
Eđitimde ve Psikolojide Ölçme ve Deęerlendirme Dergisi

Journal of Measurement
and Evaluation in
Education and Psychology

ISSN:1309-6575

Kış 2016
Winter 2016

Cilt: 7- Sayı: 2
Volume: 7- Issue: 2



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)

Owner

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

Editör

Prof. Dr. Selahattin GELBAL

Editor

Prof. Dr. Selahattin GELBAL

Yardımcı Editör

Yrd. Doç. Dr. Kübra ATALAY KABASAKAL

Assistant Editor

Assist. Prof. Dr. Kübra ATALAY KABASAKAL

Genel Sekreter

Doç. Dr. Tülin ACAR

Secretary

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

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

Dil Editörü

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

Language Reviewer

Assoc. Prof. Dr. Burcu ATAR
Assist. Prof. Dr. Derya ÇOBANOĞLU AKTAN
Dr. Ayfer Sayın

Sekreteryaya

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

Secretarait

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

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

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.

İletişim

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

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 KAN (Gazi Üni.)
Ahmet TURAN (Pearson)
Ali BAYKAL (Bahçeşehir Üni.)
Adnan ERKUŞ (Emekli Öğretim Üyesi)
Arif ÖZER (Hacettepe Üni.)
Ayfer SAYIN (Gazi Üni.)
Aylin ALBAYRAK SARI (Hacettepe Üni.)
Ayşegül ALTUN (Ondokuz Mayıs Üni.)
Bayram BIÇAK (Akdeniz Üni.)
Bayram ÇETİN (Gazi Üni.)
Bilge BAŞUSTA UZUN (Mersin Üni.)
Bilge GÖK (Hacettepe Üni.)
Burak AYDIN (Recep Tayyip Erdoğan
Üniversitesi)
Burcu ATAR (Hacettepe Üni.)
Burhanettin ÖZDEMİR (Siirt Üni.)
Cem Oktay GÜZELLER (Hacettepe Üni.)
Cindy M. WALKER (Duquesne University)
David KAPLAN (University of Wisconsin)
Deniz GÜLLEROĞLU (Ankara Üni.)
Derya ÇAKICI ESER (Kırıkkale Üni.)
Derya ÇOBANOĞLU AKTAN (Hacettepe Üni.)
Dilara BAKAN KALAYCIOĞLU (ÖSYM)
Dilek GENÇTANRIM (Kırşehir Ahi Evran Üni.)
Durmuş ÖZBAŞI (Çanakkele Onsekiz Mart Üni.)
Duygu GÜNGÖR (İzmir Üni.)
Elif Bengi ÜNSAL ÖZBERK (Adalet Bakanlığı)
Emine ÖNEN (Gazi Üni.)
Emrah GÜL (Hakkari Üni.)
Emre ÇETİN (Doğu Akdeniz Üni.)
Ergül DEMİR (Ankara Üni.)
Esin TEZBAŞARAN (İstanbul Üni.)
Esin YILMAZ KOĞAR (Hacettepe Üni.)
Esra Eminoğlu ÖZMERCAN (MEB)
Evrin ÇETİNKAYA YILDIZ (Erciyes
Üniversitesi)
Fatih KEZER (Kocaeli Üni.)
Fatih ORCAN (Karadeniz Teknik Üni.)
Fatma BAYRAK (Hacettepe Üni.)
Fazilet TAŞDEMİR (Recep Tayyip Erdoğan Üni.)
Funda NALBANTOĞLU YILMAZ (Nevşehir
Üni.)
Göksu GÖZEN (Mimar Sinan Güzel Sanatlar Üni.)
Gülden KAYA UYANIK (Sakarya Üni.)
Gülşen TAŞDELEN TEKER (Sakarya Üni.)

Hakan KOĞAR (Akdeniz Üni.)
Hakan Yavuz ATAR (Gazi Üni.)
Halil YURDUGÜL (Hacettepe Üni.)
Hatice KUMANDAŞ (Artvin Çoruh Üni.)
Hülya KELECİOĞLU (Hacettepe Üni.)
Hüseyin SELVİ (Mersin Üni.)
İbrahim Alper KÖSE (Abant İzzet Baysal Üni.)
İlker KALENDER (Bilkent Üni.)
İsmail KARAKAYA (Gazi Üni.)
Kaan Zülfikar DENİZ (Ankara Üni.)
Kübra ATALAY KABASAKAL (Hacettepe Üni.)
Mehmet KAPLAN (MEB)
Meltem ACAR GÜVENDİR (Trakya Üni.)
Mustafa ASİL (University of Otago)
Nagihan BOZTUNÇ ÖZTÜRK (Hacettepe Üni.)
Neşe GÜLER (Sakarya Üni.)
Neşe ÖZTÜRK GÜBEŞ (Mehmet Akif Ersoy Üni.)
Nuri DOĞAN (Hacettepe Üni.)
Nükhet DEMİRTAŞLI (Ankara Üni.)
Okan BULUT (University of Alberta)
Onur ÖZMEN (TED Üniversitesi)
Ömer KUTLU (Ankara Üni.)
Recep Serkan ARIK (Dumlupınar Üni.)
Sakine GÖÇER ŞAHİN (Hacettepe Üni.)
Sedat ŞEN (Harran Üni.)
Seher YALÇIN (Ankara Üni.)
Selahattin GELBAL (Hacettepe Üni.)
Sema SULAK (Bartın Üni.)
Serdar ÇAĞLAK (Osmangazi Üniveristesi)
Seval KIZILDAĞ (Adıyaman Üni.)
Sevda ÇETİN (Hacettepe Üni.)
Sevilay KILMEN (Abant İzzet Baysal Üni.)
Şeref TAN (Gazi Üni.)
Şeyma UYAR (Mehmet Akif Ersoy Üni.)
Tahsin Oğuz BAŞOKÇU (Ege Üni.)
Terry A. ACKERMAN (University of North
Carolina)
Tülin ACAR (Parantez Eğitim)
Türkan DOĞAN (Hacettepe Üni.)
Yavuz AKPINAR (Boğaziçi Üni.)
Yeşim ÖZER ÖZKAN (Gaziantep Üni.)
Zekeriya NARTGÜN (Abant İzzet Baysal Üni.)

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



İÇİNDEKİLER / CONTENTS

Concerto: Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamaları için Bir Platform Concerto: A Framework for Computerized Adaptive Test Applications Eren Can AYBEK	251
Eğitsel Yazılım Değerlendirme Formlarının ve Ölçeklerinin İncelenmesi Review of the Educational Software Evaluation Forms and Scales Ahmet ARSLAN, Ali GERİŞ, Taibe KULAKSIZ	272
Nitelik Sayısının Madde Seçme Algoritmalarının Performansı Üzerindeki Etkisi The Impact of Attribute Size on the Performance of Item Selection Algorithms Mehmet KAPLAN	285
Comparison of Classical Linear Regression and Orthogonal Regression with Respect to the Sum of Squared Perpendicular Distances Dik Uzaklıklar Kareler Toplamına Göre Klasik Doğrusal Regresyon ile Ortogonal Regresyonun Karşılaştırılması Taliha KELEŞ, Murat ALTUN	296
İlköğretim ve Ortaöğretim Başarı Ölçülerinin Yükseköğretime Geçiş Sınav Puanlarını Yordama Gücü Predictive Power of Primary and Secondary School Success Criterion on Transition to Higher Education Examination Scores Atilla ÖZDEMİR, Selahattin GELBAL	309
Sosyal Bilgiler Odaklı Başarı Yönelimleri Ölçeği (SOBYÖ): Geçerlik ve Güvenirlik Çalışması Social Studies Oriented Achievement Goal Scale (SOAGS): Validity and Reliability Study Melehat GEZER, İbrahim Fevzi ŞAHİN	335
Tam Öğrenme Yaklaşımının Öğrencilerin Akademik Başarıları Üzerindeki Etkisi: Bir Meta-Analiz Çalışması The Effect of Mastery Learning Model on Students Academic Achievement: A Meta-Analysis Study Tarık BAŞAR, İlkay AŞKIN TEKKOL, Selahattin GELBAL	355
Eğitimde Ölçme ve Değerlendirme Dersinde Başarıyı Etkileyen Faktörlerin Çoklu Göstergeler Çoklu Nedenler Modeliyle İncelenmesi The Investigation of the Factors Affecting Achievement in Measurement and Evaluation Course with Multiple Indicators Multiple Causes Model Kemal KURŞUN, Derya ÇOBANOĞLU AKTAN	372
Computer Adaptive Multistage Testing: Practical Issues, Challenges and Principles Bireye Uyarlanmış Çok Aşamalı Testler: Pratik Konular, Zorluklar ve Prensipler Halil İbrahim SARI, Hasibe YAHSI SARI, Anne Corinne HUGGINS MANLEY	388
Öğretmen – Öğrenci İlişki Ölçeği'nin (ÖÖİÖ) Uyarlanması Adaptation of Teacher – Student Relationship Scale (STRS) Selen DEMİRTAŞ ZORBAZ, Arif ÖZER, Dilek GENÇTANIRIM-KURT, Tuncay ERGENE	407

Vee Diyagramından Elde Edilen Puanların Güvenirliđinin Klasik Test Kuramı ve Genellenebilirlik Kuramına Gre İncelenmesi The Examination of Reliability of Vee Diagrams According to Classical Test Theory and Generalizability Theory Betl POLAT DEMİR	419
The Examination of Model Fit Indexes with Different Estimation Methods under Different Sample Sizes in Confirmatory Factor Analysis Dođrulayıcı Faktr Analizinde Farklı rneklem Byklklerinde Farklı Kestirim Yntemleriyle Hesaplanan Uyum İndekslerinin İncelenmesi Ayfer SAYIN	432
A Scaling of Preservice Teachers' Motivational Factors in Choosing the Teaching Profession đretmen Adaylarının đretmenlik Mesleđini Seęme Nedenlerinin lęeklenmesi Sakine GÇER ŐAHİN, Selahattin GELBAL	443
Çocuklardaki Kısa Sreli Anlık Bellek İŐlevinin Deđerlendirilmesi Evaluation of Short Term Memory Span Function In Children BarıŐ ERGL, Arzu ALTIN YAVUZ, Ebru GNDOđAN AŐIK	459

Concerto: Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamaları için Bir Platform*

Concerto: A Framework for Computerized Adaptive Test Applications

Eren Can AYBEK**

Öz

Bilgisayar ortamında bireye uyarlanmış test (BOBUT) çalışmalarında gerçek verilere dayalı olarak yürütülen post-hoc simülasyonlarının yanı sıra canlı BOBUT uygulamalarına da yer verilmektedir. Bu çalışmada, araştırmacıların canlı BOBUT uygulaması yürütebilmelerini sağlayan bir yazılım olan Concerto'nun tanıtılması, kurulumunun yapılarak işlevsel bir BOBUT örneğinin araştırmacılar tarafından çalıştırılabilmesi için gerekli olan adımların gösterilmesi amaçlanmıştır. Bu amaçla Concerto'nun bir Linux sunucuya kurulum adımları; sunucunun kurulumu, gerekli yazılımların kurulumu, Concerto'nun kurulumu ve örnek bir BOBUT uygulamasının indirilerek düzenlenmesi ve çalıştırılması için gerekli adımlar gösterilmiştir.

Anahtar Kelimeler: bilgisayar ortamında bireye uyarlanmış test, madde tepki kuramı, concerto

Abstract

In the studies of computerized adaptive tests (CAT), real life CAT applications are used besides of post-hoc simulations which held by real-data. The purpose of this study is to introduce Concerto, a CAT platform which provides practical framework for researchers, and demonstrate the installation of Concerto and run a sample CAT application. For this purpose; configuration of Concerto to Linux server, installation of server and necessary software, installation of Concerto and downloading, editing and running a sample CAT application have been demonstrated step-by-step.

Keywords: computerized adaptive testing, item response theory, concerto

Bilgisayar Ortamında Bireye Uyarlanmış Test Uygulamalarında Kullanılan Yazılımlar

Bilgisayar Ortamında Bireye Uyarlanmış Testler (BOBUT [Computerized Adaptive Tests [CAT]) ile ilgili yürütülen araştırmalar incelendiğinde, gerçek verilere dayalı olarak gerçekleştirilen post-hoc simülasyon çalışmaları (Gardner ve diğerleri, 2004; Haley, Coster, Andres, Kosinski ve Ni, 2004; Haley ve diğerleri, 2009; Smits, Cuijpers ve van Straten, 2011; Betz ve Turner, 2011; Aybek, 2016) olduğu görülmektedir. Bunun yanında canlı bir BOBUT uygulamasının yürütüldüğü çalışmaların (Hol, Vorst ve Mellenbergh, 2007; Elhan, Öztuna, Kutlay, Küçükdeveci ve Tennant, 2008; Gibbons ve diğerleri, 2012; Chien, Wang, Wang ve Lin, 2009; Chien, Lai, Lu, Wang, Chen, Wang ve Su, 2011; Aybek, 2016) sayısı da her geçen gün artmaktadır. Bu çalışmada, araştırmacıların canlı BOBUT uygulamaları için kullanabileceği yazılımlar kısaca tanıtılmış ve açık kaynaklı bir yazılım olan Concerto'nun bir GNU/Linux sunucuya kurulum adımları gösterilmiştir.

Canlı BOBUT uygulamalarında kullanılabilecek yazılımlar incelendiğinde FastTest, Pearson VUE, Prometric, McCann gibi ticari yazılımların yanında Concerto, IRT-CAT ve OSCATS gibi açık kaynaklı ve ücretsiz yazılımların da olduğu görülmektedir. Ticari yazılımlara ait bilgilere

* Bu makale, Prof. Dr. R. Nüket Demirtaşlı danışmanlığında hazırlanmış olan “Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) Olarak Uygulanabilirliğinin Araştırılması” başlıklı doktora tezi kapsamında üretilmiştir.

** Dr., Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi, Eskişehir – Türkiye. e-posta: erencan@aybek.net

yazılımların web sayfalarından ulaşılması mümkündür. Açık kaynaklı olan Concerto, IRT-CAT ve OSCATS yazılımlarına dair kısa bilgilere ise başlıklar halinde yer verilmiştir.

IRT-CAT

3 PL ve Rasch modellerini destekleyen IRT-CAT yazılımı web tabanlı bir yazılımdır. Akademik Özgür Lisans (Academic Free License) ile lisanslanmış yazılımla ilgili son güncelleme 2009 yılında yapılmıştır.

Yazılım adresi: <https://sourceforge.net/projects/irt-cat>

OSCATS

Son olarak 2011 yılında güncellenen OSCATS yazılımı, Genel Kamu Lisansı (General Public License) altında lisanslanmış ve canlı BOBUT uygulamalarına imkan sağlamasının yanında simülasyon çalışmalarında da kullanılabilir. 1 PL, 2 PL, 3 PL Modeller, Sınıflamalı Tepki Modeli (Nominal Response Model), Kısmi Puanlama Modeli (Partial Credit Model), Genelleştirilmiş Kısmi Puanlama Modeli (Generalized Partial Credit Model) ve Kademeli Tepki Modeli (Graded Response Model) ile çalışabilen yazılım, altı adet madde seçim yöntemini desteklemekle birlikte durdurma kuralı olarak yalnızca sabit sayıda madde uygulandığında testin durdurulmasına izin vermektedir. En yüksek olasılık (Maximum Likelihood) ve Beklenen Sonsal (Expected a Posteriori) yöntemleri ile θ kestirimi yapılmasına izin veren yazılım gelecekte R ve MATLAB kodlarını çalıştırmayı planlamaktadır.

Yazılım adresi: <https://code.google.com/archive/p/oscats/wikis/Features.wiki>

Concerto

Cambridge Üniversitesi Psikometri Merkezi tarafından geliştirilen Concerto, açık kaynaklı ve R temelli bir yazılımdır. Concerto web ortamında çalışmakta ve BOBUT uygulamaları için kullanıcıya bir çerçeve oluşturmaktadır. Yazılım temel olarak R'da çalıştırılan kod çıktıları MySQL veri tabanına kaydetmekte ve daha sonra veri tabanına kaydedilen bilgileri çağırarak web tabanlı olarak kullanıcıya göstermektedir. Concerto; test modülü, HTML şablon modülü ve tablo modülüne sahiptir. Test modülünde R üzerinde catR paketinin de yardımıyla BOBUT algoritması geliştirilmekte, HTML şablon modülünde testi alan bireylerin karşılaşacağı sayfalar tasarlanmakta ve tablo modülünde ise maddelerin, kullanıcıların verdiği yanıtlar ile R tarafından üretilen çıktıların kaydedildiği veri tabanları oluşturulmaktadır.

Concerto; R'dan aldığı çıktıyı veri tabanı aracılığı ile kullanıcıya göstermekte, kullanıcıdan aldığı girdiyi ise yine veri tabanı aracılığı ile R'a ulaştırmakta ve testin işleyişini devam ettirmektedir. Bu nedenle Concerto'nun kullanıcı, MySQL ve R arasında bir köprü görevini üstlendiği söylenebilir. Concerto'nun R temelli olmasının en büyük avantajı ise test geliştiricisinin R sayesinde ihtiyaçlarına uygun madde tepki kuramı (MTK) modelini, madde seçim ve θ kestirim yöntemini, durdurma kuralını belirleyerek BOBUT uygulamasını geliştirmesine olanak sağlamasıdır. Bu noktada araştırmacılar mevcut R paketleri ile sınırlıdır. Başka bir anlatımla R paketleri ile desteklenen tüm MTK modelleri, madde seçim ve özellik / yetenek kestirim yöntemleri kullanılabilir. R paketi olarak bulunmasa dahi araştırmacılar, kendi R betiklerini yazarak diledikleri algoritmaları kullanabilmektedir. Yazılımın özellikle 4.0 ve sonraki sürümlerinde test geliştiricileri diledikleri gibi R kodu çalıştırabilmektedirler. Sonraki bölümde Concerto yazılımının bir Linux sunucuya kurulum adımları gösterilmiştir.

Concerto Yazılımının Bir Linux Sunucuya Kurulumu

Bu bölümde Concerto platformunun, Ubuntu Linux 14.04 x64 işletim sistemine sahip bir sunucuya nasıl kurulacağını adımları yer almaktadır. Kurulum aşamaları temel olarak Concerto'dan (2016) yararlanılarak gerçekleştirilmiş olup diğer kaynaklardan yararlanıldığı durumlarda ilgili aşamada atf yapılmıştır. Ayrıca bu kurulum rehberi ile kurulan Concerto, yalnızca eğitim ve araştırma amacı ile kullanılmalıdır. Bu rehber, sunucu güvenliğini sağlama vb. bilgiler vermemektedir. Gerçek BOBUT uygulamalarında Concerto'nun kullanılması için sunucu güvenlik uzmanlarından yardım alınması önerilmektedir.

Rehberde yer alan komutların çalışmaması ya da *erişim engellendi, access denied, permission required vb.* hatalar alınması durumunda komutun başına `sudo` eklenerek komut yeniden çalıştırılabilir.

Gereklilikler

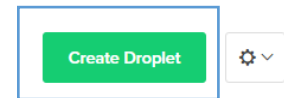
Concerto v4'ün kurulması ve çalıştırılabilmesi için gereklilikler şu şekildedir:

1. Linux işletim sistemi
2. v5.3 ya da daha yeni sürüm numarasına sahip PHP
3. php-process paketi (CentOS gibi bazı Linux dağıtımlarında gerekli olabilir)
4. PHP safe mod: OFF olmalı
5. PHP magic quotes: OFF olmalı
6. PHP short open tags: ON olmalı
7. PHP open base dir: ON olmalı
8. v5 ya da daha yeni sürüm numarasına sahip MySQL
9. v2.15 ya da daha yeni sürüm numarasına sahip R
10. concerto R paketi yüklenmeli
11. root erişimine sahip bir Linux kullanıcısı
12. Veritabanındaki tüm yetkilere sahip bir MySQL kullanıcısı
13. En az 1GB RAM

Bu raporda Concerto, Digitalocean (www.digitalocean.com) firmasına ait sunuculara kurulmuştur. Dolayısıyla Concerto'nun bu firmaya ait sunuculara nasıl kurulacağı açıklanmıştır. Ancak Ubuntu Linux 14.04 x64 işletim sistemine sahip tüm sunuculara benzer işlem adımları takip edilerek Concerto kurulumu sağlanabilir.

Sunucu Kurulumu

Digitalocean'da sanal sunucular droplet adıyla anılmaktadır. Üye girişi yapıldıktan sonra Create Droplet düğmesine tıklayarak yeni bir sunucu kurulabilir (Şekil 1). Sunucu kurulumu ve çalıştırılması ücretli bir işlem olup benzer işlem adımları kişisel bilgisayara kurulabilecek bir GNU/Linux işletim sistemi ile de gerçekleştirilebilir. Ancak işletim sistemi kurulumu bu makalenin konusu olmadığından dolayı sanal sunucu ile çalışılması tercih edilmiştir.



Şekil 1. Yeni Bir Sunucunun Kurulması

Açılan sayfada Distributions ve One-click App sekmeleri bulunmaktadır. Dileyenler Distributions sekmesinden istedikleri dağıtımı kurabilirler. Ancak bu anlatımda One-click App sekmesinde yer alan hazır kurulumlardan yararlanılmıştır. Concerto kurulumu için öncelikle One-click App sekmesinden LAMP on 14.04 seçilir. Daha sonra 512MB RAM ve sunucu konumu olarak Amsterdam seçilir. İletişim gecikmelerinin önüne geçmek için sunucu ile Türkiye arasındaki mesafenin en kısa olması yararlı olacaktır. Bu nedenle sunucu konumu olarak Amsterdam seçilmiştir. Concerto 1GB RAM'e ihtiyaç duymasına rağmen, sunucunun hafızası 512 MB olarak tanımlanacak ancak sunucu kurulumunun ardından RAM miktarı 1GB olarak artırılacaktır. Sunucunun en temel özellikler ile kurulması, kullanıcıların diledikleri zaman RAM'i istedikleri miktara artırıp, dilediklerinde en temel RAM miktarına geri dönmelerini sağlamaktadır (Şekil 2).

Distributions One-click Apps Backups

Cassandra on 14.04	Discourse on 14.04	Django on 14.04
Docker 1.10.3 on 14.04	Dokku v0.4.14 on 14.04	Drone on 14.04
Drupal 8.0.5 on 14.04	Elixir on 14.04	ELK Logging Stack on 14.04
Ghost 0.7.8 on 14.04	GitLab 8.5.6 CE on 14.04	Joomla! 3.4.8 on 14.04
LAMP on 14.04	LEMP on 14.04	Magento 2.0.2 CE on 14.04
MEAN on 14.04	MediaWiki on 14.04	MongoDB 3.2.0 on 14.04
Mumble Server (murmur) on 14.04	node v4.4.0 on 14.04	ownCloud 9.0.0 on 14.04
PHPMyAdmin on 14.04	Redis 3.0.7 on 14.04	Redmine on 14.04
Ruby on Rails on 14.04 (Postgres, Nginx, U...)	WordPress on 14.04	

Choose a size

\$5/mo \$0.007/hour 512 MB / 1 CPU 20 GB SSD Disk 1000 GB Transfer	\$10/mo \$0.015/hour 1 GB / 1 CPU 30 GB SSD Disk 2 TB Transfer	\$20/mo \$0.030/hour 2 GB / 2 CPUs 40 GB SSD Disk 3 TB Transfer	\$40/mo \$0.060/hour 4 GB / 2 CPUs 60 GB SSD Disk 4 TB Transfer	\$80/mo \$0.119/hour 8 GB / 4 CPUs 80 GB SSD Disk 5 TB Transfer
\$160/mo \$0.239/hour 16 GB / 8 CPUs 160 GB SSD Disk 6 TB Transfer	\$320/mo \$0.476/hour 32 GB / 12 CPUs 320 GB SSD Disk 7 TB Transfer	\$480/mo \$0.714/hour 48 GB / 16 CPUs 480 GB SSD Disk 8 TB Transfer	\$640/mo \$0.952/hour 64 GB / 20 CPUs 640 GB SSD Disk 9 TB Transfer	

Choose a datacenter region

New York 3 2 1	San Francisco 1	Amsterdam 3 2	Singapore 1	London 1
Frankfurt 1	Toronto 1			

Şekil 2. Sunucu Özelliklerinin Ayarlanması – 1

Daha sonra Choose a hostname kısmına sunucuya vermek istenilen isim (Concerto, CAT, Deneme vb.) yazılıp Create düğmesine tıklanır (Şekil 3).

Finalize and create

How many Droplets?

Deploy multiple Droplets with the same [configuration](#).

– 1 Droplet +

Choose a hostname


Give your Droplets an identifying name you will remember them by. Your Droplet name can only contain alphanumeric characters, dashes, and periods.

ubuntu-1gb-ams2-01

Create

Şekil 3. Sunucu Özelliklerinin Ayarlanması – 2

Kurulum başarıyla tamamlandıktan sonra sunucunun IP adresi ve sunucuya ait donanım özellikleri, sunucu durumu gibi bilgilerin yer aldığı ekran açılmaktadır. Sunucu adının sol tarafındaki yeşil nokta, sunucunun şu anda aktif olduğunu göstermektedir (Şekil 4). Şu anda 512MB RAM ile çalışan sunucunun RAM miktarını 1GB'a yükseltmek için More düğmesine tıklanarak açılan menüde Resize Droplet seçilir.

Img	Name	IP Address	Created
	Concerto 512 MB Memory / 20 GB Disk / AMS2	146.185.130.121	Let's get to work! More

Şekil 4. Sunucu Kurulumunun Tamamlanması

Açılan sayfada, sunucunun yeniden boyutlandırılması için öncelikle kapatılması gerektiğine dair bir hata mesajı görünür. Sunucuyu kapatmak için sol menüden Power bağlantısına tıklanır (Şekil 5).

- Access
- Power
- Resize**
- Snapshots
- Settings
- Backups
- Graphs
- History
- Destroy

Droplet Resize

Please power off your server from the command line first before scaling the Droplet.

Şekil 5. Sunucunun Yeniden Boyutlandırılması - 1

Açılan sayfada Power Off düğmesi tıklanır ve sunucunun kapanması beklenir. Daha sonra yeniden Resize bağlantısına tıklanır ve bu sayfada Flexible seçilir. Flexible seçildiğinde sunucuya ayrılan RAM ve trafik miktarı değişmekte ancak sabit disk büyüklüğü ve sunucuya ayrılan işlemci sayısı değişmemektedir. Kullanıcıların Concerto üzerinde çalışmadıklarında temel paket seçeneğine kolaylıkla dönebilmeleri için Flexible seçilmesi önemlidir. Daha sonra 1GB RAM özelliği olan sunucu paketi seçilerek Resize düğmesine tıklanır (Şekil 6).

Droplet Resize

<input type="radio"/> Permanent This will increase the disk size, CPU and RAM of your Droplet.	<input checked="" type="radio"/> Flexible This will increase or decrease the CPU and RAM of your Droplet.
--	---

Select Size

\$5/mo \$0.007/hour	\$10/mo \$0.015/hour	\$20/mo \$0.030/hour	\$40/mo \$0.060/hour
512 MB / 1 CPU 20 GB SSD Disk 1000 GB Transfer	1 GB / 1 CPU 20 GB SSD Disk 2 TB Transfer	2 GB / 2 CPU 20 GB SSD Disk 3 TB Transfer	4 GB / 2 CPU 20 GB SSD Disk 4 TB Transfer

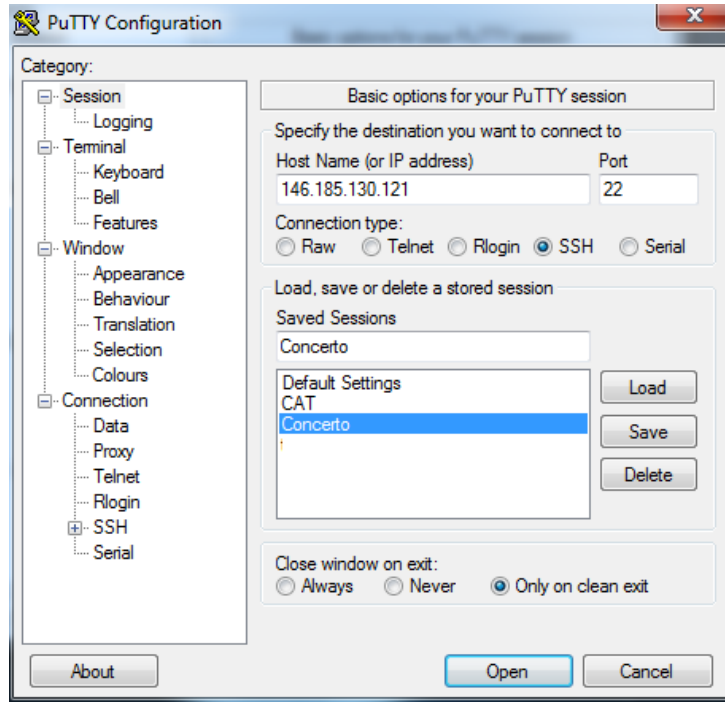
Şekil 6. Sunucunun Yeniden Boyutlandırılması – 2

Sunucunun yeniden boyutlandırılmasının ardından, Power sayfasından sunucu yeniden açılır ve sunucunun kurulumu sonlanmış olur. Sonraki aşamada, sunucuya Concerto'nun çalışması ve yönetilebilmesi için gerekli olan yazılımların kurulumu yapılmalıdır.

Sunucuya Yazılımların Kurulumu

Sunucunun kurulumu yapıldıktan sonra, sunucuya yüklü işletim sistemine erişilerek Concerto'nun gerektirdiği yazılımların kurulması ya da ayarlanması gerekmektedir. Sunucuya uzaktan erişim için Windows ortamında putty adlı yazılımdan yararlanılacaktır. Linux dağıtımı kullanan bir bilgisayardan sunucuya erişmek için Terminal kullanılabilir. Ancak bu anlatım Windows yüklü bilgisayardan sunucuya erişimi kapsamaktadır.

