorbalığa uğrama derecesi azaldıkça, öğrencilerin matematik başarılarının artma eğiliminde olduğu gösterilmiştir. Bu rapor, negatif çıkan çalışmanın sonucunu destekler niteliktedir (Büyüköztürk ve ark., 2014). Ancak, okul ortamının modelde yalnızca dört değişken ile temsil edilme azlığının yanı sıra, Bayar ve Bayar (2013) okulun eğitimsel kalitesi için dört temel unsurdan bahseder; bunlar öğretmen, öğrenci, veli ve yöneticilerdir. Bu bağlamda, öğrenci görüşüne dayalı anketlere ek olarak öğretmen ve yönetici görüşleri de alındığında, bu araştırmada gözlenen matematik başarısının okul ortamında istatistiksel olarak manidar bulunmayan etkisinin farklılaşabileceği düşünülebilir.

2. Matematik başarısının yordandığı yapısal model, 7 coğrafi bölgede ayrı ayrı test edilmiş ve modelin bütün bölgelerde geçerli olduğu saptanmıştır. Daha sonra modeldeki değişkenlerin bölgeler arasında karşılaştırılmasının anlamlı olabilmesi için, modelin coğrafi bölgelere göre çoklu grup doğrulayıcı faktör analizi ile ölçme değişmezliğine bakılmıştır. Matematik

başarısına ilişkin oluşturulan model için metrik değişmezlik seviyesinde bir eşdeğerlik elde edilmiştir. Böylece, yalnızca bu aşamada yapılan karşılaştırmalarda gruplar arası farklılıkların ölçme aracından kaynaklanmadığı söylenebilir. Yani, bu aşamada farklı coğrafi bölgelerdeki tüm öğrencilerin maddelere benzer şekilde yanıt verdiğini ve bölgelerden elde edilen puan karşılaştırmalarının anlamlı olduğunu söylemek mümkündür. Ancak, model ölçek değişmezliğini sağlamadığından, gizil değişkenlerin ortalamaları bakımından bir karşılaştırma yapılması mümkün olamamaktadır. Bu bağlamda, maddelere ilişkin faktör yükleri coğrafi bölgeler arasında değişmez iken, faktörler arası korelasyonlar eşdeğerlik gösterememektedir. Bu durum, gizil değişken ortalamalarının bölgeler arasında farklılaştığını ve coğrafi bölgelere ilişkin ortalamaları karşılaştırmanın anlamlı olamayacağını ifade eder. Böylece farklılıkların bölgelerden mi, yoksa ölçme aracından mı kaynaklandığı kestirilememektedir.

Yapılan değişmezlik testi sonuçları, modele alınan değişkenlerin matematik başarısına etkilerinin coğrafi bölgelere göre tam değişmezlik koşullarını yerine getirmediğini göstermektedir. Buna göre modele alınan değişkenlerin coğrafi bölgelere göre farklılıkları, bu bölgelerden kaynaklanabileceği gibi ölçme aracından da kaynaklanıyor olabilir. Bölgelerin farklılıklarına ilişkin yorum yapılırken bu durum unutulmamalıdır. Wu, Li ve Zumbo (2007), TIMSS 1999 değerlendirmesindeki matematik testlerini kullanarak, 7 farklı ülkeden 21 farklı çapraz karşılaştırma ile yaptıkları çalışmada; sadece yapısal ve metrik değişmezlik koşulları karşılanırken, aynı kültürlerde ise katı değişmezliğin sağlandığını belirtmektedirler. Bu bulgu, TIMSS 2011 Türkiye sekizinci sınıf örnekleminde aynı kültüre ait farklı coğrafi bölgelerden alınan öğrencilerden oluşturulan matematik başarısı modeli ile tutarlılık göstermemektedir.

3. Modele alınan matematikle ilgili duyuşsal özellikler, ev ve okul ortamı değişkenlerinin bölgelere ait regresyon katsayılarındaki farklılıklar incelendiğinde, okul gizil değişkeninin İç Anadolu ve Güneydoğu Anadolu bölgesi hariç, ülke genelindeki gibi matematik başarısını yordamadığı saptanmıştır. Duyuşsal özellikler değişkeni ise regresyon katsayılarında çok farklılık göstermediğinden, bölgelerden daha çok bireysel farklılıklarla ilişkili gözükmemektedir. Bu yüzden, matematik başarısını yordama modeline alınan değişkenlerden hangilerinin bölgeler arasında farklılık gösterdiğinin tespiti için, ev ortamına ait değişkenlerin incelemesine geçilmiştir. Ev ortamı değişkeninin bölgelere göre frekans dağılımı incelendiğinde, evdeki kitap sayısı, bilgisayar ve internet bağlantısına sahip olma en yüksek Marmara bölgesinde iken, en düşük Güneydoğu Anadolu bölgesinde ve takiben Doğu Anadolu bölgesinde gözlenmiştir. Çalışma masası ve kendisine ait bir odaya sahip olunması ise Ege bölgesinde yüksek oranda bulunurken, yine en düşük Güneydoğu Anadolu bölgesinde ve takiben Doğu Anadolu bölgesinde saptanmıştır.

Ersoy'un (2007) TIMSS 2007 çalışmasında bahsettiği gibi, öğrencilerin ev ortamı ve buradaki olanakları başarıyı etkileyen etmenlerdendir. Evlerindeki kitap sayısı arttıkça, öğrencilerin matematik başarı puan ortalamaları da artmaktadır. Çelebi ve arkadaşları (2014) da yine evde bulunan eğitimsel kaynakların başarıyı olumlu düzeyde arttırdığından bahsetmektedir. Sekizinci sınıflardan elde edilen verilere göre Türkiye, evdeki eğitimsel kaynakların azlığının oranı açısından dünyada Fas ve Gana'dan sonraki en kötü üçüncü ülkedir. Oral ve McGivney (2013) TIMSS 2011 çalışması ile ilgili yayınladıkları raporda bu duruma ek olarak, dördüncü sınıflarda gözlenen ve belli bir kitap sayısından (yaklaşık 200 kitaptan) sonra öğrencinin doyum noktasına ulaştığı durumun, sekizinci sınıf düzeyinde geçerli olmadığını vurgulamaktadırlar. Böylece, bu sonucun öğrencilerin ders başarısına, ileriki yaş ve sınıflarda ders dışı kaynakların da katkı sağlayabileceğini gösterdiğini savunmaktadırlar. Ayrıca, evde bilgisayar ve internet olması ile bunların kullanılmasının da daha yüksek öğrenci başarısı ile ilişkili olduğunu saptamışlardır.

Ersoy (2007) ise bu tip olanakların yalnızca teknolojiye sahip olma olmama gibi değerlendirilmeyerek, bu olanaklara sahip öğrencilerin sosyo-ekonomik ve kültür düzeylerinin de göreceli olarak daha iyi olduğundan bahsetmektedir. Ev ortamı ile ilişkili değişkenlerde düşük frekansa sahip olan Güneydoğu ve Doğu Anadolu bölgelerindeki bu durum, Ersoy'un (2007) görüşü

ile örtüşmektedir. Sonuç olarak, ev ortamı değişkeninin bölgelere göre frekans dağılımı incelendiğinde, ülke genelinde yüksek oranda sahip olunan bu olanakların Güneydoğu ve Doğu Anadolu bölgelerinde azaldığı ve matematik başarısına etkisinin bu bölgelerde diğerlerine kıyasla daha fazla olabileceği düşünülebilir. Alanyazındaki diğer çalışmalarla paralel olarak bu çalışmada, Türkiye'deki öğrencilerin ev ortamı gibi yaşam koşullarının halen başarılarını etkilediğini göstermektedir.

Bu çalışmanın önerileri; araştırma bulgularından çıkan ve daha sonra yapılacak araştırmalar için olmak üzere aşağıdaki gibi özetlenebilir.

Araştırma Bulgularından Çıkan Öneriler

1. Matematik başarısına en büyük etkinin bu alan ile ilgili duyuşsal özellikler olduğu düşünöldüğünde, ilgili devlet kurumlarının, okul yöneticilerinin ve öğretmenlerin öğrencilere matematięi sevdirmeye yönelik gerekli etkinlikleri planlayıp, uygulanması konusunda gerekli çalışmalar yapılmalıdır. Ayrıca, matematięe deęer verme ve matematięe yönelik kendilerine güven düzeylerini arttırmak için de yapılacak gerekli çalışmaların fayda sağlayacağı düşünölmektedir.
2. Ev ortamının matematik başarısındaki bölgesel farklılıkların nedenlerinden biri olabileceęi düşünöldüğünde, öğrencilere eğitim kaynaęı oluşturabilecek evde daha çok kitaba, bilgisayar ya da internet bağlantısına sahip olma gibi kaynakların artırımına yönelik tedbirler alınmalı, özellikle bunlar bölgeler arasındaki farklılıęı ortadan kaldıracak yönde olmalıdır. Ayrıca, ev ortamının ailelerin ekonomik durumlarıyla ilişkisi göz önüne alındığında, ekonomik açıdan dezavantajlı öğrencilerin bulunduğu okullarda ekonomik kalkınmayı sağlayacak gerekli tedbirler alınarak, öğrenciler bu bakımlardan eşitlenmelidir.

Daha Sonra Yapılacak Araştırmalar İçin Öneriler

1. Bu araştırmada öğrenci görüşlerine dayalı anketin bazı maddeleri kullanılmıştır. Başka çalışmalarda, modele alınmayan dięer deęişkenlerin de kullanılması önerilebilir.
2. Başka çalışmalarda öğrenci anketi dışında, yönetici ve öğretmen görüşlerine dayalı anketlerin bölgelere göre farklılıklarının da incelenmesi önerilebilir.
3. Matematik başarısını etkileyen deęişkenlerin ulusal düzeyde yapılacak bir sınavla veya çeşitli simölasyon çalışmalarıyla bölgelere göre farklılıklarının incelenmesi önerilebilir.
4. Bu araştırmada matematik başarısını yordayan modelin, bölgeler arası ölçme deęişmezlięi testinde ölçek deęişmezlięini neden sağlamadığının sebepleri ve uyumsuzluęa neden olan deęişkenlerin incelenmesi önerilebilir.
5. Model tam deęişmezlik koşulunu sağlamadığından, coęrafi bölgeler kısmi deęişmezlik yönünden incelenebilir ve modelde teoriye uygun modifikasyon yapılması önerilebilir.
6. Model, Türkiye örnekleminde bölgelere göre ölçme deęişmezlięini sağlamamıştır. Farklı coęrafi bölgelerin belirgin olduğu bazı ölkelerde de Türkiye ile karşılaştırmalı bir çalışma yapılması önerilebilir.
7. Öğrenci görüşüne dayalı anketlerden yararlanılarak ölçme deęişmezlięinin coęrafi bölgeler bakımından incelenecek benzer bir çalışmanın fen başarısını etkileyen deęişkenler yönünden de yapılması önerilebilir.

KAYNAKÇA

- Akyüz, G. (2014). TIMSS 2011'de öğrenci ve okul faktörlerinin matematik başarısına etkisi. *Eğitim ve Bilim, Geniş Ölçekli Test Uygulamaları Özel Sayısı*, 39 (172), 150-162.
- Başusta, N. B. (2010). Ölçme eşdeğerliği. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Kış 2010*, 1(2),58-64.
- Bayar, V. ve Bayar, S. A. (2013). *TIMSS 2011 matematik başarısı ulusal değerlendirme raporu*. Türk Eğitim Sendikası TIMSS 2011 Matematik Başarısı Ulusal Değerlendirme Raporu, Ankara.
- Brese, F. & Mirazchiyski, P. (2010). *Measuring students' family background in large-scale education studies*. Paper presented in the 4th IEA International Research Conference. July 1 – 3. Gothenburg, Sweden. Retrieved from <http://www.iea.nl/irc-2010.html>.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş. ve Atar, H. Y. (2014). *TIMSS 2011 ulusal matematik ve fen raporu. 8. sınıflar*. Ankara MEB: YEGİTEK Yayınları.
- Çelebi, N., Güner, H., Kaya G. T., ve Korumaz, M. (2014). Neoliberal eğitim politikaları ve eğitimde fırsat eşitliği bağlamında uluslararası sınavların (PISA, TIMSS ve PIRLS) analizi. *Tarih Kültür ve Sanat Araştırmaları Dergisi*, 3 (3), 33-75.
- Chiu, M. M. & Klassen, R. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20, 2-17.
- Drasgow, F. (1984) Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, 95(1), Jan 1984, 134-135.
- Dünya Bankası Raporu (2011). *Türkiye'de temel eğitimde kalite ve eşitliğin geliştirilmesi zorluklar ve seçenekler*. The World Bank 1818 H Street, NW Washington, DC 20433, USA.
- Ersoy, Y. (2007). TIMSS-2007: Uluslararası matematik ve fen araştırması-II: başarıyı etkileyen örtük değişkenler ve genel eğilimler. [Çevrim-içi:<http://www.f2e2-ogretmen.com/dagarcigimiz/f2e2-522.pdf>], Erişim tarihi: 27 Nisan 2015.
- Else-Quest, N. M., Hyde, J. S. & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: Meta-analysis. *Psychological Bulletin*, 136(1), 103–127.
- Feniger, Y., Livneh, I. & Yogev, A. (2012). Globalisation and the politics of international tests: the case of Israel. *Comparative Education*, 48 (3), 323-335.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care*, 44 (11/3), 78-94.
- Harrington, D. (2009). *Confirmatory factor analysis*. New York: Oxford University Press, Inc.
- House, J. D. (2006). Mathematical beliefs and achievement of elementary school students in Japan and United States: results from the Third International Mathematics and Science Study. *The Journal of Genetic Psychology*, 167(1), 31–45.
- Hooper, D., Coughlan, J. & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- In'nami, Y., & Koizumi, R. (2013). Structural equation modeling in educational research: A primer. In M. S. Khine (Ed.), *Applications of structural equation modeling in educational research and practice* (pp. 23–51). Rotterdam, the Netherlands: Sense Publishers.
- Kline, R. B. (2015). *Principles and practice of structural equation modelling*. New York: Guilford Publications, Inc.
- Kupari, P. & Nissinen, K. (2013). *Background factors behind mathematics achievement in finnish education context: explanatory models based on TIMSS 1999 and TIMSS 2011 Data*. Paper presented in the 5th IEA International Research Conference. 26-28 Haziran, Singapur. [Çevrimiçi:http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC2013_Kupari_Nissinen.pdf]
- Marsh, H. W. & Hau, K-T. (2004). Explaining paradoxical relations between academic selfconcept and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96, 56 – 67.
- Martin, M.O. & Mullis, I.V.S. (Eds.) (2012). *Methods and procedures in TIMSS and PIRLS 2011: characteristics of national samples*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M.O. & Mullis, I.V.S. (Eds.) (2012). *Methods and procedures in TIMSS and PIRLS 2011: stratified two-stage cluster sample design*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

- Mathforum (2006). *TIMSS 2007 (Third International Mathematics and Science Study)*. Drexel School of Education. [Çevrim-içi: <http://mathforum.org/social/timss/timss.brochure.html>], Erişim tarihi: 27 Mart 2015.
- Mullis, I. V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y. & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V.S., Martin, M.O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Oral, I. ve McGivney, E. (2013). *Türkiye'de matematik ve fen bilimleri alanlarında öğrenci performansı ve başarının belirleyicileri TIMSS 2011 analizi*. İstanbul: Eğitim Reformu Girişimi Raporu.
- Özer, Y. ve Anıl, D. (2011). Öğrencilerin fen ve matematik başarılarını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 41, 313-324.
- Öztürk, L. (2010). *TIMSS 2007 ve eğitim sistemimizin başarısı: öğretmen ve yönetici görüşleri*. Yüksek lisans tezi, Marmara Üniversitesi: İstanbul.
- Schumacher, R. E. & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling (2nd Edition)*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Singh, K., Granville, M. & Dika, S. (2002). Mathematics and science achievement: effects of motivation, interest and academic engagement. *The Journal of Educational Research*, 95, 323 – 332.
- Somer, O., Korkmaz, M., Dural, S. ve Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde cevap kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi*, Aralık 2009, 24(64), 61-75.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics (5th Edition)*. USA: Pearson Education, Inc.
- TIMSS (2011) *TIMSS 2011 international database*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands. [Çevrim-içi: <http://timss.bc.edu/timss2011/international-database.html>] Erişim tarihi: 19 Ekim 2014.
- Vandecandelaere, M., Speybroeck, S., Vanlaar, G., De Fraine, B. & Van Damme, J. (2012). Learning environment and students' mathematics attitude. *Studies in Educational Evaluation*, 38, 107–120.
- Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments; applications in the substance use domain. *The science of prevention: methodological advances from alcohol and substance abuse research*, 281-324.
- Winheller, S., Hattie, J. A. & Brown, G. T. L. (2013). Factors influencing early adolescents' mathematics achievement: high-quality teaching rather than relationships. *Learning Environments Research*, 16, 49–69.
- Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: a demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12 (3), 1-26.
- Yıldırım, H. H., Yıldırım, S., Ceylan, E. ve Yetişir, M. İ. (2013). *Türkiye perspektifinden TIMSS 2011 sonuçları*. Türk Eğitim Derneği Tedmem Analiz Dizisi I: Ankara

EXTENDED ABSTRACT

Introduction

One of the most important issues that worry Turkish citizens is education. In such an important area, the problems are inevitable. From past to present, the relations of the education and the concept of equality of opportunity is one of the most common complex issues depends on many factors in the studies of this field. In this study, using the TIMSS 2011 data, some of the variables affecting the mathematics achievement of eighth-grade students in Turkey were modeled and whether the variables showed differences among the regions according to the measurement invariance examination.

Method

In the first step, items in the student survey of TIMSS 2011 Turkey sample affecting the mathematics achievement were examined by exploratory factor analysis. After controlling assumptions, remaining 24 items were loaded into three factors including; mathematics related

affective domain, home and school environment. Then, the model was established by structural equation modeling (SEM) to test the relationship between the variables in the model and students' mathematics achievement.

In the second step of the study, the predicted structural model of mathematics achievement was tested for each of the seven geographic regions and found that the model is applicable in all regions. Then, the invariance of mathematics achievement model across regions was examined by means of multi groups confirmatory factor analysis to compare the variables in the model between geographical regions.