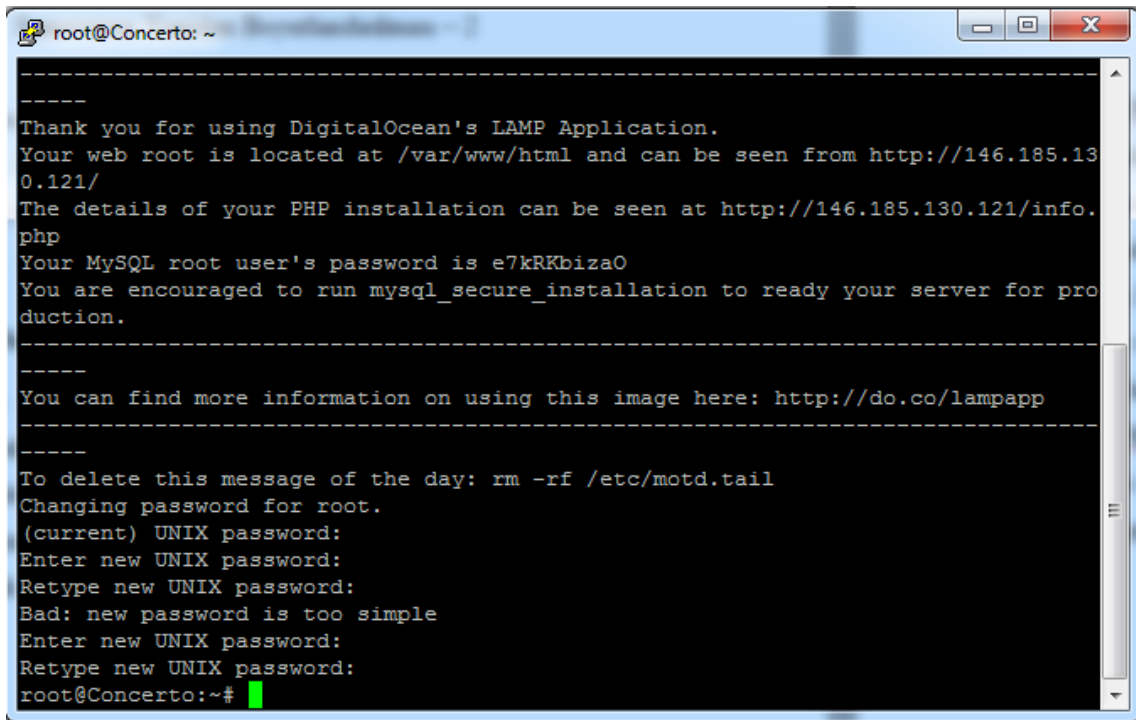
Gerekli olan putty yazılımı www.putty.org adresinden indirilebilir. Kurulum gerektirmeyen bu yazılım açıldığında Şekil 7'deki gibi bir ekran çıkacaktır.



Şekil 7. Putty Ekranı

Putty ekranında Host Name bölümüne Şekil 4'te görülen, sunucuya ait IP adresi yazılmalıdır. Saved sessions kısmına bir isim yazılıp Save düğmesine tıklanırsa, daha sonra sunucuya hızlı bir şekilde bağlanmak için sunucu bilgileri kaydedilir. Open düğmesine tıkladığında sunucuya bağlantı gerçekleştirilir. Bu aşamada ekrana gelen uyarıda Evet seçeneği seçilerek bağlanma işlemine devam edilir. Ekranda kullanıcı adı ve şifre sorulacaktır. Kullanıcı adı olarak root girilip Enter tuşuna basılır. Şifre ise firma tarafından e-posta olarak gönderilmektedir. E-posta ile gelen şifreyi kopyaladıktan sonra, Putty ekranına sağ tıkladığında otomatik olarak yapıştırılmaktadır (bu aşamada ekranda herhangi bir şifre girilmemiş gibi görünebilir). Enter tuşuna basarak sunucuya giriş yapılabilir.

Sunucuya giriş yapıldığında root şifresinin değiştirilmesi istenecektir. Önce firma tarafından gönderilen şifreyi girilip daha sonra sırasıyla yeni şifre iki defa girildiğinde root kullanıcıya yeni bir şifre atanmış olur. Bu işlemlerin ardından artık sunucuya erişim tamamlanmış olur (Şekil 8).



```
root@Concerto: ~
-----
Thank you for using DigitalOcean's LAMP Application.
Your web root is located at /var/www/html and can be seen from http://146.185.130.121/
The details of your PHP installation can be seen at http://146.185.130.121/info.php
Your MySQL root user's password is e7kRkKbiza0
You are encouraged to run mysql_secure_installation to ready your server for production.
-----
You can find more information on using this image here: http://do.co/lampapp
-----
To delete this message of the day: rm -rf /etc/motd.tail
Changing password for root.
(current) UNIX password:
Enter new UNIX password:
Retype new UNIX password:
Bad: new password is too simple
Enter new UNIX password:
Retype new UNIX password:
root@Concerto:~#
```

Şekil 8. Sunucuya Uzaktan Erişim

Bu aşamadan sonra, öncelikle Concerto'nun gerektirdiği yazılımların kurulması ya da ayarlanması gerekmektedir. Sunucu oluşturulurken LAMP on 14.04 seçeneği seçildiği için PHP ve MySQL sunucuya kurulu olarak gelmektedir. Şekil 8 incelendiğinde *Your MySQL root user's password is* satırında MySQL root kullanıcısının şifresi de yer almaktadır.

Phpmyadmin Kurulumu (Tercihe bağlı)

Phpmyadmin, Concerto için bir gereklilik değildir. Ancak phpmyadmin kurulumu, veri tabanı yönetiminde (veri tabanı oluşturma, veri tabanı inceleme, kullanıcı atama vb.) kolaylık sağlayacaktır. Phpmyadmin kurulumu, Ellingwood (2014)'un rehberinden yararlanılarak gerçekleştirilmiştir.

Öncelikle aşağıdaki komut yardımıyla phpmyadmin paketleri sunucuya kurulur.

```
sudo apt-get install phpmyadmin -y
```

Kurulum aşamasında çıkan ilk uyarıda apache2 seçilir. Kırmızı ile seçili olsa da apache2'yi seçmek için Space tuşuna basılması gerekmektedir. * işareti yoksa apache2 seçilmemiş anlamına gelmektedir. Bu nedenle apache2 Space tuşuna basarak seçildikten sonra Tab tuşuna basılarak OK'e

gidilir ve Enter'a basılır. İkinci uyarıda Yes seçilerek Enter'a basılır. Daha sonra kurulum, veri tabanlarına erişim sağlayabilmek için MySQL root şifresini ister. MySQL root şifresi, sunucuya ilk bağlanıldığında karşılama ekranında çıkmaktadır (Şekil 8). Fare ile ekran yukarı doğru kaydırılarak en üste çıkıldığında karşılama ekranı yeniden görüntülenebilir. MySQL root şifresi buradan alındıktan sonra fare yardımıyla ekran en alta kaydırılır ve MySQL root şifresi girilerek Enter'a basılır.

Kurulum daha sonra phpmyadmin ara yüzüne erişebilmek için yeni bir şifre tanımlanmasını ister. Tanımlanan şifre iki defa sorulur ve ikisine de aynı şifrenin girilmesi gerekir. Şifre tanımlaması yapıldıktan sonra kurulum sonlanır (Şekil 9).

```
Creating config file /etc/phpmyadmin/config-db.php with new version
granting access to database phpmyadmin for phpmyadmin@localhost: success.
verifying access for phpmyadmin@localhost: success.
creating database phpmyadmin: success.
verifying database phpmyadmin exists: success.
populating database via sql... done.
dbconfig-common: flushing administrative password
root@Concerto:~#
```

Şekil 9. Phpmyadmin Paketlerinin Kurulumu

Phpmyadmin kurulduktan sonra sırasıyla aşağıdaki komutlar girilir:

```
sudo php5enmod mcrypt
sudo service apache2 restart
```

Daha sonra web tarayıcısının adres satırına <http://ipnumarası/phpadmin> yazılarak phpmyadmin ekranına ulaşılır (Şekil 10).


The image shows the phpMyAdmin login interface. At the top, there is a logo for phpMyAdmin and the text 'phpMyAdmin'e Hoş Geldiniz'. Below this, there is a section for language selection with a dropdown menu currently set to 'Türkçe - Turkish'. Underneath, there is a login form with two input fields: 'Kullanıcı Adı:' (Username) and 'Parola:' (Password). A 'Git' button is located at the bottom right of the form.

Şekil 10. Phpmyadmin Giriş Ekranı

Phpmyadmin kurulumunun ardından sunucuya ip numarası üzerinden mi yoksa bir alan adı (ör: concerto.catirt.com) aracılığıyla mı erişileceğine, ayrıca sunucuda birden fazla alan adı barındırılıp barındırılmayacağına karar verilmesi gerekmektedir. Bu anlatımda sunucuda tek bir alan adı barındırılmış, bu nedenle yalnızca bir alan adı (concerto.catirt.com) ile kurulum aşamalarına yer verilmiştir. Birden fazla alan adı barındırmak isteyen araştırmacılar Sveldrov'dan (2012) yararlanabilirler.

Sunucuya Alan Adı ile Erişmek (Tercihe Bağlı)

Sunucuya, dolayısıyla Concerto'ya, bir alan adı aracılığı ile erişmek için öncelikle alan adı sunucuya tanımlanmalıdır. Bunun için sunucunun listelendiği ekranda (Şekil 11), More tuşuna tıklanarak Add Domain bağlantısına tıklanır (Şekil 11).

Img	Name	IP Address	Created
	Concerto 1 GB Memory / 20 GB Disk / AMS2	146.185.130.121	3 hours ago


[More](#) [Add a Domain](#)

Şekil 11. Sunucuya Alan Adı Eklenmesi – 1

Açılan sayfada hangi alan adı ile sunucuya erişmek istendiği belirtilerek Create Record düğmesine tıklanmalıdır (Şekil 12). Bu aşamada yalnızca sahibi olunan alan adı üzerinde işlem yapılabileceği unutulmamalıdır.

Add a Domain

Specifying an IP Address will set the default record for the domain (an A record @) to your selected IP.

Domain concerto.catirt.com	 Concerto 146.185.130.121	Create Record
-------------------------------	--	-------------------------------

Şekil 12. Sunucuya Alan Adı Eklenmesi – 2

Daha sonra açılan sayfada birtakım bilgiler yer almaktadır. Bu bilgilerin alan adı yönetim panelinden alan adına atanması gerekmektedir. Bu işlem alan adının alındığı firmanın yönetim panelinden gerçekleştirilebilir. Bu aşamada zorluk yaşayanlar alan adının alındığı firmadan yardım alabilirler (Bu işlem yapıldıktan sonra, alan adının sunucuya yönlmesi 24 saati bulabilmektedir). Alan adının sunucuya yönlendirilmesi ile birlikte artık <http://concerto.catirt.com> adresi ziyaret edildiğinde, doğrudan sunucuya erişim sağlanmaktadır.

Bu aşamadan sonra Concerto dosyaları indirilerek sunucuya yüklenmelidir. Bunun için <https://code.google.com/archive/p/concerto-platform/downloads> adresi ziyaret edilerek son sürüm indirilebilir. Ancak bu anlatım v4.00 beta8 sürümü için yapılmıştır. Gelecek sürümlerde bu anlatımın bazı kısımları geçerliliğini yitirebilir.

Concerto Dosyalarının Sunucuya Yüklenmesi

Concerto dosyaları Putty aracılığı ile doğrudan sunucuya yüklenebileceği gibi bir FTP yazılımı ile bilgisayardan sunucuya gönderilme yoluyla da yüklenebilir. Concerto dosyalarının Putty aracılığı ile sunucuya yüklenmesi için aşağıdaki komutlar kullanılır:

```
mkdir /var/www/html/test
```

Yukarıdaki kod, test adlı bir klasör oluşturmaktadır. *test* yerine, *concerto*, *bobut* vb. ifadeler yazılarak dilenen klasör adında bir klasör oluşturulabilir. Eğer Concerto <http://alanadi.com/test> gibi

bir adres yerine doğrudan <http://alanadi.com> adresinden çalıştırılmak istenirse herhangi bir klasör oluşturmaya gerek yoktur. Bu durumda dosyaların /var/www/html altına yüklenmesi gerekir.

```
cd /var/www/html/test
```

Oluşturulan test klasörüne gidilir.

```
wget https://storage.googleapis.com/google-code-archive-downloads/v2/code.google.com/concerto-platform/concerto_platform_v4.0.0.beta8.zip
```

Yukarıdaki komut ile Concerto v4.0.0.beta8 dosyaları test klasörüne indirilmiş olur. Ancak indirilen dosya .zip formatındadır ve dosyaların zip içerisinden çıkarılması gerekmektedir. Bunun için aşağıdaki kod kullanılır:

```
unzip concerto*.zip
```

Bu kod, klasör içerisinde concerto adı ile başlayan zip dosyalarını dışarı çıkarır. Bu işlemin ardından dosyalar sunucuya yüklenmiş olur.

```
dir
```

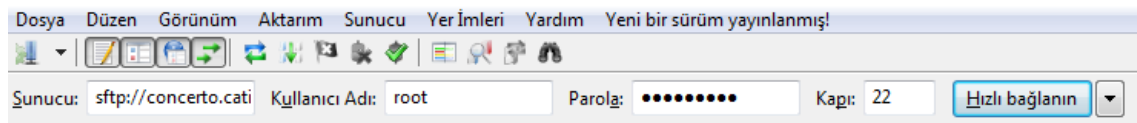
Komutu verilerek klasörün içeriği görüntülenebilir. Dileyen kullanıcılar zip dosyasını silmek için:

```
rm concerto*.zip
```

komutunu kullanabilirler. Böylece klasördeki concerto adı ile başlayan tüm zip dosyaları silinmiş olur.

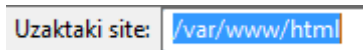
Concerto dosyaları, komut satırı yerine FTP yazılımı aracılığı ile gönderilmek istenirse Concerto bilgisayara indirildikten sonra dosyalar, zip dışına çıkarılmalıdır. Dosyaların sunucuya aktarımı için bir FTP yazılımına ihtiyaç duyulmaktadır. Bu aşamada açık kaynaklı bir yazılım olan FileZilla yazılımı tercih edilmiştir. FileZilla https://filezilla-project.org/download.php?show_all=1 adresinden indirilebilir.

Filezilla açıldığında sunucu kısmına sftp://concerto.catirt.com ya da sftp://sunucuipadresi yazılmalıdır. Kullanıcı adı root ve şifre ise sunucuya ilk erişim sağlandığında kullanıcı tarafından değiştirilen root şifresidir. Kapı ya da port ise 22 olmalıdır (Şekil 13).



Şekil 13. Dosyaların Sunucuya Aktarımı – 1

Bağlanın düğmesine tıklandığında gelen uyarıya Evet dedikten sonra Filezilla yazılımı sunucuya bağlanacaktır. Filezilla sunucuya bağlandığında sunucuda yer alan klasör ve dosyaları listeler. Başka bir deyişle uzaktaki bir bilgisayarın içerisinde gezinme imkanı sunar. Concerto dosyaları, bu uzaktaki bilgisayarın /var/www/html yoluna yüklenmelidir. Bunun için Uzaktaki site bölümüne /var/www/html yazılarak Enter'a basılabilir ya da dosya yöneticiden dosya yolu bulunabilir.



Şekil 14. Dosyaların Sunucuya Aktarımı – 2

Şekil 14'te görülen /var/www/html klasörüne yüklenen her dosyaya, internet üzerinden erişim mümkündür. Bu nedenle kullanılacak maddeler vb. bu kısma atılmamalıdır. Bir sunucuda birden fazla alan adı ya da site barındırılmak istenildiğinde bu yolun farklılaştığı unutulmamalıdır.

Concerto dosyaları, sürükle-bırak yapılarak Filezilla aracılığı ile /var/www/html yoluna gönderilir. Bu aşamada dileyenler test adında bir klasör oluşturarak dosyaları bu klasöre yükleyebilir (Şekil 15).

Dosya Adı	Boyut	Tip	Son değişiklik	İzinler	Sahip/Grup
..					
cms		Dosya klasörü	17.03.2016 14:0...	drwxr-xr-x	root root
cron		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
css		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
data		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
js		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
lib		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
media		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
query		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
setup		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
socks		Dosya klasörü	17.03.2016 14:1...	drwxr-xr-x	root root
favicon.ico	1.150	Simge	17.03.2016 14:0...	-rw-r--r--	root root
index.php	6.140	PHP Dosyası	17.03.2016 14:0...	-rw-r--r--	root root
Ini.php	25.632	PHP Dosyası	17.03.2016 14:0...	-rw-r--r--	root root
license.txt	18.429	Metin Belgesi	17.03.2016 14:0...	-rw-r--r--	root root
remote.php	4.321	PHP Dosyası	17.03.2016 14:0...	-rw-r--r--	root root
SETTINGS.php	3.657	PHP Dosyası	17.03.2016 14:0...	-rw-r--r--	root root

Şekil 15. Dosyaların Sunucuya Aktarımı – 3

Artık Concerto dosyaları sunucuda erişime açık bir şekilde yer almaktadır. Bu aşamadan sonra Concerto'nun gerekliliği olan R ve R paketleri kurulmalıdır.

R Paketlerinin Kurulumu

Concerto *catR* paketini R üzerinde çalıştırmakta ve R çıktılarını *MySQL* veri tabanına kaydetmektedir. Bu nedenle sunucuda R yazılımı, *catR* ve *concerto* paketleri ile analizlerin yapılabilmesi için gerekli olan diğer paketlerin sırası ile kurulması gerekmektedir. Ancak Ubuntu 14.04 için depolarda bulunan R yazılımı 3.0.2 sürümüne sahip olup yeni paket kurulumunda sorun yaşatabilmektedir. Hata ile karşılaşmamak için sunucuya R'in son sürümü kurulmalıdır. Kurulum için aşağıdaki adımlar izlenebilir (Attali, 2015).

Öncelikle Ubuntu'nun yazılım kaynaklarına yeni bir kaynak eklenmelidir. Bunun için

```
sudo sh -c 'echo "deb http://cran.rstudio.com/bin/linux/ubuntu
trusty/" >> /etc/apt/sources.list'
```

komutu verilir.

Daha sonra eklenen kaynaktaki dosyaların indirilerek kurulabilmesi için bu kaynağın yetkilendirilmesi gerekmektedir. Bunun için sırasıyla aşağıdaki komutlar girilir:

```
gpg --keyserver keyserver.ubuntu.com --recv-key E084DAB9
gpg -a --export E084DAB9 | sudo apt-key add -
```

Bu aşamadan sonra R'in son sürümü kurulumu hazır hale gelmiştir. R, aşağıdaki komutlarla kurulur:


```
sudo apt-get update
sudo apt-get -y install r-base
```

R kurulumunun ardından R komutu verilerek Enter'a basılır. Bu aşamada R konsolu uzak bir bilgisayarda çalışmaktadır. Çeşitli işlemler yapılarak R test edilebilir. Ancak Concerto için birtakım paketlerin yüklenmesi gerekmektedir. Bunun için R konsoluna aşağıdaki kodlar girilir.

```
install.packages("rjson")
install.packages("catR")
```

Bu komut R'a *rjson* ve *catR* paketlerini yükler. Ancak komutu verdikten sonra R, paketlerin hangi sunucudan indirileceğini sorar. Ardından 1 ya da dilenen herhangi bir sunucunun kodu yazılarak Enter'a basılabilir. Bu noktada, paketlerin kurulumunda R hata veriyorsa sunucunun RAM miktarı düşük gelmiş olabileceği unutulmamalıdır. Özellikle 512MB RAM'e sahip sunucular ile çalışılırken R'a paket kurulumu yapılamamaktadır. Bu nedenle kurulum sırasında sunucunun en az 1 GB RAM'e sahip olduğundan emin olunmalıdır. *Concerto* paketinin R'a yüklenebilmesi için *session* ve *RMySQL* paketlerine ihtiyaç duyulur.

```
install.packages("session")
```

Komutu girildikten sonra R'a *session* paketi de kurulmuş olur. *RMySQL* paketi de benzer şekilde:

```
install.packages("RMySQL")
```

Komutu ile kurulabilir ancak R'ın bu paketin kurulumunda hata vermesi olasıdır. Bunun nedeni, Ubuntu'da *RMySQL* paketi için gerekli olan bazı yazılımların yüklü olmamasıdır. Eksik yazılımları kurmak için öncelikle R'dan çıkmak gerekmektedir. Bunun için `q()` komutu verilir ve R'dan çıkılarak aşağıdaki komutlar çalıştırılır:

```
sudo apt-get install libmysqlclient-dev
sudo apt-get install libmariadbclient-dev
```

Yazılımların kurulumu tamamlandıktan sonra yeniden R komutu verilerek R'a giriş yapılır ve

```
install.packages("RMySQL")
```

komutu tekrar çalıştırılır.

Daha sonra R'dan `q()` komutu kullanılarak çıkılır. *Concerto*'nun R ile iletişim kurabilmesi için *Concerto* paketleri R'a aktarılır. Bunun için öncelikle *Concerto* dosyalarının yüklendiği dizinde yer alan *lib* klasörü altındaki R klasörüne gidilmesi gerekmektedir. Bu anlatımda *Concerto* dosyaları */var/www/html/test* klasörüne yüklenmişti. Bu nedenle aşağıdaki komut kullanılarak */var/www/html/test/lib/R* klasörüne gidilir.

```
cd /var/www/html/test/lib/R
```

Daha sonra sırasıyla aşağıdaki komutlar çalıştırılır:

```
R CMD build concerto
R CMD INSTALL -l /usr/lib/R/library concerto
```

Yukarıdaki komutların çalıştırılmasının ardından R'a *concerto* paketi yüklenmiş olur.

Concerto için Veri Tabanı Oluşturma

R, veri analizi yaparken *concerto* paketi aracılığı ile bir MySQL veri tabanından veri okumakta ve analiz sonuçlarını MySQL veri tabanına kaydetmektedir. Bu nedenle *Concerto*'nun çalışabilmesi için bir veri tabanına ihtiyaç duyulmaktadır. Veri tabanı oluşturmak için Putty ekranından aşağıdaki komutlar kullanılabilir (Kyngdon, 2015):

```
mysql -u root -p
```

Bu komutun ardından şifre sorulacaktır. Şekil 8'deki MySQL root şifresi girilerek Enter'a basılır.

```
GRANT ALL PRIVILEGES ON *.* TO 'aybek'@'localhost' IDENTIFIED BY 'sifre' WITH GRANT OPTION;
```

Bu komutlar ile *aybek* adında ve şifresi *sifre* olan bir kullanıcı oluşturularak tüm yetkiler bu kullanıcıya verilmiştir. Daha sonra yeni bir veri tabanı oluşturulmalıdır. Bunun için:

```
CREATE DATABASE bobut;
```

komutu girilir. Bu komut *bobut* adında bir veri tabanı oluşturur.

PHP Ayarları

Concerto, PHP ayarlarında *safe mode = OFF*, *magic quotes = OFF*, *short open tags = ON* olmasını gerektirmektedir. Bunun için *php.ini* dosyası düzenlenmelidir. Ancak *safe mode* ayarları PHP 5.4.0 sürümünden itibaren, *magic quotes* ise PHP 7.0.0'dan itibaren kaldırılmıştır. Dolayısıyla yalnızca *short open tags*'ın ON olarak ayarlanması yeterli olacaktır. Bunun için Putty'de

```
nano /etc/php5/apache2/php.ini
```

komutu verilerek Putty üzerinde düzenleme yapılabilir. Düzenlemeden çıkmak için sırasıyla *Ctrl + o* ve ardından *Ctrl + x* tuşlarına basılmalıdır. Başka bir yol ise Filezilla yardımıyla */etc/php5/apache2* klasörüne gitmek ve *php.ini* dosyasının bilgisayara indirilerek metin düzenleyici bir program (not defteri ya da notepad++) ile düzenlenmesidir. Dosya içerisinde *short_open_tag* aranır ve değeri *ON* olarak atanır. Dosya düzenlendikten sonra kaydedilir ve yine Filezilla yardımıyla sunucuya yüklenir. Daha sonra değiştirilen *php.ini* dosyasının aktif olması için Putty kullanılarak *php* kodlarının çalışmasını sağlayan *apache2* servisi yeniden başlatılır:

```
sudo service apache2 restart
```

Concerto Ayarları

Concerto'nun yüklenmesi için tüm hazırlıkların tamamlanmasının ardından sunucuyla ilgili bazı bilgilerin Concerto'nun ayar dosyasında tanımlanması gerekmektedir. Bunun için:

```
nano /var/www/html/test/SETTINGS.php
```

komutu ile dosya düzenlenebilir. Aşağıda bu dosyada değiştirilmesi gereken parametreler yer almaktadır. Aşağıda yer almayan parametrelerde herhangi bir değişiklik yapılmasına gerek yoktur.

```
$db_master_user = "aybek";  
$db_master_password = "sifre";  
$db_master_name = "bobut";  
$path_external = "http://concerto.catirt.com/test/";  
$path_php_exe = "/usr/bin/php5";
```

Değişiklikler yapıldıktan sonra sırasıyla *Ctrl + o* ve *Ctrl + x* tuşlarına basılarak dosya kaydedilip kapatılır. *SETTINGS.php* dosyası sunucu yerine bilgisayarda düzenlenmek istenirse, Filezilla ile */var/www/html/test* klasörüne gidilerek *SETTINGS.php* dosyası bilgisayara indirilir ve not defteri ya da notepad++ gibi bir metin düzenleyici yardımıyla düzenlenebilir. Daha sonra <http://concertonunadresi/setup> yolu izlenir. Bu anlatımda <http://concerto.catirt.com/test/setup> yolu izlenmiştir.

Açılan sayfada varsa hatalar düzeltilir. Ancak rehberdeki adımlar izlendiği takdirde klasör yazma izni hatası dışında bir hata ile karşılaşmadan kurulumun tamamlanması öngörülmektedir. Klasör yazma izni ile ilgili hatanın çözümü sonraki bölümde verilmiştir.

Concerto Kurulumu Sırasında Çıkabilecek Olası Hatalar ve Çözüm Önerileri

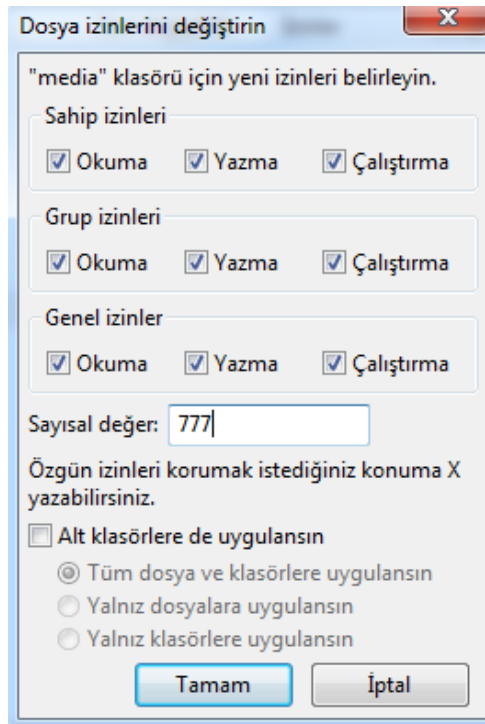
Karşılaşılabilecek ilk hata bazı klasörlerin yazılabilir olmadığına dair hatadır. Concerto, girilen bilgiler yardımıyla sunucuda dosyalar oluşturabilir. Bunun için belirli bazı klasörlerin Concerto tarafından

yazılabilir, yani değiştirilebilir olması gerekmektedir. Örneğin Şekil 16'daki gibi bir hata ile karşılaşmış olsun.

Check for the latest Concerto Platform version	your current version: v4.0.0.beta8 IS UP TO DATE	-
PHP version at least v5.3	your PHP version: 5.5.9-1ubuntu4.14 - PASSED	-
PHP 'safe mode' must be turned OFF	your PHP 'safe mode' is turned OFF - PASSED	-
PHP 'magic quotes' must be turned OFF	your PHP 'magic quotes' is turned OFF - PASSED	-
PHP 'short open tags' must be turned ON	your PHP 'short open tags' is turned ON - PASSED	-
MySQL connection test	Host: localhost, Port: 3306, Login: aybek CONNECTED - PASSED	-
MySQL database connection test	MySQL database aybek IS CONNECTABLE - PASSED	-
MySQL database tables structure test	MySQL database tables structure IS CORRECT - PASSED	-
Rscript file path must be set.	your Rscript file path: /usr/bin/Rscript EXISTS - PASSED	-
PHP executable file path must be set.	your PHP executable file path: /usr/bin/php5 EXIST - PASSED	-
R executable file path must be set.	your R executable file path: /usr/bin/R EXIST - PASSED	-
R version at least v2.15	your R version: 3.2.4 - PASSED	-
media directory path must be writable	your media directory: /var/www/html/test/media/ IS NOT WRITABLE -	Set media directory rights to 0777.

Şekil 16. Klasör Yazma İzni Hatası

Şekil 16'daki hata, /var/www/html/test/media/ klasörünün yazılabilir olmadığını göstermektedir. Hatanın sağ tarafında, klasör izinlerinin 0777 olarak ayarlanması gerektiği belirtilmektedir. Media klasörünün iznini 0777 ayarlamak için Filezilla ya da putty üzerinden komut kullanılabilir. Filezilla üzerinden klasör ayarlarını 0777 yapmak için /var/www/html/test/ klasörüne gidilir. Daha sonra media klasörüne sağ tıklanır ve *Dosya izinlerine* tıklanır. Şekil 17'deki gibi bir pencere açılacaktır. Burada sayısal değer kısmına 777 girildikten sonra Tamam düğmesine tıklanır ve daha sonra Concerto kurulum sayfası yenilenir. Concerto başka klasörler için de yazma izni hatası verebilir. Hata veren tüm klasörlerin yazma izinleri 777 yapılarak kurulum sayfası yenilenir. Araştırmacının kurulum sırasında hata ile karşılaştığı klasörler; *media*, *socks*, *data*, *files* (/var/www/html/test/cms/js/lib/fileupload/php/files), *thumbnails* (/var/www/html/test/cms/js/lib/fileupload/php/thumbnails), *cache* (/var/www/html/test/cms/lib/ckeditor/plugins/pgrfilemanager/PGRThumb/cache) klasörleri olmuştur. Kurulum başlamadan önce bu klasörlerin yazma izinleri 777 olarak ayarlandığı takdirde, kurulum aşamasında herhangi bir yazma izni hatasıyla karşılaşılacaktır.



Şekil 17. Klasör Yazma İzninin Ayarlanması

Filezilla yerine komut kullanarak yazma izinleri değiştirilmek istenirse:

```
sudo chmod 777 /var/www/html/test/media
sudo chmod 777 /var/www/html/test/socks
sudo chmod 777 /var/www/html/test/data
sudo chmod 777 /var/www/html/test/cms/js/lib/fileupload/php/files
sudo chmod 777 /var/www/html/test/cms/js/lib/fileupload/php/thumbnails
sudo chmod 777 /var/www/html/test/cms/lib/ckeditor/plugins/pgrfilemanager/PGRThumb/cache
```

komutları kullanılabilir.

Media, socks, data, files, thumbnails ve cache klasörlerine 777 yazma izni verildikten sonra sayfa yenilenir ve kurulumun başka bir hata vermemesi durumunda Şekil 18'deki sayfa ile karşılaşılır.

test description	test result	recommendation
Check for the latest Concerto Platform version	your current version: v4.0.0.beta8 IS UP TO DATE	-
PHP version at least v5.3	your PHP version: 5.5.9-1ubuntu4.14 - PASSED	-
PHP 'safe mode' must be turned OFF	your PHP 'safe mode' is turned OFF - PASSED	-
PHP 'magic quotes' must be turned OFF	your PHP 'magic quotes' is turned OFF - PASSED	-
PHP 'short open tags' must be turned ON	your PHP 'short open tags' is turned ON - PASSED	-
MySQL connection test	Host: localhost, Port: 3306, Login: aybek CONNECTED - PASSED	-
MySQL database connection test	MySQL database aybek IS CONNECTABLE - PASSED	-
MySQL database tables structure test	MySQL database tables structure IS CORRECT - PASSED	-
Rscript file path must be set.	your Rscript file path: /usr/bin/Rscript EXISTS - PASSED	-
PHP executable file path must be set.	your PHP executable file path: /usr/bin/php5 EXIST - PASSED	-
R executable file path must be set.	your R executable file path: /usr/bin/R EXIST - PASSED	-
R version at least v2.15	your R version: 3.2.4 - PASSED	-
media directory path must be writable	your media directory: /var/www/html/test/media/ IS WRITABLE - PASSED	-
socks directory path must be writable	your socks directory: /var/www/html/test/socks/ IS WRITABLE - PASSED	-
data directory path must be writable	your data directory: /var/www/html/test/data/ IS WRITABLE - PASSED	-
files directory path must be writable	your files directory: /var/www/html/test/cms/js/lib/fileupload/php/files IS WRITABLE - PASSED	-
thumbnails directory path must be writable	your thumbnails directory: /var/www/html/test/cms/js/lib/fileupload/php/thumbnails IS WRITABLE - PASSED	-
cache directory path must be writable	your cache directory: /var/www/html/test/cms/lib/ckeditor/plugins/pgrfilemanager/PGRThumb/cache IS WRITABLE - PASSED	-
concerto R package must be installed.	concerto package IS INSTALLED - PASSED	-

Test completed. Every item passed correctly.

IT IS STRONGLY RECOMMENDED TO DELETE THIS /setup DIRECTORY NOW!

Şekil 18. Concerto Kurulumunun Tamamlanması

Kurulumun tamamlandığına dair uyarı mesajının hemen altında güvenlik nedeniyle setup klasörünün silinmesi gerektiği bilgisi yer almaktadır. Setup klasörü silinebilir ya da daha sonra bazı ayarların değiştirip yeniden kurulum yapılması isterse klasörün adı değiştirilebilir.

Son olarak putty üzerinden sunucuda iki dosyada değişiklik yapılması gereklidir. İlk olarak:

```
sudo visudo
```

komutu verilir. Daha sonra açılan dosyaya aşağıdaki satırlar eklenmelidir.

```
www-data ALL=(%concerto) NOPASSWD: /usr/bin/R
www-data ALL=(%concerto) NOPASSWD: /bin/kill
```

Burada SETTINGS.php dosyasında *\$php_user* değeri değiştirilmediği için *www-data* olarak girilmektedir. Bu değer değiştirildiği takdirde *www-data* yerine *\$php_user* değerinin yazılması gerekmektedir. */usr/bin/R* sunucuda R'in çalıştırılacağı yoldur. R yukarıdaki adımlarla kurulduysa bu yolun değiştirilmesine gerek yoktur. SETTINGS.php dosyasında *\$r_users_group* değeri değiştirilmediyse *concerto* kısmı değiştirilmemelidir. Bu dosyada *Defaults requiretty* ifadesi varsa

silinmelidir. Satırlar eklendikten sonra *Ctrl + o* komutuna basılarak dosya kaydedilir ve *Ctrl + x* komutu ile metin düzenlemeden çıkılabilir.

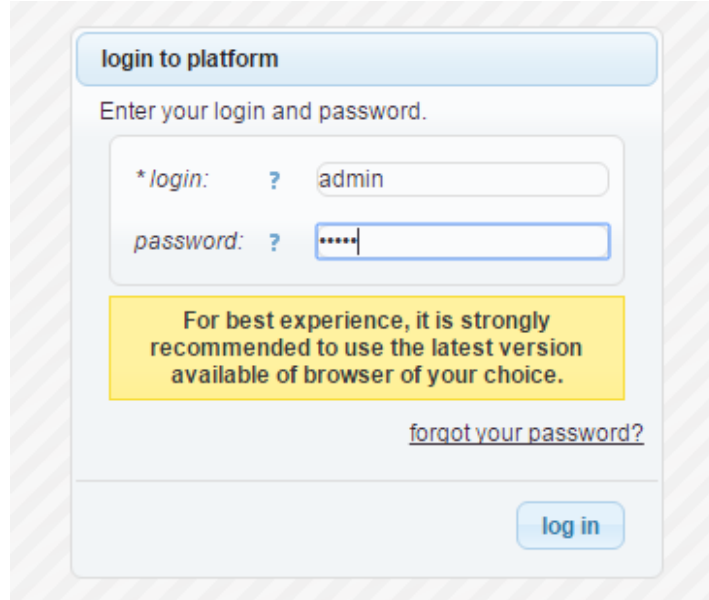
Ardından putty’de root olarak oturum açılmış durumdayken

```
crontab -e
komutu girilir. Bu dosyaya aşağıdaki satırlar eklenir.
* * * * * /usr/bin/php /var/www/html/test/cron/users.php
0 0 1 * * /usr/bin/php /var/www/html/test/autocomplete.php
```

/var/www/html/test bölümü Concerto’nun yüklediği klasör yolu ile değiştirilmelidir. Bu dosya da *Ctrl + o* kaydedilir ve *Ctrl + x* yardımıyla dosyadan çıkarılır.

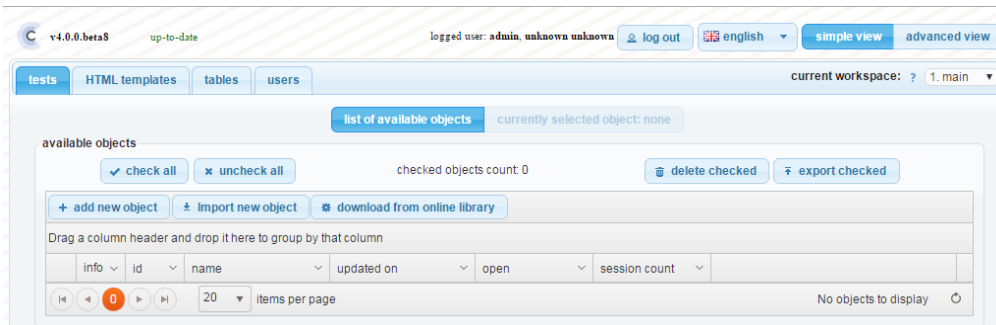
Concerto Yönetim Paneli

Concerto’nun kurulumu tamamlandıktan sonra <http://ipadresini/cms> yolu izlenebilir. Bu anlatımda <http://concerto.catirt.com/test/cms> yolu izlenmiş ve Concerto’nun yönetim paneline giriş yapılmıştır. Yönetim paneline ilk girişte kullanıcı adı: *admin* ve şifre: *admin* kullanılmalıdır. Yönetim paneli giriş ekranı Şekil 19’da yer almaktadır.



Şekil 19. Concerto Yönetim Paneli Giriş

Concerto yönetim paneline giriş yapıldıktan sonra tests, HTML templates, tables ve users sekmeleri ile karşılaşmaktadır. Tests sekmesinden R kodları ile test algoritması oluşturulmakta, HTML templates sekmesinden R ile kullanıcı arasındaki arayüz tasarlanmakta, tables sekmesinden veritabanları yönetilmekte ve Users sekmesinden kullanıcılar ile ilgili işlemler yapılmaktadır (Şekil 20).



Şekil 20. Concerto Yönetim Paneli Anasayfası

İlk kullanım için basit bir BOBUT uygulaması Concerto sunucularından alınmıştır. Bunun için download from online library düğmesine tıklanır. Açılan pencereden çeşitli test dosyaları indirilebilir (Şekil 21). Bu test dosyaları, test geliştiricileri tarafından kütüphaneye yüklenmiş olup bazı test dosyaları doğru şekilde çalışmayabilir. Bu çalışmada, yazar tarafından denenmiş ve çalıştığına emin olunan test dosyası kullanılmıştır. Bu dosyalar araştırmacılara örnek oluşturması bakımından önemli olup hazır test paketleri olarak anlaşılmalıdır. Araştırmacılar bu dosyayı örnek alarak kendi canlı BOBUT uygulamalarını geliştirmelidir.

info	id	name	author	revision	upload date	download count
?	306	cat	審秋山 / Concerto CAT コミュニティ	1	2014-10-22 08:20:11	120
?	222	survey	Arthur Halma / none	1	2014-03-10 12:35:17	37
?	297	cat	審秋山 / Concerto CAT コミュニティ	1	2014-09-20 20:16:59	37
?	352	Big Five	Ahmed El-Boukhary / ASSESS TM	1	2015-04-27 06:09:55	19
?	207	Test 1	unknown unknown /	1	2014-02-04 08:10:13	18
?	344	survey	super user / UB	1	2015-04-01 19:43:58	17
?	278	survay_tutorial	審秋山 / Concerto CAT コミュニティ	1	2014-08-15 17:14:44	12
?	291	cat	審秋山 / Concerto CAT コミュニティ	1	2014-09-09 15:11:05	11
?	273	survay_tutorial	審秋山 / Concerto CAT コミュニティ	1	2014-08-15 17:08:14	10
?	360	PHQ	unknown unknown /	1	2015-06-04 13:41:36	9

Şekil 21. Concerto Çevrim İçi Kütüphanesi'ndeki Test Örnekleri

Download count sekmesine tıklanarak örnek testler indirilme sayısına göre sıralanabilir. En çok indirilen 306 id numaralı testi indirmek için indirilme sayılarının sağ tarafındaki sütunda yer alan çark simgesine tıklanır ve test dosyası indirilir. Çevrim içi kütüphaneye yüklenen dosyalara sabit bir id numarası atanmaktadır. Bu işlem yalnızca teste ait R kodunu indirmektedir. Testin bir bütün olarak çalışabilmesi için Concerto Çevrim İçi Kütüphanesi'nden HTML Templates ve tables sekmelerine giderek uygun olan dosyaların indirilmesi gerekir. Bunun için hem HTML Templates sekmesinden hem de Tables sekmesinden download from online library düğmesine ayrı ayrı tıklanması gerekmektedir.

HTML Templates için 298, 299 ve 300 id numaralı dosyalar indirilmelidir. Tables içinse 301 ve 302 id numaralı veri tabanları indirilmelidir. İndirme işlemi bittiğinde tests sekmesine dönülür ve test algoritmasında küçük değişiklikler yapılır (Şekil 22).

```

1 library(catR)
2 responses<-NULL
3 qasked<-NULL
4 theta<-runif(1,min=-1,max=1)
5 result1<-concerto.template.show(templateID=1,params=list(),finalize=F,workspaceID=concerto$workspaceID)
6 studentID<-result1$studentID
7 session_id<-result1$session_id
8 result2<-concerto.table.query(sql=paste("SELECT * FROM concerto_1.`itembank`",sep=""),params=list())
9 a<-as.numeric(result2$a)
10 b<-as.numeric(result2$b)
11 ibank<-cbind(a,b,0,1)
12 question<-result2$question
13 correct<-as.numeric(result2$correct)
14 opt1<-result2$opt1
15 opt2<-result2$opt2
16 opt3<-result2$opt3
17 opt4<-result2$opt4
18 for(j in 1:20){
19   item<-nextItem(ibank,model=NULL,theta=theta,out=qasked,criterion="MEI")
20   i<-item$item
21   qasked<-c(qasked,i)
22   result3<-concerto.template.show(templateID=2,params=list(question=question[i],correct=correct[i],opt1=opt1[i],opt2=opt2[i],opt3
=opt3[i],opt4=opt4[i]),finalize=F,workspaceID=concerto$workspaceID)
23   score<-as.numeric(result3$radio)==correct[i]
24   responses<-c(responses,score)
25   theta<-thetaEst(ibank[qasked,],responses,method="EAP")
26   concerto.table.query(sql=paste("INSERT INTO concerto_1.`responses` SET `studentID`=",dbEscapeStrings(concerto$db$connection,
toString(studentID)),"`,`responses`=",dbEscapeStrings(concerto$db$connection,toString(responses)),"`,`session_id`=",dbEscapeStr
ings(concerto$db$connection,toString(concerto$test_session_id)),"`,`items_asking`=",dbEscapeStrings(concerto$db$connection,toString
(qasked)),"`,`response_time`=",dbEscapeStrings(concerto$db$connection,toString(concerto$time_taken)),"`,`theta`=",dbEscapeStr
ings(concerto$db$connection,toString(theta)),"`,`sep`=",sep=""),params=list())
27 }
28 result4<-concerto.template.show(templateID=3,params=list(studentID=studentID,theta=theta),finalize=F,workspaceID=concerto$workspa
ceID)

```

Şekil 22. Örnek Test Algoritması

Yapılması gereken değişiklikler sırasıyla şöyledir:

1. 5. satırda yer alan templateID numarası, HTML templates sekmesinde yer alan start şablonunun ID numarası olmalıdır.
2. 8. satırda yer alan 'cd_3'.itembank' ifadesinde cd_3; concerto_1 ile değiştirilmelidir. Bunun nedeni, SETTINGS.php'de aksini belirtmedikçe Concerto'nun otomatik olarak concerto_ ön eki ile veri tabanlarını oluşturması ve ilk veri tabanı olarak concerto_1'i oluşturmuş olmasıdır.
3. 22. satırda yer alan templateID, HTML templates sekmesinde yer alan question şablonunun ID numarası olmalıdır.
4. 26. satırda yer alan 'cd_3'.itembank' ifadesinde cd_3, concerto_1 ile değiştirilmelidir.
5. 28. satırda yer alan templateID, HTML templates sekmesinde yer alan result şablonunun ID numarası olmalıdır.

Bu değişikliklerin yapılmasının ardından begin düğmesine tıklanarak ilk BOBUT örneği çalıştırılabilir. Bu örneğe ait ekran görüntüsü Şekil 23'te yer almaktadır. Ayrıca verilen yanıtlar incelenmek istenirse *phpmyadmin* üzerinden ya da *tables* sekmesinden *results* tablosu incelenerek yanıtlar ve kestirimler incelenebilir.

<p>Let's Start CAT</p> <p>Enter your student ID <input type="text"/></p> <p><input type="button" value="START"/></p> <p>1</p>	<p>Answer the question:</p> <p>[a=1.444] [b=0.2] 73+10=?</p> <p><input type="radio"/> 83</p> <p><input type="radio"/> 85</p> <p><input type="radio"/> 86</p> <p><input checked="" type="radio"/> 88</p> <p><input type="button" value="SUBMIT"/></p> <p>2</p>	<p>Your test result</p> <p>Your student ID : eren</p> <p>Your ability : 3.50130599870637</p> <p><input type="button" value="FINISH"/></p> <p>3</p>
---	---	--

1. Giriş; 2. Madde; 3. Sonuç Sayfaları

Şekil 23. Canlı BOBUT Uygulaması Ekran Görüntüsü Örneği

Şekil 23'te görülebileceği üzere BOBUT uygulaması R kodları ile geliştirilmekte ve çalıştırılmaktadır. Her araştırmacı, kendi durumuna uygun olan kodları, örnek testlerden de yararlanarak geliştirebilir. Teste başlama ve testi bitirme zamanı, IP numarası, her maddede geçirilen süreler vb. ek bilgiler veri tabanına kaydedilebilir. Bu noktada testin sahip olacağı özellikler, test geliştiricisinin R bilgisi ile sınırlı olacaktır.

Programın Ulaşılabilirliği

Program açık kaynaklı olarak sunulmakta ve <http://www.psychometrics.cam.ac.uk/newconcerto> adresinden indirilebilmektedir. Dileyen araştırmacılar site üzerinden demo hesabı açarak Concerto'yu deneyebilirler ya da bu raporda gösterilen adımları kullanarak kendi sunucularına kurarak BOBUT uygulamalarını gerçekleştirebilirler.

KAYNAKÇA

- Attali, D. (2015). How to set up R on Ubuntu 14.04. 15 Mart 2015 tarihinde <https://www.digitalocean.com/community/tutorials/how-to-set-up-r-on-ubuntu-14-04> adresinden alınmıştır.
- Aybek, E.C. (2016). *Kendini Değerlendirme Envanteri'nin bilgisayar ortamında bireye uyarlanmış test (BOBUT) olarak uygulanabilirliğinin araştırılması*. (Doktora Tezi, Ankara Üniversitesi, Ölçme ve Değerlendirme Anabilim Dalı, Ankara).
- Betz, N. E., & Turner, B. M. (2011). Using item response theory and adaptive testing in online career assessment. *Journal of Career Assessment, 19*(3), 274–286. doi:10.1177/1069072710395534
- Chien, T.W., Lai, W.P., Lu, C.W., Wang, W.C., Chen, S.C., Wang, H.Y., & Su, S.B. (2011). Web-based computer adaptive assessment of individual perceptions of job satisfaction for hospital workplace employees. *BMC medical research methodology, 11*, 47.
- Chien, T.W., Wang, W.C., Wang, H.Y., & Lin, H.J. (2009). Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC health services research, 9*, 135.
- Concerto. (2016). Concerto v4 installation and updating. 19 Mart 2016 tarihinde <http://concertoplatform.com/legacy/installation4.html> adresinden alınmıştır.
- Elhan, A. H., Öztuna, D., Kutlay, Ş., Küçükdeveci, A. A., & Tennant, A. (2008). An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC musculoskeletal disorders, 9*(1). doi: 10.1186/1471-2474-9-166
- Ellingwood, J. (2014). How to install and secure phpMyAdmin on Ubuntu 14.04. 15 Mart 2015 tarihinde <https://www.digitalocean.com/community/tutorials/how-to-install-and-secure-phpmyadmin-on-ubuntu-14-04> adresinden alınmıştır.
- Gardner, W., Shear, K., Kelleher, K. J., Pajer, K. A., Mammen, O., Buysse, D., & Frank, E. (2004). Computerized adaptive measurement of depression: a simulation study. *BMC Psychiatry, 4*(13). doi: 10.1186/1471-244X-4-13
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2012). Development of a computerized adaptive test for depression. *Archives of General Psychiatry, 69*(11), 1104–12.
- Haley, S. M., Coster, W. J., Andres, P. L., Kosinski, M., & Ni, P. (2004). Score comparability of short forms and computerized adaptive testing: simulation study with the activity measure for post-acute care. *Archives of physical medicine and rehabilitation, 85*(4), 661-666.
- Haley, S. M., Ni, P., Jette, A. M., Tao, W., Moed, R., Meyers, D., & Ludlow, L. H. (2009). Replenishing a computerized adaptive test of patient-reported daily activity functioning. *Quality of Life Research, 18*(4), 461-471.
- Hol, A. M., Vorst, H. C. M., & Mellenbergh, G. J. (2007). Computerized adaptive testing for polytomous motivation items: Administration mode effects and a comparison with short forms. *Applied Psychological Measurement, 31*(5), 412–429.
- Kyngdon, A. (2015). How to install Concerto v4.0.0 for computer adaptive testing. 10 Mart 2016 tarihinde <https://www.youtube.com/watch?v=GallSyKkjdo> adresinden alınmıştır.
- Smits, N., Cuijpers, P., & van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry research, 188*(1), 147–55. doi:10.1016/j.psychres.2010.12.001
- Sveldrov, E. (2012). How to set up Apache virtual hosts on Ubuntu 12.04 LTS. 20 Mart 2015 tarihinde <https://www.digitalocean.com/community/tutorials/how-to-set-up-apache-virtual-hosts-on-ubuntu-12-04-lts> adresinden alınmıştır.

EXTENDED ABSTRACT

Introduction

When the studies about Computerized Adaptive Testing (CAT) are examined, real data based post-hoc simulations (Gardner et.al, 2004; Haley, Coster, Andres, Kosinski & Ni, 2004; Haley et.al, 2009; Smits, Cuijpers & van Straten, 2011; Betz & Turner, 2011) are found. In addition to post-hoc simulations; the number of the studies with live CAT applications (Hol, Vorst & Mellenbergh, 2007; Elhan, Öztuna, Kutlay, Küçükdeveci & Tennant, 2008; Gibbons et.al, 2008; Chien, Wang, Wang & Lin, 2009; Chien, Lai, Lu, Wang, Chen, Wang & Su, 2011) increases day-by-day. In this report, a brief introduction to various live CAT software has been presented and installation steps of Concerto, an open source live CAT software, to a Linux server have been shown.

Besides the commercial live CAT software like FastTest, Pearson VUE, Prometric and McCann, there are also open source and free alternatives like Concerto, IRT-CAT and OSCATS. Additional information about commercial software can be obtained from the official websites. Information of open source live CAT software like Concerto, IRT-CAT and OSCATS are presented below.

Practical CAT Software

IRT-CAT

IRT-CAT is a web-based software that supports 3PL and Rasch models. It is licensed under Academic Free License and latest update of the software was in 2009.

Web: <https://sourceforge.net/projects/irt-cat>

OSCATS

OSCATS' latest update was in 2011 and it's licensed under General Public License (GPL). It can be used for post-hoc simulations and live CAT applications. It supports 1PL, 2PL, 3PL, Nominal Response Model, Partial Credit Model, Generalized Partial Credit Model and Graded Response Model, also it supports six item selection method but only let fixed-item-length as a stopping rule. Users can use Maximum Likelihood and Expected a Posteriori ability estimation methods, and developers plan to support R and Matlab syntax in the future.

Web: <https://code.google.com/archive/p/oscats/wikis/Features.wiki>

Concerto

Concerto is developed by Cambridge University, Psychometrics Center and it's an open-source, R-based software. Concerto runs on web environment and provides a real life CAT framework to test developers. It is basically saves the R output to MySQL database and shows the user by recall the output from MySQL database. Concerto has test, HTML template and table modules. Users can develop their CAT algorithm with R script in test module. HTML templates module has HTML layouts which test-takers see; and table module has the databases that keep items, responses and R outputs.

Concerto shows the R outputs to test-takers via database; and transmit the test-takers' input to R via database and run the test process. That's why, Concerto basically a bridge between MySQL, R and test-takers. The biggest advantage of Concerto is to let the test developer, develop his/her test with various item response theory models, item selection and ability estimation methods and stopping rule. In this paper the steps of installation of Concerto to a Linux server have been shown.

Requirements

Requirements for install and run Concerto v4 are listed below:

1. Linux operating system
2. PHP v.5.3 or newer
3. php-process package (It may be needed for some Linux distros like CentOS)
4. PHP safe mod: OFF

5. PHP magic quotes: OFF
6. PHP short open tags: ON
7. PHP open base dir: ON
8. MySQL v.5 or newer
9. R v.2.15 or newer
10. concerto R package must be installed
11. a Linux user with root privileges
12. a MySQL user with all permissions granted
13. Minimum 1GB RAM

In this report, Concerto have been installed to servers of Digitalocean (www.digitalocean.com) company and explained according to Digitalocean servers. But Concerto may be install with same steps to server which run Ubuntu Linux 14.04 x64 operating system.

Accesibility off the program

Concerto is open source and downloadable from <http://www.psychometrics.cam.ac.uk/newconcerto> URL. Researchers can create a demo account on Concerto website to evaluate the software or install their own servers by following the steps presented this paper.

Results and Discussion

Concerto provides practical and flexible solution to researchers who wants to develop real life CAT applications. The biggest advantages of Concerto are being open-source and letting researchers develop their tests as they wish.

Eğitsel Yazılım Değerlendirme Formlarının ve Ölçeklerinin İncelenmesi*

Review of the Educational Software Evaluation Forms and Scales

Ahmet ARSLAN **

Ali GERİŞ ***

Taibe KULAKSIZ****

Öz

Bu çalışmanın amacı eğitsel yazılımları değerlendirmek için oluşturulan ölçeklerin ve formların incelenmesini ve değerlendirmesini yaparak gelecek çalışmalar için önerilerde bulunmaktır. Bu amaç doğrultusunda içerik analizi yöntemi kullanılarak “Education Resources Information Center (ERIC)”, “Marmara Üniversitesi E-Kütüphane” “YÖK Ulusal Tez Merkezi” ve “ScienceDirect” veri tabanlarında “Eğitsel Yazılım Değerlendirme”, “Yazılım Değerlendirme” gibi anahtar kelimeler ile tarama yapılmış ve içeriğinde eğitsel yazılım değerlendirme aracı bulunan 42 çalışmaya ulaşılmıştır. Belirlenen ölçütler neticesinde bu çalışmalardan 29 tanesi incelemeye alınmış ve değerlendirmelerde bulunulmuştur. Bu incelemelerin neticesinde eğitsel yazılım değerlendirme araçlarının geliştirilmesinde 2006 – 2010 yılları arasında önceki yıllara göre artış olduğu buna karşın eğitsel oyunlar gibi önemli bir alanda yeterli değerlendirme aracının bulunmadığı dikkati çekmektedir. Eğitsel yazılım değerlendirme araçlarının geliştirilmesindeki geçerlik güvenirlik çalışmalarında da ortak bir yöntem anlayışının olmaması göz önüne alındığında gelecek çalışmalarda söz konusu hususa dikkat edilmesi gerektiği sonucuna varılmıştır.

Anahtar Kelimeler: yazılım değerlendirme, eğitsel yazılım değerlendirme, yazılım değerlendirme formları

Abstract

The main purpose of this study is to review existing evaluation forms and scales that have been prepared for educational software evaluation. In addition to this purpose, the study aims to provide insight and guidance for future studies in this context. In total, forty-two studies that including evaluation forms and scales have been taken into consideration. “Educational software evaluation”, “Software evaluation”, “Educational software evaluation forms/scales” were searched as keywords in the: “Education Resources Information Centre (ERIC)”, “Marmara University e-Library”, “National Thesis Center” and “Science Direct” databases. Twenty-nine of them have met the review selection criteria and been evaluated. There is an increase in the number of evaluation tools between 2006 – 2010. However, it was noticed that there is no sufficient number of evaluation tools targeting “educational games”. It was concluded that reliability and validity studies are very important part of developing educational software evaluation tools and this is a matter that should be considered in future studies.

Keywords: key software evaluation, educational software evaluation, software evaluation forms

GİRİŞ

Bilgisayarların birçok alanda kullanımı ile önemli noktaya gelen yazılımların eğitim amaçlı hazırlanmış olanlarına “Eğitsel Yazılım” adı verilmektedir (Kelleci, 2010). Avcı (2010) da eğitim hizmetlerinde kullanılması amacıyla geliştirilen tüm yazılımların eğitsel yazılım olarak nitelendirildiğini belirtmiştir. Bilgisayar temelli öğrenme ortamları ve eğitsel yazılımlar; ilköğretim, ortaöğretim ve lise eğitimi seviyesinde olduğu kadar uzaktan eğitim veren yükseköğretim gibi kurumlar tarafından da sıklıkla kullanılmaktadır (Ateş, 2011). Kullanılan eğitsel yazılımların, uygulanacak programın hedeflerine yönelik olarak ve öğretme - öğrenme ilkeleri temel alınarak geliştirilen standartlara uygun olarak hazırlanmaları gerekmektedir (Güzeller ve Korkmaz, 2007). Kara

*Bu çalışmanın bir kısmı 4-6 Aralık 2015 tarihinde Prag’da düzenlenen “Multidisciplinary Academic Conference on Education, Teaching and Learning” isimli konferansta sunulmuştur.

**Yard. Doç. Dr. Marmara Üniversitesi, Atatürk Eğitim Fakültesi, BÖTE, İstanbul, Türkiye, aarslan@marmara.edu.tr

*** Arş. Gör. Marmara Üniversitesi, Atatürk Eğitim Fakültesi, BÖTE, İstanbul, Türkiye, ali.geris@marmara.edu.tr

**** Arş. Gör. İstanbul Aydın Üniversitesi, Eğitim Fakültesi, BÖTE, İstanbul, Türkiye, taibekulaksiz@gmail.com

(2009), yaptığı çalışmada yetersiz düzeyde olan yazılımlar ile yapılan öğretim faaliyetlerinden olumlu sonuçlar elde etmenin beklenmemesi gerektiğini belirtmiştir.

Eğitsel yazılımlar için gerekli olan geliştirme unsurları göz önünde bulundurulmadan hazırlanan yazılımların öğretim amacına, öğrenciye, öğretmene ve öğretme süreçlerine karşı birçok açıdan olumsuz sonuçlar gösterebileceği de dikkate alınmalıdır (Tankut, 2008). Potansiyel olumsuz sonuçlar göz önüne alındığında, özellikle eğitsel ders yazılımlarının çok geniş kapsamlı olarak değerlendirilmeleri ve kullanılmadan önce bu değerlendirmelerin göz önünde bulundurulmalarının gerekliliği ön plana çıkmaktadır (Gül ve Yeşilyurt, 2011). S.H. Lee, Choi ve Byun (1996) da eğitsel yazılımların doğru değerlendirilmesinin ve seçilmesinin, yazılımların başarılı olabilmelerinin en büyük etkeni olduğunu vurgulamıştır.