When the differences among the regions were examined according to regression coefficients, except for Central and Southeast Anatolia school latent variables is not one of the predictors of mathematics achievement. Regression coefficient of the mathematics related affective domain variables did not show much difference. As a result, frequency distribution of the home environment was analyzed in accordance with regions by considering that the biggest differences between the regions may be due to items of this variable.

Results and Discussion

According to the result of the analysis performed on the model, the measuring and the structural model fit indices that predictors of mathematics achievement were determined to test goodness of fit. Moreover, when the regression equation of the model was examined, the coefficient of determination was found 0.36. It was seen that mathematics related affective behaviors variable shows positively the highest effect in defining mathematics achievement. The second variable affecting achievement also positively was the home environment variable. Lastly, a negative non-statistically significant relationship was determined between the school environment and mathematics achievement. While these results were consistent with some studies on the variables affecting the math success, they can also be varied with some of them. Many studies mention that the most critical concept about learning is the attitude against mathematics. Attitude against mathematics is multidimensional structure Vandecandelaere and friends (2012) grouped it into three section. In many countries, the success of students with positive attitudes towards mathematics were found to be high (Else-Quest et al., 2010; House, 2006; Shen and full, 2008; Singh et al., 2002; Winhel et al., 2013). Kupari and Nissinen (2013) also emphasized that students' home environment is associated with mathematics achievement in school (Chiu & Klassen, 2010; Marsh & Hau, 2004; Singh et al., 2002) As well as school environment was represented by only four variables in the studied model, Bayar and Bayar (2013) mentioned about four basic elements for the educational quality of the school which are teachers, students, parents and administrators. If teachers' and administrators' surveys could be taken in addition to the opinions of the students, the relation between school environment and mathematics achievement which was found non-significant in this study could be changed.

Based on the result of the analysis compared the variables in the model between geographical regions, only the configural and weak invariance conditions in the sub-groups are met by the model. Therefore, it was concluded that intergroup differences do not result from the assessment and evaluation instrument, but among the regions. Moreover, because strong invariance (and thus strict invariance) condition cannot be met, measurement invariance cannot be exactly determined. Thus, whether the differences are result of the regions' differences or of the assessment and evaluation instrument could not be specified. By using TIMSS 1999 mathematics assessment tests, Wu, Li and Zumbo (2007) found that only structural and metric invariance conditions were met in their study conducted with 21 different cross comparisons from 7 different countries, while at the same cultures indicated that solid constancy is achieved. These findings did not show any consistency with the model of the mathematics achievement of TIMSS 2011 Turkey students established from different geographical regions of the same culture.

According to the result of last analysis about the frequency distribution of the home environment variable, it was seen that the number of books at home, computer and internet connection were the

highest in Marmara, the lowest has been identified in the Southeastern Anatolia and following in the Eastern Anatolia region. Having study desk for their use and own room at a high rate in the Aegean region, however, were the lowest again in Southeast Anatolia and Eastern Anatolia regions followed. Consistent with other studies in the literature (Çelebi ve ark., 2014; Ersoy, 2007) Oral ve McGivney, 2013), this study showed that students' living conditions such as home environment in Turkey still affect their success.

In terms of these results, concluding remark of this study is that when we make comparison among the regions, we should be cautious about interpretations of math achievement scores because as stated earlier, the differences can be result of the regions' differences as well as the assessment and evaluation instrument.

Web-tabanlı Öz-Değerlendirme Sisteminde Öğrenenlerin Öz-Müdahale Algısı ve Test Alma Davranışlarının Başarı Üzerine Etkisi*

The Effect of Self-Intervention Perception and Test Taking Behaviour on Success in Web-Based Self-assessment System

Fatma BAYRAK **

Halil YURDUGÜL ***

Öz

Günümüzde eğitimin her kademesinde öğrenci merkezli değerlendirme türlerinden olan öz-değerlendirme yaklaşımı yaygınlaşmaktadır. Biçimlendirmeye dönük değerlendirme (*formative*) kapsamında ele alınan bu yaklaşımda dönüt ön plana çıkmaktadır. Öğrenenler bu değerlendirmeler aracılığıyla ölçütlere (*criterion-referenced*) ve gruba (*norm-referenced*) göre var olan performans durumları hakkında; kendilerini yeniden değerlendirecek de gelişim süreçleri hakkında (*self-referenced*) dönütler alabilirler. Öğrenenler aldıkları bu dönütlere dayalı olarak öğrenmelerini iyileştirmek için öğrenme süreçlerine kendileri müdahalede bulunabilirler. Bu süreçte dönütleri ele almaları, müdahaleye karar vermeleri ve müdahaleyi uyguladıktan sonra etkililiğini belirlemeleri öz-müdahale algısı olarak isimlendirilmiştir. Çalışma kapsamında, öğrenenlerin öz-değerlendirmelerini desteklemek için geliştirilen web-tabanlı sistemde bir testteki sorulara cevap vererek, test sonunda performansı ile ilgili dönütlerine ulaşması “*test alma davranışı*” olarak isimlendirilmiş ve test alma davranışının hem başarı hem de öz-müdahale algısı üzerine etkisi incelenmiştir. Aynı zamanda öğrenenlerin öz-müdahale algılarının başarıları üzerine etkisi olup olmadığını araştırılmıştır. Bu amaç için Hacettepe Üniversitesi Eğitim Fakültesinde Eğitimde Ölçme ve Değerlendirme dersini alan öğrenciler çalışmaya katılmış ve web-tabanlı bir öz-değerlendirme sistemini kullanmıştır. Başarı üzerine etkisi incelenen faktörler arasındaki bağıntılar yapısal eşitlik modeli ile sınanmıştır. Yapısal örüntülere göre; öz-müdahale algısının test alma davranışından bağımsız olduğu bulgusuna ulaşılmıştır. Bununla beraber hem test alma davranışının hem de öz-müdahale algısının başarı üzerine istatistiksel olarak anlamlı etkisi olduğu belirlenmiştir.

Anahtar Kelimeler: E-değerlendirme, öz-değerlendirme, öz-müdahale algısı, test alma davranışı

Abstract

Self-assessment approach which is a type of student-centered assessment have become more widespread at all levels of education. This approach is the scope of formative assessment and feedback is a crucial part of it. Learners are able to obtain feedback on their present performance in reference to criterion (criterion-referenced) and the group (norm-referenced) by means of these assessments as well as feedback on their advancement (self-referenced) through self-re-assessment. Based on these feedbacks, learners may interfere their own learning process to improve their learning. In this process students can decide and plan intervention and after implementation of the intervention they can evaluate the effectiveness of intervention. This whole process is referred to as self-intervention perception. Within the scope of this study, “test-taking behavior” was defined as the obtaining of feedback on performance of a test. The effect of test-taking behavior on success and self-intervention perception was examined. We simultaneously examined the effect of self-intervention perception on success. For this purpose, students attending the Measurement and Evaluation in Education class in Faculty of Education at Hacettepe University participated in repetitive testing on a web-based evaluation

* Bu çalışmanın bir kısmı 9. Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu’nda sözlü olarak sunulmuş ve sunulan çalışma bildiri özet kitapçığında yayınlanmıştır.

** Dr. Hacettepe Üniversitesi, Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara-Türkiye, e-posta: fbayrak@hacettepe.edu.tr

*** Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara-Türkiye, e-posta: yurdugul@hacettepe.edu.tr

system. Links relating the factors that affect success were examined using a structural equation model. According to the structural patterns, it was found that perceived self-intervention is independent of the number of tests taken. However, it was found that both the number of tests and the learner's self-intervention perception has an effect on success.

Keywords: E-assessment, self-assessment, self-intervention perception, test taking behaviour

GİRİŞ

Amacına göre öğretimsel değerlendirmeler, değer biçmeye dönük (summative) ve biçimlendirmeye dönük değerlendirme (formative) (Özçelik, 2011; Sadler, 1998; Scriven, 1967; Akt: Dunn ve Mulvenon, 2009) şeklinde sınıflandırılmaktadır. Değer biçmeye dönük değerlendirmenin “öğrenmenin değerlendirilmesi” (assessment of learning) olarak da isimlendirildiği ve düzey belirleme ve/veya öğrenmenin belgelendirilmesini amaçlamakta olduğu ifade edilmektedir. Biçimlendirmeye dönük değerlendirmenin ise “öğrenme için değerlendirme” (assessment for learning) olarak da isimlendirildiği ve öğretimi ve/veya öğrenmeyi yeniden şekillendirmeyi amaçladığı ifade edilmektedir.

Öğrenme için değerlendirme, öğrenenlerin öğrenme eksikliklerini tanımlamak ve öğrenmeyi arttırmak amacıyla öğretmenlerin çeşitli araç ve stratejileri kullandığı bir süreç olarak tanımlanabilir (Pinchok ve Brandt, 2009). Ayrıca öğrenme için değerlendirmenin, öğretmen veya öğrenci tarafından üstlenilmiş etkinlikleri kapsadığı ve meşgul olunan öğretme ile öğrenme etkinliklerini iyileştirmek için *dönüt* olarak iş gören bilgiyi sağladığı da ifade edilmektedir (Black ve William, 1998).

Günümüzde ise yapılandırmacı öğrenme yaklaşımına dayalı öğrenmelerde *öğrenci merkezli değerlendirmelerin* (öz-değerlendirme, akran değerlendirme, işbirlikli değerlendirme vb.) teknolojik uygulamaların getirdiği avantajlarla birlikte daha çok önem kazandığı görülmektedir. Biçimlendirmeye dönük değerlendirmenin özel bir biçimi olan bu tür değerlendirmeler de ise “öğrenme için değerlendirme”den farklı olarak dönütler öğrenene sunulur ve öğrenenlerin öğrenmelerini iyileştirmek için kendi sorumluluğunu alması beklenmektedir. Bu amaçla yapılan değerlendirmeler biçimlendirmeye dönük değerlendirme kapsamında olup “öğrenme olarak değerlendirme” (assessment as learning) şeklinde nitelendirilmektedir (Earl ve Katz, 2006).

Öğrenci merkezli değerlendirme yaklaşımının içerisinde yer alan öz-değerlendirme, son dönemlerde sıkça ele alınan ve öğrenenlerin kendi güçlü/zayıf yönlerini tespit edebildikleri, öğrenme farkındalıklarını artırabildikleri ve öğrenme yaşantılarını şekillendirebilmek için gerekli kararları verebildikleri bir değerlendirme etkinliğini ifade etmektedir. Ancak bilme hissinden dolayı yapılan değerlendirmelerinin yanlış olma ihtimalinin de olduğu ifade edilmektedir (Terry, 2011). Buna bağlı olarak öğrenenin var olan durumu ile istenen durum arasındaki farkı objektif olarak görmesi için desteğe ihtiyaç duyacağını söylemek yanlış olmayacaktır.

Öz-Değerlendirme, Dönüt Ve Öz-Müdahale

Eğitsel bağlamda öz-değerlendirme, öğrencilerin yerine getirdiği (öğrenme ve/veya değerlendirme görevlerine yönelik) görev-performansları hakkında yargıda bulunması olarak tanımlanmaktadır (Andrade ve Valtcheva, 2009; Terzis ve Economides, 2011). Öz-değerlendirmenin asıl amacının öğrenenin kendi performansındaki güçlü ve zayıf yönleri tespit ederek, öğrenmesini geliştirebilmesi olduğu söylenebilir (McMillan, 2007). Bu yüzden süreçte, dönüt anahtar bir role sahiptir. Öğrenenlerin var olan durumları hakkında dönütler alabilmesi için kendini test etme (self-testing) etkinliklerinin öz-değerlendirmede yaygın olarak kullanılan uygulamalar olduğunu ifade edilmektedir (Brew, 1999). Öğrenen bu süreçte kendi oluşturduğu ya da eğitimcinin hazırladığı sorulara cevap verebilir. Öğrenenlerin kendi hazırladığı soruların kapsamının ve düzeyinin yetersiz olabileceği için bu süreçte eğitimcinin hazırladığı sorular ön plana çıkmaktadır. Ancak yükseköğretimdeki sınıf mevcutlarının fazla olması, eğitimcilerin hem soru hazırlaması hem de eksik olduğu noktalarla ilgili öğrenene dönüt vermesi için yeterince zaman ayıramamasına neden olmaktadır. Bu süreçte bilgisayarların kullanılmasıyla, büyük miktarda soru veri tabanlarında

saklanabilmekte ve seçme algoritmaları ile etkili biçimde performans gösterilmektedir (Conejo, Guzmán, Millán, Pérez-de-la-Cruz ve Trella, 2004). Bu anlamda da teknoloji ile bireyin öz değerlendirme ve öğrenme süreçlerine destek sağlanabilmektedir.

Daha önce de ifade edildiği gibi bu süreçte dönüt anahtar bir role sahiptir ve farklı dönüt türleri söz konusudur. Öğrenenler ölçütlere (*criterion-referenced*) ve gruba göre (*norm-referenced*) var olan performans durumları hakkında; aynı zamanda kendilerini yeniden değerlendirerek gelişim süreçleri hakkında öz-referanslı (*self-referenced*) dönütler alabilirler (Brookhart, 2008; Hattie ve Gan, 2011). Öz referanslı karşılaştırmada öğrenenin daha önceki performansları temel alınır. Bu anlamda geçmiş performansa dayalı dönüt türü öğrenenin öğrenme sürecini şekillendirmesi açısından yarar sağlar. Benzer şekilde Bangert-Drowns, Kulik, ve Kulik (1991) ile Leeming (2002) tekrarlı ölçümler (frequent testing) ile sağlanan dönütlerin, öğrenenlerin kendi öğrenme süreçlerini daha iyi izlemesine, eksik ve güçlü yanlarını görmesine yardımcı olacağını ifade etmiştir. Öz-değerlendirme açısından her bir dönüt türü öğrenmenin iyileştirmesi için ayrı bir öneme sahiptir. Ancak Sadler (1989) dönütlerin öğrenen tarafından kullanılmadığı sürece öğrenmede iyileşmenin sağlanamayacağını ifade etmiştir. Buna bağlı olarak öğrencilerin aldıkları dönütleri inceleyerek eksik yönlerini tamamlayabilmeleri için gerekli müdahaleleri uygulamaları beklenmektedir. Uygulanan müdahalenin amacının büyüme/gelişme olduğu söylenebilir. Bu bağlamda Lentz, Allen ve Ehrhardt (1996) de müdahaleyi algılanan eksikliğin tamamlanabilmesi için değişik bir şey yapma olarak tanımlamaktadır.

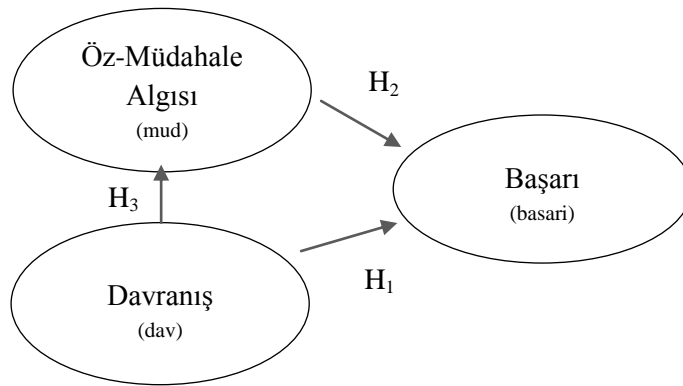
Eğitsel bağlamda öğrencilerin ölçülmesi amaçlanan özelliklere ilişkin tekrarlı ölçümler/testler ile ilgili yapılan çalışmalara bakıldığında ise tekrarlı ölçümlerin özel eğitimde yaygın olarak kullanıldığı ve öğretimsel müdahalelere temel oluşturduğu görülmektedir (Vaughn ve Fuchs, 2003). Bununla ilgili olarak 1970'li yıllarda Stanley Deno tarafından müfredata dayalı değerlendirme (*curriculum based assessment*) kavramı kullanılmaya başlanmıştır. Buna göre öğretmen, özel eğitime ihtiyaç duyan öğrencilerin performanslarına yönelik bir ölçüm yapar. Bu ölçüm üzerine öğretimsel müdahaleye yönelik karar verir ve bunu uygular. Bir sonraki ölçümde performans artışını inceler. Bu yaklaşım öğrenme güçlüğü olan öğrenciler için müdahaleye yanıt kuramı (response to intervention-RTI) olarak ortaya çıkmıştır. Ancak özel eğitimde müdahaleyi yapan bireyin süreci yürüten öğretmen olduğu görülmektedir. Yükseköğretimde ise öğrencilerin kendi öğrenme sorumluluklarını alarak, eksikliklerini tamamlamak için öğrenme süreçlerine müdahalede bulunmaları beklenmektedir. Buna bağlı olarak bu müdahale bir öz-müdahaledir (self-intervention). Deno tarafından ortaya konulan *özel eğitim problem çözme modeli* temel alındığında (1998) bireylerin uygun müdahalelere karar vermesi için a) *eksik yönlerin* belirlenmesi, b) bu eksik yönlerin tamamlanması için *plan* yapılması, c) plan dahilinde gelişimin *izlenmesi*, d) *değerlendirme* yapılarak etkili ve etkili olmayan yönlerin/kararların belirlenmesi ve buna bağlı olarak eksiklerin tamamlanması için tekrar sürece başlanması adımlarını takip etmelidir. Bu süreçte tekrarlı ölçümlere dayalı olarak aldıkları bireyselleştirilmiş dönütler öğrenenlere yol göstermektedir. Öz-müdahale sürecinde de öğrenenin bu karar sürecinden geçerek müdahalesini belirlemesi ve uyguladıktan sonra müdahalenin etkililiğine yönelik değerlendirme yapması gerekmektedir. Bu değerlendirmeyi yapabilmesi için gerekli verileri toplamasına ihtiyaç vardır. Süreçte dönütleri ele almaları, müdahaleye karar vermeleri ve müdahaleyi uyguladıktan sonra etkililiğini belirlemeleri öz-müdahale algısı olarak isimlendirilmiştir. Dönütler sayesinde öğrenenlerin eksik olduğu noktalara yönelik farkındalıkları artar ve öğrenenler daha iyi öğrenmeler gerçekleştirebilmek için konuyu tekrar etme, kullandığı çalışma stratejisini değiştirme veya çözemediği soruları inceleme gibi müdahalelerde bulunabilir. Buradan hareketle alınan test ve dönüt sayısının öz-müdahale algısını arttıracığı ve öz-müdahale algısı yüksek olan bireylerin de daha başarılı olmaları beklenmektedir.

Özellikle kalabalık sınıflar göz önüne alındığında, bireyselleştirilmiş dönütler öğrenene verilemeyebilir. Ancak gelişen internet tabanlı teknolojiler bu eksikliği giderecek niteliklere sahiptir. İlgili alanyazında ele alınan çalışmalara bakıldığında web tabanlı öz-değerlendirme sistemlerine ilişkin uygulamaların son yıllarda yoğunluk kazanmaya başladığı görülebilir.

Web tabanlı sistemler aracılığıyla büyük miktarda veri saklanıp işlenerek dönütler oluşturulması bu tarz sistemlerin bir avantajıdır. Web tabanlı öz-değerlendirme uygulamalarının bir diğer olumlu özelliği ise öğrencilerin istediği zaman ve yerde bu sistemlere erişip kendilerini değerlendirebilmeleridir. Bu olumlu özelliklerin yanı sıra alanyazın incelendiğinde, genellikle çalışmalarda tek test uygulamaları yapıldığı ve bu uygulamalarda dönütlerden yararlanma durumunun incelenmemiş olduğu (Van der Kleij, Eggen, Timmers ve Veldkamp, 2012; Wang, 2007; Wang, Wang, Wang, Huang ve Chen, 2004; Timmers, Braber-van den Broek, van den Berg, 2013) görülmüştür. Tek bir öz-değerlendirme öğrencinin yalnızca ilgili öğrenmelerine ilişkin bir defalık dönüt verebilmektedir. Bu noktadan hareketle öğretimsel müdahalelerin temelini oluşturmuş olduğu müfredata dayalı değerlendirmelerdeki gibi bir değerlendirme sistemi öz-müdahaleler için de yarar sağlayacaktır. Bunun içinde aynı öğrenmeleri değerlendirmeye yönelmiş benzer özellikteki değerlendirme görevlerine (tekrarlı ölçüm) ve bu ölçümlere dayalı gelişimlerine yönelik dönütler verebilen sistemlerin öğrenenlerin öz-müdahale algılarını nasıl etkilediğinin belirlenmesine yönelik araştırmalara ihtiyaç olduğu görülmektedir.

Araştırmanın Amacı

Öğrenenin bu çalışma kapsamında geliştirilen web tabanlı değerlendirme sisteminde bir testteki sorulara cevap vererek, test sonunda performansı ile ilgili dönütlerine ulaşması “*test alma davranışı*” olarak isimlendirilmiş ve test alma davranışının hem başarı hem de öz-müdahale algısı üzerine etkisi incelenmiştir. Aynı zamanda öğrenenlerin öz-müdahale algılarının başarı üzerindeki etkisi incelenmiştir. Şekil 1’de araştırmanın hipotez örüntüsü verilmiştir.



Şekil 1. Araştırmanın hipotez örüntüsü

Araştırmada test edilecek hipotezler şu şekildedir:

H₁: " WT-ÖS’de alınan test sayısının (davranış) öğrenci başarısı üzerinde etkisi vardır."

H₂: " WT-ÖS’de öz-müdahale algısının öğrenci başarısı üzerinde etkisi vardır."

H₃: " WT-ÖS’de alınan test sayısının (davranış) WT-ÖS’de öz-müdahale algısı üzerinde etkisi vardır."

YÖNTEM

Çalışma Grubu

Çalışma kapsamında, Hacettepe Üniversitesi Eğitim Fakültesinde okuyan ve Eğitimde Ölçme ve Değerlendirme dersini alan öğrencilerin (N=59) web tabanlı bir değerlendirme sisteminde tekrarlı uygulamalara dayalı olarak değerlendirme yaşantıları geçirmeleri sağlanmıştır.

Veri Toplama Araçları

Eğitimde Ölçme ve Değerlendirme Başarı Testi

Başarı testi Doğan ve İnal (2012) tarafından geliştirilmiştir. Test çoktan seçmeli beş seçenekli 35 sorudan oluşmaktadır. Başarı testinin KR-20 (Kuder-Richardson) formülü ile hesaplanan güvenilirlik katsayısı 0.896'dır. Testin kullanımı için yazarlardan izin alınmıştır.

Web-Tabanlı Öz-Değerlendirme Sisteminde Bulunan Ölçme ve Değerlendirme Testleri

Web-tabanlı öz-değerlendirme sisteminde Eğitimde Ölçme ve Değerlendirme dersine yönelik 4 adet çoktan seçmeli test bulunmaktadır. Bu testlerin ilki öğrencilerin sistemi tanımalarına yönelik olarak 15 adet çoktan seçmeli sorudan oluşurken, diğer 3 test 35 adet çoktan seçmeli maddeden oluşmaktadır. Bahsedilen son 3 test Doğan ve İnal (2012) tarafından geliştirilmiş başarı testi temel alınarak hazırlanmıştır ve uzman görüşü alınarak her bir testteki sorular ve testlerin kapsamlarının da denk olduğu belirlenmiştir. Sistemde tutulan iz kayıtları (log) ile öğrenenlerin kaç test aldığı belirlenmiştir.

Web-tabanlı Öz-değerlendirme Sisteminde Öz Müdahale Algısı Ölçeği

Ölçek, öğrenenlerin web tabanlı öz değerlendirme sisteminde test aldıktan sonra sunulan dönütlere dayalı kendi oluşturdukları müdahaleleri ile ilgili algısını belirlemek amacıyla geliştirilmiştir. Ölçeğin geçerlik ve güvenilirlik çalışmaları yapılmıştır (Bayrak, 2014). Ölçek maddeleri müdahale kavramı üzerine çalışan Deno'nun özel eğitim problem çözme modeli (1998) temel alınarak oluşturulmuştur. Ölçeğin faktöriyel geçerliğini belirlemek için doğrulayıcı faktör analizi yapılmıştır. Tablo 1'de faktör analizi sonuçları sunulmuştur.

Tablo 1. WT-ÖS'de Öz Müdahale Algısı Ölçek Verilerinin Model-Veri Uyum Değerleri

| Model | χ^2 /sd | Uyum iyiliği indeksi (GFI) | Ortalama Karekök Hata Tahmini (RMSEA) | Görelî Uyum İndeksi (CFI) | Normlaştırılmamış uyum indeksi (NNFI) |
|---------------------|--------------|----------------------------|---------------------------------------|---------------------------|---------------------------------------|
| Model I Tek boyutlu | 1.42 | 0.86 | 0.05 | 0.99 | 0.99 |

Ölçme sonuçlarının güvenilirliğini belirlemek için hem Cronbach Alfa katsayısı hem de yapı güvenirliliği hesaplanmıştır. Her iki güvenilirlik düzeyi 0.70'den büyüktür (Tablo 2). Bu anlamda ölçme sonuçlarının güvenilir olduğu söylenebilir.

Ölçeğin yapı geçerliği; yakınsama geçerliği (convergent validity) teknikleri ile incelenmiştir. Buna bağlı olarak ölçek maddelerinin boyuta ilişkin standartlaştırılmış faktör yükleri incelenmiş ve faktör yüklerinin 0.66 - 0.89 arasında değiştiği görülmüştür. Ayrıca ortalama açıklanan varyans

değerlerinin 0.50'den büyük olduğu belirlenmiştir. Yakınsama geçerliği için bu değerler ölçütleri karşılamaktadır.

Tablo 2. WT-ÖS'de Öz-Müdahale Algısı Ölçek Verilerinin Güvenirlik ve Ortalama Açıklanan Varyans değerleri (AVE)

| Boyut | Madde | Alfa Güv. | Yapı Güv. | OAV (AVE) |
|--------------------|----------|-----------|-----------|-----------|
| Öz-Müdahale Algısı | 10 madde | 0.94 | 0.94 | 0.60 |

Bu bulgulara bağlı olarak geliştirilen ölçeğe ilişkin; ölçek puanlarının güvenirlik katsayı değerlerinin kritik eşik olan 0.70 değerinden (Nunnally ve Bernstein, 1994) yüksek olduğu, diğer taraftan da geçerlik bulguları doğrultusunda madde puanlarının toplanıp faktör/yapı puanları olarak nitelendirilmesinin (geçerlik sonuçları doğrultusunda) olanaklı olduğu gözlenmiştir.

Web Tabanlı Öz-değerlendirme Sistemi ve Uygulama Süreci

Sistemde ilgili derse yönelik 4 adet çoktan seçmeli test bulunmaktadır. İlk test öğrencilerin sistemi tanımalarına yönelik açılmıştır. Her bir test 5 gün için erişime açılmıştır ve erişime açılan test öğrencilere duyurulmuştur. Öğrenci, testin erişime açık olduğu 5 gün içinde, istediği bir zaman sisteme kendi kullanıcı adı ve şifresi ile giriş yaparak test alabilmiştir. Öğrencilerin bir testi bir kez alma hakkı vardır. Belirtilen zaman içerisinde (bu çalışma kapsamında bir teste erişim 5 gün ile sınırlandırılmıştır) sisteme giriş yaptığında sınav süresi, soru türü ve soru sayısı ile ilgili bilgileri içeren yönergeyi görmekte ve "Sınav yönergelerini dikkatlice okudum." ifadesini işaretlemişse testte başlayabilmektedir (Şekil 2.).

XXX Sınavına Hoşgeldiniz !

Değerli _____ ;

Sınava başlamadan önce, aşağıdaki sınav yönergelerini lütfen dikkatlice okuyunuz, ardından da Sınavı Başlat butonuna tıklayınız.

Sınav Yönergeleri

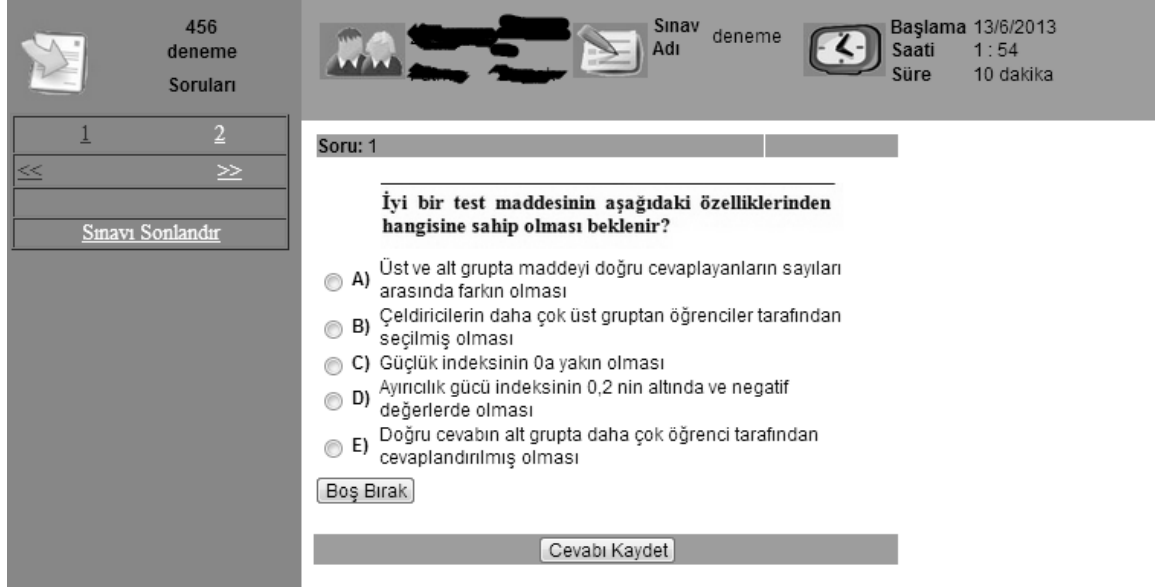
- 1- Sınav 30 / 03 / 2014 tarihinde saat 23 : 10 başlayabilirsiniz.
- 2- Sınavda süre kısıtlaması bulunmaktadır. Tüm soruları 10 dakika içinde cevaplamanız gerekmektedir.
- 3- Sınavda toplam 10 soru bulunmaktadır. Bunlardan, 10 tanesi çoktan seçmeli (beş seçenekli) sorularından oluşmaktadır.

Sınav Yönergelerini Dikkatlice Okudum.

Sınavı Başla

Şekil 2. Test yönergesi

Öğrenci yönergeyi okuduktan sonra Şekil 3'de gösterilen sistem arayüzünü görmektedir. Bu ekranda soruya verdiği cevabı kaydedebilmekte, istediği soruya geçebilmekte ve istediği zaman sınavı sonlandırabilmektedir.



Şekil 3. Sistem arayüzü

Çalışmaya katılım gönüllülük temelinde gerçekleştirilmiş ve testleri almak öğrenciler için zorunlu tutulmamış; test alan öğrencilere ayrıca bir puan verilmemiştir. Test erişime kapatıldıktan sonra, testi alan öğrenci verileri araştırmacı tarafından incelenmiştir ve bireyselleştirilmiş dönüt kartları hazırlanmıştır. Dönüt kartları ölçüt, norm ve öz-referanslı dönütlerini içermiştir. Dönüt kartları e-posta aracılığıyla öğrenenlere gönderilmiştir. Bu kapsamda öğrencilere sunulan dönütler ertelenmiş dönütler kapsamında ele alınmaktadır. Ekte örnek bir dönüt kartı bulunmaktadır. Her testten sonra ilgili testin bilgileri girilmiştir (Ek 1 ve Ek 2).

Öğrenciler 4 hafta sonunda başarı testi ve web-tabanlı öz-değerlendirme sistemlerinde öz-müdahale algısı ölçüğüne cevap vermişlerdir.

Verilerin Analizi

Çalışma kapsamında alınan test sayısı ile öğrencilerin öz-müdahale algılarının öğrencilerin başarısı üzerine etkisi olup olmadığı incelenecektir. Ayrıca öz-müdahale algısının alınan test sayısından etkilenip etkilenmediği belirlenecektir. Bu kapsamda yapısal eşitlik modeli ile bağıntılar sınanmıştır.

BULGULAR

Yapısal eşitlik modeli ilgili alanyazına bağlı olarak ortaya konulmuş gizil değişkenler arasında ilişkilerin ortaya konulmasını sağlayan bir yaklaşımdır. Bu araştırma kapsamında oluşturulan hipotezler de Şekil 1'de gösterilmiştir.

Sistem, biçimlendirmeye dönük değerlendirme kapsamında ele alınan öğrenme olarak değerlendirme amacıyla hazırlanmış ve öğrenenlerin öğrenme gelişimlerine destek olmasına yönelik dönütler öğrenenlere sağlanmıştır. İki kişi sistemi hiç kullanmadığı için analiz yapılmadan önce veri kümesinden çıkarılmıştır. Bu aşamada normallik varsayımı için aykırı ve uç değerler belirlenmiş ve veri kümesi temizlenmiştir.

Yapısal eşitlik modelinde bağıntıların ortaya konmasından önce gizil değişkenlere yönelik ölçme modellerinin veri-model uyumlarının test edilmesi gerekmektedir. Bununla ilgili olarak kurulacak olan yapısal eşitlik modelinde bulunan bütün ölçme modellerinin tek bir modelde incelenmesi de mümkündür. Ayrıca ölçmeye çalışılan yapıların ölçülüp ölçülmediğine karar vermek için yapı

geçerliği, ölçme araçlarından elde edilen iç tutarlılık anlamındaki güvenilirlik katsayıları da belirlenmelidir. Bu çalışmada yapı geçerliği Fornell ve Larcker (1981) tarafından önerilen bağıntılardan yararlanılmıştır. Buna göre yapı geçerliği yakınsama (convergent validity) ve iraksama geçerliği (divergent validity) olarak iki alt bölümde incelenmektedir. Yakınsama geçerliği ilişkili yapılar arasındaki yüksek bağıntıyı ifade ederken, farklı yapıların ayrıştığı bulgu ise iraksama geçerliği olarak nitelendirilir (Kline, 2005). Ancak iraksama geçerliği günümüzde ayırteci geçerlik (discriminant validity) olarak kullanılan yöntemler ile test edilmektedir. Yakınsama ve ayırteci geçerlik için önemli bir kavram olan faktörlerin ortalama açıklanan varyans (Average Extraction Variance-AVE) (OAV) kullanılmıştır. OAV aynı zamanda yapı değişkenlerinin varyansını ifade etmektedir.

Bunlara bağlı olarak araştırmanın hipotezlerini test etmek için yapısal eşitlik modelini kurmadan önce ölçme modelleri için tek bir model kurularak, sırasıyla ölçme modellerinin veri-model uyumları, yakınsama geçerlik, ayırteci geçerlik ve güvenilirlik bulguları raporlanmış, daha sonrada model parametrelerinin istatistiksel anlamlılıkları tartışılmıştır.

Ölçme modellerinin veri-model uyumunu incelemek için kurulan modelin standartlaştırılmış faktör yükleri ile boyutlar arası ilişkiler Ek 3'te gösterilmiştir.

Standartlaştırılmış etki katsayıları Pearson korelasyon katsayısı olarak da nitelendirilir. Bu regresif katsayı standartlaştırıldığından dolayı birimden bağımsız hale getirilir. Bu nedenle çalışmada öncelikle standartlaştırılmış etki katsayılarının anlamlılığı test edilmiş; daha sonra katsayının büyüklüğü birimden bağımsız olarak yorumlanmıştır (Ek 3). Standartlaştırılmış etki düzeyleri incelendiğinde öğrencilerin aldığı test sayısı olarak tanımlanan davranışın öğrenci başarısı üzerine etki katsayısının 0.52 olduğu; öğrencilerin öz-müdahale algısının başarı üzerine etki katsayısının ise 0.31 olduğu belirlenmiştir. Bu değerlerin t sınama sonuçlarına göre istatistiksel olarak anlamlı olduğu belirlenmiştir. Ancak test alma davranışlarının öğrencilerin öz-müdahale algıları üzerine istatistiksel olarak anlamlı bir etkisi olmadığı belirlenmiştir.

Aynı zamanda kurulmuş olan modelin veri-model uyumu için Tablo 3'te verilen indeks değerleri incelenmiştir. Bu değerlerden sadece GFI değerinin kabul edilen değerler arasında olmadığı görülmüştür. Bununla ilgili olarak Breivik ve Olsson (2001; Akt: Kline, 2011) GFI değerinin gözlem sayısından etkilendiği ifade etmiştir. Diğer indeksler (hata ve uyum) temel alındığında model-veri uyumunun olduğu söylenebilir.