Eğitsel yazılım değerlendirme, yazılımlarının öğrencilere verdiği öğrenme imkânlarının ne düzeyde etkili olduğunu incelenmesidir (Kelly, 2008). Çeliköz ve Erişen (2010) ise eğitsel yazılım değerlendirmeyi, yazılımların belirli niteliklerdeki mevcut düzeyleri ile olması beklenen düzey arasındaki ilişkinin tespiti olarak tanımlamışlardır. Bu ifadelerden yola çıkarak eğitsel yazılımların öğrencilerin öğrenme gereksinimlerini karşılama, belirlenen öğrenme amacına uygun olarak deneyimler sunmasının ve bunu etkili olarak gerçekleştirebilmesinin değerlendirilmede önemli olduğu düşünülmektedir (Ateş, 2011).

Eğitsel yazılım değerlendirme alanında yapılmış birçok çalışmada farklı unsurlar üzerinde durulmuş ve bu unsurlar üzerinden değerlendirmeler yapılmıştır. Şimşek (1998), yazılım değerlendirme kriterlerini; kapsamlılık, ayırıcılık, nesnellik, esneklik, gereksinimlere duyarlılık ve koşullara uygunluk olarak sıralamıştır. Eğitsel yazılımlarda uygunluk, işlerlik, etkililik ve verimlilik ölçütlerinin birer kalite göstergesi olduğunu ise 2010 yılında yaptıkları çalışmada Çeliköz ve Erişen söylemiştir. Arslan (2006) gerçekleştirdiği çalışmada eğitsel yazılımların “Tasarım Değerlendirmesi” ve “Bütünselliğin Değerlendirilmesi” olarak iki ana ekseninde toplandığı sonucuna varmıştır.

Yapılan farklı çalışmaların sonucunda oluşturulan değerlendirme formlarının veya ölçeklerinin değerlendirme aşamasında olduğu gibi farklı kriterleri temel aldığı belirlenmiştir. Ateş (2011) oluşturduğu eğitsel yazılım değerlendirme formunu altı boyut kapsamında ele alırken Gülbahar ve Tınmaz (2006) dört boyut kapsamında kendi değerlendirme formlarını oluşturmuşlardır. Alan yazında bir çok farklı boyutta ve kapsamda oluşturulan eğitsel yazılım değerlendirme formlarına ulaşmak mümkündür.

Cohen (1983), eğitsel ders yazılımlarını değerlendirmek için alan yazın üzerinde henüz ortak bir karara varılmamış olduğunu belirtmiştir. Sim ve diğerleri (2006) bu kadar çeşitli ve farklı yaklaşımların olmasının, eğitsel yazılım değerlendirmede ortak bir metodolojinin tam olarak geliştirilemediği şeklinde yorumlanabileceği gibi eğitsel yazılım değerlendirmede farklı odak noktalarının olduğu ve farklı bakış açılarının değerlendirmeye çeşitlilik kazandırdığı gibi de yorumlanabileceğini belirtmişlerdir. Değerlendirmenin yapılacağı eğitsel yazılımların amaçlarına yönelik olarak alan yazın taraması yapılarak, uzman görüşleri temel alınarak vb. çeşitli yöntemler kullanılarak hazırlanan bu formlarda var olan mevcut durumun incelenmesi, eksiklerin ortaya konulması ve gelecekte yapılacak olan eğitsel yazılım değerlendirme araçlarına yönelik önerilerde bulunulması amacıyla erişilebilen eğitsel yazılım değerlendirme formlarının ve ölçeklerinin incelenmeleri araştırmaya değer bir konu olarak görülmüştür.

Araştırmanın Amacı

Alan yazın taraması neticesinde ve ele alınan problem ışığında çalışmanın amacı, “1996 – 2015 yılları arasında ülkemizde yapılan eğitsel yazılım değerlendirme araçlarının değerlendirilmesi ve incelenmesi” olarak belirlenmiştir. Bu esas amaç doğrultusunda eğitsel yazılım değerlendirme araçlarının,

1. Geliştirildiği yıllara göre dağılımı nasıldır?
2. Türleri nelerdir?

3. Hazırlanma süreçleri nasıldır?
4. Eğitsel yazılım değerlendirme araçlarında geçerlik ve güvenilirlik çalışması yapılmı durumu nasıldır?
5. Eğitsel yazılım değerlendirme araçlarında puanlama türleri nasıldır?
6. Eğitsel yazılım değerlendirme araçlarının hazırlandığı yazılım türleri nelerdir?
7. Eğitsel yazılımı değerlendirme araçları hangi kullanıcı kitlesinin değerlendirme yapmasına uygun olarak hazırlanmıştır?
8. Eğitsel yazılımı değerlendirmede araçlarındaki boyut sayısına göre dağılımları nasıldır?
9. Eğitsel yazılımı değerlendirmede araçlarındaki madde sayısına göre dağılımları nasıldır?

sorularına yanıt aranmıştır.

YÖNTEM

Bu araştırmada eğitsel yazılım değerlendirme formlarını incelemek ve önerilerde bulunmak amacıyla içerik analizi yöntemi kullanılmıştır. İçerik analizi sözel, yazılı ve diğer materyallerin nesnel olarak değerlendirilmesini, sistematik bir inceleme yoluyla ele alınmasını sağlayan bilimsel bir yaklaşımdır (Sert, Kurtoğlu, Akıncı ve Seferoğlu, 2012). İçerik analizinde, toplanan kaynaklardan elde edilen nitel araştırma verilerinin işlenmesi, verilen kodlanması ve temaların bulunması, elde edilen temaların düzenlenmesi ve yorumlanması olarak dört temel aşama bulunmaktadır (Yıldırım ve Şimşek, 2006). Bu çalışma kapsamında da içerik analizinin temelleri esas alınmış ve sonuçlara ulaşılmıştır.

Tarama ve Seçim Ölçütleri

Çalışma kapsamında kullanılacak değerlendirme araçlarını belirlemek amacıyla internet üzerinden erişimle ilgili arama ortamı ve anahtar kelimelere ilişkin ölçütler belirlenmiştir. Buna bağlı olarak “Education Resources Information Center (ERIC)”, “Marmara Üniversitesi E-Library”, “ScienceDirect”, “Google Scholar” ve “Ulusal Tez Merkezi” veri tabanlarında “Eğitsel Yazılım Değerlendirme”, “Web Yazılım Değerlendirme”, “Yazılım Değerlendirme”, “Eğitsel Yazılım Değerlendirme Ölçeği/Formu/Anketi” anahtar kelimeleri kullanılarak tarama yapılmıştır.

Belirtilen arama kriterlerine göre yazarlar tarafından bireysel olarak toplanan çalışmalar ortak bir incelemeye alınmıştır. Bu inceleme neticesinde toplanan çalışmalardan aynı olanlar elenmiştir. Bu işlemler neticesinde toplam 42 çalışmaya ulaşılmıştır ve bunlar içerisinde eğitsel yazılım değerlendirme formu bulunmayan 11 ve dili Türkçe olmayan 2 tane olmak üzere toplam 13 çalışma kapsam dışı tutulmuştur. Ardından yapılan inceleme nihayetinde çalışmada kullanılan 5 formun ikincil kaynak (özgün halinden alıntılanan veya uyarlanan) olduğu tespit edilmiştir. Ancak birincil kaynaklara (çalışmanın özgün hali) tarama sonucunda ulaşılamadığı için formlar var olan çalışmalar üzerinden değerlendirmeye alınmıştır. Bu incelemeler neticesinde Tablo 1’de tam listesi verilen toplam 29 eğitsel yazılım değerlendirme aracı çalışma kapsamına dâhil edilmiştir. Geçerlik ve güvenilirliği sağlamak amacıyla çalışmanın yazarları tarafından akran değerlendirmesi, uzman incelemesi, derinlemesine veri toplama ve analiz sürecinin anlatılması işlemleri gerçekleştirilmiştir.

Tablo 1. İncelemeye Alınan Çalışmalar

Sıra	Deęerlendirme Aracının İsmi	Çalışma Türü	Referans
1	Ders Yazılımları Deęerlendirme Formu	Yüksek Lisans Tezi	Çeliköz (1996)
2	Yazılım Kimlik ve Deęerlendirme Formu	Bildiri	Öztürk ve İnan (1998)
3	Öğretimsel Web Sitesi Analiz Formu	Yüksek Lisans Tezi	Arslan (2002)
4	Öğretimsel Web Sitesi Analiz Formu	Bildiri	Arslan (2004)
5	Eğitsel Yazılım Ölçeęi	Yüksek Lisans Tezi	Demir (2004)
6	Web Yazılımı Deęerlendirme Formu	Doktora Tezi	Erdoğan (2005)
7	Yazılım Deęerlendirme Kontrol Listesi	Kitap	Kaya (2006)
8	Eğitsel Yazılımı Deęerlendirme Anketi	Yüksek Lisans Tezi	Güneş (2007)
9	Öğretim Amaçlı Web Sitelerini Deęerlendirme Ölçeęi	Bildiri	Yiğit, Altun, Alev, Dertlioęlu, ve Bülbül (2007)
10	BDÖ Materyal Deęerlendirme Formu	Makale	Kutluca ve Birgin (2007)
11	Eğitim Yazılımı Deęerlendirme Formu	Makale	Güzeller ve Korkmaz (2007)
12	Eğitim Yazılımı Deęerlendirme Formu A	Yüksek Lisans Tezi	Tankut (2008)
13	Eğitim Yazılımı Deęerlendirme Formu B	Yüksek Lisans Tezi	Tankut (2008)
14	Örütbaę Sitelerini Deęerlendirme Ölçeęi	Makale	Yiğit, Bütüner, ve Dertlioęlu (2008)
15	BDÖ Materyali Deęerlendirme Formu	Bildiri	Akbulut, Akdeniz, ve Dinçer (2008)
16	Bilgem Öğretim Yazılımı Deęerlendirme Formu	Yüksek Lisans Tezi	Şakar (2008)
17	Eğitim Yazılımları Deęerlendirme Ölçeęi	Makale	Kara (2009)
18	Memnuniyet Kontrol Listesi	Yüksek Lisans Tezi	Avcı (2010)
19	Eğitsel Yazılım Deęerlendirme Formu - I	Yüksek Lisans Tezi	Kelleci (2010)
20	Eğitsel Yazılım Deęerlendirme Formu - II	Yüksek Lisans Tezi	Kelleci (2010)
21	Yazılım Deęerlendirme Formu	Makale	Gül ve Yeşilyurt (2011)
22	E-Öğrenme Materyalinin Kullanılıřılıęı Anketi	Doktora Tezi	Özgür (2011)
23	E-Öğrenme Materyali Uzman Görüşü Formu	Doktora Tezi	Özgür (2011)
24	Eğitsel Yazılım Deęerlendirme Formu	Makale	Ateş (2011)
25	WTEM Deęerlendirme Ölçeęi	Makale	Erümit (2013)
26	Eğitsel Yazılımların Program Deęerlendirme Yaklaşımına Göre Olusturulan Deęerlendirme Kriterleri	Bildiri	Yıldız ve Sarıtepeci (2013)
27	Eğitsel Yazılımların Deęerlendirilme Kriterleri	Bildiri	Yıldız ve Sarıtepeci (2013)
28	Öğrenme Nesnesi Deęerlendirme Ölçeęi	Makale	Gürer ve Yıldırım (2014)
29	Açık Ders Malzemeleri Platformları Deęerlendirme Ölçeęi	Yüksek Lisans Tez	Aksaya (2014)

Toplam 29 makale ve tez ile çalışma devam ettirilerek bu çalışmaların “Yapıldığı Yıl”, “Deęerlendirme Aracı Türü”, “Yazılım Türü”, “Hazırlık Süreci”, “Deęerlendirme Aracı Puanlama Türü”, “Deęerlendiren Hedef Kitlesi”, “Geçerlik Güvenirlik”, “Boyut Sayısı”, “Madde Sayısı” bağlamında dağılımları incelenmiştir.

BULGULAR

Araştırma kapsamında incelenen 29 çalışma ile ilgili belirtilen problem durumlarına göre tablolar oluşturulmuş ve deęerlendirilmelerde bulunulmuştur.

Yapıldıkları Yıllara Göre Çalışmaların Dağılımı

Eğitsel yazılım deęerlendirme aracı geliştirme çalışmalarının dağılımları yıllarına göre gruplandırılarak Tablo 2’de verilmiştir.

Tablo 2. Yapıldıkları Yıllara Göre Çalışmaların Dağılımı

Yıl Aralığı	Çalışma Sayısı	%
1996 – 2000	4	13,8
2001 – 2005	6	20,7
2006 – 2010	13	44,8
2011 – 2015	6	20,7

Değerlendirmeye alınan çalışmaların yıllara göre dağılımına bakıldığında %44,8'lik oran ile çalışmaların çoğunluğunun 2006 – 2010 yılları arasında yapılmış olduğu gözlemlenmektedir. Bunun yanı sıra eğitsel yazılım değerlendirme araçlarının ortaya konulduğu ilk on yıllık dönemde ise toplam 10 adet değerlendirme aracının ortaya konulduğu tespit edilmiştir. Eğitsel yazılımların popülerliğinin arttığı son dönemlerde ise (2011 – 2015) ülkemizde yapılan çalışma sayısında yaşanan düşüş ile bu dönemde yapılmış çalışmaların toplama oranı %20,7 olarak tespit edilmiştir.

Değerlendirme Aracı Türüne Göre Çalışmaların Dağılımı

Alan yazın kapsamında ulaşılabilen ve bu çalışma kapsamında ele alınan eğitsel yazılım değerlendirme araçlarının isimlendirme türlerine göre dağılımları Tablo 3'te verilmiştir.

Tablo 3. Değerlendirme Aracı Türüne Göre Çalışmaların Dağılımı

Ölçek Türü	Çalışma Sayısı	%
Anket	2	6,9
Form	16	55,1
Kriter Listesi	4	13,8
Ölçek	7	24,2

Araştırmacılar eğitsel yazılım değerlendirme araçlarını geliştirirken bu araçları kendi amaçlarına ve belirledikleri kriterlere göre isimlendirmişlerdir. Tablo 3'e bakıldığında incelemeye alınan bu değerlendirme araçlarından %55,1'nin "Form" olarak isimlendirildiği, %24,7'sinin ise "Ölçek" olarak isimlendirildiği görülmektedir.

Yazılım Türüne Göre Çalışmaların Dağılımı

Eğitsel yazılım değerlendirme araçlarının değerlendirmeyi hedef aldıkları yazılım türlerine göre dağılımları Tablo 4'te verilmiştir.

Tablo 4. Yazılım Türüne Göre Çalışmaların Dağılımı

Yazılım Türü	Çalışma Sayısı	%
Eğitsel Oyun	1	3,5
Eğitsel Yazılım	18	62,0
Web Tabanlı Eğitsel Yazılım	10	34,5

Tablo 4'e bakıldığında eğitsel yazılım değerlendirme araçlarının %62,0'lık bir oran ile büyük çoğunluğunun "Eğitsel Yazılımları" değerlendirmeyi temel aldıkları görülmektedir. On adet değerlendirme aracı ise "Web Tabanlı Eğitsel Yazılımları" değerlendirmeyi hedef alarak

geliştirilmiştir. Bunun yanı sıra “Eğitsel Oyunları” değerlendirmek amacıyla geliştirilmiş sadece bir adet değerlendirme aracı olması ise dikkat çekici bir veri olarak görülmektedir.

Hazırlık Sürecine Göre Çalışmaların Dağılımı

Eğitsel yazılım değerlendirme araçlarının geliştirilmesi aşamasında çalışmalarda kullanılan temel süreçlerin dağılımları Tablo 5’te ele alınmıştır.

Tablo 5. Hazırlık Sürecine Göre Çalışmaların Dağılımı

Hazırlık Süreci	Çalışma Sayısı	%
Belirtilmemiş	5	17,2
İkincil Kaynak	5	17,2
Alan yazın	10	34,5
Alan yazın + Uzman Görüşü	5	17,2
Uyarlama	4	13,9

Tablo 5’te görüldüğü üzere çalışmalarının %17,2’sinin yani 5 adet çalışmanın hazırlık süreci ile ilgili herhangi bir açıklamaya rastlanmadığından ötürü süreç bu çalışma kapsamında “Belirtilmemiş” olarak nitelendirilmiştir. %17,2’lik orana sahip olan ve “İkincil Kaynak” olarak ifade edilen 5 değerlendirme aracının asıl çalışmalarına ulaşamadığından hazırlık süreçleri ile ilgili bilgiye ulaşamamıştır. %34,5’luk orana sahip 10 çalışmanın ise sadece alan yazın taramasıyla elde edilen veriler ışığında araştırmacılar tarafından çalışmanın amacına uygun olarak hazırlandığı görülmektedir. “Alan yazın + Uzman Görüşü” olarak belirtilen ve %17,2’lik orana sahip 5 çalışma da ise alan yazın ile birlikte uzman görüşlerine de başvurularak değerlendirme araçları oluşturulmuştur. Son olarak ise “Uyarlama” başlığı altında verilen ve %13,9’luk orana sahip 4 adet değerlendirme aracı araştırmacılar tarafından var olan bir çalışma üzerinden kendi çalışma amaçlarına uygun olarak düzenlenerek hazırlanmıştır.

Değerlendirme Aracı Puanlama Türüne Göre Çalışmaların Dağılımı

Alan yazın kapsamında ulaşılabilen ve bu çalışma kapsamında ele alınan eğitsel yazılım değerlendirme araçlarının puanlama türlerine göre dağılımları Tablo 6’da verilmiştir.

Tablo 6. Değerlendirme Aracı Puanlama Türüne Göre Çalışmaların Dağılımı

Puanlama Türü	Çalışma Sayısı	%
3’lü Likert	4	13,8
5’li Likert	19	65,5
7’li Likert	1	3,5
Karma	3	10,3
Olumlu / Olumsuz	2	6,9

Tablo 6’da görüldüğü üzere çalışmaların %1,8’inde “3’lü Likert”, %65,5’inde “5’li Likert” ve %3,5’inde “7’li Likert” tipinde derecelendirilen ölçme araçları kullanılmıştır. Bunun yanı sıra %10,3’lük orana sahip olan üç çalışmada ise iki farklı puanlama türü kullanıldığı türü tespit edilmiş ve bunlar “Karma” başlığı altında ifade edilmiştir. “Olumlu / Olumsuz” kategorisinde verilen puanlama türü ise “Evet/Hayır” ve “Katılıyorum/Katılmıyorum” şeklindeki ifadeleri temsil etmektedir.

Değerlendiren Hedef Kitesine Göre Çalışmaların Dağılımı

Belirlenen kriterler çerçevesinde çalışma için seçilen değerlendirme araçlarının hedef kitlelerine ait verilerin bulunduğu bilgilerin dağılımları Tablo 7’de verilmiştir.

Tablo 7. Değerlendiren Hedef Kitesine Göre Çalışmaların Dağılımı

Hedef Kitle	Çalışma Sayısı	%
Belirtilmemiş	4	13,8
İlk ve Ortaöğretim Öğrencisi	5	17,2
Karma	1	3,5
Uzman	11	37,9
Üniversite Öğrencisi	8	27,6

İncelemeye alınan değerlendirme araçlarının hedef kitlesinin 4 tane çalışmada “Belirtilmemiş” olduğu, 5 tanesinin de ise “İlk ve Ortaöğretim Öğrencisi” olduğu tespit edilmiştir. “Karma” başlığı altında verilen değerlendirme grubunu ilk ve ortaöğretim öğrencileri, öğretmenler, akademisyenler ve uzmanlar oluşturmaktadır. 11 tanesinde “Uzman” şeklinde ifade edilen gruba ise sadece öğretmenler ve akademisyenler oluşturmaktadır. Üniversite öğrencilerine yönelik olarak hazırlanan 8 tane çalışma ağırlıklı olarak eğitim fakültesinde öğrenim gören öğrencilerle yürütülmüştür.

Geçerlik Güvenirlik Durumlarına Göre Çalışmaların Dağılımı

İncelenen çalışmaların değerlendirme araçlarına yönelik olarak yapılan geçerlik güvenirlik durumlarına Tablo 8’de yer verilmiştir.

Tablo 8. Geçerlik Güvenirlik Durumlarına Göre Çalışmaların Dağılımı

Geçerlik Güvenirlik	Çalışma Sayısı	%
Belirtilmemiş	2	6,9
Geçerlik Var-Güvenirlik Yok *	2	6,9
Güvenirlik Var-Geçerlik Yok **	6	20,8
Geçerlik Güvenirlik Var	6	20,8
Geçerlik Güvenirlik Yok	13	44,6

* Faktör analizi veya kapsam geçerliği yöntemleri kullanılmıştır.

** Ki-Kare testi, cronbach alpha değeri hesaplanması ve puanlayıcılar arası güvenirlik yöntemleri kullanılmıştır.

Tablo 8’e göre iki çalışmada (%6,9) geçerlik ve güvenirliğe dair herhangi bir bilginin “Belirtilmemiş” olduğu saptanmıştır. İki tanesinde sadece “Geçerlik” ve altısında ise sadece “Güvenirlik” olduğu görülmüştür. Ayrıca hem geçerliği hem de güvenirliği olan altı adet (%20,8), ikisinin de olmadığı ise 13 tane (% 44,6) değerlendirme aracı olduğu tespit edilmiştir.

Boyut Sayılarına Göre Çalışmaların Dağılımı

Eğitsel yazılım değerlendirme araçlarının sahip oldukları boyut sayılarının dağılımlarına ait bilgiler Tablo 9’da verilmiştir.

Tablo 9. Boyut Sayılarına Gre alıřmaların Dađılımları

Boyut Sayısı	alıřma Sayısı	%
	9	31,1
Drt	6	20,7
Beř	3	10,3
Altı ve zeri	5	17,2
Yok	6	20,7

Tablo 9 incelendiđinde “” boyut sayısına sahip olan alıřmaların oranı %31,1, “Drt” boyutlu olanların oranı %20,7 ve “Beř” boyutlu olanların oranı %10,3 olarak belirlenmiřtir. “Altı ve zeri” olarak belirtilenler ise 6, 9, 10 ve 15 boyuta sahip olan alıřmalardır. Bunların yanı sıra boyutlandırılmadan ele alınan %20,7’lik orana sahip altı deđerlendirme aracı bulunduđu tespit edilmiřtir.

Madde Sayılarına Gre alıřmaların Dađılımları

Deđerlendirmeye alınan alıřmaların madde sayılarına gre gruplandırılmıř haline Tablo 10’da yer verilmiřtir.

Tablo 10. Madde Sayılarına Gre alıřmaların Dađılımları

Madde Sayısı	alıřma Sayısı	%
0 – 25	11	37,9
26 – 50	11	37,9
51 – 75	2	6,9
76 - 100	2	6,9
101 ve zeri	3	10,4

Arařtırma kapsamında incelenen aralarda madde sayısı “0 - 25” ve “26 - 50” arasında olan 11’er (%37,9), “51 - 75” ve “76 - 100” arasında olan 2’řer (6,9) ve “101 ve zerinde maddesi olan 3 (%10,4) deđerlendirme aracı bulunduđu sonucuna varılmıřtır.

SONULAR ve TARTIřMA

Ulařılabilen alan yazın dâhilinde bu alıřmada incelenen eđitsel yazılım deđerlendirme aralarına 1992 yılından itibaren rastlanmaktadır. Numanoglu'nun (1992) geliřtirdiđi “Yazılım Deđerlendirme Formu” eđitsel yazılım deđerlendirme araları konusunda ikincil kaynak aracılıđıyla ulařılabilen ilk rnek olarak belirlenmiřtir. 2000’li yılların bařlarına kadar olan srete yapılan arařtırmaların azlıđı dikkat ekmekle birlikte kullanılan eđitsel yazılım deđerlendirme aralarının da genellikle uyarlama ve alıntı řeklinde olduđu tespit edilmiřtir. 2006-2010 yılları arasında ise arařtırmacıların eđitsel yazılım ve deđerlendirilmesi alanında yaptıkları alıřmaların yođunluđunun arttıđı ve bunun yanı sıra yeni bir veri toplama aracı geliřtirme alıřmalarında da buldukları gze arpmaktadır. Sim, Macfarlane, ve Read’in (2006) eđitsel yazılımların deđerlendirilmesine ihtiya duyulduđu ve bu konuda aıka belirlenmiř yntemlerin geliřtirilemediđi grř 2006-2010 yılları aralıđında grlen yođunluđun nedeni olabileceđi dřnlmektedir. Buna karřılık her geen gn eđitimde teknoloji entegrasyonunun ve ađırlıklı olarak eđitsel yazılım kullanımının arttıđı 2011-2015 yılları arasında ise eđitsel yazılım deđerlendirme konusunda lkemizde yapılan arařtırma sayısında bir dřř olduđu gzlemlenmektedir. Bunun sebebinin ise belirtilen srete lkemizde yapılan alıřmaların odak noktalarının deđiřmesi ve eđitsel yazılım deđerlendirme alanına olan ilginin azalması olduđu dřnlmektedir.

Web hizmetlerinin gelişmesiyle birlikte günümüzde nesne ambarları, web destekli eğitim, web tabanlı eğitim, açık ders malzemeleri gibi eğitim ortamları yaygınlaşmaktadır (Yiğit ve diğerleri, 2008). Buna bağlı olarak geliştirilen web tabanlı eğitim yazılım değerlendirme araçlarının da toplam değerlendirme araç sayısının %34,5'lük bir bölümünü oluşturduğu görülmektedir. Bunun yanı sıra teknolojinin eğitim dünyasına getirdiği bir diğer kavram olan eğitsel oyunların öğrencilerin derse olan ilgilerini arttırdığı sonucuna ulaşılmıştır (Ateş, 2011). Ancak eğitsel oyunların değerlendirilmesi ile ilgili ulaşılabilen alan yazında sadece bir adet ölçme aracının olması bu alandaki eksikliğin bir göstergesi olarak düşünülebilir. Bu ortamların eğitsel ve içerik olarak kalitesinin değerlendirilmesi, bu sayede eksik bulunan noktalarda gerekli iyileştirmelerin yapılmasının eğitime olumlu yansımaları olacağı düşünülmektedir.

Hazırlık süreçlerine dair değerlendirme araçları göz önüne alındığında ise araçların en çok alan yazın taraması ile oluşturulduğu görülmüştür. Buna ek olarak uzman görüşlerinin alındığı veya var olan bir değerlendirme aracının araştırmanın amacına uygun olarak uyarlanmasının yapıldığı da görülmektedir. İncelenen çalışmalara geçerlik ve güvenilirlik durumlarına göre bakıldığında ise bazı çalışmalarda salt geçerlik bazılarında ise salt güvenilirlik çalışmalarının yapıldığı, kimilerinde ikisinin de olduğu ve ya da hiçbirinin olmadığı çalışmaların olduğu saptanmıştır. Geçerlik ve güvenilirlik çalışmaları kapsamında başvurulan yöntemlerde ise alan uzmanı görüşüne başvurma, faktör analizi çalışmaları, dil uzmanına veri toplama aracını inceletme gibi yöntemler kullanılmıştır.

Alan yazın taraması sonucu ulaşılabilen değerlendirme araçları neticesinde eğitim yazılımlarını değerlendirmek üzere hazırlanan araçların yöntem ve içerik olarak ortak bir noktada toplanmadığı görülmektedir. Diğer yandan geliştirilen değerlendirme araçları madde ve boyut bağlamında içerik, tasarım, öğretimsel gibi kısıtlı sayıda noktalar etrafında toplanmaktadır. Bu durum da farklı değerlendirme araçlarında tekrar eden boyutlar ve maddeler olarak görülmektedir. Aksi yönden düşüldüğünde ise çok fazla farklı boyut ve maddenin olması ise henüz eğitsel yazılım değerlendirme kriterlerinin ortak bir çerçevede buluşmadığını da göstermektedir. Bu değerlendirme araçları için belirlenmiş veya temellendirilmiş bir geçerlik güvenilirlik kriterlerinin de olmaması bir diğer sorunu oluşturmaktadır. Bu alanda yapılabilecek çalışmaların değerlendirme araçlarının geliştirilmesinde geçerlik ve güvenilirlik çalışmalarını artırma konusunda faydalı olacağı düşünülmektedir. Aynı zamanda eğitsel oyun gibi bazı alanlarda da değerlendirme araçlarının eksikliğini giderilmesi gerekli görülmektedir. Son 2-3 yıl dikkate alındığında ise bu alanda geliştirilen veri toplama aracı sayısında bir düşüş vardır. Ancak gelişen teknoloji ve eğitime yansımaları düşünüldüğünde ölçme ve değerlendirme araçlarının güncel ve teknolojik gelişmeleri takip eder vaziyette olmasının, geliştirilebilir ve güncellenebilir olmasının eğitim-öğretim süreçleri açısından olumlu olacağı düşünülmektedir.

KAYNAKÇA

- Akbulut, Ö. E., Akdeniz, A. R., ve Dinçer, G. T. (2008). *Bilgisayar destekli bir öğretim materyalinin tasarlanması ve değerlendirilmesi*. 8th International Educational Technology Conference, Eskişehir, Türkiye.
- Aksaya, H. (2014). *Açık ders malzemeleri platformlarının yaşam boyu öğrenme perspektifinde incelenmesi*. (Yüksek Lisans Tezi, Marmara Üniversitesi, Gazetecilik Anabilim Dalı, İstanbul). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Arslan, A. (2002). *Web destekli bilgisayar öğretiminin tasarım kriterlerinin incelenmesi*. (Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul).
- Arslan, A. (2004). *Eğitsel web ortamı değerlendirme formu geliştirmeye yönelik örnek bir çalışma*. IV. Uluslararası Eğitim Teknolojileri Sempozyumu ve Fuarı, Sakarya, Türkiye.
- Aslan, Ö. (2006). *Lise matematik dersinde kullanılan eğitsel yazılım içeriklerinin öğrenme başarıları yönünden karşılaştırılması*. (Yüksek Lisans Tezi, Marmara Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, İstanbul), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Ateş, A. (2011). Eğitsel yazılım değerlendirme ölçeği: geçerlik ve güvenilirlik çalışması. *Eğitim Teknolojileri Araştırmaları Dergisi*, 2(1).
- Avcı, B. (2010). *Eğitsel yazılımların kullanılabilirliğinin göz izleme ve sesli düşünme metotlarıyla incelenmesi*. (Yüksek Lisans Tezi, Marmara Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, İstanbul), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.

- Cohen, V. B. (1983). Criteria for the evaluating of microcomputer courseware. *Educational Technology*, 23(1), 9-14.
- Çeliköz, N. (1996). *Bilgisayar destekli öğretim için özel ders türünde bir ders yazılımı hazırlanması, uygulanması ve deęerlendirilmesi*. (Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Programları ve Öğretimi Anabilim Dalı, Ankara), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Çeliköz, N., ve Erişen, Y. (2010). *Eğitimde bilgisayar kullanımı*. Ö. Demirel ve E. Altun (Ed.), Öğretim Teknolojileri ve Materyal Tasarımı. Ankara: Pegem A Yayıncılık.
- Demir, Ü. (2004). *İlköğretim 7. sınıf öğrencilerinin eğitsel yazılım ekran tasarım seçimlerinin ve ekran tasarımında dikkat ettikleri noktaların deęerlendirilmesi*. (Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Eğitim Programları ve Öğretimi Anabilim Dalı, İzmir), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Erdoğan, Y. (2005). *Web tabanlı yüksek öğretimin öğrencilerin akademik başarıları ve tutumları doğrultusunda deęerlendirilmesi*. (Doktora Tezi, Marmara Üniversitesi, İstanbul).
- Erümit, S. F. (2013). Web tabanlı uzaktan eğitimde biyoloji dersi için ders materyali tasarımı: kriterler, uygulama ve deęerlendirme. *Journal of Instructional Technologies ve Teacher Education*, 2(1), 86-111.
- Gül, Ş., ve Yeşilyurt, S. (2011). Yapılandırmacı öğrenme yaklaşımına dayalı bir ders yazılımının hazırlanması ve deęerlendirilmesi. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 1(40), 19-36.
- Gülbahar, Y., ve Tınmaz, H. (2006). Implementing project-based learning and e-portfolio assessment in an undergraduate course. *Journal of Research on Technology in Education*, 38(3), 309-327.
- Güneş, B. (2007). *Psikolojik danışmanlık ve rehberlik hizmetlerine yönelik bir eğitim yazılımının geliştirilmesi*. (Yüksek Lisans Tezi, Çukurova Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, Adana), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Gürer, M. D., ve Yıldırım, Z. (2014). Öğrenme nesnesi deęerlendirme ölçeęi'nin (öndö) geliştirilmesi, geçerlik ve güvenilirlik çalışması. *Eğitim ve Bilim*, 39(176), 121-130.
- Güzeller, C., ve Korkmaz, Ö. (2007). Bilgisayar destekli öğretimde bir ders yazılımı deęerlendirmesi. *Kastamonu Eğitim Dergisi*, 15(1), 155-168.
- Kara, Y. (2009). Biyoloji öğretimi için hazırlanmış eğlenceli eğitim yazılımı deęerlendirmesi. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 27, 17-30.
- Kaya, Z. (2006). *Öğretim teknolojileri ve materyal geliştirme*. Ankara: Pegem A Yayıncılık.
- Kelleci, Ö. (2010). *Bir eğitsel yazılım deęerlendirme formunun geliştirilmesi ve uygulanması*. (Yüksek Lisans Tezi, Marmara Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, İstanbul), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Kelly, G. (2008). A collaborative process for evaluating new educational technologies. *Campus-Wide Information Systems*, 25(2), 105-113.
- Kutluca, T., ve Birgin, O. (2007). Doğru denklemi konusunda geliştirilen bilgisayar destekli öğretim materyali hakkında matematik öğretmeni adaylarının görüşlerinin deęerlendirilmesi. *Gazi Eğitim Fakültesi Dergisi*, 27(2), 81-97.
- Lee, S. H., Choi, W. ve Byun, H. (1996). *Criteria for evaluating and selecting multimedia software for instruction*. Proceedings of selecting research and development presentations at the 1996 National Convention of the Association for Educational Communications and Technology, Indianapolis, USA.
- Numanoęlu, M. (1992). *MEB BDE projesi bilgisayar destekli öğretim ders yazılımlarında bulunması gereken eğitsel özellikler*. (Yüksek Lisans Tezi, Ankara Üniversitesi, Ankara).
- Özgür, H. (2011). *Syracuse modeli ile e-öğrenme ortamı için tasarlanmış bir dersin öğrencilerin başarısına etkisi: "trakya üniversitesi eğitim fakültesi örneęi"*. (Doktora Tezi, Trakya Üniversitesi, Bilgisayar Mühendislięi Anabilim Dalı, Edirne), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Öztürk, C., ve İnan, N. U. (1998). *İlköğretim sosyal bilgiler derslerinde kullanılabilecek bazı bilgisayar yazılımlarının deęerlendirilmesi*. IV. Ulusal Sınıf Öğretmenlięi Sempozyumu, Denizli, Türkiye.
- Sert, G., Kurtoęlu, M., Akıncı, A., ve Seferoęlu, S. S. (2012). *Öğretmenlerin teknoloji kullanma durumlarını inceleyen araştırmalara bir bakış: bir içerik analizi çalışması*. XVII. Akademik Bilişim Konferansı Bildirileri. Uşak Üniversitesi, Uşak, Türkiye.
- Sim, G., Macfarlane, S., ve Read, J. (2006). All work and no play: measuring fun, usability, and learning in software for children. *Computers ve Education*, 46, 235-248.
- Şakar, Ç. (2008). *Otistik öğrencilere yönelik eğitsel yazılım tasarlama, geliştirme ve deęerlendirme sürecinin incelenmesi*. (Yüksek Lisans Tezi, Marmara Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, İstanbul), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Şimşek, N. (1998). *Öğretim amaçlı bilgisayar yazılımlarının deęerlendirilmesi: kavramlar teknikler, araçlar ve uygulama*. Ankara: Siyasal Kitabevi.
- Tankut, Ü. S. (2008). *İlköğretim 7. sınıf sosyal bilgiler dersinde bilgisayar destekli öğretimin akademik başarıya ve kalıcılıęa etkisi*. (Yüksek Lisans Tezi, Çukurova Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Anabilim Dalı, Adana), <http://tez2.yok.gov.tr/> adresinden edinilmiştir.

- Yiğit, N., Altun, T., Alev, N., Dertlioğlu, K., ve Bülbül, M. Ş. (2007). *Öğretim amaçlı web sitelerinin değerlendirilmesine yönelik bir ölçek tasarımı*. 1. Ulusal İlköğretim Kongresi, Ankara, Türkiye.
- Yiğit, N., Bütüner, S. Ö., ve Dertlioğlu, K. (2008). Öğretim amaçlı örütbağ sitesi değerlendirme ölçeği geliştirme. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 2(2), 38-51.
- Yıldırım, A., ve Şimşek, H. (2006). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Yıldız, H., ve Sarıtepeci, M. (2013). *Program değerlendirme modelleri ışığında eğitsel yazılımlar üzerine bir inceleme*. Akademik Bilişim Konferansı, Antalya, Türkiye.

EXTENDED ABSTRACT

Introduction

Computer software that is created and used for educational purposes is called educational software (Kelleci, 2010). As with most educational materials, educational software also needs to be developed in accordance with standards based on certain teaching and learning theories; in addition to specific requirements of the target educational setting where the material will be implemented (Güzeller ve Korkmaz, 2007). According to Kara (2009) otherwise, positive results should not be expected from teaching activities conducted with low-quality educational software. In order to assess what counts as high-quality software, methods on educational software evaluation have been developed (Kelly, 2008). Similarly to examine effectiveness of a given learning software and the opportunities it offers to students Çeliköz and Erişen (2010) defined educational software evaluation as a way to determine the relationship between current and the expected level of quality of educational software.

It was Cohen (1983) who first stated that there was no common ground about the definition of educational software evaluation in the literature. Since then, many studies have emphasized different elements of evaluation and suggested software evaluation be based on these particular elements. Şimşek (1998) has arranged evaluation criteria in his book as (a) sophistication, (b) resolution, (c) objectivity, (d) flexibility, (e) sensitivity and (f) compliance with conditions. Çeliköz and Erişen (2010) have also stated their opinions on elements of educational software evaluation, defining compatibility, interoperability, effectiveness and efficiency as quality indicators. Aslan (2006) has concluded in his study that educational software evaluation can be carried out in two ways which he refers to as “Design Review” and “Evaluation of Wholeness”.

A short review of related literature may reveal numerous educational software evaluation forms of different sizes and scopes. However, a lack of research focusing on software evaluation forms is also present and this paper makes an effort of examining educational software evaluations that have been encountered during a systematic literature review.

The purpose of the study has therefore been formulated as “A review of educational software evaluation forms and scales in Turkey.” The followings are our research questions:

1. What is the distribution of educational software evaluation tools according to their respective years of development?
2. What types of educational software evaluation tools exist?
3. How are the development processes of educational software evaluation tools?
4. How is the status of development work on educational software evaluation tools in terms of relevant validity and reliability studies conducted?
5. What kind of scoring systems do educational software evaluation tools use?
6. From a technical standpoint, what types of software tools have been employed for the development of software evaluation tools?
7. What kind of practitioner groups have been considered as target users in the existing educational software evaluation tools?

8. How are the distributions of the number of sub dimensions in existing educational software evaluation tools?
9. How are the number of items distributed in the existing educational software evaluation tools?

Research Method

In this study, the method of content analysis was used for the purpose of reviewing educational software evaluation tools, in order to provide information on the state of a given topic and to interpret this information to come up with recommendations for future endeavours. Content analysis method is a scientific approach where oral, written, and other materials are evaluated objectively and handled through systematic review (Sert, Kurtoęlu, Akıncı, ve Seferoęlu, 2012). In the content analysis method, there are four main phases: the processing of qualitative research data obtained from collections of resources, coding of the data, identifying the themes and the organization and interpretation of results (Yıldırım ve řimřek, 2006). This study is based on the fundamentals of the content analysis method and its results have been achieved in this manner.

Results and Discussion

From a review of the related literature that is accessible by the authors, it has been observed that educational software evaluation tools in Turkish language are under development since the year 1992. Hence, the "Software Evaluation Form" developed by Numanoęlu (1992) is the first example of secondary sources about educational software evaluation tools. The lack of research on the educational evaluation is remarkable in the early 2000s. The tools used for educational software in these years seem to be either adapted or borrowed from other contexts. Between 2006 and 2010, educational software and educational software evaluation studies increase intensely and efforts on developing new educational software evaluation tools seem to have grown. It is believed that a statement made by Sim, Macfarlane, and Read (2006) regarding the insufficient methods in the development of educational software evaluation could be the main trigger for the increase of popularity in this field. However, between the years 2011 and 2015, there was a decrease in the development of educational software evaluation tools, no matter how the use of computer technology in education kept on increasing.

With the advances in the Web technologies, educational environments like learning object repositories, or concepts such as web-based learning and open courseware have become increasingly widespread (Yięit, Bütüner, ve Dertlioęlu, 2008). For this reason, it has been observed that Web-based educational software evaluation tools comprise 34.5 % of all the evaluation tools that have been reviewed.

It has been found that there is only one evaluation form regarding educational games, a major category of educational software, as laid out by Ateř (2011), who has solidified how educational games increase student motivation towards lessons. This situation points at a lack of interest in the area of educational games, as far as educational software evaluation is concerned. Evaluation of the quality and content of these educational environments should help improve their effectiveness.

Considering the development processes of educational software evaluation tools, it has been observed that most of them were developed after an initial review of scientific literature. In order to answer questions regarding validity and reliability studies during tool development, a review has been carried out and it has been observed that researchers generally use methods such as consulting expert opinion, conducting statistical factor analyses, or asking linguists to review translated versions of evaluation forms. Educational software evaluation tools show similarity to one another in terms of number of items and number of dimensions. This causes redundancy in items and dimensions of different evaluation tools. On the other hand, it has been seen that there is a general lack of common methods or opinions in the creation of evaluation tools because there are also many different item themes and dimensions present. Another problem is that, there is no common method for establishing reliability and validity in studies.

Future studies carried out in the field of educational software evaluation could be helpful for tools development by concretely assessing reliability and validity of existing tools. Also, it is necessary to make up for the lack of evaluation tools targeting educational games. Considering the last two or three years, there seems to be a decrease in the number of educational evaluation tools being developed. However, it is believed that having evaluation and assessment tools that are up to date and that follow new technology closely, could prove very effective in terms of enhancing teaching and learning processes.

Nitelik Sayısının Madde Seçme Algoritmalarının Performansı Üzerindeki Etkisi

The Impact of Attribute Size on the Performance of Item Selection Algorithms

Mehmet KAPLAN*

Öz

Bilişsel tanı modelleri (BTM) ve bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamaları hem eğitim alanında hem de psikoloji gibi diğer alanlarda hızla artmaktadır. Günümüze kadar BOBUT uygulamaları üzerine yapılan araştırmaların çoğunluğu, genellikle özetleyici bir değerlendirme sağlayan madde tepki kuramına dayalı modeller ile gerçekleştirilmiştir. Oysa BTM gibi bireylerin uzmanlaştığı ya da yetersiz kaldığı konularda daha ayrıntılı sonuç sağlayan biçimlendirici değerlendirme teknikleri son zamanlarda giderek önem kazanmıştır. BTM'nin BOBUT (BiTBOBUT) uygulamalarındaki kullanımı bireylerin yetenek düzeyleri hakkında daha ayrıntılı ve etkin bir değerlendirme sağlamada etkili bir yöntemdir. Bu çalışmanın amacı, BiTBOBUT uygulamalarındaki nitelik sayısının madde seçme algoritmalarının performansları üzerindeki etkisini araştırmaktır. Simülasyon çalışmasında farklı nitelik sayısı ve BTM'nin, madde seçme algoritmalarının performansları üzerindeki etkisi, ortalama test uzunlukları gibi betimsel istatistik değerlerine bakılarak araştırılmıştır. Elde edilen sonuçlara göre farklı nitelik sayısının ortalama test uzunlukları üzerinde önemli değişimlere sebep olduğu fakat farklı BTM kullanımının algoritmalar üzerinde test uzunluklarına herhangi bir etkisinin olmadığı gözlemlenmiştir.

Anahtar Kelimeler: bilişsel tanı modelleri, bilgisayar ortamında bireyselleştirilmiş testler, nitelik sayısı.

Abstract

The use of cognitive diagnosis models (CDMs) and computerized adaptive testing (CAT) has been increasing in both education and other fields such as psychology. To date, most of the research in CAT has been done using item response theory models which provide summative scores. However, formative assessment techniques (e.g., CDMs) that provide more detailed information about individuals' strengths and weaknesses have become popular in the recent years. The use of cognitive diagnosis computerized adaptive testing (CD-CAT) can produce more diagnostic information with an efficient testing design. This paper aims to investigate the impact of attribute size on the performance of item selection indices in terms of average test lengths. The result of this study showed that increasing the attribute size resulted in longer average test lengths; however, using different CDMs did not change the average test lengths.

Keywords: cognitive diagnosis models, computerized adaptive testing, attribute size.

GİRİŞ

İzleme veya ünite testleri gibi öğrenme niteliğinin izlenmesinde kullanılan biçimlendirici değerlendirme tekniklerinin önemi son zamanlarda eğitim literatüründe giderek artmıştır. Biçimlendirici değerlendirme teknikleri, özetleyici değerlendirme tekniklerine göre özellikle öğretme ve öğrenme stratejilerini geliştirmek için bireylerin yetenek düzeyleri hakkında daha açıklayıcı bilgi sağlayabilir (DiBello & Stout, 2007). Örneğin dönem sonunda açıklanan karne notları; bireylerin hangi konuları iyi derecede öğrendiği, hangi konularda ise yetersiz kaldıkları

* Araş. Gör. Dr., Artvin Çoruh Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Artvin-Türkiye, e-posta: mehmetkaplan@artvin.edu.tr

hakkında detaylı bir bilgi sağlamayabilir. Oysa etkinlik sonunda biçimlendirici değerlendirme yöntemleriyle ilgili kazanımlara dayalı ve detaylı bir şekilde öğrencilere dönüt vermek, hem öğrenci hem de öğretmen için daha iyi öğrenme ve öğretme ile sonuçlanabilir. Bu bağlamda bilişsel tanı modelleri (BTM; *cognitive diagnosis models*) bireylerin yetenek düzeyleri hakkında daha ayrıntılı bir değerlendirme sağlayabilmek için geliştirilmiş psikometrik modellerdir (de la Torre, 2009). BTM'nin en önemli amacı, belirli bir alanda bireylerin uzman ve yetersiz oldukları *noktaları*, *özellikleri* ya da *nitelikleri* tespit etmektir. Bu çalışmada *nitelik* kelimesi terim olarak bireyin uzman ve yetersiz olduğu *nokta* ya da *özellik* için kullanılmıştır. Türkiye'de de BTM ile ilgili çalışmalar gün geçtikçe literatürde yerini almaya başlamaktadır (Başokçu, 2014).

Gerek eğitim alanında gerekse psikoloji gibi diğer alanlarda kullanımı giderek artan diğer bir uygulama ise bilgisayar ortamında bireye uyarlanmış test (BOBUT; *computerized adaptive testing*) uygulamalarıdır. BOBUT uygulamaları, kağıt ve kalem testlere alternatif olarak geliştirilmiş olup daha kısa test uzunluğu ile birlikte daha iyi yetenek kestirimine olanak sağlamaktadır (Meijer & Nering, 1999; van der Linden & Glas, 2000). Her ne kadar madde tepki kuramı (MTK; *item response theory*) modelleri kullanılarak gerçekleştirilen BOBUT uygulamaları, bireylerin yetenek seviyeleri hakkında özetleyici değerlendirme sağlayarak mevcut test sisteminin ihtiyaçlarını karşılamış gibi görünse de BOBUT uygulamalarının BTM ile birlikte kullanımıyla bireylerin bilişsel seviyeleri hakkında daha etkin bir tanısal değerlendirme elde edilebilir. Son zamanlarda bilişsel tanı modelleri kullanılarak bilgisayar ortamında bireye uyarlanmış test (BiTBOBUT) uygulamalarının kullanımı biçimlendirici değerlendirme çalışmalarında hızla artmaktadır.

BTM

Daha önce de değinildiği gibi BTM'nin amacı bireyleri sıralama amaçlı bir puanlama yapmak yerine, bir konuyu öğrenmek için sahip olunması gereken niteliklerde bireylerin uzmanlaşması ya da yetersiz kalmasına göre sınıflandırma yapmaktır (DiBello & Stout, 2007). Bunun için her bir bireye nitelik vektörü olarak adlandırılan çok boyutlu α tayin edilir. Bu vektör genellikle 0 ve 1 olmak üzere iki değerli değişkenlerden oluşur. Bireyin nitelik vektörü her bir nitelik $k = 1, 2, \dots, K$ için $\alpha = \{\alpha_k\} = \{\alpha_1, \alpha_2, \dots, \alpha_K\}$ şeklinde ifade edilir. Eğer birey k sıradaki nitelik için uzmanlaşmışsa $\alpha_k = 1$, uzmanlaşmamış ve yetersiz kalmışsa, $\alpha_k = 0$ şeklinde gösterilir. Örneğin beş niteliğin bulunduğu bir konu alanında, eğer birey birinci ve üçüncü niteliklerde uzmanlaşıp diğerlerinde yetersiz kalmışsa bireyin nitelik vektörü $\alpha = (1,0,1,0,0)$ şeklinde gösterilir. Ayrıca BTM'de her bir madde için hangi niteliğin kullanılıp kullanılmayacağı Q-matris yardımıyla belirlenebilir (Tatsuoka, 1983). Bireylere tayin edilen nitelik vektörü gibi maddenin doğru cevaplanması için hangi niteliği gerektirdiğine göre her bir maddeye q-vektörü tayin edilir. Bu vektörler her bir madde için birleştirildiğinde Q-matris elde edilmiş olur. Toplam madde sayısı (J) ve nitelik sayısının (K) boyutlarını oluşturduğu Q-matris genellikle iki değerli değişkenlerden oluşur. Eğer j sırasındaki madde k niteliğini gerektiriyorsa, Q-matrisin bu madde ve nitelik için olan elementi $q_{jk} = 1$ değerini alır. Aksi halde bu element sıfır değerini alır. Örneğin yine beş niteliğin bulunduğu bir konu alanında, j sıradaki madde sadece ikinci ve dördüncü nitelikleri gerektiriyorsa, bu maddeye ait q-vektörü $q_j = (0,1,0,1,0)$ şeklinde gösterilir.

Günümüze kadar BTM'nin uygulanabilirliğini arttırmak için birçok model geliştirilmiştir. Bu modeller, üzerindeki varsayımlar düşünülerek *kısıtlayıcı* (*restricted*) ve *genel* (*general*) modeller olarak iki gruba ayrılabilir. Kısıtlayıcı modellere örnek olarak; *the deterministic inputs, noisy "and" gate* (DINA; de la Torre, 2009; Haertel, 1989; Junker & Sijtsma, 2001) modeli ve *the deterministic input, noisy "or" gate* (DINO; Templin & Henson, 2006) modeli verilebilir. Kısıtlayıcı modeller nitelik vektörleri ile görev performansı arasında güçlü varsayımlar gerektirir. Fakat ürettikleri sonuçlar genel modellere nazaran daha kolay ve anlaşılırdır. Örneğin DINA modelinde varsayım gereği bir maddenin doğru cevaplanabilmesi için bireyin o maddenin gerektirdiği tüm niteliklere sahip olması gerekmektedir. DINO modelinde ise bireyin maddenin gerektirdiği en az bir niteliğe sahip olması yeterlidir. Bu bağlamda DINA modeli telafi edici olmayan bir model iken DINO modeli telafi edici özellikte bir modeldir. Genel modellere örnek olarak ise *the log-linear CDM* (Henson, Templin, & Willse, 2009), *general diagnostic* (von Davier, 2008) modeli ve *the*

generalized DINA (G-DINA; de la Torre, 2011) modeli gösterilebilir. Kısıtlayıcı modellerin aksine genel modellerde güçlü varsayımlar yoktur. Genel modellerin en önemli özelliği, verinin modele uyumunun kısıtlayıcı modellere göre her zaman daha iyi sağlanmasıdır. Ancak genel modeller daha çok parametreyle çalıştıkları için bu modellerden çıkan sonuçları yorumlamak kısıtlayıcı modellere göre daha zordur. Verinin modele uyumu sağlanması koşuluyla kısıtlayıcı modelleri tercih etmek, analiz sonuçlarının daha anlaşılır yorumlanması açısından faydalıdır.

BiTBOBUT

BOBUT uygulamalarının en önemli amacı, test ilerledikçe bireylerin yetenek seviyeleri hakkında daha fazla bilgi elde etmektir. BOBUT uygulamalarını bireye özgü yapan unsur ise maddelerin bir algoritma tarafından bireylerin yetenek seviyelerine göre uyarlanmış olarak uygulama esnasında seçilmesidir. Maddelerin bu şekilde seçilmesi, düşük yetenek düzeyindeki bireylere daha az sayıda zor madde sorulmasına ve aynı şekilde, yüksek yetenek düzeyindeki bireylere daha az sayıda kolay madde sorulmasına neden olur. Bu durum, hem test uzunluğunun kısalmasını hem de bireylerin yetenek kestirimlerinin daha doğru yapılmasını sağlar (Meijer & Nering, 1999).

BOBUT uygulamaları; madde havuzu, başlangıç noktası, madde seçme algoritması, puanlama ve testi sonlandırma kuralı olmak üzere beş ana bileşenden oluşur (Thissen & Mislevy, 2000). BOBUT uygulamaları genellikle MTK modelleri üzerine kurulmuştur ve bilindiği gibi bu modellerde kullanılan yetenek değişkeni θ süreklidir. Ancak BTM’de kullanılan nitelik vektörü α sürekli olmayıp kategorik değişken olduğundan, BOBUT uygulamalarında bulunan bazı kavramlar BiTBOBUT uygulamalarında kullanılamaz. Örneğin BOBUT uygulamalarında, en yüksek Fisher bilgi fonksiyonu madde seçme algoritması olarak yaygın bir şekilde kullanılmakta iken BiTBOBUT uygulamalarında bu fonksiyon kullanılamaz çünkü Fisher bilgi fonksiyonu sadece sürekli değişkenlerde çalışabilmektedir. Bu soruna çözüm olarak BiTBOBUT uygulamalarında ilk olarak, hem sürekli hem de kategorik değişkenlerle çalışabilen *Kullback-Leibler* (K-L) *bilgi istatistiği* ve *Shannon entropi* yöntemi, kullanılmıştır (Xu, Chang, & Douglas, 2003). K-L bilgi istatistiği, t sırasındaki maddenin uygulanması sonucunda elde edilen nitelik vektörü kestiriminin, diğer nitelik vektörleri ile uzaklığıdır. Shannon entropi ise rastgele değişkenle ilişkilendirilmiş belirsizlik ölçüsü olarak tanımlanmıştır. Bu çalışmanın sonucunda her iki algoritmanın performansı karşılaştırıldığında, Shannon entropi yönteminin, K-L bilgi istatistiğine göre nitelik vektörlerinin sınıflandırmasında daha iyi sonuçlar sergilediği gösterilmiştir.

Daha sonra Cheng (2009), K-L bilgi istatistiğini geliştirerek BiTBOBUT uygulamalarında kullanılmak üzere yeni bir madde seçme algoritması önermiştir. Önerilen bu algoritmada bir önceki algoritmadan farklı olarak K-L bilgi istatistiği, nitelik vektörlerinin sonsal dağılımları ile ağırlıklandırılmıştır. Diğer bir ifadeyle bireyin nitelik vektörü kestiriminde, t ’nci madde uygulandıktan sonra bazı nitelik vektörlerinin diğer vektörlere göre olma olasılığının daha fazla olmasından dolayı, vektörler arası uzaklığın sonsal dağılıma göre ağırlıklandırılması daha doğru sonuçlar doğuracaktır. Yapılan simülasyon çalışmasının sonucunda, *sonsal ağırlıklandırılmış Kullback-Leibler indeksi* (PWKL; *Posterior-Weighted Kullback-Leibler*), K-L bilgi istatistiği ve Shannon entropi yöntemine göre daha yüksek nitelik vektör sınıflandırılma değerleri vermiştir. PWKL algoritmasının matematiksel gösterimi Eşitlik 1’de yer almaktadır.

$$PWKL_j(\hat{\alpha}_i^{(t)}) = \sum_{c=1}^{2^K} \left[\sum_{x=0}^1 \log \left(\frac{P(X_j = x | \hat{\alpha}_i^{(t)})}{P(X_j = x | \alpha_c)} \right) P(X_j = x | \hat{\alpha}_i^{(t)}) \pi_i^{(t)}(\alpha_c) \right] \quad (1)$$

Eşitlik 1’de, $\hat{\alpha}_i^{(t)}$, t sayıda madde uygulandıktan sonraki; i bireyinin j sırasındaki madde için nitelik vektörü kestirimi; $P(X_j = x | \alpha_c)$, 2^K adet nitelik vektörünün herbirinin (α_c) verilen yanıt örüntüsüne göre başarı olasılık değerleri ve $\pi_i^{(t)}(\alpha_c)$; t sayıda madde uygulandıktan sonraki her bir 2^K nitelik vektörünün sonsal dağılım değerleridir.

BiTBOBUT uygulamalarında kullanılmak üzere iki yeni madde seçme algoritması daha yakın zamanda geliştirilmiştir (Kaplan, de la Torre, & Barrada, 2015). Bu algoritmalarından ilki K-L bilgi istatistiğine dayalı olan *değiştirilmiş sonsal ağırlıklandırılmış K-L* indeksi (MPWKL; *modified PWKL*), ikincisi ise *G-DINA modeli ayırım* indeksidir (GDI; *G-DINA model discrimination index*). Yapılan simülasyon çalışması sonucuna göre her iki madde seçme algoritması, benzer şekilde sonuç sergilemiş olup her ikisi de PWKL algoritmasından daha yüksek nitelik vektör sınıflandırması sağlamıştır. Ayrıca GDI algoritması, indirgenmiş nitelik vektörleri ile çalıştığından diğer algoritmalara göre daha hızlı uygulama süresine, MPWKL ise en yavaş uygulama süresine sahip olmuştur. Kaplan ve diğerlerinin (2015) belirttiği gibi yeni önerilen madde seçme algoritmalarının daha iyi sonuç sağlaması, BiTBOBUT uygulama esnasında nitelik vektör kestirimine ihtiyaç duymamasından kaynaklanmaktadır. Özellikle testin başlangıcında kestirilen nitelik vektörlerinin genellikle doğru olmaması ya da diğer bir ifadeyle yeterlik düzeyinin doğru kestirilememesi, BiTBOBUT uygulamalarındaki madde sayısının artmasına neden olmaktadır. Oysa yeni önerilen algoritmalarda test süresince bireylerin nitelik vektör kestirimi yapılmadığından böyle bir sorun yoktur. GDI, her bir maddenin uygulanmasından sonra hesaplanan sonsal ağırlıklandırılmış 2^K nitelik vektörlerinin basit varyans hesabıdır. GDI madde seçme algoritmasının matematiksel gösterimi Eşitlik 2’de verilmiştir.

$$\zeta_j^2 = \sum_{c=1}^{2^{K_j^*}} \left[P(X_{ij} = 1 | \alpha_{cj}^*) - \bar{P}_j \right]^2 \pi(\alpha_{cj}^*) \quad (2)$$

Eşitlik 2’de \bar{P}_j , bireyin yanıt örüntüsüne göre her bir 2^K nitelik vektörü için verilen başarı olasılık değerlerinin ortalamasıdır. α_c^* ise indirgenmiş nitelik vektörüdür. İndirgenmiş nitelik vektörü şu şekilde açıklanabilir: Örneğin beş niteliğin bulunduğu bir konu alanında, maddenin doğru yanıtlanması ikinci, dördüncü ve beşinci nitelikleri gerektirsin. Yukarıda da gösterildiği gibi bu madde için q-vektörü $\mathbf{q}_j = (0,1,0,1,1)$ şeklinde gösterilir. İndirgenmiş nitelik vektörü ise $\mathbf{q}_j^* = (1,1,1)$ şeklinde gösterilir. GDI ile ilgili daha ayrıntılı bilgi için de la Torre (2011) çalışmasına, madde seçme algoritmaları için ise Kaplan ve diğerlerinin (2015) araştırmasına bakılabilir. Bu çalışmada, madde seçme algoritması olarak PWKL ve GDI kullanılmış olup nitelik sayısının bu iki algoritma üzerindeki etkisi araştırılmıştır.