Tablo 3. Ölçme modellerine yönelik uyum ve hata indeksleri ile bu indekslerin kabul edilebilir değer aralığı

| | Kabul Edilebilir Değerler | Bu çalışmada hesaplanan değer |
|---------------------------------------|---------------------------|-------------------------------|
| Uyum iyiliği indeksi (GFI) | GFI \geq 0,90 | 0.82 |
| Ortalama Karekök Hata Tahmini (RMSEA) | RMSEA \leq 0,08 | 0,08 |
| Normlaştırılmamış uyum indeksi (NNFI) | NNFI \geq 0,90 | 0.97 |
| Görelî Uyum İndeksi (CFI) | CFI \geq 0,90 | 0.97 |

Yakınsama geçerliği kapsamında; faktör yükleri ve ortalama açıklanan varyans (OAV) değerleri incelenmiştir. Boyutlara ilişkin standartlaştırılmış faktör yükleri 0.65 ile 0.94 arasında değişmektedir ve bu değerler istatistiksel olarak anlamlıdır. Ayrıca ortalama açıklanan varyans değeri algılanan öz-müdahale yapısı için 0.50'den büyük olduğu belirlenmiştir. Bu bulgulara dayanarak ölçme modelinin yakınsama geçerliği için belirlenen ölçütleri karşıladığı söylenebilir.

Ayırteci geçerlik kapsamında OAV'nin karekökünün (0.78) boyutlar arasındaki korelasyonlardan (0.14 ve 0.31) büyük olduğu belirlenmiştir. Aynı zamanda bu değer Fornell ve Larcker (1981)'in

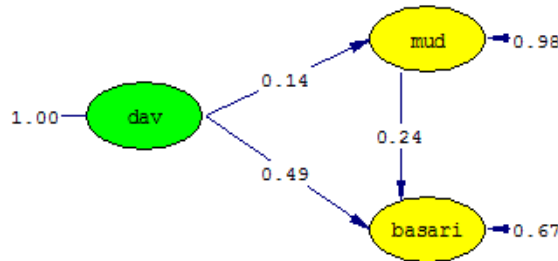
belirttiği ölçüt olan 0.50'den büyük olduğu görülmüştür Buna bağlı olarak ayırtecdici geçerliğin sağlanmış olduğu ifade edilebilir.

Şekil 1'de verilen araştırma modelinin parametreleri yapısal eşitlik çözümlemesiyle kestirilmiş ve modelin uyum indeksleri elde edilmiştir (Tablo 4). Bununla ilgili olarak Breivik ve Olsson (2001; Akt: Kline, 2011) GFI değerinin gözlem sayısından etkilendiği ifade etmiştir. Diğer indeksler temel alındığında model-veri uyumunun olduğu söylenebilir.

Tablo 4. Araştırma modeline yönelik oluşturulacak yapısal eşitlik modeline yönelik hesaplanan uyum ve hata indeksleri ile bu indekslerin kabul edilebilir değer aralıkları

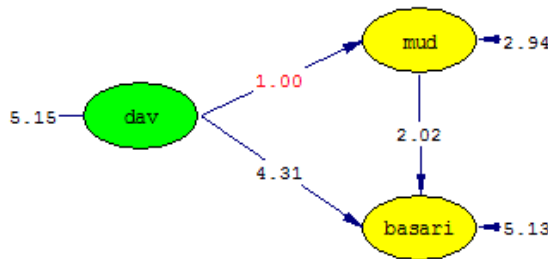
| | Kabul Edilebilir Değerler | Bu çalışmada hesaplanan değer |
|---------------------------------------|---------------------------|-------------------------------|
| Uyum iyiliği indeksi (GFI) | GFI \geq 0,90 | 0.80 |
| Ortalama Karekök Hata Tahmini (RMSEA) | RMSEA \leq 0,08 | 0,06 |
| Normlaştırılmamış uyum indeksi (NNFI) | NNFI \geq 0,90 | 0.97 |
| Görel Uyum İndeksi (CFI) | CFI \geq 0,90 | 0.97 |

Model-veri uyumu tespit edildikten sonra bağıntı katsayıları ve katsayıların t değerleri incelenmiştir. Tablo 5'te bağlantı katsayıları ve bu katsayılara ilişkin t değerleri ve hipotez sonuçları sunulmuştur (Şekil 4 ve Şekil 5).



Chi-Square=67.94, df=52, P-value=0.06803, RMSEA=0.076

Şekil 4: Yapısal eşitlik modeline ilişkin standart katsayılar



Chi-Square=67.94, df=52, P-value=0.06803, RMSEA=0.076

Şekil 5: Yapısal eşitlik modeline ilişkin t değerleri

Tablo 5. Hipotez sonucuna göre hesaplanmış olan bağıntıların β katsayıları ve bu katsayıya yönelik t değeri

| Hipotezler | Bağıntılar | β katsayısı | t değeri | Sonuç |
|------------|--------------------|-------------------|----------|-------|
| H1 | DAVRANIS -> BASARI | 0.49 | 4.31 | KABUL |
| H2 | MUD -> BASARI | 0.24 | 2.02 | KABUL |
| H3 | DAVRANIS -> MUD | 0,14 | 1.00 | RED |

Yapısal eşitlik modeli sonuçlarına göre;

H₁: " WT-ÖS'de alınan test sayısının (davranış) öğrenci başarısı üzerinde etkisi vardır." hipotezi kabul edilmiştir ($\beta = 0.49$; $p > 0.01$, $t = 4.31$).

H₂: " WT-ÖS'de öz-müdahale algısının öğrenci başarısı üzerinde etkisi vardır." hipotezi kabul edilmiştir ($\beta = 0.24$; $p > 0.01$, $t = 2.02$).

H₃: " WT-ÖS'de alınan test sayısının (davranış) WT-ÖS'de öz-müdahale algısı üzerinde etkisi vardır." hipotezi red edilmiştir ($\beta = 0.14$; $p < 0.01$, $t = 1.00$).

Sisteme ilişkin öğrenenlerin kullanma davranışlarına yönelik yapısal örüntüleri çözümlenmiştir. Elde edilen bulgulara göre; öz-müdahale algısının alınan test ve dönüt sayısından bağımsız olduğu bulgusuna ulaşılmıştır. Gizil değişkenler arasında yapısal eşitlikler aşağıda verilmiştir:

$$\text{Başarı Performansı} = 0.71 * \text{Öz-Müdahale} + 1.96 * \text{Test Alma Davranışı}$$

Bu kestirim değerlerine göre test alma davranışının ve öz-müdahale algısının istatistiksel olarak anlamlı bir etkisi söz konusu iken, başarı üzerinde özellikle test ve dönüt alma sıklığının daha etkin bir gösterge olduğu anlaşılmıştır. Bununla birlikte, bu iki değişken birlikte başarı performansının %33'ünün ($R^2 = 0.33$) açıklamaktadır. Geri kalan %67'si ise modele alınmayan diğer etkenler ile açıklanmaktadır.

SONUÇLAR ve TARTIŞMA

Öz-değerlendirme günümüzde öğrenenlerin öğrenme farkındalıklarını arttırdığı için eğitimin her kademesinde yaygınlaşmaktadır. Özellikle yükseköğretimde, bireylerin iş yaşamlarında kişisel gelişimlerine devam edebilmeleri için öz-değerlendirmelerine destek olacak etkinliklerin ön planda tutulması ayrıca önerilmektedir (Boud, 2000). Bu kapsamda hem etkinlikler hem de dışsal dönütler (objektif oldukları için) ön plana çıkmaktadır. Sınıf mevcutlarının fazla olması bu sürecin sağlıklı işlemesine engel olabilmektedir (Boud, 2000).

Teknoloji destekli ortamlar sayesinde hem bu süreçler daha sağlıklı yönetilebilmekte hem öğrenenler kendilerini istediği kadar test edebilmekte hem de bireysel dönütler alabilmektedirler. Dönüt bu süreçte önemli bir bileşendir. Bu dönütler farklı türde olabilmektedir. Öğrenenler değerlendirme yaşantıları aracılığıyla ölçütlere (criterion-referenced) ve gruba göre (norm-referenced) var olan performans durumları hakkında; aynı zamanda kendilerini yeniden değerlendirerek gelişim süreçleri hakkında (self-referenced) dönütler alabilirler. Öğrenenler aldıkları bu dönütlere dayalı olarak daha iyi öğrenmeler gerçekleştirebilmek için öğrenme süreçlerine kendileri müdahalede bulunabilirler. Bu süreçte dönütleri ele almaları, müdahaleye karar vermeleri ve müdahaleyi uyguladıktan sonra etkililiğini belirlemeleri öz-müdahale algısı olarak isimlendirilmiştir.

Bu noktadan hareketle çalışma kapsamında Eğitimde Ölçme ve Değerlendirme dersini alan öğrenciler, web tabanlı bir öz-değerlendirme sistemini kullanmıştır. Öğrencilerin sistemde bir testteki sorulara cevap vererek, test sonunda performansı ile ilgili dönütlerine ulaşması "test alma davranışı" olarak isimlendirilmiş ve test alma davranışının hem başarı hem de öz-müdahale algısı üzerine etkisi incelenmiştir. Aynı zamanda öğrenenlerin öz-müdahale algılarının başarı üzerinde etkisi olup olmadığı incelenmiştir. Buradan hareketle hipotez edilen araştırma modelinin yapısal eşitlik yaklaşımıyla çözümlenmesi sonucu elde edilen parametrelerin anlamlılığı test edilmiştir.

Elde edilen bulgulara göre; test alma davranışının başarı üzerinde anlamlı bir etkisi olduğu sonucuna ulaşılmıştır. Benzer olarak alanyazında test alma sıklığı ile öğrencilerin başarılarında artış olduğu çeşitli araştırmalarda raporlanmıştır (Brown, Andrade ve Chen, 2015, Dobson, 2008; Geist ve Soehren, 1997; Johnson, 2006).

Geliştirilen bu sistem ile öğrencilerin kendilerini test ederek, öz-değerlendirme süreçleri kapsamında öğrenmelerini iyileştirmek için öz-müdahalelerini planlamaları, müdahaleyi uygulayarak süreci izlemeleri ve müdahalenin etkililiğini belirlemelerine destek olacak dönütün sağlanması amaçlanmıştır. Bu araştırma sonuçlarına bağlı olarak öz-müdahale algısı yüksek olan öğrencilerin başarılarının da yüksek olduğu görülmüştür. Diğer taraftan öğrencilerin test aldıkça gelişimleri hakkında daha fazla bilgi içeren dönütler alabilmektedirler. Buna bağlı olarak özellikle öğrencilerin müdahalelerini izlemeleri ve değerlendirmeleri ile ilgili öz-müdahale algılarının artması ve dolayısıyla başarılarının artması beklenmekteydi. Bununla ilgili olarak Bangert-Drowns, Kulik, Kulik (1991) ve Leeming (2002) tekrarlı ölçümler (frequent testing) ile sağlanan dönütlerin, öğrencilerin kendi öğrenme süreçlerini daha iyi izlemesine, eksik ve güçlü yanlarını görmesine yardımcı olacağını ifade etmiştir. Ancak öz-müdahale algısının test alma davranışından bağımsız olduğu görülmektedir. Bununla ilgili olarak alanyazında öğrencilerin sunulan dönütleri nasıl yorumlayacaklarını bilmedikleri ifade edilmektedir (Sadler, 1998; Wang, 2014). Bu noktadan hareketle öğrencilerin kendi eksik yönlerini belirleme ve eksiklerini tamamlamak için öz-müdahale süreçlerinde ölçüt, norm ve özellikle öz referanslı dönütleri nasıl kullandıklarına ve sistem ile ilgili gerekli müdahalelerin belirlenmesine yönelik çalışmalara ihtiyaç vardır.

KAYNAKÇA

- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12-19. doi: 10.1080/00405840802577544
- Bangert-Drowns, R. L. Kulik, J. A. & Kulik, G. L. C. (1991). The effects of frequent class-room testing. *Journal of Educational Research* , 85, 89-99.
- Bayrak, F. (2014). *Web tabanlı öz-değerlendirme sisteminde algılanan öz müdahalenin etkililiği*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), 151-167.
- Brew, A. (1999) Towards autonomous assessment: using self-assessment and peer assessment. In: S. Brown & A. Glasner (Eds), *Assessment matters in higher education: choosing and using diverse assessment*, (pp. 159–171). Buckingham, Open University Press/SRHE.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: ASCD
- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444-457. doi: : 10.1080/0969594X.2014.996523
- Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J. L., & Ríos, A. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education*, 14(1), 1-33
- Deno, S. L. (1998). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn, (Eds.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). Guilford Press.
- Doğan, N., & İnal, H. (2012). *Eğitim Fakültesi öğrencilerinin öğrenme stilleri ve ölçme ve değerlendirme dersi başarısı arasındaki ilişki*. Paper presented at Eğitimde ve Psikolojide Ölçme ve Değerlendirme III. Ulusal Kongresi, Bolu, September, 19-21.
- Dobson, J. L. (2008). The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education*, 32(4), 297-302.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1-11.
- Earl, L. M., & Katz, M. S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship & Youth
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.

- Geist, J. R. & Soehren, S. E. (1997). The Effect of Frequent Quizzes on Short- and Long-Term Academic Performance. *Journal of Dental Education*, 61(4), 339-345.
- Hattie, J., & Gan, M., (2011). Instruction Based on Feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning* (pp. 249-271), New York, New York, USA: Routledge
- Johnson, G. M. (2006). Optional online quizzes: College student use and relationship to achievement. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 32(1).
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Publication, Inc.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Publication, Inc.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, 29(3), 210-212.
- Lentz, F. E., Allen, S. J., & Ehrhardt, K. E. (1996). The conceptual elements of strong interventions in school settings. *School Psychology Quarterly*, 11(2), 118-136.
- McMillan, J. H. (Ed.). (2007). *Formative classroom assessment: Research, theory and practice*. New York: Teacher's College Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGrawHill.
- Özçelik, D. A., (2011). *Ölçme ve Değerlendirme* (3. Baskı). PEGEM Akademi, Ankara
- Pinchok, N., & Brandt, W. C. (2009). *Connecting formative assessment research to practice: An introductory guide for educators*. New York, NY: Learning Point.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2), 119-144.
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84
- Terry, W. S. (2011). Öğrenme ve bellek. (Çev. Ed. B. Cangöz). Ankara: Anı Yayıncılık. (Orijinal eserin yayın tarihi 2009).
- Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education*, 56(4), 1032-1044. doi:http://dx.doi.org/10.1016/j.compedu.2010.11.017
- Timmers, C. F., van den Broek, J. B., & van den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education*, 60(1) , 25-31. doi:10.1016/j.compedu.2012.07.007
- Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58, 263 – 272. doi:10.1016/j.compedu.2011.07.020.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, 18, 137-146.
- Wang, T. H. (2007). What strategies are effective for formative assessment in an e-learning environment?. *Journal of Computer Assisted Learning*. 23(3), 171–186.
- Wang, T. H. (2014). Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Computers & Education*, 73, 189-203. doi:10.1016/j.compedu.2013.12.002
- Wang, T. H., Wang, K. H., Wang, W. L., Huang, S. C., & Chen, S. Y. (2004). Web-based Assessment and Test Analyses (WATA) system: development and evaluation. *Journal of Computer Assisted Learning*, 20(1), 59-71.

EXTENDED ABSTRACT

Introduction

Self-assessment has become more common at every educational level due to its role in increasing learners' awareness and determining their own strengths and weaknesses. In higher education in particular, it is recommended to prioritise activities supporting self-assessment in order to promote the ability for personal development in the future professional life (Boud, 2000). In this context, activities and external feedback are important. Large class sizes can prevent the proper continuation of this process (Boud, 2000). In relation, learners can obtain impartial feedback for their actual performance by taking tests prepared on web-based environments as well as longitudinal feedback by testing themselves on a subject repeatedly. Through this feedback, learners are equipped with the

information needed to intervene in their learning in order to make improvements. In addition, feedback allows learners to evaluate the effectiveness of these interventions in following measurement. This process, as performed by the learner, can be defined as self-intervention perception. It is of note that previous studies have used off-the-shelf software and are limited to a single test while almost never using feedback about learners' past performances. Therefore, it should be determined whether systems that do frequent testing and can provide different types of feedback (criteria, norm, and self-referenced) change learners' success. In this way inferences can be made regarding the design of these types of systems.

Method

This study included students enrolled in the Faculty of Education at Hacettepe University that went through an assessment process based on repetitive applications in a web-based assessment system. Parameters in the hypothesised model were analysed using the structural equation approach.

Results and Discussion

Results showed that a) frequency of test and b) self-intervention perception significantly affect learner success. Previous studies have similarly shown that the frequency of test increases learners' success. (Brown, Andrade and Chen, 2015, Dobson, 2008; Geist and Soehren, 1997; Johnson, 2006).

On the other hand, as learners took a single test and did not obtain self-referenced feedback, it is necessary to determine whether this changed learners' self-intervention perception. From this point of view, learners who used repetitive web based assessment systems were expected to have a high self-intervention perception and be more successful. Although self-intervention perception is independent of test-taking behaviour, results show that it affects success. However, Bangert-Drowns, Kulik, Kulik (1991) and Leeming (2002) reported that feedbacks provided by frequent testing help learners better monitor their own learning process and identify their strengths and weaknesses. The current study shows that while both the number of tests and a learner's self-intervention perception affect success, the effect of the former is greater than that of the latter. Correspondingly, previous studies reported that learners do not know how to interpret the feedback (Sadler, 1998; Wang, 2014). From this point of view, studies are needed to determine how learners use criteria, norm and self-referenced feedbacks in their self-intervention process to identify their own weaknesses and to overcome them.

EKLER

EK 1: Dönüt Kartı Ekran Görüntüsü 1



Sevgili

Eğitimde ölçme ve değerlendirme konu alanından ikinci kez girmiş olduğun sınavdan sana özel bulgulara ulaşılmıştır. Buna göre test genelinden X puan almış gözükyorsun. Ancak hangi konularda güçlü ve hangi konularda zayıf olduğuna ilişkin ayrıntılara ekteki 2 nolu tablodan ulaşabilirsin.

Bununla birlikte; birlikte sınava girdiğin arkadaşların arasında sınav başarısına göre X. sırada olduğun gözükmektedir. Başarı dilimine göre ise üst/orta/alt grup sınıflamasında X grupta olduğun görülmektedir.

Diğer taraftan; sınavda gruba göre zor ve kolay sorulara göre yapılan sınıflamada ise senin X grubunda olduğun ortaya çıkmıştır. Bu konudaki ayrıntıları 1 nolu tabloda inceleyebilirsin.

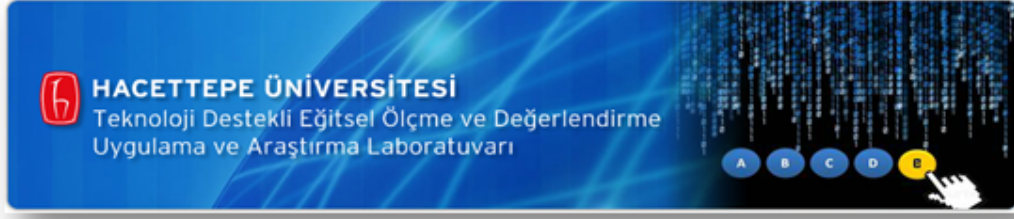
! Sevgili

Bu sınavların amacı; eğitimde ölçme ve değerlendirme konu alanındaki durumunu, gelişimini ve güçlü/zayıf olduğun konular ile grup içerisindeki konumunu belirlemektir. Özellikle zayıf/eksik olduğun konulardaki önlemlerinin işe yarayıp yaramadığını kontrol etmek, başarılı olduğun konularda da başarının tutarlı olduğunu görebilmek için bir sonraki sınavın gerekliliklerini yerine getirmeni rica ediyoruz.

Sevgiler ve başarılar

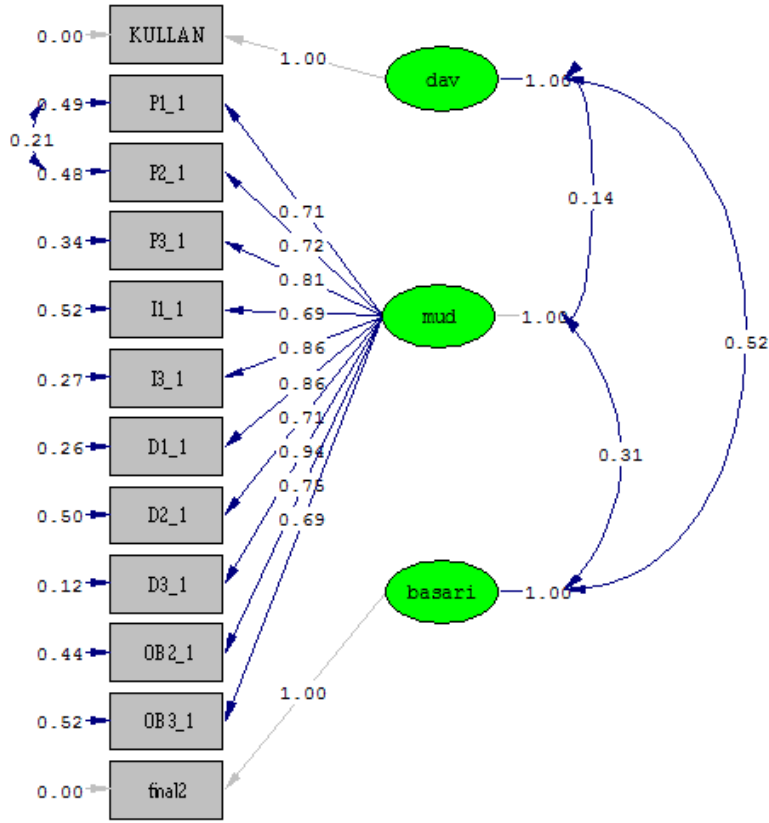
| | |
|------------------------|---|
| A | Etkili bir öğrenme gerçekleştirmişsiniz. |
| A' | Etkili bir öğrenme gerçekleştirmişsiniz. Ancak dikkatsizlikle kaçırılmış sorular var. |
| B | Etkili bir öğrenme gerçekleştirmişsiniz. Ancak çalışmanız gereken konularınız da bulunmaktadır. |
| B' | Konulara yeterince çalışmamışsınız. |
| C | Öğrenme düzeyiniz çok yetersiz. |
| C' | Öğrenme gerçekleştirmemişsiniz. |
| Sizin sınıfınız | |

EK 2: Dönüt Kartı Ekran Görüntüsü 2



| Tablo 2. Konular | TEST 2 | | Cevabınız* | MDCY | Cevabınız* | MDCY |
|--|------------|------|------------|------|------------|------|
| | Cevabınız* | MDCY | | | | |
| Temel Kavramlar | | 48 | | 61 | | |
| Ölçme ve Değerlendirmenin Amaçları | | 88 | | 24 | | |
| Temel Kavramlar | | 96 | | 98 | | |
| Ölçüte göre değerlendirme türleri | | 72 | | 80 | | |
| Ölçek türleri | | 86 | | 98 | | |
| Ölçme Türleri | | 42 | | 87 | | |
| Ölçek Türleri | | 60 | | 67 | | |
| Değişken | | 56 | | 65 | | |
| Temel Kavramlar | | 72 | | 96 | | |
| Puanlama | | 24 | | 41 | | |
| Ölçek türleri | | 56 | | 22 | | |
| Korelasyon | | 62 | | 67 | | |
| Hata türleri | | 88 | | 83 | | |
| Geçerlik Türleri | | 38 | | 33 | | |
| Geçerlik Türleri | | 80 | | 80 | | |
| Güvenirlilik | | 86 | | 83 | | |
| Güvenirlilik | | 60 | | 70 | | |
| Ölçme Aracında Aranılan Nitelikler | | 86 | | 93 | | |
| Hata türleri | | 68 | | 91 | | |
| Güvenirlilik | | 64 | | 89 | | |
| Ölçme araçları | | 88 | | 93 | | |
| Ölçme araçları | | 90 | | 93 | | |
| Şans başarısı | | 74 | | 76 | | |
| Hedeflerin Sınıflandırılması (Bilişsel Alan) | | 42 | | 37 | | |
| Birim | | 44 | | 85 | | |
| Davranış | | 70 | | 85 | | |
| Hedeflerin Sınıflandırılması (Duyuşsal Alan) | | 88 | | 91 | | |
| Test Geliştirme Süreci | | 80 | | 76 | | |
| Dereceli Puanlama Anahtarı | | 12 | | 43 | | |
| Dağılım Ölçüleri | | 28 | | 11 | | |
| Standart Puanlar | | 26 | | 63 | | |
| Merkezi Eğilim Ölçüleri | | 70 | | 74 | | |
| Madde İstatistikleri | | 22 | | 72 | | |
| Toplam Doğru Cevap Sayınız | | | | | | |

EK 3. Ölçme modelleri ve aralarındaki ilişkiler (standartlaştırılmış faktör yükleri)



Chi-Square=67.94, df=52, P-value=0.06803, RMSEA=0.076

dav: test alma davranışı

mud: öz-müdahale algısı

Ergenler için Yeterlik Algısı Ölçeğinin (YAÖ) Geliştirilmesi*

Developing Perceived Competence Scale (PCS) for Adolescents

Arif ÖZER** Dilek GENÇTANIRIM KURT*** Seval KIZILDAĞ****
Selen DEMİRTAŞ ZORBAZ***** Fatma ARICI ŞAHİN*****
Tülin ACAR***** Tuncay ERGENE*****

Öz

Bu çalışmada, lise öğrencilerinin yeterlik algılarını ölçmek amacıyla Yeterlik Algısı Ölçeği geliştirilmiştir. Ölçek geliştirme çalışmaları üç ayrı örneklem üzerinde gerçekleştirilmiştir. Ankara ilindeki liselerde 2011 - 2012 eğitim ve öğretim yılında okullarına devam eden 372 öğrenci betimleyici faktör analizi, 668 öğrenci doğrulayıcı faktör analizi ve 75 öğrenci test tekrar test güvenilirlik çalışmaları katılmışlardır. İç tutarlık katsayıları (Cronbach ve tabakalı α) her grup için ayrı ayrı hesaplanmıştır. Verilerin çözümlenmesinde Factor 8.02 ve LISREL 8.70 paket programları kullanılmıştır. Analiz sonuçlarına göre; betimleyici ve doğrulayıcı faktör analizlerinin yapıldığı örneklerde iç tutarlık (Cronbach α) katsayıları akademik yeterlik boyutu için sırasıyla .90 - .93, sosyal yeterlik boyutu için .82-.86'dır. Ölçeğin bütünü için iç tutarlık (tabakalı α) katsayısı ise .91 olarak hesaplanmıştır. Test tekrar test güvenilirlikleri için düzeltilmiş korelasyon katsayıları (r) sosyal yeterlik için .94; akademik yeterlik için .90'dır. Faktör analizleri sonucunda elde edilen regresyon ağırlıkları ve genel uyum katsayılarına ek olarak, benzeme (convergent) ve ayırma (discriminant) geçerliklerine ilişkin bulgular yeterliğin, akademik (16 madde) ve sosyal (14 madde) olmak üzere iki alt boyutta (toplam 30 madde) ele alınabileceğini göstermektedir.

Anahtar Kelimeler: Algılanan yeterlik, akademik yeterlik, sosyal yeterlik, benzeme geçerliği, ayırma geçerliği

Abstract

In this study, Perceived Competence Scale was developed to measure high school students' perceived competence. Scale development process was verified on three different samples. Participants of the research are some high school students in 2011-2012 academic terms from Ankara. Participants' numbers incorporated in exploratory factor analysis, confirmatory factor analysis and test-retest reliability are respectively as follows: 372, 668 and 75. Internal consistency coefficients (Cronbach's and stratified α) are calculated separately for each group. For data analysis Factor 8.02 and LISREL 8.70 package programs were used. According to results of the analyses, internal consistency coefficients (α) are .90 - .93 for academic competence, .82 - .86 for social competence in the samples that exploratory and confirmatory factor analysis performed. For the whole scale internal consistency coefficient (stratified α) is calculated as .91. As a result of test-retest reliability, adjusted correlation coefficients (r) are .94 for social competence and .90 for academic competence. In addition to fit indexes and regression weights obtained from factor analysis, findings related convergent and discriminant validity indicate that competence can be addressed in two dimensions which are academic (16 items) and social (14 items).

Keywords: Perceived competence, academic competence, social competence, convergent validity, discriminant validity

* Bu makale, TÜBİTAK tarafından desteklenen Prof. Dr. Tuncay Ergene'nin yürütücülüğünü yaptığı "Lise Öğrencilerinde Riskli Davranışların İncelenmesi" başlıklı ve 110K430 numaralı SOBAG projesinin bir kısmından oluşturulmuştur. Dr. Tuncay ERGENE, Boston Children's Hospital'da bulunan ve Principle Investigator (PI): Kerim M. Munir, MD, MPH, DSC.'in olduğu NIMH ICORTHA Fogarty International Mental Health and Developmental Disabilities (MH/DD) Research Training Program'ı (D43TW05807) tarafından desteklenmektedir.

**Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, arifozer@hacettepe.edu.tr:

***Yard. Doç. Dr., Ahi Evran Üniversitesi, Eğitim Fakültesi, Kırşehir – Türkiye, digenc@gmail.com

****Arş. Gör. Dr., Adıyaman Üniversitesi, Eğitim Fakültesi, Adıyaman-Türkiye, e-posta: sevalpdr@gmail.com

*****Arş. Gör., Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, e-posta: selenpdr@gmail.com

*****Arş. Gör., Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, e-posta: farici@hacettepe.edu.tr

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

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

GİRİŞ

Son dönemde Psikolojik Danışma ve Rehberlik alanında ergen ve çocuklarla yapılan çalışmalarda riskli davranışların sıklıkla konu edinildiği dikkat çekmektedir. Bu durumdaki artışın temel nedeni ise bu tür davranışların nasıl önlenebileceği sorusuna yanıt vermektedir. Dolayısıyla alanda çalışan profesyonellerin önemle üzerinde durduğu bir konu da riskli davranışlara ilişkin koruyucu faktörlerin belirlenmesidir. Çoğunlukla riskli davranışlarla çalışan araştırmacılar çocuklarda ve ergenlerde riskli davranışlarla baş etmede akademik ve sosyal yeterlik alanları üzerinde durmakta; bu özelliklerin koruyucu bir faktör olduğunu dile getirmektedirler (Achenbach ve Edelbrock, 1979; Achenbach ve Zigler, 1963; Achenbach ve McConaughy, 1987). Bu bağlamda öncelikle sosyal ve akademik olmak üzere iki boyutta incelenen yeterliğin, çocuk ve ergenlerin ruh sağlığı, sosyal ve kişisel gelişimi açısından koruyucu faktör olduğu düşünülebilir.

Akademik yeterlik, literatürde çoğunlukla öğretmenlerin verdikleri ödev ve yaptıkları sınavlara göre ölçülmektedir ancak akademik yeterlik öğrencilerin akademik performanslarından daha fazlasını ifade etmektedir. Başka bir ifade ile akademik yeterliğin sadece başarı ile ölçülmesi ya da başarı ile açıklanması mümkün görülmemektedir, çünkü akademik yeterlilik performanstan çok daha fazlasını içeren bir kavramdır. Anderman ve Midgley (1997) öğrencilerin sadece kendi yeteneklerini değerlendirmenin ötesinde, güçlük derecesine bağlı olarak derslerdeki yaptıklarını değerlendirerek akademik yeterlik algılarını oluşturduklarını ifade etmektedirler. Dolayısıyla bu noktada ölçülen başarıdan çok çocuk ya da ergenin kendisini o konuda nasıl bulduğu önem kazanmaktadır. Akademik yeterliğe sahip çocuk ya da ergenlerin daha mutlu oldukları ve yeterlik algıları ile benlik saygılarının daha yüksek olduğu ileri sürülmektedir (Blechman, Tinsley, Carella ve McEnroe, 1985). Akademik yeterlik algısının düşük olması ise, genel olarak problem çözme, sonuçları hakkında başarıya ilişkin işlevsel olmayan nedensel çıkarımlar ve zayıflatılmış başarı amaçlarıyla ilişkilidir (Phillips, 1984). Parsons (1982) çocukların yeteneklerine ilişkin benlik kavramlarını geçmiş başarılarından çok, gelecek ders planları, beklentileri ve değerlerinin belirlediğini ifade etmektedir. Harter (1984) düşük yeterlik algısına sahip çocuk ve ergenlerin zorlu görevlerden kaçındıklarını belirtmektedir. Bu bilgiler ışığında akademik yeterlik algısı yüksek olan öğrencilerin dersleri anlamakta ve katılmakta zorlanmadıkları, okul dışında da derslerle uğraşmaktan zevk aldıkları, akademik ortalamalarını ölçüt almadan başarılarını değerlendirebildikleri, ödevlerini yapabileme potansiyeline sahip ve yeteneklerinin farkında oldukları söylenebilir. Akademik yeterlik sadece bireyin akademik hayatını değil, sosyal ve duyuşsal birtakım özelliklerini de etkilemektedir. Örneğin yapılan bir çalışmada (Ryan, Gheen ve Midgley, 1998) öğrencilerin sahip olduğu akademik yeterlikleri ile yardım aramadan kaçınma arasında negatif bir ilişki bulunmuştur. Bu durum ergenlerin akademik alanda kendilerini yeterli bulmalarının kişisel ve sosyal gelişimleri üzerinde de olumlu etkileri olabileceğini ortaya koymaktadır. Bu bağlamda da sosyal ve psikolojik bir takım risklerden ya da olumsuzluklardan ergenleri korumaktadır.

Yeterliğin diğer bir boyutu olan sosyal yeterlik ise, bir bireyin sosyal çevresi ile başa çıkmadaki genel eğiliminin göstergesi olarak anlaşılabilir. Ayrıca araştırmacılar akademik ve sosyal yeterliliğin birbirini tamamlayan iki faktör olduğunu belirtmektedirler (Wentzel, 1993). Sosyal yeterlik hem bireyin edindiği bilgileri kullanarak uygun davranışı seçmesi hem de bunları değişen sosyal çevrede kişisel amaçlarına yönelik kullanmasından oluşmaktadır. Bu nedenle bireyler her yerde herkes tarafından yeterli görülmebilirler. Başka deyişle, bireyler belli etkinlikleri yapmakta ya da belli çevrelerde performans göstermede yeterlidirler ve başka alanlarda yetersizlik gösterebilirler. Bu noktada kişinin kendi yeterliğine ilişkin algısı önem kazanmaktadır. Sosyal yeterlik, birey ve diğerleri arasındaki ilişkide gözlenmekle birlikte, bireyin kendisine ilişkin farkındalığıyla da ilişki içerisindedir. Bu özelliklerin çocukluk ve ergenlik döneminde kazanılmış olması önemli görülmektedir. Aşağıda sosyal yeterliğe ilişkin literatürde yer alan farklı tanımlara ve açıklamalara yer verilmiştir.

Sosyal yeterliğin basit bir tanımı yapılacak olursa; sosyal yeterlik, olumlu sonuçlar elde ederek çevreyi kontrol eden becerilerin toplamıdır (Connolly ve Bruner, 1974; Foster ve Ritchey, 1979; Gresham, 1981; Hops, 1983; Shure, 1981; akt. Blechman ve diğ., 1985). Sosyal yeterlik, bir yandan taleplere karşı esnek, uyumlu tepkiler geliştirebilme bir yandan da içinde bulunduğu çevrede fırsatlar yaratabilme ve onlardan yararlanabilme becerisi (Waters ve Sroufe, 1983) olarak tanımlanırken Rose-Krasnor (1997) sosyal yeterliği, kısa ve uzun vadeli gelişimsel ihtiyaçları gidermek üzere geliştirilmiş

organize davranışlar bütünü olarak tanımlamaktadır (akt. Evirgen, 2010). Başka bir açıklamada da sosyal yeterlik, diğer insanların duygularını anlama, yetişkin müdahalesi olmadan yaşlılarıyla problemlerini çözebilme, ihtiyaç ve duygularını uygun bir şekilde ifade edebilme yeteneği olarak ifade edilmektedir (Kellam ve diğ., 1975; Macgowan, Nash ve Fraser, 2002; Werthamer-Larsson, Kellam ve Wheeler, 1991; akt. Williams 2004). Sosyal olarak yeterli olan çocukların iyi düzeyde sözel becerileri, geçici durumlara uyum sağlama ve duygu düzenleme becerilerine sahip oldukları da belirtilmektedir (Mendez, Fantuzzo ve Cicchetti, 2002; akt. Oades-Sese, Esquivel, Kaliski, ve Maniatis, 2011). Owens ve Johnston-Rodriguez (2010) ise, sosyal yeterliği, ergenlerin okulda sosyal etkileşimde başarılı olabilmeleri için gerekli olan sosyal, duygusal, bilişsel beceriler ve davranışlar olarak tanımlamaktadırlar.