Araştırmanın Amacı

Bu çalışmanın amacı, BiTBOBUT uygulamalarındaki nitelik sayısının (*attribute size*) madde seçme algoritma performansı üzerindeki etkisini araştırmaktır. Bu bağlamda aşağıdaki sorulara yanıt aranmıştır:

1. Nitelik sayısının artması, iki farklı madde seçme algoritması düşünüldüğünde, test uzunluklarında ne gibi değişikliğe yol açmıştır?
2. Yanıt örüntülerinin oluşturulmasında kullanılan BTM’nin, farklı nitelik sayısı göz önüne alındığında, test uzunluklarında ne gibi değişikliğe yol açmıştır?
3. Test sonlandırma kriterinin, yine farklı nitelik sayısı göz önüne alındığında, test uzunluklarında ne gibi değişikliğe yol açmıştır?

YÖNTEM

Araştırmanın Türü

Bu çalışma, BTM’de genel bir model olan G-DINA modeli çatısı altında bir simülasyon çalışması olup BiTBOBUT uygulamalarında kullanılan madde seçme algoritmalarının farklı nitelik sayısı

kullanıldığındaki performansı, ürettiği farklı test uzunluk değerlerine göre incelenmiştir. Bu sebepten dolayı bu çalışma temel bir araştırmadır.

Madde Havuzu Üretimi

Bu çalışmanın amacı olan nitelik sayısının madde seçme algoritmaları üzerindeki etkisi, $K = 4, 5, 6$ ve 7 şeklinde dört farklı değer kullanılarak incelenmiştir. Madde havuzunun oluşturulması için nitelik sayısı da göz önüne alınarak bütün olası $2^K - 1$ tane q-vektörü düşünülmüş ve her bir q-vektör için 20 madde üretilerek madde havuzu oluşturulmuştur. Sonuç olarak $K = 4, 5, 6$ ve 7 için madde havuz büyüklükleri sırasıyla 300, 620, 1260 ve 2540 olarak belirlenmiştir. Nitelik sayısı artınca madde havuzunun büyüklüğünün de artması, nitelik kestiriminde madde seçme algoritmalarının ihtiyacı kadar olan maddeyi havuzda bulmalarında sorun yaşamaması için gerekli görülmüştür.

Yanıt Örüntülerinin Üretimi

Bireylerin yanıt örüntülerinin üretiminde bazı kısıtlamalarla G-DINA modelinden de elde edilebilen DINA ve DINO modelleri kullanılmıştır. Ayrıca bu modellerde kullanılan tahmin (*guessing*) ve kaydırma (*slip*) parametreleri sırasıyla $g = s = 0,1$ olarak sabitlenmiştir. Diğer bir deyişle parametre değerleri sıfıra yakın olduğu için bu simülasyon çalışmasında ayırt ediciliği yüksek maddeler kullanılmıştır. Örneklem büyüklüğü olarak 3.000 bireyin yanıt örüntüleri bu iki modelin varsayımlarına göre nitelik sayısına göre oluşan farklı havuz büyüklükleri için ayrı ayrı üretilmiştir. Yanıt örüntülerinin üretimi Ox (Doornik, 2011) yazılımı kullanılarak kodlanmıştır.

Test Başlangıç Kuralı ve Madde Seçme Algoritmaları

Daha önceden BiTBOBUT uygulamalarında tanımlanmış iki farklı madde seçme algoritması olan PWKL ve GDI kullanılmıştır. Her iki madde seçme algoritması nitelik vektörlerinin sonsal dağılım değerlerini kullandığından test başlamadan önce madde seçme algoritmalarının değerlerinin hesaplanmasında 2^K nitelik vektörü için tekbiçimli (*uniform*) dağılım kullanılmıştır. Bu şekilde her bir durum için uygulanacak ilk madde rastgele bir şekilde seçilmiştir.

Test Sonlandırma Kuralı ve Nitelik Kestirme Yöntemi

Bu simülasyon çalışmasında BiTBOBUT uygulamaları, nitelik sayısına göre oluşan (2^K) nitelik vektörlerinin sonsal dağılım değerlerinin en büyük-en küçük (EBEK; *minimax*) değerleri baz alınarak sonlandırılmıştır. Bu şekildeki sonlandırma kuralı ile her birey için farklı test uzunlukları elde edilmiştir. Bu sonlandırma kuralıyla ilgili diğer bir nokta ise madde havuzunda yeterli madde bulunduğu, bu sonlandırma kuralı ile hedeflenen nitelik vektörlerinin doğru sınıflandırılma (NVDS; *correct attribute vector classification*) değerleri EBEK değerlerinden yüksek olacak şekilde elde edilir. Sonlandırma kuralı olarak ayrıca en uzun testin de 80 maddeyi geçmemesi koşul olarak simülasyon çalışmasına tanımlanmış ve üç farklı EBEK değeri olan $P = 0,65, 0,75$ ve $0,85$ BiTBOBUT uygulamasını sonlandırmak için ölçüt olarak alınmıştır.

PWKL algoritması yukarıda da bahsedildiği gibi her madde uygulanmasından sonra nitelik vektörünün kestirimine ihtiyaç duyduğundan, bireylerin nitelik vektörlerinin kestirimi her madde uygulamasından sonra en yüksek sonsal (MAP; *maximum a posteriori*) yöntemi kullanılarak elde edilmiştir. Bireylerin BiTBOBUT uygulaması sonucunda oluşan en son nitelik vektörlerinin kestirimi için de yine MAP yöntemi her iki madde seçme algoritması için kullanılmıştır. GDI madde seçme algoritması bireylerin nitelik vektörlerinin kestirimini sadece BiTBOBUT uygulamasının en sonunda gerçekleştirdiğini belirtmek yerinde olacaktır.

Verilerin Analizi

Verilerin analizinde her durumda bireyler için elde edilen farklı test uzunluklarının en kısa, en uzun, ortalama ve değişim katsayısı değerleri hesaplanmıştır. Değişim katsayısı test uzunluklarının standart sapmalarının ortalama değerlerine oranlanmasıyla bulunur. BiTBOBUT uygulaması Ox (Doornik, 2011) yazılımı kullanılarak kodlanmıştır ve işlemci hızı 2,5 olan bir bilgisayarda simülasyon çalışması gerçekleştirilmiştir.

BULGULAR

Veri analizi sonucunda elde edilen bulgular simülasyon çalışmasında kullanılan her bir etmen için ayrı ayrı incelenmiş ve sonuçlar; DINA ve DINO modelleri için sırasıyla Tablo 1 ve 2’de gösterilmiştir. Test sonlandırma kuralı bölümünde de değinildiği gibi bulguların ilk incelenmesinde elde edilen NVDS oranlarının her bir durum için EBK değerlerinden yüksek olup olmadığı kontrol edilmiştir. Tablo 1 ve 2’den de görüleceği gibi NVDS oranları her bir durum için EBK değerlerinden yüksek çıkmıştır. Örneğin $K = 4$, EBK değeri 0,65 ve DINA modeli kullanıldığında, PWKL ve GDI için NVDS oranları sırasıyla 0,66 ve 0,74 çıkmıştır. Bu da hedeflenen NVDS oranının (0,65) elde edildiğini göstermiştir. Diğer bir ifadeyle genel olarak BiTBOBUT uygulamasından önce hedeflenen bireylerin nitelik vektörlerinin doğru sınıflandırılmasına, uygulama sonrasında en az EBK değeri kadar ulaşılmıştır. Elde edilen bulgulardaki diğer ilginç bir nokta ise GDI madde seçme algoritması, $K = 4$ olduğu zamanda gerçekleşmiştir. Bu durumda kullanılan modelden bağımsız olarak test uzunluğu tüm bireyler için dört (test uzunluklarının standart sapması da sıfır) olarak gerçekleşmiştir. Yani PWKL madde seçme algoritmasından farklı olarak nitelik sayısı dört olduğunda GDI madde seçme algoritması her bir tek nitelikli maddeleri kullanarak ve başka madde tipini kullanmaya ihtiyaç duymadan hedeflenen NVDS oranlarına ulaşmıştır.

Tablo 1. DINA Modeli İçin Elde Edilen Sonuçlar

K	EBEK	GDI				PWKL					
		NVDS	En Kısa	En Uzun	Ortalama	DK	NVDS	En Kısa	En Uzun	Ortalama	DK
4	0,65	0,66	4	4	4,00	0,00	0,74	2	19	5,64	0,37
	0,75	0,82	5	18	6,32	0,29	0,81	2	21	6,84	0,40
	0,85	0,88	5	19	7,36	0,30	0,88	3	29	8,12	0,39
5	0,65	0,74	6	21	7,32	0,27	0,73	2	28	8,55	0,42
	0,75	0,79	6	24	8,20	0,30	0,80	3	29	9,83	0,42
	0,85	0,86	6	28	9,55	0,31	0,89	3	34	11,31	0,41
6	0,65	0,72	7	21	8,86	0,26	0,74	3	44	12,16	0,47
	0,75	0,77	7	36	10,02	0,29	0,81	3	44	13,49	0,46
	0,85	0,91	7	46	12,40	0,25	0,91	3	46	15,10	0,44
7	0,65	0,72	8	31	10,78	0,28	0,73	3	58	17,23	0,50
	0,75	0,78	8	32	12,04	0,28	0,80	3	60	18,73	0,49
	0,85	0,90	8	36	14,43	0,25	0,89	4	64	20,53	0,47

Not: K = Nitelik sayısı, EBK = En büyük-en küçük, DK = Değişim katsayısı.

Nitelik Sayısı

Tahmin edildiği gibi nitelik sayısı madde seçme algoritmalarının performansını ciddi bir şekilde etkilemiştir. Örnek olarak EBK değeri 0,65 seçildiğinde nitelik sayısının ortalama test uzunluklarına etkisi DINA ve DINO modelleri için sırasıyla Şekil 1 ve 2’de gösterilmiştir. Elde edilen sonuçlarda dikkati çeken ilk nokta ortalama test uzunluklarının nitelik sayısına göre artışın, PWKL madde seçme algoritması için GDI algoritmasına göre daha hızlı olmasıdır. Örneğin $K = 4$ olduğunda, her iki madde seçme algoritması için ortalama test uzunlukları PWKL ve GDI için sırasıyla 4 ve 5,64 iken $K = 7$ olduğunda ortalama test uzunlukları arasındaki fark açılarak sırasıyla 10,78 ve 17,23 olmuştur. Benzer durum en uzun test değerleri için de gözlenmiştir. Örneğin $K = 4$ ve

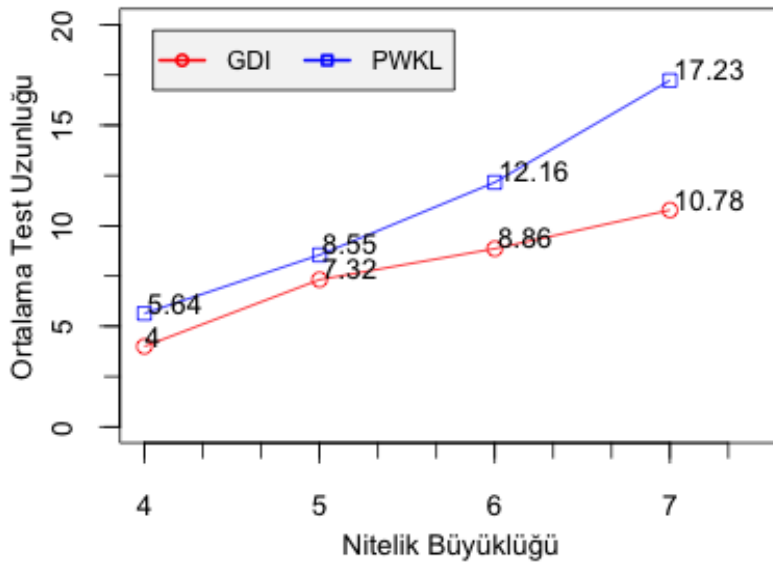
DINA modeli kullanıldığında, en uzun test Tablo 1'den de görüleceği gibi sırasıyla 4 ve 19 iken $K = 7$ olduğunda test uzunlukları, en uzun 31 ve 58 olmuştur.

PWKL ve GDI madde seçme algoritmalarından elde edilen test uzunlukları karşılaştırıldığında ise en kısa test PWKL madde seçme algoritmasından elde edilmiştir. Örneğin PWKL algoritması için en kısa test uzunluk değerleri iki ile dört arasında alırken GDI için bu değerler dört ile sekiz arasında değişmiştir. Değişim katsayı değerleri PWKL algoritması için nitelik sayısı ile artmakta iken GDI için böyle bir örüntü ortaya çıkmamıştır.

Tablo 2. DINO Modeli İçin Elde Edilen Sonuçlar

K	EBEK	GDI				PWKL					
		NVDS	En Kısa	En Uzun	Ortalama	DK	NVDS	En Kısa	En Uzun	Ortalama	DK
4	0,65	0,65	4	4	4,00	0,00	0,73	2	18	5,85	0,40
	0,75	0,82	5	17	6,38	0,30	0,81	2	21	7,01	0,43
	0,85	0,88	5	22	7,44	0,31	0,88	3	23	8,29	0,41
5	0,65	0,75	6	21	7,30	0,28	0,75	2	30	8,89	0,46
	0,75	0,79	6	25	8,16	0,30	0,82	3	38	10,17	0,46
	0,85	0,86	6	26	9,47	0,30	0,89	3	40	11,54	0,44
6	0,65	0,70	7	22	8,93	0,27	0,72	3	46	12,95	0,51
	0,75	0,76	7	30	10,09	0,30	0,80	3	55	14,27	0,50
	0,85	0,89	7	33	12,46	0,26	0,89	3	56	15,83	0,47
7	0,65	0,71	8	26	10,79	0,28	0,75	3	67	18,34	0,54
	0,75	0,77	8	28	12,08	0,28	0,82	3	71	19,79	0,52
	0,85	0,90	8	36	14,50	0,25	0,90	4	75	21,52	0,50

Not: K = Nitelik sayısı, EBK = En büyük-en küçük, DK = Değişim katsayısı.



Şekil 1. DINA Modeli ve 0,65 EBK Değeri İçin Ortalama Test Uzunlukları

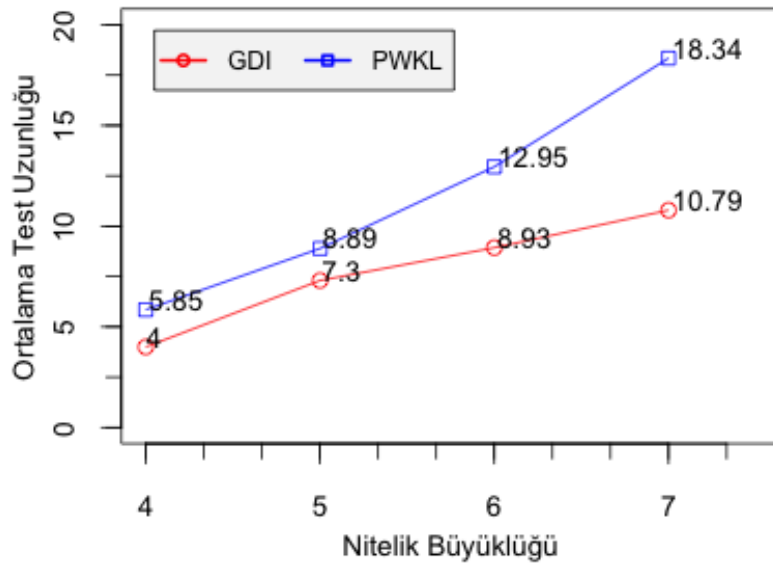
EBEK Değeri

Daha iyi NVDS değerlerinin elde edilmesi hedeflendiğinde test sonlandırma kuralındaki EBK değerinin artırılması gerekmektedir. EBK değerinin artırılması da her iki madde seçme algoritması için tahmin edildiği gibi en uzun ve ortalama test değerlerinin artmasına neden olmuştur. Örneğin $K = 4$ ve DINA modeli GDI madde seçme algoritması ile birlikte kullanıldığında, EBK değerinin 0,65'ten 0,85'e çıkarılması en uzun test değerlerinin 4'ten 19'a ve ortalama test uzunluklarını da 4'ten 7,36'ya yükseltmiştir. Fakat her bir durum için hesaplanan en kısa test

değerleri bazı durumlarda artış gösterirken bazı durumlarda sabit kalmıştır. Özellikle $K = 4$ olduğunda ve GDI madde seçme algoritması kullanıldığında, en kısa test değerleri modelden bağımsız olarak EBK değerlerinin artmasıyla artarken nitelik sayısı dörtten fazla olduğu durumlarda (örneğin $K = 5, 6$ ve 7) ise en kısa test değerleri nitelik sayısı artarken sabit kalmıştır. Aynı durum PWKL madde seçme algoritması kullanıldığında gözlemlenmemiştir. PWKL kullanıldığında en kısa test değerleri bazen artarken bazen sabit kalmıştır. Test uzunluklarının değişim katsayısı incelendiğinde EBK değerlerine göre belirli bir örüntü ortaya çıkmamıştır.

Bilişsel Tanı Modeli

Bu simülasyon çalışmasında bireylerin yanıt örüntülerinin oluşturulmasında kullanılan iki farklı modelin madde seçme algoritmaları üzerindeki performansları da incelenmiştir. Tablo 1 ve 2’de elde edilen sonuçlardan da görüleceği gibi kullanılan modeller elde edilen test uzunluklarının en kısa, ortalama ve değişim katsayısı değerlerinde gözle görülür bir etkisi olmamıştır. Ancak istisnai olarak PWKL algoritması kullanıldığında ve $K = 7$ olarak alındığında, EBK değerlerinden bağımsız olarak iki model arasındaki ortalama test uzunluğu farkı sadece bir madde olarak çıkmıştır. Örneğin sözü edilen durumdaki EBK değeri 0,65 için; DINA modeli kullanıldığında ortalama test uzunluğu 17,23 olarak; DINO modeli kullanıldığında da test uzunluğu 18,34 olarak elde edilmiştir. Ayrıca kullanılan model en uzun test değerlerinin üzerinde etkisi olmuştur fakat genel bir örüntü ortaya çıkmamıştır. Bu sonuç genel olarak bu iki model için ileride yapılacak olan gerçek BiTBOBUT uygulamalarında model uyumunun ortalama test uzunluklarında bir etkisinin olmayacağını göstermiştir.



Şekil 2. DINO Modeli ve 0,65 EBK Değeri İçin Ortalama Test Uzunlukları

SONUÇLAR ve TARTIŞMA

MTK ve BOBUT uygulamaları eğitimde ölçme ve değerlendirme alanında üzerinde çalışılmaya devam edilen en önemli konulardandır ve iki konu hakkında da birçok araştırma yapılmıştır. Her ne kadar MTK modelleri kullanılarak uygulanan BOBUT uygulamaları var olan sınav sisteminin ihtiyaçlarını gideriyor olsa da BOBUT uygulamalarının BTM ile kullanılması öğrenme ve öğretme becerilerinin geliştirilmesinde daha etkili olacağı, araştırmacıların ilgisini çekmektedir. Bu bağlamda BiTBOBUT uygulamalarının önemi biçimlendirici değerlendirme tekniklerinin arasında giderek artmaktadır.

Bu çalışmada farklı nitelik sayısının madde seçme algoritmalarının performansı üzerindeki etkileri, test uzunluklarına bakılarak araştırılmıştır. Ayrıca test sonlandırma için kullanılan farklı EBK

değerlerinin ve bireylerin yanıt örüntülerinin oluşturulmasında kullanılan farklı BTM'nin algoritmalar üzerindeki etkileri de incelenmiştir. Yapılan simülasyon çalışması sonucunda nitelik sayısının artırılması, ortalama test uzunluklarının artmasıyla sonuçlanmıştır. Daha önemlisi nitelik sayısının artmasıyla ortalama test uzunluk artışı, PWKL algoritması kullanıldığında GDI algoritmasına göre daha hızlı olmuştur. EBK değerlerinin artırılması, her iki madde seçme algoritması için ortalama test uzunluklarını artırırken yanıt örüntülerinin oluşturulmasında kullanılan modellerin ortalama test uzunluklarında herhangi bir etkisi gözlemlenmemiştir. Bu çalışmanın sonucu olarak nitelik sayısının etkisi GDI algoritması için daha az olduğundan ileride yapılacak olan pratik uygulamalarda özellikle fazla nitelik kullanıldığında PWKL algoritmasının kullanılmamasıdır.

Her ne kadar bu çalışma nitelik sayısının algoritmalar üzerindeki etkisini araştırmada bir ön çalışma olarak kabul edilebilse de simülasyonda kullanılan etmenler sonuçları genelleylemek için yeterli değildir. Örneğin yanıt örüntülerinin oluşturulmasında kullanılan madde parametreleri her bir madde için sabitlenmiştir. Halbuki gerçek BTM uygulamalarında bu durum söz konusu olmayacaktır. Ayrıca yanıt örüntülerinin oluşturulmasında sadece iki kısıtlayıcı model kullanılmıştır. Oysa literatürde bulunan diğer modeller kullanıldığında madde seçme algoritmalarının performansı merak konusudur. Bu çalışmadaki diğer kısıtlı durum ise madde havuzunun oluşturulmasında her bir q-vektörün kullanılmasıdır. Yine gerçek BTM uygulamalarında her bir q-vektör için madde yazımı mümkün olmayabilir.

KAYNAKÇA

- Başoğlu, T. O. (2014). Öğrenci yeteneğinin kestiriminde bilişsel tanı modelleri ve uygulamaları. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*.
<http://www.efdergi.ibu.edu.tr/index.php/efdergi/article/view/1341> adresinden erişildi.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74, 619-632.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- DiBello, L. V., & Stout, W. (2007). Guest editors introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 44, 285-291.
- Doornik, J. A. (2011). Object-oriented matrix programming using Ox (Versiyon 6.21) [Computer software]. London, England: Timberlake Consultants Press.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 333-352.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.
- Hsu, C.-L., Wang, W.-C., & Chen, S.-Y. (2013). Variable-length computerized adaptive testing based on cognitive diagnosis models. *Applied Psychological Measurement*, 37, 563-582.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, 39, 167-188.
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, 23, 187-194.
- Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.
- Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer et al. (Eds.). *Computerized adaptive testing: A primer* (ss. 101-133). Hillsdale, NJ: Lawrence Erlbaum Associates.
- van der Linden, W. J., & Glas, C. A. W. (Eds.). (2000). *Computerized adaptive testing: Theory and practice*. Boston, MA: Kluwer.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *The British Journal of Mathematical and Statistical Psychology*, 61, 287-307.
- Xu, X., Chang, H.-H., & Douglas, J. (2003, Nisan). *Computerized adaptive testing strategies for cognitive*

diagnosis. Paper presented at the annual meeting of National Council on Measurement in Education, Montreal, Quebec, Canada.

EXTENDED ABSTRACT

Introduction

Cognitive diagnosis models (CDMs) have been developed to detect mastery and nonmastery of attributes in a particular area. In contrast to unidimensional item response models (IRTs), CDMs provide more detailed evaluation about the strengths and weaknesses of students (de la Torre, 2009). A general CDM called generalized deterministic inputs, noisy “and” gate (G-DINA; de la Torre, 2011) model, a generalization of the deterministic inputs, noisy “and” gate (DINA; de la Torre, 2009; Haertel, 1989; Junker & Sijtsma, 2001) model, relaxes some of the strict assumptions of the DINA model. The G-DINA model considers the effects of all the possible interactions among the required attributes, which can contribute to the probability of success both in positive and negative ways. Several reduced models can be obtained from the G-DINA model by adding some constrictions to the model.

Computerized adaptive testing (CAT) has been developed as an alternative to paper-and-pencil tests, and it provides more accurate ability estimations of examinees with shorter and tailored tests (Meijer & Nering, 1999; van der Linden & Glas, 2000). Because attribute vectors in CDMs are discrete, some of the concepts in traditional CAT such as Fisher information are not applicable in cognitive diagnosis CAT (CD-CAT). Fortunately, several indices have been proposed for CD-CAT. First, the efficiency of the Kullback-Leibler (K-L) information was investigated as an item selection method in CD-CAT, and the results showed that it outperformed random selection in terms of attribute classification accuracy (Xu, Chang, & Douglas, 2003). Later, another index, namely, the posterior-weighted K-L (PWKL), was proposed, and it is a modified version of the K-L information in which it is weighted by the posterior distribution of the latent classes, and the simulation study showed that PWKL outperformed K-L information (Cheng, 2009). Recently, two new item selection indices, namely, the modified PWKL (MPWKL) and the G-DINA discrimination index (GDI) were proposed for CD-CAT. The results showed that both of the indices performed similarly and had higher classification accuracy compared to the PWKL (Kaplan, de la Torre, & Barrada, 2015).

Method

This simulation study was conducted under the G-DINA model framework. The design of the simulation study consisted of several factors. First, item pools were created considering different levels of attribute size (i.e., $K = 4, 5, 6,$ and 7). For each of $2^K - 1$ q-vectors, 20 items were created. Therefore, item pool sizes were 300, 620, 1260, and 2540 for $K = 4, 5, 6,$ and 7 , respectively. Second, examinees’ response patterns were generated using two reduced models, namely, the DINA and DINO models, and the guessing and slip parameters were fixed at 0.1 to eliminate the impact of item quality. Also, the number of examinees was set to 3,000. Third, two item selection indices, namely, the PWKL and GDI, were used in the CD-CAT administration. Uniform distribution was used in the calculation of the indices at the beginning of the test. This also means that the first item was selected randomly for each condition. Last, the CD-CAT administration was terminated after the largest posterior probability of an attribute vector was at least as large as prespecified minimax values, 0.65, 0.75, and 0.85 (Hsu, Wang, & Chen, 2013). This termination rule resulted in different test lengths for different examinees. In addition, maximum a posteriori (MAP) was used to estimate examinees’ attribute vectors. Different test lengths for different examinees were examined to analyze the results using the minimum, maximum, mean, and coefficient of variation of the test lengths.

Results and Discussion

Several results can be gleaned from this simulation study. First, as expected, correct attribute vector classification (CVC) rates were always higher than the prespecified minimax values for each condition. For example, using $K = 4$, the minimax of 0.65 and DINA model, the CVC rates were higher than the minimax, and they were 0.66 and 0.74 for the PWKL and GDI, respectively. Second, interestingly, the test lengths were the same for each examinee (and therefore the standard deviation of the test lengths was zero) when the GDI and $K = 4$ were used regardless of the model. Third, as expected, increasing the attribute size resulted in longer test lengths regardless of the other factors. In addition, the PWKL yielded even longer test lengths compared to the GDI when the attribute size was larger. For example, when $K = 4$, the mean test lengths were 4 and 5.64 for the PWKL and GDI, respectively; however, when $K = 7$, they were 10.78 and 17.23, respectively. The maximum test lengths showed similar patterns as the mean test lengths; however, the PWKL yielded smaller minimum test lengths compared to the GDI. Increasing the attribute size resulted in larger CV rates; however, there was no clear pattern for the GDI. Forth, increasing the prespecified minimax values resulted in larger maximum and mean test lengths for each item selection index regardless of the model. For example, when $K = 4$ and the DINA model was used, increasing the minimax value from 0.65 to 0.85 increased the maximum test lengths from 4 to 19, and increased the mean test lengths from 4 to 7.36. However, increasing the the prespecified minimax values did not show a clear pattern for the minimum and CV of test lengths. Last, using different CDMs in generating response patterns did not have an impact on the performance of the indices except that when $K = 7$ and the PWKL was used, the difference of the mean test lengths between two models was only one item. However, using different CDMs resulted in different maximum test lengths for different conditions, but there was no clear pattern to generalize the results.

Compared to traditional unidimensional IRT models, CDMs provide more information that can be used to inform instruction and learning. These models can reveal examinees' strengths and weaknesses by diagnosing whether they have mastered a specific set of attributes. CAT is a tool that can be used to create tests tailored for different examinees. This allows for a more efficient determination of what students know and do not know. In this study, the impact of different levels of attribute size on the item selection indices was examined using different factors in the simulation study. Although this study showed that the GDI is promising when larger attribute sizes to be used, more research needs to be done to determine its viability. For example, it would be interesting to investigate the impact of different CDMs, item quality, and item pool design on the performance of item selection indices.

Comparison of Classical Linear Regression and Orthogonal Regression with Respect to the Sum of Squared Perpendicular Distances

Dik Uzaklıklar Kareler Toplamına Göre Klasik Doğrusal Regresyon ile Ortogonal Regresyonun Karşılaştırılması*

Taliha KELEŞ**

Murat ALTUN***

Abstract

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. The purpose of this study was the trivial presentation of the equation for orthogonal regression (OR) and the comparison of classical linear regression (CLR) and OR techniques with respect to the sum of squared perpendicular distances. For that purpose, the analyses were shown by an example. It was found that the sum of squared perpendicular distances of OR is smaller. Thus, it was seen that OR line has appeared to present a much better fit for the data than CLR line. Depending on those results, the OR is thought to be a regression technique to obtain more accurate results than CLR at simple linear regression studies.

Keywords: orthogonal regression equation, perpendicular distance, the sum of squared perpendicular distances

Öz

Regresyon analizi, değişkenler arasındaki ilişkiyi inceleme ve modellemede kullanan istatistiksel bir tekniktir. Bu çalışmanın amacı ortogonal regresyon (OR) eşitliğini açık bir şekilde sunmak ve dik uzaklıklar kareleri toplamına göre klasik doğrusal regresyon (KDR) ile ortogonal regresyonu karşılaştırmaktır. Bu amaçla analizler bir örnek üzerinden gösterilmiştir. Araştırma sonucunda ortogonal regresyonun dik uzaklıklar kareler toplamının daha küçük olduğu bulunmuştur. Buradan bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkiyi, ortogonal regresyon doğrusunun klasik doğrusal regresyon doğrusundan daha iyi temsil ettiği görülmüştür. Bu sonuçlara bağlı olarak basit doğrusal regresyon çalışmalarında OR tekniğinin KDR tekniğinden daha doğru sonuçlar elde etmede kullanılabilecek bir regresyon tekniği olduğu düşünülmektedir.

Anahtar Kelimeler: ortogonal regresyon eşitliği, dik uzaklıklar, dik uzaklıklar kareler toplamı

INTRODUCTION

Regression analysis is a statistical technique for investigating and modeling the relationship between dependent and independent variables (Montgomery, Peck & Vining, 2012; Sykes, 1993). Applications of regression analysis exist in almost every field, including engineering, physical and chemical sciences, education, economics, management, astronomy, medical sciences, political science, life and biological sciences, and social sciences. In fact, regression analysis may be the most widely used statistical technique (Montgomery et al., 2012).