Semrud-Clikeman'a (2007) göre sosyal yeterlik, bir durumla ilgili olarak başkasının bakış açısını alabilme becerisidir. Sosyal yeterlik kavramı sosyal beceriler, sosyal etkileşim ve kişilerarası iletişimi kapsar iken sosyal beceriler, tekrarlanabilen ve amaç yönelimli davranışları varsayar ve sosyal etkileşim, sözel olmayan iletişim ve diğer kişiyle olan etkileşimini ifade eder. Harvilchuck-Laurenson (1997) de buna benzer olarak sosyal yeterliğin bir bireyin amaçlarına ulaşma ya da bunun için çaba gösterme yeteneğine işaret ettiğini ve kişinin sosyal yeterlik düzeyinin, büyük ölçüde sosyal problem çözme becerileri ve algılanan sosyal destek özelliklerine dayandığını ifade etmiştir. Sosyal problem çözme ise, birtakım yetenekleri gerektirir. İlk olarak birey, problemleri potansiyel olarak kendi kontrolünde görmelidir (Rhodes ve Jason, 1990). Daha sonra birey, problemleri çözmek için farklı çözüm yolları düşünebilmelidir. Bireyin sonuçları değerlendirebilmesi ve kendi özellikleriyle paralel olarak çözüm adımlarını belirlemesi de bu süreç dahilindedir.

Kişilerarası görevleri başarmada ve diğerlerinin yanıtlarını işlemede yetenekli olma (Weinstein 1969); kültürel olarak tanımlanmış görevleri yerine getirmede yetenekli olma (Ogbu 1981); eş zamanlı olarak zamanla ve düzeni bozmadan diğ.,yle olumlu ilişkilerini korurken sosyal etkileşim içinde kişisel hedeflerini başarma (Robin ve Rose-Krasnor, 1992); belirli sosyal beceriler temelinde başkalarıyla etkili bir şekilde etkileşim içinde olma (Baron ve Markman, 2000); başkalarıyla sosyal olarak kabul edilen ve değerli olan, aynı zamanda kişisel olarak yararlı veya ilk olarak başkalarına yararlı olan özel yollarla verilen sosyal bir bağlam içinde etkileşim içinde olabilme yeteneği (Ryan ve Shim, 2006; akt. Pavliga, 2008) sosyal yeterliğin göstergeleri olarak tanımlanmaktadır. Goldberg (1977) sosyal yeterlik gelişiminin başlangıcıyla ilgili bir model önermiştir. Bu modele göre; çocuğun akranlarıyla ve yetişkin bakım verenlerle olası deneyimler yaratmadaki becerisi, sosyal yeterlik gelişiminin merkezidir. Goldberg 'olasılık' (contingency) kavramını çocukların diğerlerine olan tepkilerini okuyabilmesi, okunabilir ve tahmin edilebilir davranışsal ipuçlarını görebilmesi, diğerlerinden tepki alması (ilgi, yardım ve oyun) ve tatmin edici duyguları ve deneyimleri başlatması olarak tanımlamaktadır (akt. Carson, Wagner ve Schultz, 2001).

Bunlara ek olarak daha kapsamlı bir açıklama (Craig, 2000; akt. Evirgen, 2010) yapılmış ve sosyal yeterliği yüksek olan kişilerin, sosyal değerlerin benimsenmesi (ilgi gösterme, yardımseverlik, dürüstlük, eşitlik, sosyal adalet, sorumluluk); kişiler arası beceriler (karşılıklı sosyal destek, işbirliği içeren ilişkiler, duygu düşünce ihtiyaç temelinde iletişim kurma, sosyal ortama uygun davranışlar sergileme, empati, çatışmalara uzlaşmacı bir tavır sergileme); kendini düzenleme (dürtü kontrolü, kendini gözlemleme, zevk veren durumları erteleyebilme); planlama ve karar verme becerileri (seçimler yapabilme, problemleri çözebilme, plan yapabilme, planları uygulayabilme, sosyal hedeflere uygun harekete geçebilme); kültürel yeterlik (farklı ırk ve etnik kökenlere saygı gösterme, onlarla iletişime geçebilme, sosyal adalete uygun davranma, haksız davranışların farkında olma); olumlu kimlik duygusu kazanımı (kişisel güç, yeterlik duygusu, amaca sahip olma, kişisel değerlere sahip olma, olumlu bir gelecek beklentisi) gibi becerileri taşımaları gerektiğini ifade etmişlerdir.

Araştırmanın Amacı

Yukarıda verilen bilgiler ışığında sosyal yeterlik algısı yüksek olan ergenlerin kolay ilişki kurabilen ve sürdürebilen, empatik olan, diğer insanlara karşı duyarlı, yakın ilişkileri olan ve bunları sürdürebilen, duygularının farkında olan, sözel ve sözel olmayan mesajlara duyarlı olan, kendini kabul düzeyleri

yüksek olan, sosyal becerilerde yeterli olan vb. özelliklere sahip olan gençler oldukları söylenebilir. Sonuç olarak, olumlu nitelikler olarak ele alınan akademik ve sosyal yeterlik algısı, ergenleri birçok risk faktörüne karşı da koruyucu bir nitelik taşımaktadır. Bu bağlamda yeterlik özelliğinin anlaşılması önemli görülmektedir. Türkiye'deki literatüre bakıldığında bireylerin yeterlik algılarını ölçmeye yarayan ölçeklerin sınırlı sayıda olduğu görülmüştür. Bunlardan biri olan Öz-Etkililik-Yeterlik Ölçeği, Gözüm ve Aksayan (1999) tarafından öğretmenler üzerinde Türkiye'ye uyarlanmıştır. Ölçek yetişkinlere uygulanabilmekte ve bireyin genel öz-etkililik-yeterlik algısını ölçmek amaçlanmaktadır. Aynı amaçla uyarlanan Genel Özyeterlik Ölçeği de 18 yaş üstü bireylere uygulanabilen bir ölçektir (Yıldırım ve İlhan, 2010). Aypay (2010) tarafından uyarlanan Genel Özyeterlik Ölçeği de diğerleri gibi benzer şekilde yetişkinlere uygulanmaktadır. Bu ölçeklerin yanı sıra bireylerin fen, bilgisayar, matematik gibi tek bir alana (Işıksal ve Aşkar, 2003; Akkoyunlu, Orhan ve Umay, 2005; Ekici, 2005; Günhan ve Başer, 2007) veya öğretmenlik gibi tek bir mesleğe (Çapri ve Kan, 2006) yönelik yeterlik algılarını ölçmeye yarayan ölçeklerin uyarlandığı veya geliştirildiği görülmektedir. Bununla birlikte ergenlerin akademik ve sosyal yeterliklerini ölçen bir ölçeğe ise rastlanmamıştır.

Bu bilgiler ışığında bu çalışmada geliştirilen Yeterlik Algısı Ölçeği'nde (YAÖ) bireylerin kendilerine yönelik algılarını içeren maddelere yer verilmiş ve YAÖ'nin kuramsal yapısı Achenbach ve Edelbrock'un (1979) açıklamalarına dayanarak oluşturulmuştur. Bu çalışmada ergenlik dönemini kapsayan lise öğrencilerine yönelik olarak bu özelliklerin ölçülebilmesi amacıyla Likert tipi bir ölçeğin geliştirilmesi, bu yolla literatüre ve eğitim alanındaki uygulayıcıların çalışmalarına katkı sağlanması amaçlanmıştır.

YÖNTEM

Çalışma Grupları

Yeterlik Algısı Ölçeğini geliştirme çalışmaları 1) ölçeğin hazırlandığı, 2) ölçek revizyonunun yapıldığı ve 3) test – tekrar test güvenilirliğinin incelendiği olmak üzere üç çalışma grubu üzerinde yapılmıştır. Aşağıda bu üç çalışma grubuna ait demografik bilgiler verilmiştir.

Çalışma grubu 1: YAÖ'nin geliştirilmesi

Yeterlik Algısı Ölçeği taslak formu, 2011-2012 eğitim ve öğretim yılında Ankara ili Milli Eğitim Müdürlüğü'ne bağlı üç lisede toplam 372 öğrenciye uygulanmıştır. Öğrencilerin seçiminde uygun örnekleme yöntemi kullanılmıştır. Öğrencilerin sınıf düzeylerine ve cinsiyetlerine ilişkin bilgiler Tablo 1'de verilmiştir.

Tablo 1. Katılımcıların Demografik Özellikleri

| Kişiler | 9.Sınıf | 10.Sınıf | 11.Sınıf | 12.Sınıf | Toplam |
|---------|---------|----------|----------|----------|--------|
| Kadın | 34 | 69 | 44 | 48 | 195 |
| Erkek | 44 | 59 | 29 | 45 | 177 |
| Toplam | 78 | 128 | 73 | 93 | 372 |

Tablo 1'de görüldüğü gibi, katılımcıların 195'i kız (% 52), 177'si erkektir (% 48). Dokuzuncu (% 21) ve on birinci (% 20) sınıflara devam edenlerin sayısı yaklaşık 70; onuncu sınıfa devam edenlerin 128 (% 34), on ikinci sınıfa devam edenlerin 93'tür (% 25). Tablodan da izlendiği gibi, öğrencilerin sınıf ve cinsiyetleri bakımından yeterli düzeyde çalışma grubunda temsil edildikleri söylenebilir. Örneklem büyüklüğünün belirlenmesinde faktör yükleri ve faktörlerin maddelerdeki açıkladıkları toplam varyanslar (communality) ve ölçekteki madde sayısı göz önünde bulundurulmuştur (Costello ve Osborne, 2005; Winter, Dodou ve Wieringa, 2009; MacCallum, Wideman, Zhang ve Hong, 1999; Stevens, 2009; Tabachnick ve Fidell, 2007).

Çalışma grubu 2: YAÖ'nin gözden geçirilmesi (revizyon)

Yeterlik Algısı Ölçeği'nin gözden geçirilmiş versiyonu, 2012-2013 eğitim ve öğretim yılında Ankara ili Milli Eğitim Müdürlüğü'ne bağlı ve ilk çalışmaya katılan okullardan ayrı, dört lisede toplam 668 öğrenciye uygulanmıştır. Araştırma grubuna katılan öğrencilere ait bilgiler Tablo 2'de verilmiştir.

Tablo 2. Katılımcıların Demografik Özellikleri

| Kişiler | 9.Sınıf | 10.Sınıf | 11.Sınıf | 12.Sınıf | Toplam |
|---------|---------|----------|----------|----------|--------|
| Kadın | 82 | 105 | 79 | 94 | 360 |
| Erkek | 70 | 104 | 74 | 60 | 308 |
| Toplam | 152 | 209 | 153 | 154 | 668 |

Tablo 2'de görüldüğü gibi, katılımcıların 360'ı kız (% 53.9), 308'i erkektir (% 46.1). Farklı sınıflara devam edenlerin oranları % 22.8 ile % 31.3 arasındadır. Bu gruptan elde edilen veriler rastgele yöntemle ikiye bölünmüş, 147 kişilik bir grupta (83 kız, 64 erkek) betimleyici faktör analizi; geriye kalan 521 kişilik grupta (277 kız, 244 erkek) ise doğrulayıcı faktör analizi yapılmıştır.

Çalışma grubu 3: YAÖ Test tekrar test güvenilirliği

Test tekrar test güvenilirliği, 2011- 2012 eğitim ve öğretim yılında Ankara ili Mamak ilçe Milli Eğitim Müdürlüğüne bağlı bir liseden seçilen üç sınıfta yapılmıştır. Veriler iki hafta ara ile toplanmıştır. Uygulamaya katılanların 38'i kız, 37'si erkektir. Ayrıca öğrencilerin 53'ü lise iki, 22'si ise lise üçüncü sınıfa devam etmektedir.

Veri Toplama Aracı

Bu çalışmada veri toplama aracı olarak çalışmada geliştirilen Yeterlik Algısı Ölçeği ile Kişisel Bilgi Formu kullanılmıştır. Makalenin genelinde Yeterlik Algısı Ölçeği'ne ilişkin bilgiler sunulmaktadır. Araştırmacılar tarafından geliştirilen Kişisel Bilgi Formu aracılığıyla çalışma gruplarına ait demografik veriler elde edilmiştir.

Verilerin Analizi ve İşlem Yolu

Yeterlik Algısı Ölçeği rasyonel yaklaşıma dayalı olarak iki aşamada geliştirilmiştir. İlk aşamada literatür taraması yapılmış, madde havuzu meydana getirilmiştir. Araştırma ekibi yeterlik algısı literatürüne uygun maddeleri bireysel olarak hazırlamış, daha sonra bu ekip bir araya gelerek madde havuzundan aynı özelliğe yönelik en uygun ifadeleri seçerek ya da ifade düzeltmesi yaparak 43 maddeden oluşan taslak formu oluşturmuşlardır. Bu form beşli Likert derecelendirmesine uygun şekilde ("Tamamen Uygun", "Uygun", "Kısmen Uygun", "Uygun Değil", "Hiç Uygun Değil") düzenlenmiştir. Daha sonra Psikolojik Danışma ve Rehberlik alanından üç, Eğitimde Ölçme ve Değerlendirme alanından bir uzmandan maddelerin uygunluğu konusunda görüş alınmıştır. Uzmanların önerileri doğrultusunda forma üç yeni madde daha eklenerek, 46 maddelik form ifadelerin anlaşılabilirliğini ve yanıtlanmanın kolaylığını kontrol etmek amacıyla 26 lise öğrencisine uygulanmıştır. Maddelerin öğrenciler tarafından kolaylıkla anlaşılabilirdiği görülmüş, ardından uygulamaya geçilmiştir. Çalışmanın birinci grubu olan 372 öğrenciden elde edilen ölçümlerin geçerliği betimleyici (exploratory) faktör analizi yapılarak incelenmiştir. Ölçümlerin güvenilirliği ise iç tutarlık katsayıları (Cronbach alfa ve tabakalanmış alfa) hesaplanarak test edilmiştir. 372 kişilik veri setinde, betimleyici faktör analizi yapılmadan önce normallik, doğrusallık, aykırı gözlem, çoklu bağlantı, örneklem büyüklüğü ve eksik değer varsayımları gözden geçirilmiş, verilerin bu varsayımları karşıladığı gözlemlenmiştir (Costello ve Osborne 2005; Winter, Dodou ve Wieringa, 2009; Fabrigar, Wegener, MacCallum ve Strahan, 1999; Hair, Black, Babin ve Anderson, 2010; Kline, 2010;

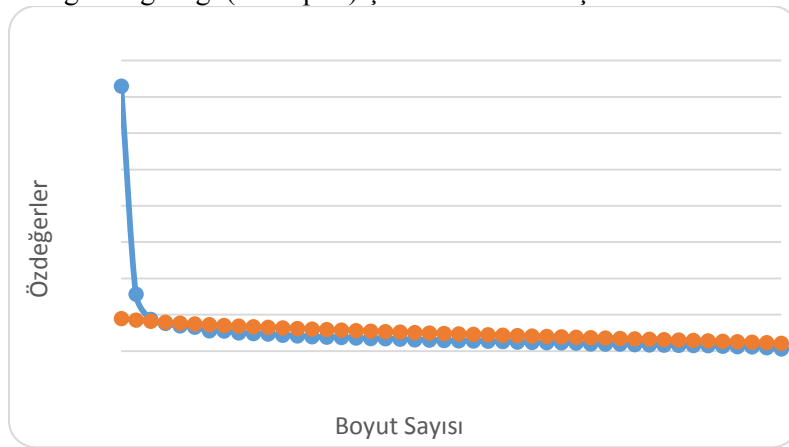
MacCallum, Wideman, Zhang ve Hong, 1999; Pedhazur ve Schmelkin, 1991; Stevens, 2009; Tabachnick ve Fidell, 2007). Madde puanlarının hiçbirinde eksik değer oranı %5'i aşmadığı; eksik değerlerin maddeler ve kişiler arasında rasgele dağıldıkları görülmüştür. Madde puanlarının her birinde (univariate) normallik, eğiklik ve basıklık katsayılarıyla ve P-P grafiklerle; maddelerin doğrusal bileşimlerindeki (multivariate) normallik Mardia'nın asimetric eğiklik ve basıklık katsayılarıyla incelenmiştir (Kline, 2010). Maddeler arasındaki doğrusallık rasgele seçilen maddeler arasında çizilen saçılma grafikleriyle değerlendirilmiştir. Madde dağılımlarının her birinde aykırı değerler (outlier) z puanlarıyla; doğrusal bileşimlerindeki aykırı puanlar Mahalanobis Uzaklık Testiyle (χ^2) gözden geçirilmiştir. Bu inceleme sonunda 9 univariate, 3 de multivariate aykırı değer veri setinden çıkarılmıştır. Ayrıca univariate çoklu bağlantı (Collinearity) Pearson korelasyon; multivariate çoklu bağlantı (singularity) Tolerans ve VIF değerleri incelenerek araştırılmıştır. Madde puanları 1-5 arasında derecelendirildiğinden, betimleyici faktör analizi polychoric korelasyon katsayıları üzerinde paralel ve map testler yoluyla principal axis faktör elde etme yöntemi (multivariate normallik sayıltısı ihlal edildiğinden (Fabrigar, Wegener, MacCallum ve Strahan, 1999), direct oblimin eksen döndürme tekniği kullanılarak gerçekleştirilmiştir.