The most fundamental aim of simple linear regression analysis is to predict the values of the dependent variable, Y, when the values of the independent variable, X, is known in order to write line, curve and surface equations which represent a cloud of the data. This prediction is made according to the regression line and the curve (Lane, 2016).

* This study was presented at The 2nd International Researches Statisticians and Young Statisticians Congress on May 4-8 2016, Ankara-Turkey.

** Dr., Provincial Directorate of National Education, Strategy Development Service, R&D Department, Bursa-Turkey, e-mail: talihak@hotmail.com

*** Prof. Dr., Uludag University, Faculty of Education, Department of Elementary Mathematics Education, Bursa-Turkey, e-mail: maltun@uludag.edu.tr

The equation of $\hat{Y} = f(X)$ is used to find the predicted values of dependent variable \hat{Y} from the values of independent variable (X). The formula for linear regression line is

$$\hat{Y} = a + bX$$

Where a is the intercept, b is the slope of the line.

As illustrated in Figures 1a, classical linear regression (CLR) equation is attained by minimizing the sum of squares of vertical distances from the data points to the regression line (Ortiz, Pogliani & Besalú, 2010; Ding, Chu, Jin & Zhu, 2013; Elfessi & Hoar, 2001; Isobe, Feigelson, Akritas & Babu, 1990; Kane & Mroch, 2010; Leng, Zhang, Kleinman & Zhu, 2007; Ludbrook, 2010).

The sum of squares of vertical distances from data points to the regression line, $E = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - bx_i - a)^2$ is to be minimized, where e_i is called the i^{th} residual (Calzada & Scariano, 2003; Ortiz et al., 2010; Glaister, 2005; Scariano & Barnet, 2003).

As illustrated in Figures 1b, orthogonal regression (OR) equation is attained by minimizing the sum of the squares of perpendicular distances between the data points and the regression line (Carr, 2012; Ortiz et al., 2010; Ding et al., 2013; Elfessi & Hoar, 2001; Isobe et al., 1990; Kane & Mroch, 2010; Leng et al., 2007; Li, 1984; Ludbrook, 2010; Nievergelt, 1994; Scariano & Barnet, 2003). The sum of the squares of perpendicular distances between the data points and the regression line,

$$E_{\perp} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - bx_i - a)^2}{b^2 + 1}$$

is to be minimized, where d_i is perpendicular distance (Calzada & Scariano, 2003; Ortiz et al., 2010; Glaister, 2005; Li, 1984; Scariano & Barnet, 2003).

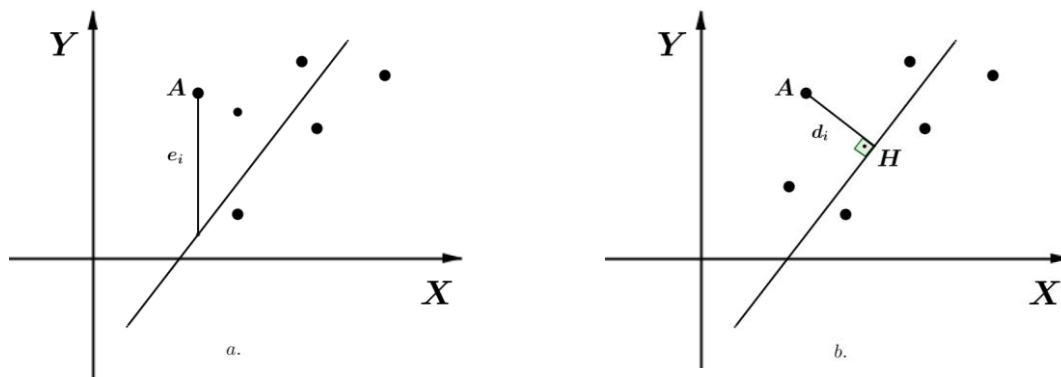


Figure 1. Classical Linear Regression (a), Orthogonal Regression (b)

OR identifies the line that minimizes the sum of squared perpendicular distances of the data points from the line (Carroll & Ruppert, 1996; Isobe et al., 1990; Kane & Mroch, 2010). OR has been known by various names; for example, it is known as the errors-in-variables (Carroll & Ruppert, 1996; Fišerová & Hron, 2010; Markovsky & Van Huffel, 2007), total least squares (Calzada & Scariano, 2003; Elfessi & Hoar, 2001; Golub & Van Loan, 1980; Markovsky & Van Huffel, 2007; Nievergelt, 1994), least squares (Adcock, 1878), major axis (Carr, 2012; Isobe et al., 1990; Kermack & Haldane, 1950; Ludbrook, 2010), measurement errors (Markovsky & Van Huffel, 2007), orthogonal error regression, and orthogonal least squares regression (Carr, 2012). In addition, OR is called a method of moments estimator in 1987 by Fuller (Carroll & Ruppert, 1996).

The method of OR was first discovered by Adcock in 1878 (Adcock, 1878) and later by Person (Isobe et al., 1990; Pearson, 1901). The method of OR has been rediscovered many times, often

independently (Ding et al., 2013). This method has been understood for well over a hundred years, but CLR technique is preferred to orthogonal regression technique in the statistical community because of its easy computation (Ludbrook, 2010). CLR is the most popular method for linear regression (Ding et al., 2013). Computational formulas for orthogonal regression line can be somewhat quite complicated (Calzada & Scariano, 2003; Carr, 2012; Carroll & Ruppert, 1996; Ortiz et al., 2010; Isobe et al., 1990; Kane & Mroch, 2010; Scariano & Barnet, 2003). Most literature on this method is necessarily brief and heavily symbolic (Calzada & Scariano, 2003; Carr, 2012; Kermack & Haldane, 1950; Nievergelt, 1994).

Simple linear regression analysis is used to define the nearest line to the data points (Akdeniz, 2013; Genç, Sertkaya & Demirtaş, 2003; Saraçlı, Doğan & Doğan, 2009). The shortest distance from a point to a line is perpendicular distance (Warton, Wright, Falster & Westoby, 2006). Thus, it is the perpendicular distance that is important, and the statistical distance in this case is the shortest distance to the line. The orthogonal regression is the line that minimizes the sum of squares of the shortest distances from the data points to the line.

One of the assumptions of classical linear regression, is that the independent variable can not include any measurement error, indicating that the only source of the error term is the dependent variable (Glaister, 2005; Scariano & Barnet, 2003). However, Deming (1943), indicated in his book “Statistical Adjustment of Data” in that independent variable may include measurement error in practice (Saraçlı, Doğan & Doğan, 2009). Besides, Glaister (2005) and Stöckl, Dewitte and Thienpont (1998) stated that both dependent and independent variables are subject to measurement error in practice. In this case, the estimates in CLR are no longer accurate (Glaister, 2005) in practice. The estimates in OR technique, when there is a measurement error in both variables (Carr, 2012; Glaister, 2005; Scariano & Barnet, 2003), are more accurate. Since the independent variable may include measurement error in measurement error models, OR technique may be more appropriate and may give better results in those situations (Calzada & Scariano, 2003; Carr, 2012; Warton et al., 2006).

A review of international literature indicated that considerable researches have been conducted analyses to compare OR with other regression techniques (Calzada & Scariano, 2003; Carr, 2012; Ding et al., 2013; Elfessi & Hoar, 2001; Glaister, 2005; Leng et al., 2007; Lolli & Gasperini, 2012; Lundbrook, 2010; Ortiz et al., 2010). Calzada and Scariano (2003) had studied on contrasting the ordinary and orthogonal (or total least squares) regression techniques using real data. They found that OR is better than ordinary least squares regression in this example. They suggested that there is no clear cut answer as to whether the ordinary least squares or orthogonal regression technique should be preferred in a given application. The method chosen by the researcher must be based on a keen understanding of the data as well as the sources of the errors present in the observations. Carr (2012) had compared three of such linear regression methods (classical linear regression, orthogonal regression, and reduced major axis) using geyser eruption data. He found that mean square error of classical linear regression is the smallest. He stated that classical linear regression has better than orthogonal regression and reduced major axis regression. Ding et al. (2013) had investigated and compared least squares and orthogonal regression in measurement error modeling for prediction of material property. They stated that orthogonal regression has better performance than classical linear regression with respect to standard deviation of residuals. Ortiz et al. (2010) had reviewed classical linear regression and orthogonal regression techniques in two variable linear regressions. They used an example with 18 data. They stated that standard deviation of the residues of orthogonal regression is smaller than the classical linear regression implying that it is a better method than classical linear regression.

The method comparison studies carried out in Turkey were analyzed. It is remarkable that very few studies have been done in this regard. Bland-Altman and OLS Bisector (Saraçlı & Çelik, 2012), Deming regression technique and ordinary least squares (Saraçlı, Doğan & Doğan, 2009), least squares and robust M estimation methods (Coşkuntuncel, 2013), least squares and least median squares estimation methods (Alma & Vupa, 2008), and simple linear regression and hierarchical linear model (Atar, 2010) techniques of method comparison studies have been conducted.

In Turkey, there are few studies about orthogonal regression in geophysics (Öztürk, 2012) and statistic (Saraçlı, 2011). Öztürk (2012) had made an analysis on the application of four different statistical regression methods for different data sets with respect to the determination of coefficient (R^2). It was showed that the representation of empirical relationships will be more suitable and reliable by Least Sum of Absolute Deviations or Robust regressions for clustered samples whereas by Least Squares or Orthogonal regressions for scattered data. Saraçlı (2011) had examined that data sets with different sample sizes simulated via Monte Carlo simulation were used to see the performances of Type II regression techniques (Ordinary least square (OLS) bisector, major axis, reduced major axis, Deming regression and passing bablok regression techniques) by mean square error. The result of his study concluded that the OLS bisector technique gave the best results in all conditions with different distribution types, different sample sizes, and the data set with or without an outlier.

In comparison studies published both in national and international literature, it has been seen that CLR (Carr, 2012) and ordinary least square (OLS) bisector (Saraçlı, 2011) demonstrated the superiority with respect to mean square error. In addition to those studies, OR technique demonstrated the superiority with respect to determination of coefficient (R^2) (Calzada & Scariano, 2003) or standard deviation of residuals (Ding et al., 2013; Ortiz et al., 2010).

Although the benefits of OR technique has been discussed, it has been seen that researches in social sciences and education field are using only the CLR technique in Turkish literature (Bağçeci, Döş & Sarıca, 2011; Bahar, 2006; Bayat, Şekercioğlu & Bakır, 2014; Doruk, Özdemir, & Kaplan, 2015; İlğan, Erdem, Yapar, Aydın & Aydemir, 2012; Kesicioğlu & Güven, 2014; Üredi & Üredi, 2005;). In Turkey, there is no study about orthogonal regression in social sciences and education. In this respect, it is expected that this study sheds light on studies that can be done in the future about this subject.

This Study Aim

The purpose of this study is the trivial presentation of the equation for orthogonal regression (OR) and the comparison of classical linear regression (CLR) and OR according to the sum of squared perpendicular distances.

METHOD

Classical linear regression and orthogonal regression were compared with respect to the sum of squared perpendicular distances. For this purpose, this analysis was shown by an example. For this example, data was taken from Akdeniz's book (Akdeniz, 2013).

We restricted our study to CLR and OR. Classical linear regression and orthogonal regression are compared with respect to the sum of squared perpendicular distances which is defined with the following formulation,

$$E_{\perp} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - bx_i - a)^2}{b^2 + 1}$$

Microsoft Excel 2010 was used for data organization and calculation of OR equation and sum of the squared perpendicular distances of OR and CLR. SPSS 17.0 was used for the calculation of CLR equation.

The equations for the slope and intercept of orthogonal regression were calculated in detail. The derivation of these equations is presented as follows.

Equation Derivation for OR Line

Herein, OR is explained in simple, straightforward way to facilitate easy understanding for equations.

$$\text{From Figure 1b, we see that } |AH| = d_i = \frac{|y_i - bx_i - a|}{\sqrt{b^2 + 1}}$$

$$\text{The sum of squares of the perpendicular distances is equal to } E_{\perp} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - bx_i - a)^2}{b^2 + 1}$$

where n is the sample size for the data. Derivatives with respect to a and b are calculated and equaled to zero.

$$\frac{\partial E_{\perp}}{\partial a} = 0 \text{ and } \frac{\partial E_{\perp}}{\partial b} = 0$$

The first step is taking the partial derivative of E_{\perp} with respect to the regression intercept, a,

$$\frac{\partial E_{\perp}}{\partial a} = \sum_{i=1}^n \frac{-2(y_i - bx_i - a)(b^2 + 1) - 0}{(b^2 + 1)^2} = 0$$

Multiplying both sides of this equation $(b^2 + 1)^2$ clears the denominator,

$$-2(b^2 + 1) \sum_{i=1}^n (y_i - bx_i - a) = 0$$

$$\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i - na = 0$$

$$\sum_{i=1}^n y_i = b \sum_{i=1}^n x_i + na \quad \text{and} \quad a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$

Similarly, next step is putting this result aside and taking the partial derivative of E_{\perp} with respect to the slope of regression, b,

$$\frac{\partial E_{\perp}}{\partial b} = \sum_{i=1}^n \frac{-2x_i(y_i - bx_i - a)(b^2 + 1) - 2b(y_i - bx_i - a)^2}{(b^2 + 1)^2} = 0$$

Multiplying both sides of this equation $(b^2 + 1)^2$ clears the denominator,

$$\sum_{i=1}^n [(b^2 + 1)(y_i x_i - bx_i^2 - ax_i) + b(y_i^2 + b^2 x_i^2 + a^2 - 2bx_i y_i - 2ay_i + 2abx_i)] = 0$$

$$\sum_{i=1}^n [(b^2 + 1 - 2b^2)x_i y_i + (-b^3 - b + b^3)x_i^2 + a(b^2 - 1)x_i + by_i^2 - 2aby_i + a^2b] = 0$$

$$(-b^2 + 1) \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n y_i^2 + a(b^2 - 1) \sum_{i=1}^n x_i - 2ab \sum_{i=1}^n y_i + na^2b = 0 \text{ Multiplying each side}$$

of the equation by -1

$$(b^2 - 1) \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n y_i^2 - a(b^2 - 1) \sum_{i=1}^n x_i + 2ab \sum_{i=1}^n y_i - na^2b = 0$$

$$(b^2 - 1) \sum_{i=1}^n x_i y_i + b(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2) - a(b^2 - 1) \sum_{i=1}^n x_i + 2ab \sum_{i=1}^n y_i - na^2b = 0$$

$$\sum_{i=1}^n y_i = b \sum_{i=1}^n x_i + na \text{ and } a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$

Substituting the solution for a into the equation for above equation gives

$$(b^2 - 1) \sum_{i=1}^n x_i y_i + b \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) - (b^2 - 1) \sum_{i=1}^n x_i \left[\frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \right] + 2b \sum_{i=1}^n y_i \left[\frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \right] - nb \left[\frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \right]^2 = 0$$

Rearrangement of terms yields a quadratic function of b,

$$b^2 \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right] + b \left[n \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2 + \left(\sum_{i=1}^n y_i \right)^2 \right] - \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right] = 0$$

$$b_{1,2} = \frac{- \left[n \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2 + \left(\sum_{i=1}^n y_i \right)^2 \right] \mp \sqrt{\left[n \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2 + \left(\sum_{i=1}^n y_i \right)^2 \right]^2 + 4 \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right]^2}}{2 \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right]}$$

$$b_{1,2} = \frac{\left[\left(\sum_{i=1}^n x_i \right)^2 - \left(\sum_{i=1}^n y_i \right)^2 - n \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) \right] \mp \sqrt{\left[\left(\sum_{i=1}^n x_i \right)^2 - \left(\sum_{i=1}^n y_i \right)^2 - n \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) \right]^2 + 4 \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right]^2}}{2 \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right]}$$

We can write regression equation $\hat{Y} = a + bX$. Value b is the slope of the regression equation. There are two different real roots of b. Scatter diagrams are drawn. Value b which is the same sign of slope of draft line is selected.

FINDINGS

For comparison of OR and CLR techniques, the analyses were conducted using an example.

A simple example

The relationship between the hardness and the durability of a material produced in a production center was investigated. 10 pieces of substances produced for this purpose were selected. Then, the hardness and durability test was conducted. The data obtained is as in Table 1.

Table 1. Example Data

Piece No	X (Hardness)	Y (Durability)
1	7	10
2	9	12
3	5	6
4	8	9
5	6	8
6	9	11
7	7	10
8	4	5
9	8	10
10	7	9

Table 1 displays data. Table 2 shows quantities needed for calculation of regression equation.

Table 2. Quantities Needed for Calculation of Regression Equation

Piece No	X(Hardness)	Y(Durability)	X ²	Y ²	XY
1	7	10	49	100	70
2	9	12	81	144	108
3	5	6	25	36	30
4	8	9	64	81	72
5	6	8	36	64	48
6	9	11	81	121	99
7	7	10	49	100	70
8	4	5	16	25	20
9	8	10	64	100	80
10	7	9	49	81	63
Total	70	90	514	852	660

We want to predict the values of durability from the values of hardness according to Table 2 based on the following equation.

$$\hat{Y} = a + bX$$

$$90 = 10a + 70b \text{ or } 9 = a + 7b$$

$$b_{1,2} = \frac{180 \mp \sqrt{(180)^2 + 4(300)^2}}{2(300)} = \frac{180 \mp \sqrt{392400}}{600} \quad b_1 = 1.34 \text{ and } b_2 = -0.74$$

Table 3 shows the regression equations.

Table 3. Regression Equations

Regression Equations	Slope, b	Intercept, a	$\hat{Y} = a + bX$	Correlation Coefficient, r
OR line	$b_1 = 1.34$	$a_1 = -0.38$	$\hat{Y} = -0.38 + 1.34X$	0.945
	$b_2 = -0.74$	$a_2 = 14.18$	$\hat{Y} = 14.18 - 0.74X$	
CLR line	$b = 1.25$	$a = 0.25$	$\hat{Y} = 0.25 + 1.25X$	0.945

Table 3 displays the Pearson correlation coefficient ($r=0.945$, $r^2=0.893$) indicating a good linear fit. The Pearson correlation between Hardness and Durability is 0.945, indicating that Hardness explains about 89.3 % of the variance in Durability.

OR line has equation $\hat{Y} = -0.38 + 1.34X$, whereas CLR line has equation $\hat{Y} = 0.25 + 1.25X$.

OR line is $Durability = -0.38 + 1.34 * Hardness$ while the CLR line is $Durability = 0.25 + 1.25 * Hardness$.

OR and CLR lines are shown in Figure 2 for data.

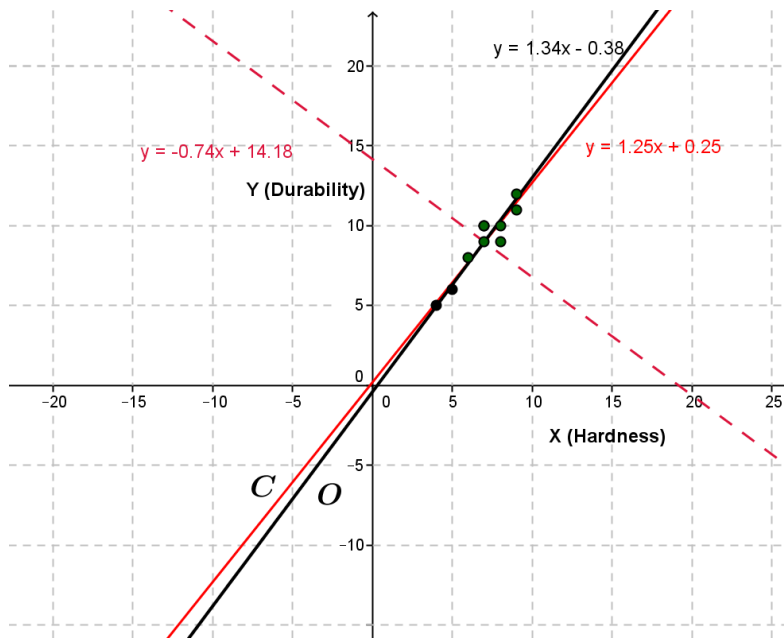


Figure 2. CLR Line (C: $\hat{Y} = 0.25 + 1.25X$) and OR Line (O: $\hat{Y} = -0.38 + 1.34X$) Applied to the Hardness and Durability of Substances Data

Table 4 shows the sum of squared perpendicular distances of regression lines.

Table 4. The Sum of Squared Perpendicular Distances of Regression Lines

Regression Lines	CLR Line c: $\hat{Y} = 0.25 + 1.25X$	OR Line o: $\hat{Y} = -0.38 + 1.34X$
$E = \sum_{i=1}^n (y_i - bx_i - a)^2$	4.50	4.69
Sum of Squares Perpendicular Distances $E_{\perp} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - bx_i - a)^2}{b^2 + 1}$	$\frac{4.5}{(1.25)^2 + 1} = 1.75$	$\frac{4.69}{(1.34)^2 + 1} = 1.67$

Presented in Table 4, CLR line and OR line give the values of E_{\perp} , 1.75 and 1.67 respectively.

We can see that OR line has smaller sum of squared perpendicular distances than CLR line which implies it has better performance than CLR in the prediction of durability from hardness of substances in this example.

RESULTS AND DISCUSSION

Regression analysis is expressed in a mathematical equation representing the relationship between variables. The effect of independent variable on dependent variable can be predicted via that mathematical equation (Büyüköztürk, Çokluk & Köklü, 2012). The fundamental purpose of simple linear regression analysis is to determine the best method that predicts the dependent variable. The

purpose of this study was the trivial presentation of the equation for OR and the comparison of OR and CLR according to the sum of squared perpendicular distances.

As Akdeniz (2013), Genç et al. (2003) and Saraçlı et al. (2009) stated, simple linear regression analysis is used to define the nearest line to the data points. As a result of comparison of the value of E_{\perp} attained from two regression equations, it is found that value of E_{\perp} for OR line is smaller. This result shows that OR has better performance than CLR in the prediction of durability from hardness of substances. OR line is closer to points than CLR line. OR line appears to present a much better fit for the data. This result is also similar to the finding from Calzada and Scariano (2003), Ding et al. (2013), Glaister (2005) and Ortiz et al. (2010). Although, OR is a better method than classical linear regression with respect to the sum of squared perpendicular distances, this result contradicts with the results of Carr (2012) and Saraçlı (2011).

Glaister (2005), Scariano and Barnet (2003), Deming (1943), Stöckl et al. (1998) stated that both dependent variable and independent variable are subject to measurement error in practice. Then, the result of our study indicates that it is precisely these circumstances for which the statistical distance to be minimized is the shortest distance to the line, and the appropriate regression line is OR line as shown in Table 4. This finding is also similar to the finding from Glaister (2005).

This research has revealed that OR is a better method than CLR with respect to the sum of squared perpendicular distances. It is more accurate to use OR technique, when there are measurement errors in both dependent variable and independent variable.

Depending on these results, the OR is thought to be a regression technique to obtain more accurate results than CLR at simple linear regression studies.

REFERENCES

- Adcock, R.J. (1878). A Problem in least squares, *Annals of Mathematics*, 5(2), 53-54.
- Akdeniz, F. (2013). *Olasılık ve İstatistik*, Nobel Kitabevi, 18. Press, Ankara, 459-492.
- Alma, Ö.G., & Vupa, Ö. (2008). Regresyon Analizinde kullanılan En Küçük Kareler ve En küçük Medyan Kareler Yöntemlerinin Karşılaştırılması, *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi*, 3(2), 219-229.
- Atar, B. (2010). Basit Doğrusal Regresyon Analizi ile Hiyerarşik Doğrusal Modeller Analizinin Karşılaştırılması, *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 78-84.
- Bahar, H.H. (2006). An Evaluation of KPSS Scores According to Grade Point Average and Gender, *Education and Science*, 31(140), 68-74.
- Bağçeci, B., Döş, B., & Sarıca, R. (2011). An Analysis of Metacognitive Awareness Levels and Academic Achievement of Primary School Students, *Mustafa Kemal University Journal of Social Sciences Institute*, 8(16), 551-566.
- Bayat, N., Şekercioğlu, G., & Bakır, S. (2014). Okuduğunu Anlama ve Fen Başarısı Arasındaki İlişkinin Belirlenmesi, *Eğitim ve Bilim, (tedmem)*, 39(176), 457-466.
- Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2012). *Sosyal bilimler için İstatistik*. Ankara: Pegem Akademi.
- Calzada, M.E., & Scariano, S.M. (2003). Contrasting total least squares with ordinary least squares part II: Examples and Comparisons, *Mathematics and Computer Education*, 37(2), 159-174.
- Carr, J.R. (2012). Orthogonal regression: A Teaching perspective, *International Journal of Mathematical Education in Science and Technology*, 43(1), 134-143.
- Carroll, R.J., & Ruppert D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models, *The American Statistician*, 50 (1), 1-6.
- Coşkuntuncel, O. (2013). The Use of alternative Regression Methods in Social Sciences and the Comparison of Least Squares and M Estimation Methods in Terms of the Determination of Coefficient, *Educational Sciences: Theory & Practice*, 13(4), 2139-2158.
- Deming, E.W. (1943). *Statistical Adjustment of Data*, Dover Publications, Inc., New York.
- Ding, G., Chu, B., Jin, Y., & Zhu, C. (2013). Comparison of orthogonal regression and least squares in measurement error modeling for prediction of material property, *Nanotechnology and Material Engineering Research. Advanced Materials Research*, 661, 166-170.

- Doruk, M., Özdemir, F., & Kaplan, A. (2015). The Relationship Between Prospective Mathematics Teachers' Conceptions on Constructing Mathematical Proof and Their Self-Efficacy Beliefs Towards Mathematics, *Kastamonu Üniversitesi Kastamonu Eğitim Dergisi*, 23(2), 861-874.
- Elfessi, A., & Hoar, R.H. (2001). Simulation study of a linear relationship between two variables affected by errors, *Journal of Statistical Computation and Simulation*, 71(1), 29-40.
- Fišerová, E., & Hron, K. (2010). Total least squares solution for compositional data using linear models, *Journal of Applied Statistics*, 37(7), 1137-1152.
- Genç, Y., Sertkaya, D., & Demirtaş, S. (2003). Klinik Araştırmalarda İki Ölçüm Tekniğinin Uyumunu İncelemede Kullanılan İstatistiksel Yöntemler. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 56(1), 1-6.
- Glaister, P. (2005). The use of orthogonal distances in generating the total least squares estimate, *Mathematics and Computer Education*, 39(1), 21-30.
- Golub, G.H., & Van Loan, C.F. (1980). An Analysis of the total least squares problem, *SIAM Journal on Numerical Analysis*, 17(6), 883-893.
- İlğan, A., Erdem, M., Yapar, B., Aydın, S., & Aydemir, Ş.Ş. (2012). Parents Interest and Regression Level of Primary State School Students Level Determination Exam (SBS) Score, *Journal of Educational Sciences Research*, 2(2), 1-17.
- Isobe, T., Feigelson, E.D, Akritas, M.G., & Babu, G.J. (1990). Linear regression in astronomy I, *The Astrophysical Journal*, 364, 104-113.
- Kane, M.T., & Mroch, A.A. (2010). Modeling group differences in OLS and Orthogonal Regression: Implications for differential validity studies, *Applied Measurement in Education*, 23, 215-241.
- Kermack, K.A., & Haldane, J.B.S. (1950). Organic correlation and allometry, *Biometrika Trust*, 37(1), 30-41.
- Kesicioğlu, O.S., & Güven, G. (2014). Investigation of the Correlation Between Preservice Early Childhood Teachers' Self-Efficacy Levels and Problem Solving, Empathy and Communication Skills, *Turkish Studies, International Periodical For the Languages, Literature and History of Turkish or Turkic*, 9(5), 1371-1383.
- Lane, D. M. (2016). "Introduction to Linear Regression", <http://onlinestatbook.com/2/regression/intro.html> Accessed 21 February 2016.
- Leng, L., Zhang, T, Kleinman, L., & Zhu, W. (2007). Ordinary Least Square Regression, Orthogonal Regression, Geometric Mean Regression and Their Applications in Aerosol Science, *Journal of Physics, Conference Series* 78(1), 1-5, <http://iopscience.iop.org/1742-6596/78/1/012084> Accessed 17 March 2016.
- Li, H.C. (1984). Generalized problem of least squares, *The American Mathematical Monthly*, 91(2), 135-137.
- Lolli, B., & Gasperini, P. (2012). A comparison among general orthogonal regression methods applied to earthquake magnitude conversions, *Geophysical Journal International*, 190(2), 1135-1151.
- Ludbrook, J. (2010). Linear regression analysis for comparing two measurers or methods of measurement: But which regression?, *Clinical and Experimental Pharmacology and Physiology*, 37, 692-699.
- Markovsky, I., & Van Huffel, S. (2007). Overview of total least- squares methods, *Signal Processing*, 87, 2283-2302.
- Montgomery D.C., Peck E.A., & Vining G.G. (2012). Introduction to Linear Regression Analysis, 5th Edition, 1-11. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470542810.html> Accessed 11 February 2016.
- Nievergelt, Y. (1994). Total least squares: State-of-the-art regression in numerical analysis, *SIAM Review (Society for Industrial and Applied Mathematics)*, 36(2), 258-264.
- Ortiz, J.V., Pogliani, L., & Besalú, E. (2010). Two-variable linear regression: Modeling with orthogonal least-squares analysis, *Journal of Chemical Education*, 87(9), 994-995.
- Öztürk, S. (2012). İstatistiksel Regresyon Yöntemlerinin Farklı Veri Gruplarına Uygulanması Üzerine Bir Analiz, *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2(2), 55-67.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2, 559-572.
- Saraçlı, S. (2011). Tip II Regresyon Tekniklerinin Monte-Carlo Simülasyonu ile Karşılaştırılması, *e-Journal of New World Sciences Academy*, 6(2), 26-35.
- Saraçlı, S., & Çelik, H.E. (2012). Metod Karşılaştırma Çalışmalarında Bland-Altman ve Tip II Regresyon Analizinin Karşılaştırılması, *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, 2(1), 11-14.
- Saraçlı, S., Doğan, İ., & Doğan, N. (2009). Medikal Metod Karşılaştırma Çalışmalarında Deming Regresyon Tekniği, *Türkiye Klinikleri J Biostat*, 1(1), 9-15.
- Scariano, S.M., & Barnett, II.W. (2003). Contrasting total least squares with ordinary least squares part I: Basic ideas and result, *Mathematics and Computer Education*, 37(2), 141-158.
- Sykes, A.O. (1993). An Introduction to Regression Analysis, *Coase-Sandor Institute for Law & Economics Working*, 20, 1-34.