İlk aşamada betimleyici faktör analizi sonucu, madde kompleksliği, düşük faktör yükleri ve maddenin ölçtüğü özellik bakımından ait olması beklenen boyutta yer almaması nedenleriyle taslak formda yer alan 46 maddeden 14'ü çıkarılmıştır. Bu işlem sırasında kapsam geçerliği göz önünde bulundurularak, sosyal boyutta yer alan, özellikle faktör yükü düşük sekiz madde daha net ifade edilmiştir. Yeterlik Algısı Ölçeği, ikinci aşamada 668 öğrenciden oluşan, ilk uygulamadaki katılımcılara benzer başka bir örnekleme uygulanmıştır. İlk uygulamadan elde edilen sonuçlara dayalı olarak sonradan düzenlenen yeni maddelerle ölçeğin faktör yapısının kısmen değişmiş olabileceği düşünülmüş, bu örneklemeden rastgele seçilen 147 gözlem (yaklaşık madde sayısı * 5 kişi) üzerinde betimleyici faktör analizi tekrarlanmıştır. Betimleyici faktör analizinden elde edilen bulguların genellenebilirliği ise kalan 521 kişilik veri setinde doğrulayıcı faktör analizi yapılarak test edilmiştir. Verilerin çözümlenmesinde Factor 8.02 ve Lisrel 8.70 paket programları kullanılmış, araştırmada hata payı .05 olarak alınmıştır.

BULGULAR

YAÖ'nin Geliştirilmesi

Yeterlik Algısı Ölçeği taslak formu oluşturulurken sosyal ve akademik yeterlik alanlarına ilişkin maddeler yazılmış ve başlangıçta ölçeğin iki boyutlu olabileceği öngörülmüştür. Boyut sayısı sınırlanmadan yapılan temel bileşenler analizi sonucunda, gerçek verilerden elde edilen özdeğerlerin mavi çizgi, paralel analiz sonucunda elde edilen random özdeğerlerin ise kırmızı çizgi ile gösterildiği özdeğerler grafiği (scree plot) Şekil 1'de verilmiştir:



Şekil 1. Scree Plot

Şekil 1’de, gerçek veri setinde I. özdeğer 15.49, II. özdeğer 3.13 olup random sayılarla üretilen özdeğerlerden yüksektir. Gerçek veri setinde özdeğeri birden büyük (Kaiser kuralı) dokuz boyut bulunmasına karşın, 3. boyutta yaklaşık olarak gerçek ve random özdeğerler aynı düzeydedir. Benzer şekilde, Minimum Avarage Partial Test (MAP) sonucunda ortalama kısmi değerler I. boyut için .011, II. boyut için .007 olarak elde edilmiş, scree plot ve paralel analiz sonuçlarında olduğu gibi 2 boyut önerilmiştir. Buna göre, hem scree plot ve paralel analiz hem de MAP test sonuçlarının, öngörülen boyut sayısını destekledikleri kabul edilmiştir. Bunun üzerine, principal axis faktör elde etme yönteminin ve direct oblimin eksen döndürme tekniğinin kullanıldığı betimleyici faktör analizi yapılmış, Kaiser-Meyer-Olkin (KMO) örnekleme yeterlik değeri .94 olarak önerilen kesme noktasından (.60) yüksek bulunmuştur. Benzer şekilde, Bartlett küresellik testi de istatistiksel bakımdan önemlidir ($\chi^2_{1035} = 7306.2$, $p = .00$ bulunmuştur. Madde–faktör korelasyonlarının (yük) belirlenmesinde, çeşitli kaynaklarda verilen kesme değerlerinin ortalaması alınarak, ölçekteki madde sayısının yanı sıra örneklemden kişi sayısı da göz önünde bulundurularak, önemli yüklerin ait oldukları boyutta .30’dan büyük, diğer boyutta ise .25’ten küçük olması kriteri benimsenmiştir (Hair, Black, Babin ve Anderson 2010; Stevens 2009; Tabachnick ve Fidell 2007; akt. Raunberheim, 2004). Buna göre, I. boyutta 24 (yükler: .40 ile .74); II. boyutta sekiz (yükler: .35 ile .54) olmak üzere 32 maddenin ölçekte kalması gerektiği gözlemlenmiştir. Bu iki boyutun, toplam varyansın %38’ini açıkladıkları görülmüştür.

Güvenirlilik çalışması

Faktör analizi sonucunda elde edilen boyutlar için iç tutarlık katsayıları (Cronbach alfa) hesaplanmıştır. I. boyutun iç tutarlık katsayısı .94, ortalaması 41.46, standart sapması 14.21’dir. Madde-toplam korelasyon katsayıları .37-.63 arasındadır. Ortalama madde –madde korelasyonu .43’tür. II. boyutun iç tutarlık katsayısı .82, ortalaması 24.26, standart sapması 6.49’dur. Madde - toplam korelasyonu .35 -.54 arasındadır. Ortalama madde –madde korelasyonu .31’dir. Ölçeğin bütünü için tabakalanmış alfa güvenirlilik katsayısı ise .94 olarak hesaplanmıştır.

YAÖ’nin Gözden Geçirilmesi (Revizyon Çalışması)

Betimleyici faktör analizi sonuçları

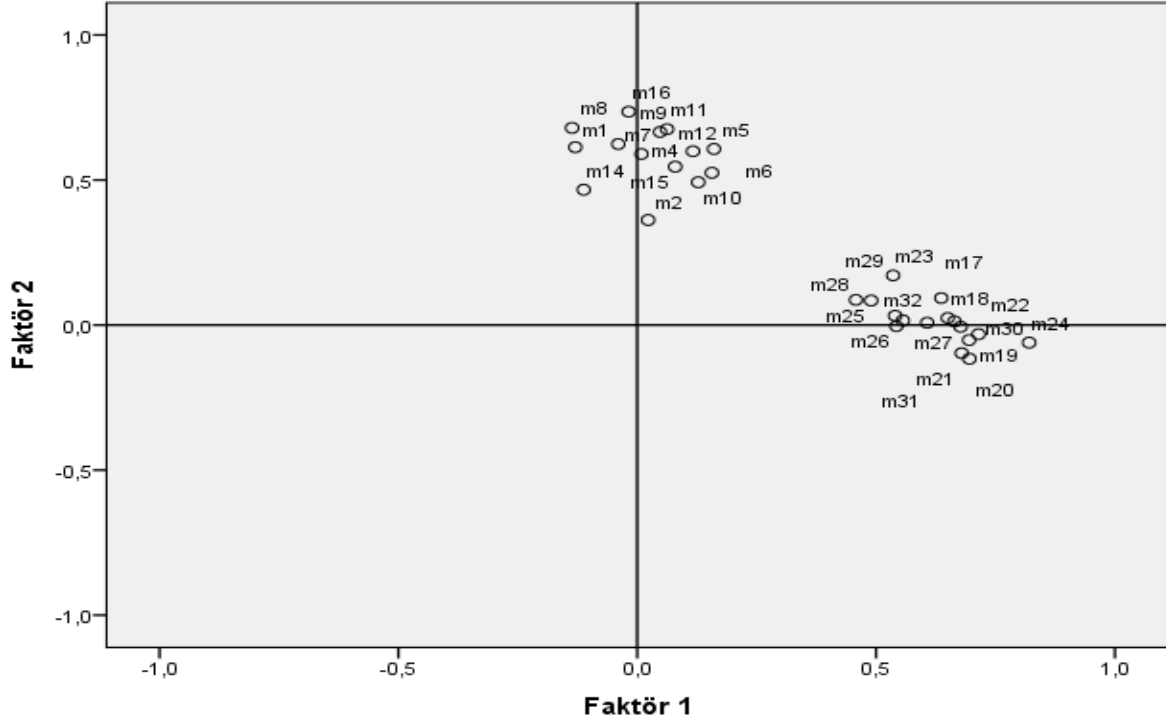
İlk çalışmada, Yeterlik Algısı Ölçeği için madde havuzu oluşturulmuş, bu havuz kullanılarak maddeler arasındaki ilişkiler, betimleyici faktör analizi ile incelenmiştir. Yapılan ilk çalışmada, her iki boyut için yaklaşık eşit sayıda, toplam 46 maddeden oluşan bir değişken seti kullanılmış ancak I. faktörde 25, II. faktörde ise 8 madde .30’dan büyük yüke sahip olarak bulunmuştur. Her ne kadar bu haliyle maddeler kuramsal olarak öngörülen boyutlarla yüksek korelasyonlara sahip olsalar da, boyutların madde sayılarında büyük oranda dengesizlik görülmüş, ayrıca sosyal ve akademik boyutlar arasında kuramsal olarak beklenenden daha yüksek korelasyonlar elde edilmiştir. Bu nedenle ölçek geliştirme çalışmasının ikinci aşamasında, bazı maddeler yeniden ifade edilip, yeni maddeler de eklenerek akademik yeterlik boyutu tekrar düzenlenmiştir. İlk çalışmada yüksek yüke sahip sosyal yeterlik maddeleriyle birlikte oluşturulan 32 maddelik yeni form, ilk uygulamadaki örnekleme benzer özellikler taşıyan, ancak önceki çalışma gruplarından başka 668 kişilik bir örnekleme uygulanmıştır. İlk çalışmada, Yeterlik Algısı Ölçeğinin sosyal ve akademik olmak üzere iki yeterlik alanından oluştuğu ortaya konduğundan, ikinci aşamada ilk olarak örneklemin bir bölümünde ($n = 147$) iki boyutlu betimleyici faktör analizi yapılmıştır. Örneklemin geriye kalanı ($n = 521$) için ise doğrulayıcı faktör analizi uygulanmıştır. Betimleyici faktör analizi sonucunda $KMO = .87$; Bartlett küresellik testi $\chi^2_{496} = 2925.94$, $p = .000$ olarak elde edilmiştir. Bu katsayılar maddeler arasındaki ilişkinin faktör analizi için uygun olduğunu göstermektedir. Faktör analizi sonucunda elde edilen faktör yükleri Tablo 1’de verilmiştir.

Tablo 3. Yeterlik Algısı Ölçeği Madde – Faktör Yükleri

| Maddeler | Faktör | |
|--|----------|--------|
| | Akademik | Sosyal |
| 1. Yeni insanlar tanımaktan hoşlanırım. | | .61 |
| 2. Duygularımı rahatça ifade edebilirim. | | .37 |
| 3. Başkalarının duygularını kolaylıkla anlayabilirim. | | |
| 4. Sosyal etkinlikler düzenlemede başarılıyım. | | .59 |
| 5. Arkadaşlarım arasında sevilen biriyim. | | .61 |
| 6. Grup çalışmalarında işbirliği yapabilirim. | | .52 |
| 7. İnsanlarla uzun süreli ilişkiler kurabilirim. | | .62 |
| 8. Arkadaşlarımla zaman geçirmekten hoşlanırım. | | .67 |
| 9. Başkalarıyla kolayca iletişim kurabilirim. | | .67 |
| 10. Başkalarının duygularına önem veririm. | | .50 |
| 11. Sıkı dostluklar kurabilirim. | | .68 |
| 12. Arkadaşlarım tarafından etkinliklere davet edilirim. | | .60 |
| 13. Arkadaşlarımın isteklerine gerektiğinde hayır diyebilirim. | | |
| 14. Sinema, tiyatro gibi etkinliklere arkadaşlarımla giderim. | | .46 |
| 15. Arkadaşlarım ihtiyaç duyduklarında bana danışırlar. | | .56 |
| 16. Arkadaşlık ilişkilerime önem veririm. | | .73 |
| 17. Derslerle ilgili soruları yanıtlamakta istekli davranırım. | .63 | |
| 18. Yeni ders konularını zorlanmadan öğrenebileceğimi düşünürüm. | .65 | |
| 19. Başarılı bir öğrenciyim. | .70 | |
| 20. Sınavlara iyi hazırlanırım. | .70 | |
| 21. Ödevlerimi öğrenme hevesiyle yaparım. | .68 | |
| 22. Ders sonrasında da öğrendiklerim üzerinde düşünürüm. | .67 | |
| 23. Derste öğrendiklerimi arkadaşlarımla tartışırım. | .53 | |
| 24. Dersleri dikkatle dinlerim. | .82 | |
| 25. Başkalarının yardımı olmadan ödevlerimi yapabilirim. | .54 | |
| 26. Ödevlerimle gurur duyarım. | .54 | |
| 27. Dersi derste öğrenirim. | .68 | |
| 28. Akademik başarımdan memnunum. | .45 | |
| 29. Derslerde sorulan zor soruları yanıtlayabilirim. | .49 | |
| 30. Derslerde yeni bir şeyler öğrenmeyi önemli bulurum. | .71 | |
| 31. Derslerle ilgili bilgilerimin yeterliğine güvenirim. | .61 | |
| 32. Ders konularını çeşitli kaynaklardan araştırırım. | .56 | |

Madde - Faktör ilişkisi (yük) kesme noktası .36

Tablo 3’de görüldüğü gibi, akademik yeterlik boyutunda yer alan maddelerin yükleri .43 - .75; sosyal yeterlik boyutundaki maddelerin yükleri .37-.65 arasındadır. Stevens (2009) 200 kişilik bir örnekleme, madde- faktör korelasyonları için kesme noktasının .36 olması gerektiğini belirtmektedir. Buna göre, 3. ve 13. maddelere ilişkin faktör yükleri bu sınırın altında kalmaktadır. Bununla birlikte, madde 3 ve 13 için ilk uygulamadaki faktör yükleri sırasıyla .57 ve .54’tür. İki uygulamada faktör yükleri arasındaki bu değişkenlik, örneklem farklılığına bağlanmış ve bu durum bulguların genellenebilirliğini sınırladığından veri setinden çıkarılmaları düşünülmüştür. İki boyutlu yapıya ilişkin açıklanan toplam varyans .37 (özdeğerler sırasıyla 7.99 ve 3.81); söz konusu maddelerin veri setinden çıkarılmasıyla tekrarlanan faktör analizi sonucunda açıklanan toplam varyans .39 (özdeğerler sırasıyla 7.83 ve 3.76) bulunmuştur. m3 ve m13 veri setinden çıkarıldıklarında, sosyal yeterlik boyutunda yer alan maddeler için madde toplam korelasyonları .30 - .67 arasında olup, iç tutarlık katsayısı .87’dir. Bu katsayılar, iki maddenin veri setinden çıkarılmasının güvenilirliği azaltmadığını göstermektedir. Bu bulgulardan hareketle ilgili maddeler veri setinden çıkarılmaya karar verilmiştir. Tekrarlanan betimleyici faktör analizine ilişkin madde-faktör ilişkileri Şekil 1’de sunulmuştur.



Şekil 2. Madde – Faktör Yükleri Saçılma Grafiği

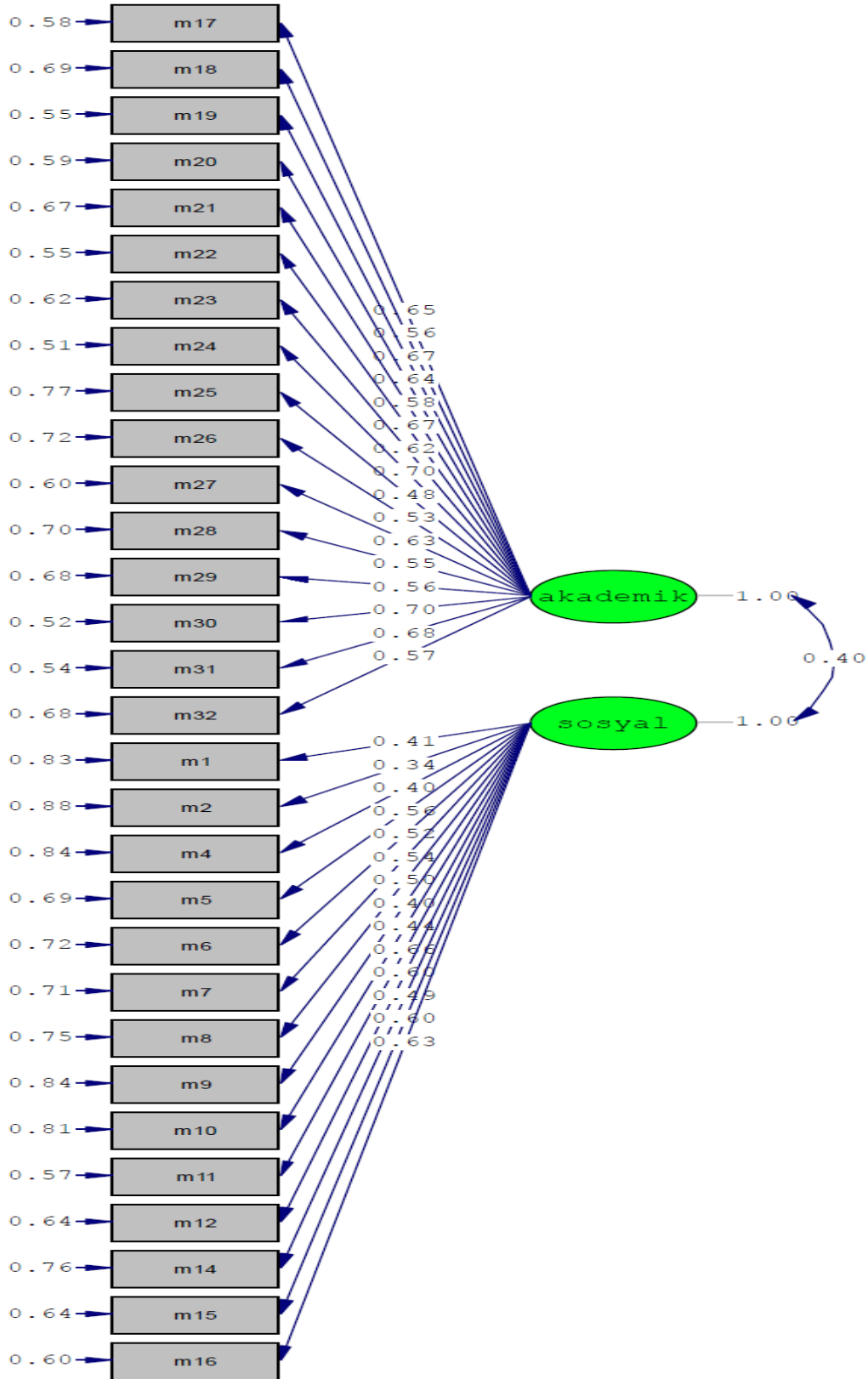
Şekil 2’de görüldüğü gibi, maddeler eksen çizgileri üzerinde kümelenmektedir. Bu durum boyutlar arasındaki ilişkilerin düşük olduğunu göstermektedir (faktörler arası $r = .30$). Ayrıca maddeler orijin noktasından (0,0) uzakta yer almaktadır. Bu da maddelerin ait oldukları faktörde yüksek yüke sahip oldukları anlamına gelmektedir. Bu sonuçlar ışığında, Yeterlik Algısı Ölçeği akademik boyutta 16 madde, sosyal boyutta 14 madde olmak üzere toplam 30 maddeden oluşturulmuştur.

Güvenirlilik çalışması

Bir bütün olarak ölçeğin iç tutarlık katsayısı (tabakalanmış alfa) .91 olarak hesaplanmıştır. Akademik yeterlik boyutunda yer alan maddeler için madde toplam korelasyonları .47-.73 arasında olup, iç tutarlık katsayısı (Cronbach alfa) .90’dır. Sosyal yeterlik boyutunda yer alan maddeler için madde toplam korelasyonları .11 - .67 arasında olup, iç tutarlık katsayısı (Cronbach alfa) .86’dır.

Doğrulayıcı faktör analizi sonuçları

Betimleyici faktör analiziyle elde edilen sonuçların benzer başka bir örnekteki ($n=521$) geçerliği doğrulayıcı faktör analizi yapılarak incelenmiş, sonuçlar Şekil 3’te verilmiştir.



Şekil 3. İki Boyutlu Yeterlik Algısı Ölçeği Doğrulayıcı Faktör Analizi Sonuçları

Şekil 3'te de görüldüğü gibi, standartlaştırılmış regresyon katsayıları akademik yeterlik için .48-.70; sosyal yeterlik için .34 - .66 arasındadır. I. boyutun açıkladığı ortalama varyans (AVE) .38; II. boyutun

açıkladığı .25'tir. I. boyutta standartlaştırılmış regresyon ağırlıkları madde 25 dışında > .50 iken, II. boyutta 6 maddenin standartlaştırılmış ağırlığı .50'nin altındadır. Bileşik (composite) güvenirlik I. boyut için .91; II. boyut için .83 olup, her ikisi de kendi AVE değerlerinden daha yüksektir. AVE değerleri < .50 olmasına karşın, Kline (2010) .30 civarındaki standartlaştırılmış ağırlıkların orta; .50 ve üzerindeki ağırlıkların ise büyük etki büyüklüklerini yansıttığını belirtmektedir. Buna göre, her iki boyutta yer alan maddeler, ait oldukları boyutların oldukça geçerli göstergeleridir. Dolayısıyla bu sonuçlar benzeme (convergent) geçerliğinin kısmen karşılandığını göstermektedir. Ayrıca faktörler arasındaki varyans .16 olup AVE değerlerinden düşüktür. Buna ek olarak, sosyal ve akademik yeterlik boyutları arasındaki korelasyon .40 olup, bu katsayıya ilişkin standart hata .04'tür. Buna göre, iki değişken arasındaki ilişkinin güven aralığı $.40 \pm 1.96 \cdot .04 = .40 \pm .08 = .32 - .48$ 'dir. Bu aralığın 1'i kapsamaması, Yeterlik Algısı Ölçeği'nin tek boyutlu bir yapıya sahip olmadığı, başka deyişle ayırma (discriminant) geçerliğine sahip olduğu düşüncesini desteklemektedir. Bu açıklamalardan ayrı olarak, modelin genel uyum katsayıları, hipotez modelin kabul edilebilir düzeyde uyuma sahip olduğunu göstermektedir ($\chi^2_{404} = 1873.78$, $p = .00$, Normed $\chi^2 = 4.64$; CFI = .92; RMSEA = .08 (.07-.09); SRMR = .07). Büyük örneklerde χ^2 testi önemli bulunduğundan, ek uyum istatistikleri verilmektedir. χ^2 'nin serbestlik derecesine bölünmesiyle elde edilen Normed χ^2 testinin 2-5 arasında olması (Tabachnick ve Fidell 2007); CFI'nin .90'dan büyük, RMSEA'nın da .07'den küçük olması istenmektedir (Bollen 1989; Hair, Black, Babin ve Anderson 2010; Kline 2010; MacCallum, Browne ve Sugawara 1996). Dolayısıyla bu görüşler, modele ilişkin katsayıların yeterli olduğuna kanıt sayılabilir.

Güvenirlik çalışması

Doğrulamalı faktör analizinin uygulandığı veri setinde iç tutarlık katsayıları (Cronbach alfa) sosyal yeterlik için .82, akademik yeterlik için .90, testin bütünü için tabakalanmış alfa .91 bulunmuştur. Madde toplam korelasyonları sırasıyla .37 - .53 ve .46 - .65 arasındadır.

Bu çalışmadan ayrı olarak, test tekrar test güvenirlik katsayıları sosyal yeterlik için (r) .79, akademik yeterlik için (r) .82 olarak bulunmuştur. Bununla birlikte, her iki ölçüm de ölçme hataları içermektedir. Buna göre, test tekrar test güvenirlikleri için düzeltilmiş korelasyon katsayıları (r) sosyal yeterlik için .94; akademik yeterlik için .90'dır. I ve II. uygulamada sosyal yeterlik için iç tutarlık katsayıları .82 ve .86; akademik yeterlik için .90 ve .93'tür.

SONUÇ ve TARTIŞMA

Bu çalışmada lise öğrencilerinin akademik ve sosyal yeterlik algısını ölçmek amacıyla Yeterlik Algısı Ölçeği geliştirilmiştir. Ölçekten elde edilen ölçümlerin geçerlik çalışmaları için betimleyici ve doğrulamalı faktör analizi; güvenirlik çalışması için ise Cronbach ve tabakalı alfa iç tutarlılık katsayıları ile test – tekrar test yöntemleri kullanılmıştır.

Yeterlik Algısı Ölçeği taslak formu oluşturulurken sosyal ve akademik yeterlik alanlarına ilişkin maddeler yazılmış ve başlangıçta ölçeğin kuramsal açıklamalara dayalı olarak (Achenbach ve Edelbrock, 1979) iki boyutlu olabileceği öngörülmüştür. İlk araştırma grubundan (n=372) elde edilen veriler üzerinde çeşitli faktör belirleme yöntemlerinin (özdeğer grafiği, paralel ve map test vb.) denendiği betimleyici faktör analizi yapılmıştır. Analiz sonuçları öngörülen yapıyla uyumlu iki boyutun, maddeler arası ilişkileri açıklamakta yeterli olduğunu göstermiştir. Bununla birlikte, boyutların madde sayıları arasında büyük oranda dengesizlik (25'e 11) görülmüş, ayrıca her iki boyut arasında kuramsal olarak beklenenden daha yüksek korelasyonlar elde edilmiştir. Bu gerekçe ile ölçek geliştirme çalışmasında revizyona gidilmiştir. Revizyon aşamasında akademik yeterlik boyutunda bazı maddeler yeniden ifade edilip, yeni maddeler de eklenerek YAÖ araştırmacılar tarafından tekrar düzenlenmiştir. Daha sonra, 32 maddeden oluşan yeni form kullanılarak geçerlik ve güvenirlik çalışmaları tekrarlanmıştır. YAÖ'nin yapısı değiştiğinden, ikinci araştırma grubundan (n=668) elde edilen veriler ikiye bölünmüş; bu gruplardan ilkinde (n=147) betimleyici faktör analizi tekrarlanmış, ikincisinde ise (n=521) doğrulamalı faktör analizi yapılmıştır. Her iki analiz sonucunda, kuramsal

yapının veri setindeki gözlenen ilişkileri yeterince açıkladığı ve modellerin kabul edilebilir düzeyde geçerlik katsayılarına sahip oldukları görülmüştür (Bollen 1989; Hair, Black, Babin ve Anderson, 2010; Kline, 2010; MacCallum, Browne ve Sugawara, 1996; Tabachnick ve Fidell, 2007). Buna göre, YAÖ kullanılarak araştırma gruplarından elde edilen ölçümlerin geçerli oldukları, farklı örneklemlerde akademik ve sosyal yeterlik boyutlarının tutarlı bir şekilde elde edildiği, dolayısıyla bu bulgunun genellenebileceği ifade edilebilir.

Ölçümlerin güvenilirlikleri Cronbach alfa katsayısı ve test-tekrar test güvenilirlik katsayıları hesaplanarak incelenmiştir. Yapılan bu analizlerde elde edilen katsayılar .79 ile .94 arasında değişmiştir ve literatüre bakıldığında güvenilirlik katsayısının ne kadar olması gerektiğinin nasıl kullanılacağına bağlı olduğu bununla birlikte, yordama ve yapı geçerliği çalışmalarında .70 düzeyinde güvenilirliğe sahip olan araçların kullanılmasının uygun olduğu belirtilmektedir (Nunnally ve Berstein, 1994). Bu bağlamda YAÖ'den elde edilen ölçümlerin güvenilir olduğu ileri sürülebilir.

Sonuç olarak, 30 maddeden oluşan Yeterlik Algısı Ölçeği, lise öğrencilerinin akademik ve sosyal yeterlik algılarını ölçmek amacıyla kullanılabilir düzeydedir. Ölçekten alınan yüksek puan, öğrencilerin kendilerini yeterli algıladığını, alınan düşük puan ise öğrencilerin yeterlik algılarının düşük olduğuna işaret etmektedir.

Çalışma kapsamında elde edilen bulgular çerçevesinde uygulayıcı ve araştırmacılar, Yeterlik Algısı Ölçeği'nden çeşitli amaçlarla kullanabilirler. Lisede çalışan psikolojik danışmanlar, öğrencilerin sosyal ve akademik yeterlik algılarını belirlemek amacıyla yapacakları bireysel ve grup değerlendirmelerinde ölçeğin tümünden ve alt boyutlarından yararlanabilirler. Ölçek aracılığıyla yeterlik algısı düşük öğrenciler belirlenerek bu öğrencilerin yeterlik algılarını arttırmaya yönelik uygulamalar yapılabilir. Ancak geliştirilen ölçme aracının öğrencilerin yeterliklerine ilişkin kendi algılarını belirlemeye yönelik olduğu, dolayısıyla ortaya çıkan sonuçların öznellik içerebileceği dikkate alınmalıdır. Araştırmacılar tarafından Yeterlik Algısı Ölçeğinin cinsiyete ve sınıf düzeyine göre norm tabloları oluşturulabilir. Ölçeğin geçerlik ve güvenilirlik çalışmaları ortaokul ve üniversite öğrencilerinde, çeşitli özel gruplarda (yetiştirme yurdunda kalanlar, farklı lise türlerinde, psikolojik danışma merkezine başvuran danışanlarda vb.) tekrarlanabilir. Ayrıca, bu çalışmada ölçeğin geçerlik çalışmaları kapsamında açıklayıcı ve doğrulayıcı faktör analizi yöntemlerine başvurulmuştur. Farklı geçerlik yöntemleri kullanılarak ölçeğin geçerliğine ilişkin sağlanan kanıtlar artırılabilir.

KAYNAKÇA

- Achenbach, T. M., & Edelbrock, C. S. (1979). The child behavior profile: II. boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology, 47* (2), 223-233.
- Achenbach, T. M., & McConaughy, S. H. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101* (2), 213-232.
- Achenbach, T., & Zigler, E. (1963). Social competence and self-image disparity in psychiatric and non-psychiatric patients. *Journal of Abnormal and Social Psychology, 67* (3), 197-205.
- Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemporary Educational Psychology, 22*(3), 269-298.
- Blechman, E. A., Tinsley, B., Carella, E. T., & McEnroe, M. J. (1985). Childhood competence and behavior problems. *Journal of Abnormal Psychology, 94* (1), 70-77.
- Bollen, K. A. (1989). *Structural equations with latent variables*. NY: Wiley.
- Carson, D. K., Wagner, B. S., & Schultz, N. W. (2001). Temperament and gender: Correlates of Toddler social competence. *Journal of Genetic Psychology, 148* (3), 289-302.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10* (7), 1-9.
- Evirgen, N. (2010). *Aile içi örüntülerin çocukların algıları açısından incelenmesi*. Yüksek lisans tezi, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4* (3), 272-299.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010) *Multivariate data analysis* (7th ed.). Pearson Company.

- Harter, S. ve Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development*, 55, 1969-1982.
- Harvilchuck-Laurenson, J. D. (1997). *The relationships of stress, social competence, and health status with adolescent academic achievement*. Doctorate thesis, The Johns Hopkins University, Maryland.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Nunnally, J. C. & Berstein, I. H. (1994). *Psychometric Theory*. 3. Edit. McCraw-Hill, Inc. New York.
- Oades-Sese, G. V., Esquivel, G. B., Kaliski, P. K., & Maniatis, L. (2011). A longitudinal study of the social and academic competence of economically disadvantaged bilingual preschool children. *Developmental Psychology*, 47 (3), 747-764.
- Owens, L. A., & Johnston-Rodriguez, S. (2010). Social competence. *International Encyclopedia of Education*. [Çevrim-içi: <http://www.sciencedirect.com/science/article/pii/B9780080448947011593>], Erişim tarihi: 6 Ekim 2012.
- Parsons, J. E. (1982). The development of expectancies, attributions, and persistence. University of Michigan Documentaries.
- Pavliga, G. K. (2008). *Toward a conceptual definition for social competence: An explanatory study*. Doctorate thesis, The University of Akron, Ohio, USA.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.
- Phillips, D. (1984). The illusion of incompetence among academically competent children. *Child Development*, 55, 2000-2016.
- Raubenheim, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, 30 (4), 59-64.
- Rhodes, J. E., & Jason, L. A. (1990). A social stress model of substance abuse. *Journal of Consulting and Clinical Psychology*, 58, 395-401.
- Semrud-Clikeman, M. (2007). *Social competence in children*. USA: Springer Science_Business Media, LLC.
- Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York: Routledge Taylor & Francis Group.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson, Allyn & Bacon.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, 85(2), 357.
- Williams, S. A. (2004). *The academic achievement gap: linkages among behavior, social and academic competencies*. Doctorate thesis, University of North Carolina, Chapel Hill.
- Winter, J. C. F. ve Dodou, D., Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44 (2), 147-181.

EXTENDED ABSTRACT

Introduction

Perceived competence is an important driving force underlying the behavior of individuals. It is stated that competence includes two dimensions as academic and social competence in literature. Academic competence is associated with features such as related not compelling to understand and attend lessons, being pleasure to deal with the lessons outside of school, evaluating the success without criteria of academic averages, expressing his/her potential to be able to do homework, becoming aware of the capabilities as well as school success (Blechman, Tinsley, Carella & McEnroe, 1985; Harter, 1983; Phillips, 1984). On the other hand, social competence is explained with features such as establishing and maintaining relationships easily, being empathic, sensitivity to other people, emotional awareness, sensitivity to verbal and nonverbal messages, self acceptance, being competent social skills (Owens & Johnston-Rodriguez, 2010; Pavliga, 2008; Williams, 2004; Semrud-Clikeman, 2007). Therefore, evaluation of the academic and social dimensions of individuals' perceived competence might be deemed necessary. Based on this need, Perceived Competence Scale (PCS) has

been developed to measure high school students' perceived academic and social competence in Turkey, and the scale development process is explained in this research.

Method

Scale development studies have been conducted on three separate samples. First study has been conducted on 372 high school students in Ankara. Validity of the measurements, obtained from draft form (consisted of 46 items) prepared by the research group, have been examined by using exploratory factor analysis. The reliability of the measurements have been tested through calculating internal consistency coefficients (Cronbach's alpha and stratified alpha). Factor 8.02 package program has been used to analyze the data.

New form of the scale, consisted of 32 items, which was revised with exploratory factor analysis at the first stage and after that new items were added or some items were taken out, has been carried out to 668 students as an other sample. Basing on the idea that factor structure of new form of the scale could change, exploratory factor analysis has been repeated on 147 data that was chosen randomly from this sample. The generalizability of the findings from exploratory factor analysis has been tested with other 521 persons' data set through confirmatory factor analysis.

Results and Discussion

As a result of exploratory factor analysis, KMO=.87, Barlett's sphericity test $\chi^2_{496} = 2925.94$, $p = .000$ has been found. Factor loads of the items in academic competence and social competence dimensions are between respectively as follows: .43-.75, .37-.65. Factor loadings are under .40 two items have been taken out of the data set. After the repeated exploratory factor analysis, perceived competence scale has been formed with 16 items for academic dimension, 14 items for social dimension, and 30 items for the whole scale.

Confirmatory factor analysis has been used in the second study to examine the validity of the results that were found in exploratory factor analysis in an other similar sample (n=521). After all, fit indexes have showed that hypothesis model is in an acceptable level ($\chi^2_{404} = 1873.78$, $p = .00$, Normed $\chi^2 = 4.64$; CFI= .92; RMSEA= .08 (.07-.09); SRMR= .07).

Cronbach's alpha internal consistency coefficient and test-retest methods have been used for reliability analyses. Internal consistency coefficient (stratified alpha) of the scale has been found .91 in the data set which exploratory factor analysis was done with. Item-total correlations for the items in academic competence dimension are between .47 and .73, and internal consistency coefficient (Cronbach's alpha) of the dimension is .90. Item-total correlations for the items in social competence dimension are between .11 and .67, and internal consistency coefficient (Cronbach's alpha) of the dimension is .86. Internal consistency coefficients (Cronbach's alpha) are .82 for social competence, .90 for academic competence, and stratified alpha is .91 for the whole scale in the data set which confirmatory factor analysis was done with. Corrected correlation coefficients (r) for test-retest reliability are .94 for social competence and .90 for academic competence.

As a result, academic competence dimension of the scale has 16 items and social competence dimension has 14 items. Perceived Competence Scale is developed to measure high school students' perceptions about their academic and social competence. If the scores are high, it means students perceive themselves more competent; on the other hand, if the scores are low, it means students' perceptions about themselves less competent.

In the light of this information, Perceived Competence Scale can be used to determine high school students' perceived competence, and assess counseling practices for raising lower perceived competence. In the future, validity and reliability studies can be repeated with adding emotional competence dimension to Perceived Competence Scale.