- Stöckl, D., Dewitte, K., & Thienpont, L.M., (1998). Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data?, *Clinical Chemistry*, 44(11), 2340-2346.
- Üredi, I., & Üredi, L. (2005). İlköğretim 8. Sınıf Öğrencilerinin Öz-düzenleme Stratejileri ve Motivasyonel İnançlarının Matematik Başarısını Yordama Gücü, *Mersin University Journal of the Faculty of Education*, 1(2), 250-260.
- Warton, D.I., Wright, I.J., Falster, D.S. & Westoby, M. (2006). Bivariate line-fitting methods for allometry, *Biological Reviews*, 81(2), 259-291.

GENİŞ ÖZET

Giriş

Regresyon analizi, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi inceleme ve modellemede kullanan istatistiksel bir tekniktir (Montgomery vd., 2012; Sykes, 1993). Regresyon analizin uygulamaları tıpta, mühendislikte, ekonomide, astronomide, sosyal ve biyolojik bilimlerde, fen bilimlerinde ve eğitim bilimlerinde, kısacası hemen hemen her alanda görülmektedir. Aslında regresyon analizi en yaygın kullanılan istatistiksel tekniklerden biridir (Montgomery vd., 2012).

İstatistiğin önemli tahminleme (yordama) tekniklerinden biri olan regresyon analizinin en temel amacı; veri bulutunu temsil edebilecek bir doğru, eğri ya da yüzey denklemi yazmak için bu denklem üzerinden bağımsız değişkenin bilinmesi halinde bağımlı değişkenin ne olabileceğini tahmin etmektir. Bu tahminleme (yordama) regresyon doğrusu veya eğrisi üzerinden yapılmaktadır (Lane,

2016). Gözlenen x (bağımsız) değişkeniyle eşlenen \hat{Y} (bağımlı) değerinin bulunabilmesi için $\hat{Y} = f(X)$ eşitliği kullanılmaktadır. Doğrusal regresyon doğrusu $\hat{Y} = a + bX$ ile gösterilir ki

buradaki \hat{Y} , tahmin edilen değerdir. Klasik doğrusal regresyon (KDR) denklemi gözlenen verilerin regresyon doğrusuna olan düşey uzaklıklarının kareleri toplamı minimize edilerek elde edilir (Ortiz vd., 2010; Ding vd., 2013; Elfessi ve Hoar, 2001; Isobe vd., 1990; Kane ve Mroch, 2010; Leng vd., 2007; Ludbrook, 2010). Şekil 1a'da görüldüğü gibi verilerin regresyon doğrusuna olan düşey uzaklıklarının kareler toplamı,

$$E = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - bx_i - a)^2 \text{ dir.}$$

Ortogonal regresyon (OR) ise gözlenen noktaların tahmin edilen regresyon doğrusuna olan dik uzaklıklarının kareleri toplamı minimize edilerek bulunur (Carr, 2012; Ortiz vd., 2010; Ding vd., 2013; Elfessi ve Hoar, 2001; Isobe vd., 1990; Kane ve Mroch, 2010; Leng vd., 2007; Li, 1984; Ludbrook, 2010; Nievergelt, 1994; Scariano ve Barnet, 2003).

Şekil 1b'de görüldüğü gibi verilerin regresyon doğrusuna dik uzaklıklarının kareleri toplamı,

$$E_{\perp} = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \frac{(y_i - bx_i - a)^2}{b^2 + 1} \text{ dir.}$$

OR farklı isimlerle bilinmektedir. Bunlardan bazıları, toplam en küçük kareler (total least squares) (Calzada ve Scariano, 2003; Elfessi ve Hoar, 2001; Golub ve Van Loan, 1980; Markovsky ve Van Huffel, 2007; Nievergelt 1994), değişkenlerde hatalar (errors-in-variables) (Carroll ve Ruppert, 1996; Fišerová ve Hron, 2010; Markovsky ve Van Huffel, 2007) ve major eksen (major axis) (Carr, 2012; Isobe vd., 1990; Kermack ve Haldane, 1950; Ludbrook, 2010) dir.

OR metodu ilk olarak 1878 yılında Adcock tarafından keşfedilmiştir (Adcock, 1878). Bu metot bir yüzyıldır bilinmesine rağmen klasik doğrusal regresyon kullanılmaktadır. Klasik regresyonun tercih edilmesinin temel nedeni hesaplanmasının kolaylığıdır (Ludbrook, 2010).

Doğrusal regresyon analizi noktalara en yakın doğruyu tanımlamak için kullanılır (Akdeniz, 2013; Genç vd., 2003; Saraçlı vd., 2009). Bir noktadan bir doğruya en kısa uzaklık dik uzaklıktır (Warton vd., 2006). OR noktalardan doğruya en kısa uzaklıkların kareleri toplamını minimize eden doğrudur. Üstelik KDR bağımsız değişkenin (X) ölçüm hatası içermediğini, hatanın kaynağının yalnızca bağımlı değişken (Y) olduğu varsayılır. Deming, 1943 yılında yazdığı “Statistical Adjustment of Data” isimli kitabında, En Küçük Kareler (EKK) regresyon tekniğinin varsayım hatalarına dikkat çekmiş ve bağımsız değişkenin (X) de pratikte hata içerebileceğini belirtmiştir (Saraçlı vd., 2009). Bunun yanında Glaister (2005) ve Stöckl vd. (1998) pratikte her iki değişkende ölçme hatası olduğunu belirtmiştir. Bu durumda KDR’deki tahminler pratikte artık doğru değildir (Glaister, 2005). Her iki değişkende ölçme hatası içeren OR tekniğinin (Carr, 2012; Glaister, 2005; Scariano ve Barnett, 2003) tahminleri daha doğrudur. Bağımsız değişkenin de ölçme hatasını içerdiği OR tekniği, hesaplamalarda daha iyi sonuçlar verebilir (Calzada ve Scariano, 2003; Carr, 2012; Warton vd., 2006).

Uluslararası literatür incelendiğinde OR ile diğer regresyon tekniklerini karşılaştıran oldukça çok araştırmanın olduğu görülmektedir (Calzada ve Scariano, 2003; Carr, 2012; Ding vd., 2013; Elfessi ve Hoar, 2001; Glaister, 2005; Leng vd., 2007; Lolli ve Gasperini, 2012; Lundbrook, 2010; Ortiz vd., 2010). Ülkemizde yapılan metot karşılaştırma çalışmaları incelendiğinde ise bu konuda yapılan çalışmaların oldukça az olduğu dikkat çekmektedir. Bland Altman ile EKK Açığortay Type II tekniğini (Saraçlı ve Çelik, 2012), Deming regresyon tekniği ile EKK tekniğini (Saraçlı vd., 2009), EKK ve robust M tahmin yöntemini (Coşkuntuncel, 2013), EKK ve en küçük medyan kareler yöntemlerini (Alma ve Vupa, 2008) ve basit doğrusal regresyon ile hiyerarşik doğrusal tekniklerini (Atar, 2010) karşılaştıran metot karşılaştırma çalışmaları yapılmıştır. Türkiye’de OR tekniği üzerine jeofizikte (Öztürk, 2012) ve istatistikte (Saraçlı, 2011) yapılan çalışmalar da azdır.

Hem ulusal hem de uluslararası bu karşılaştırma çalışmalarında, hata kareler ortalamasına göre KDR (Carr, 2012) ve En Küçük Kareler Açığortay (Saraçlı, 2011) tekniklerinin üstün olduğu gösterilmiştir. Bunların yanında OR tekniğinin; belirleyicilik katsayısına (R^2) göre (Calzada ve Scariano, 2003) ve hataların standart sapmasına göre (Ding vd., 2013; Ortiz vd., 2010) üstünlüğünü gösteren çalışmalar da vardır.

OR tekniğinin yararlarından bahsedilmesine rağmen teknik; basit doğrusal regresyon hakkında Türk literatüründe sosyal bilimler ve eğitim alanındaki araştırmalarda yalnızca KDR tekniği ile karşılaşılmıştır (Bağçeci vd., 2011; Bahar, 2006; Bayat vd., 2014; Doruk vd., 2015; İlğan vd., 2012; Kesicioğlu ve Güven, 2014; Üredi ve Üredi, 2005). Yapılan literatür çalışmasında, OR üzerine sosyal bilimler ve eğitim alanında ülkemizde yapılmış bir çalışmaya rastlanmamıştır. Bu doğrultuda bu çalışmanın bundan sonra yapılacak çalışmalara ışık tutması beklenmektedir.

Bu araştırmanın amacı OR eşitliğini açıkça sunmak ve klasik regresyon ile ortogonal regresyonu dik uzaklıklar kareler toplamına göre karşılaştırmaktır.

Yöntem

Kolay anlaşılması için OR eşitliği yeniden literatürden bağımsız olarak sunulmuştur. Çünkü bu metot üzerine çalışmakta olan çoğu literatür kısa ve yoğun sembol içermektedir (Calzada ve Scariano, 2003; Carr, 2012; Kermack ve Haldane, 1950; Nievergelt, 1994). Klasik doğrusal regresyon ile OR’yi dik uzaklıklar kareler toplamına göre karşılaştırmak için analizler bir örnek üzerinden gösterilmiştir. Örnek, Akdeniz’in (2013) kitabından alınmıştır.

Verilerin analizinde, verilerin düzenlenmesinde, OR eşitliğinin hesaplanmasında ve OR ve KDR’nin dik uzaklıklar kareler toplamının hesaplanmasında Microsoft Excel 2010 kullanılmıştır. KDR eşitliğinin hesaplanmasında SPSS 17.0 programı kullanılmıştır.

Sonuç ve Tartışma

Regresyon analizinde değişkenler arasındaki ilişki bir matematiksel eşitlikle ifade edilir. Bağımsız değişkenin bağımlı değişken üzerindeki etkileri bu matematiksel eşitlik yoluyla tahmin edilebilir (Büyüköztürk vd., 2012).

Bu araştırmada, OR eşitliğini açıkça sunmak ve klasik regresyon ile OR tekniğini dik uzaklıklar kareler toplamına göre karşılaştırmak amaçlanmıştır. Bu karşılaştırma bir örnek üzerinden gerçekleştirilmiştir.

Akdeniz (2013), Genç vd. (2003) ve Saraçlı vd. (2009)'nin belirttiği gibi basit doğrusal regresyon analizi veri noktalarına en yakın doğruyu tanımlamak için kullanılır. KDR doğrusunun ve OR doğrusunun dik uzaklıklar kareler toplamına göre karşılaştırılması sonucunda, OR'nin dik uzaklıklar kareler toplamının daha küçük olduğu görülmüştür. Bu sonuç, maddelerin setliklerinden dayanıklılığın tahmininde OR tekniğinin performansının KDR tekniğinin performansından daha iyi olduğunu göstermiştir. Buradan OR doğrusu, verilere KDR doğrusundan daha yakındır. Bu bağlamda OR doğrusunun verileri daha iyi temsil ettiği görülmüştür. Bu sonuç Calzada ve Scariano (2003), Ding vd. (2013), Glaister (2005) ve Ortiz vd. (2010)'nin bulgularıyla benzerdir. OR tekniği dik uzaklıklar kareler toplamına göre KDR tekniğinden daha iyi bir metot olmasına karşın Carr (2012) ve Saraçlı'nın (2011) sonuçları ile çelişmektedir. Bağımlı ve bağımsız değişkenin her ikisinin de ölçme hatasını içeren ve doğruya en kısa uzaklığı minimize eden istatistik uzaklık koşulları altında OR doğrusunun Tablo 4'te görüldüğü gibi uygun bir regresyon doğrusudur.

Bu araştırma OR tekniğinin dik uzaklıklar kareler toplamına göre KDR tekniğinden daha iyi olduğunu, bağımlı ve bağımsız değişkenin her ikisinin de ölçme hatası içerdiği durumda OR tekniğinin kullanılmasının daha doğru olduğunu ortaya koymuştur.

Bu sonuçlara bağlı olarak OR tekniğinin basit doğrusal regresyon çalışmalarında KDR tekniğinden daha doğru sonuçlar elde etmede kullanılabilecek bir regresyon tekniği olduğu düşünülmektedir.

İlköğretim ve Ortaöğretim Başarı Ölçülerinin Yükseköğretime Geçiş Sınav Puanlarını Yordama Gücü

Predictive Power of Primary and Secondary School Success Criterion on Transition to Higher Education Examination Scores

Atilla ÖZDEMİR *

Selahattin GELBAL **

Öz

Ülkemizde sınıf atlamanın bir yolu olarak görülen eğitimin, bireylerin hayatını değiştiren önemli bir etkiye sahip olduğu bilinmektedir. Bu bağlamda ortaöğretim sonrası eğitim hayatına devam etmek içinse üniversite giriş sınavlarından başarılı olmak gereklidir. Üniversiteye giriş sınavının geleceği belirlemedeki en önemli faktör olduğu ülkemizde, üniversiteye yerleştirme puanına eklenen ortaöğretim başarı puanı da belirleyici bir faktör olmaktadır. Eğitim programları ve YGS içerikleridüşünüldüğünde ortaokuldaki ilgili derslerin başarı puanlarının öğrencilerin yükseköğretime girişteki birinci aşama sınavı olan YGS’yi yordaması beklenen durumdur. Üniversiteye giriş puanına eklenerek yerleştirme puanını oluşturan ortaöğretim başarı puanı düşünülduğünde, ortaöğretimdeki ilgili derslerin başarı puanlarının YGS alt testi ham puanlarını ne ölçüde yordadığının bilinmesi eğitim sistemimizin geri dönüşü ve bütünlüğü açısından önemlidir. Sonuç olarak aynı öğrencinin ortaokul ve lise yılları boyunca ilgili derslerdeki başarı puanlarının boylamsal olarak izlenip bu başarının üniversiteye giriş sınavının birinci aşaması olan YGS alt testi ham puanlarını yordama gücüyle ilgili elde edilen bulgular eğitim sistemimize sağlayacağı geri bildirim açısından önemlidir. Bu sebeplerden dolayı araştırmada; 2013 yılında YGS’ye giren öğrencilerin bu sınavdaki alt testlerle ilgili 7-12 sınıf yıl sonu başarı puanlarının YGS alt testi ham puanlarını yordama gücüne bakılmıştır. Araştırmanın hedef evrenini 2012-2013 öğretim yılında Ankara ilinde YGS’ye giren öğrenciler, çalışma evrenini ise Altındağ ilçesinde 2013 yılı YGS’ye giren 533 öğrenci oluşturmaktadır. Çalışmanın verileri Ankara ili Altındağ ilçesinde bulunan üç farklı okuldaki 533 öğrenciden elde edilmiştir. Araştırma sorularını yanıtlamak üzere adimsal çoklu regresyon analizi kullanılmıştır. Araştırmada elde edilen bulgulara göre YGS Türkçe alt testini en iyi yordayan değişkenlerin 11. sınıf dil ve anlatım, 8. sınıf Türkçe, 12. sınıf dil ve anlatım, 7. sınıf Türkçe ve 12. sınıf edebiyat yıl sonu başarı puanları olduğu ve bu değişkenlerin YGS Türkçe alt testindeki başarının %58’ini açıkladığı belirlenmiştir. YGS Matematik alt testini en iyi yordayan değişkenlerin 12. sınıf matematik, 8. sınıf matematik, 12. sınıf geometri, 10. sınıf matematik ve 7. sınıf matematik yıl sonu başarı puanları olduğu ve bu değişkenlerin YGS matematik alt testindeki başarının %71’ini açıkladığı tespit edilmiştir. YGS Fen bilimleri alt testini en iyi yordayan değişkenlerin 12. sınıf seçmeli kimya, 11. sınıf seçmeli kimya, 12. sınıf seçmeli biyoloji ve 11. sınıf seçmeli fizik derslerine ait yıl sonu başarı puanları olduğu ve bu değişkenlerin YGS Fen bilimleri alt testindeki başarının %50’sini açıkladığı saptanmıştır. YGS Sosyal bilimler alt testini en iyi yordayan değişkenlerin 12. sınıf seçmeli coğrafya, 11. sınıf T.C. inkılap tarihi ve Atatürkçülük ve 10. sınıf coğrafya derslerine ait yıl sonu başarı puanları olduğu ve bu değişkenlerin YGS Sosyal Bilimler alt testindeki başarının %50’sini açıkladığı belirlenmiştir.

Anahtar Kelimeler: çoklu regresyon, yükseköğretime geçiş sınavı, ilköğretim başarı puanı, ortaöğretim başarı puanı

Abstract

It is seen that education has a significant effect that changes an individual’s life in our country in which education is a way of moving up the social ladder. In order to continue to a higher education program after

* Yrd. Doç. Dr. Sinop Üniversitesi, Sinop-Türkiye, e-posta: atillaozdemir@sinop.edu.tr

** Prof. Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Bilim Dalı, Ankara-Türkiye, e-posta: sgelbal@gmail.com

graduating from high school, students have to succeed in transition to higher education examination. Thus, the entrance exam is an important factor to determine the future of the students. In our country, middle school grades and high school grade point average that is added to university placement score are also determinants. When spiral structure of our curriculum is considered, it is expected that related courses' grades at middle school will predict the scores obtained from the first stage of transition to higher education exam (YGS). Since high school grade point average forms university placement score, being aware of how related courses' achievement scores at secondary school predict raw scores of YGS subtests is significant in terms of our education system's feedback and integrity. As a result, observing students' achievement scores in related courses during their middle and high school education longitudinally and predicting raw scores on the subtests of the first stage of university entrance exam, YGS, from middle school and high school achievement scores are substantial with regards to provide feedback to our education system. Because of those reasons, the predictive power of 7th - 12th grade year-end grade point averages of students who took YGS in 2013 on their 2013 YGS subtests' raw scores is examined. Students who took YGS exam in Ankara province at 2012-2013 school year formed the aimed population of this study and 533 students who took YGS exam in 2013 in Altındağ district of Ankara formed target population of the study. Data was obtained from 533 students at three different schools in Altındağ district of Ankara province. Stepwise multiple regression analysis was used to answer research questions. According to the findings of the research; it was found that the best variable that predict YGS Turkish subtest are 11th Grade's Language and Expression, 8th Grade's Turkish, 12th Grade's Language and Expression, 7th Grade's Turkish, and 12th Grade's Literature's year-end grade point averages and those variables explain about 58% of the variance in YGS Turkish subtests scores. The best variables which predict YGS Mathematics subtest scores are 12th Grade's Math, 8th Grade's Math, 12th Grade's Geometry, 10th Grade's Math, and 7th Grade's Math's year-end grade point averages and those variables explain 71% of variance in YGS Mathematics subtests scores. The best variables which predict YGS Physical Sciences subtest scores are 12th Grade's elective Chemistry, 11th Grade's elective Chemistry, 12th Grade's elective Biology, and 11th Grade's elective Physics' year-end grade point averages and those variables explain about 50% of YGS Physical Sciences subtests scores. The best variable which predict YGS Social Sciences subtest scores are 12th Grade's elective Geography, 11th Grade's Republic of Turkey Revolution History and Kemalism, 10th Grade's Geography's year-end grade point averages and those variables explain about 50% of YGS Social Sciences' subtests scores.

Keywords: Multiple regression, the Transition to Higher Education Examination, YGS, Primary School Grade Point Average, High School Grade Point Average.

GİRİŞ

İnsanoğlu eski zamanlardan beridoğayı keşfetmek ve yaşam kalitesini artırmak için çalışmaktadır. Böylece her gelen nesil bir önceki neslin bilgi birikiminden faydalanarak bu çabaya katkı sağlamaktadır. Bu bilgi birikimini sistemli bir şekilde aktarmanın yolu olarak okullar görülmüş ve bu okullarda uygulanan eğitim sistemi geliştirilmiştir. Eğitim, bireylerin davranışlarında istenilen değişiklikler gelmesini amaçlayan bir sistemdir (Turgut ve Baykul, 2012). Her ne kadar bu amaç tek hedef gibi görünse de olayın birden fazla boyutu olduğu aşikârdır. Bunlar eğitimin sadece bireyi geliştirme değil aynı zamanda toplumsal, siyasal ve ekonomik işlevleri olarak karşımıza çıkmaktadır (Sönmez, 2010). Yirminci yüzyılın başları ve ortalarındaki eğitim anlayışında, yetişecek bireylerin dört işlem yapabilmesi, okuma ve yazma becerilerinin üst düzeyde olması, eğitim açısından yeterli sayılmıştır (Erdem, 2005; Gürdal, 2000). Ancak 1900'lerin sonlarına doğru yalnızca bu becerileri kazanmış bireyler, çağın artan taleplerine uyum sağlamakta zorluk çekmiş ve çağı yakalamada yetersiz kalmıştır (Karip ve Köksal, 1996; Saban, 2007). Bu neden, ülkelerin eğitim politikalarında ve felsefelerinde değişiklikler yapılmasını zorunlu kılmıştır çünkü içinde bulunulan yüzyıl, bilgi ve teknoloji çağıdır. Bu değişimin aktarılması için de eğitim felsefelerinin yeniden düzenlenmesine ihtiyaç duyulmuştur.

Eğitim bir ülkenin ekonomik, sosyal ve siyasi alanlarda ileriye dönük uzak hedeflerinin belirlenmesinde en önemli etkenlerden biri olmakla birlikte aynı zamanda o ülkenin gelişmişlik düzeyini de göstermektedir (Yıldırım, 1999). Eğitimin kalitesinin ölçülmesinde ise sınavlar sıklıkla kullanılmaktadır. Öğrenci başarılarının izlenilmesi ve değerlendirilmesi sonucu birçok kararlar alınabilmektedir. Bu değerlendirmeler ulusal ve uluslararası düzeyde olmaktadır. Öğrenme düzeylerinin belirlenmesine yönelik olarak izleme ve değerlendirme sınavları da yapılmaktadır. Bu uygulamalar okullar, ilçeler ve iller düzeylerinde yapılmakla beraber ulusal

düzeyleerde de rutin aralıklarla gerçekleştirilmektedir. Öğrenci Başarısı Belirleme Sınavları (ÖBBS) Milli Eğitim Bakanlığı tarafından ilköğretim ve ortaöğretim düzeylerinde yapılmıştır. Uluslararası düzeylerde yapılan PISA, TIMMS ve PIRLS gibi tarama sınavlarına da ülkemiz katılmaktadır. Bu sınav sonuçlarına göre ülkemizdeki eğitim programlarının güncellenmesi, ülkelere göre öğrenci özelliklerinin karşılaştırılması gibi değerlendirmeler yapılabilmektedir. Uluslararası yapılan sınavlar üzerinde farklı araştırmacılar da çalışmalar yapmaktadırlar.

Aydın, Sarier ve Uysal (2012) tarafından yapılan bir çalışmada PISA sonuçları bağlamında Türkiye ve en başarılı beş ülkenin (Finlandiya, Kore, Hollanda, Japonya ve Kanada) sosyoekonomik ve sosyokültürel değişkenler açısından matematik başarı düzeylerini karşılaştırmışlardır. Araştırmada, PISA sonuçlarındaki başarısızlığın nedeni olarak Türk Eğitim Sistemi gösterilmektedir. Dünya Bankası Eğitim Raporu'nda (2005) Türk okul sisteminin, çok az öğrenciyi iyi eğittiği ve öğrencilerin çoğunu başarısız kıldığı vurgulanmıştır. Ayrıca raporda eğitim sisteminin en temel amaçları arasında yer alan etkin düşünme, algılama ve problem çözme yeteneklerini geliştirmede yetersiz olduğu belirtilmektedir. Bu nedenle milli eğitimin temel öğelerini kazandırmaya yönelik eğitim faaliyetlerinin gözden geçirmesi ihtiyacı olduğu belirtilmektedir. Ulusal düzeyde yapılan değerlendirmelerin birçoğu bir üst öğrenim düzeyine öğrenci seçmek amacıyla yapılmaktadır. Bu amaçla temel eğitimden ortaöğretime geçişte “Temel Eğitimden Orta Öğretime Geçiş Sistemi” (TEOG) kullanılmaktadır. Ortaöğretimden yükseköğretime geçişte de “Yükseköğretime Geçiş Sınavı” (YGS) ve “Lisans Yerleştirme Sınavı” (LYS) kullanılmaktadır.

Ortaöğretim zorunlu olmakla beraber öğrenci seçimi seçme sınavı ile yapılmaktadır. İlk defa 2013-2014 eğitim öğretim yılında uygulanmaya başlanan temel eğitimden ortaöğretime geçiş sisteminde (TEOG), matematik, Türkçe ve fen bilgisi derslerinin ikinci sınavları ile din kültürü ve ahlak bilgisi, T.C. İnkılap tarihi ve Atatürkçülük ve yabancı dil derslerinin birinci sınavları ortak sınav olarak yapılmaktadır. Yerleştirmeye esas puan olarak ortak sınavlardan alınan puanların yüzde yetmiş ile altı, yedi ve sekizinci sınıf not ortalamalarının yüzde otuzu alınmaktadır. Böylece öğrencilerin ortaöğretimde kayıt yaptıracakları lise türüne yerleşme puanları elde edilmektedir.

Yükseköğretime geçişte günümüze kadar farklı seçme siteleri uygulanmıştır. Üniversiteye öğrenci seçiminde ilk uygulamalar üniversitelerin kendilerine bırakılmıştır. Daha sonra merkezi sınavlar yapılarak yerleştirmeler tek merkezden yapılmaya başlanmıştır. Bu amaçla 1974'te ÜSYM kurulmuş ve halen bu görevi Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) yürütmektedir. Farklı dönemlerde öğrenci seçme süreci tek ve iki aşamalı olarak yapılmıştır. Halen iki aşamalı olarak yapılan sınav uygulaması dersler bazında birden çok sınavla yapılmaktadır.

Yapılan bu seçme sınavlarının amacı, öğrencilerin bir üst öğrenime devam ederken geçiş yapacağı bölümü belirlemek ve sınava katılan adaylar arasından en iyi olanlar ile yeterli niteliğe sahip olanları seçmektir. Sınavların, amaçlarını gerçekleştirebilmesi için bazı özelliklere sahip olması gerekir. Bu özellikler arasında en önemlileri, uygulanan ölçme aracının geçerliği, güvenilirliği ve kullanışlılığıdır. ÖSYM tarafından hazırlanan YGS (Yükseköğretime Geçiş Sınavı) testleri de bir seçme sınavında kullanılan birer ölçme aracı olduğundan bunların da yukarıda belirtilen özelliklere sahip olmaları beklenir (Turgut ve Baykul, 2012).

Turgut ve Baykul (2012) güvenilirliği, ölçme sonuçlarının tesadüfi hatalardan arınlık derecesi, geçerliği ise bir ölçme aracının hangi özellik ya da özellikleri ölçmeyi amaçladıysa, başka özelliklerle karıştırmadan, o özelliği ölçebilme derecesi olarak tanımlamaktadır. Başka bir anlatımla geçerlik, ölçme aracının geliştirilme amacına hizmet etme derecesidir. Üç çeşit geçerlik türü vardır; kapsam geçerliği, yapı geçerliği ve ölçüt bağıntılı geçerliktir. Ölçüt bağıntılı geçerlik, ölçüt ölçüleri elde edilmiş zamanına göre uygunluk, hem zaman geçerliği ve yordama geçerliği olmak üzere ikiye ayrılmaktadır (Aiken, 1971; Anastasi, 1988; Baykul, 2000; Crocker ve Algina, 1986; Tekin, 2007). Test geliştiriciler, testin geliştirme amaçlarına uygun güvenilirlik ve geçerlik belirleme yollarını kullanmaktadırlar. Seçme ve yerleştirme sınavlarındaki en önemli geçerlik türü yordama geçerliğidir (Baykul, 2000; Crocker ve Algina, 1986).

Eğitimde önemli kararlara dayanak olan seçme sınavlarının geçerlik kanıtlarından biri, bu sınavların öğrencilerin gelecekteki başarılarını yordama gücüdür. Diğerleri ise sınavın, sınava giren öğrencilerin

önceki öğrenmeleriyle olan ilişki düzeyidir (Aiken, 1971). Yordama, eldeki bilgiler ışığında gelecek ile ilgili tahminde bulunmaktır. İstatistiksel açıdan bakıldığında ise yordama, istatistiksel teknikler kullanılarak ve bilinenlerden yararlanarak bilinmeyen durumlar hakkında yapılan geleceğe yönelik tahminlerde bulunma işlemidir (Arıcı, 2001).

Yordama geçerliği testin bireylerin ileride devam edeceği eğitim programı ya da işteki başarısını tahmin etmedeki isabetlilik derecesini gösteren geçerliktir. Yordama geçerliği çalışmaları, karar verilmenden önce elde edilen test puanları ile karar verildikten sonra elde edilen ölçüt puanları arasındaki ilişkiyi belirlemek için kullanılmaktadır. Dolayısıyla testten yüksek puan alan bireylerin işte ya da belirli bir eğitim programında başarılı olacağı, testten düşük puan alanların ise başarılı olamayacağı yorumu yapılır. Bireyin gelecekte ne kadar iyi performans göstereceği bilinemez ancak iyi bir test doğru yordama yapılmasına olanak verir. Sonuç olarak test puanları ile ölçüt puanları arasındaki ilişki, seçme ve yerleştirme kararlarının geçerliğinin bir ölçüsü olarak kullanılır (Thorndike, 1982). Verilen tanımlardan da anlaşılacağı gibi yordama geçerliği iki değişken arasındaki ilişki ile açıklanmaktadır. Hazırlanan bir testin yordama geçerliği ise test puanları ile bu test puanlarına ölçüt olabilecek başka başarı ölçüsü arasındaki korelasyon ile bulunabilir. Bunlardan birincisine yordayıcı puan, ikincisine ise ölçüt puan adı verilir (Crocker ve Algina, 1986).

Bu araştırmada, öğrencilerin YGS puanlarını yordadığı düşünülen ortaokul başarı puanları ve lise başarı puanları birer yordayıcı değişken olarak ele alınmıştır. Bu değişkenlerin, YGS puanlarını ne derece yordadığının belirlenmesi ise araştırmanın temel problemini oluşturmaktadır. Araştırma da 2013 yılı YGS'deki alt testlerin ham puanları da ölçüt olarak kabul edilmiştir.

Araştırmanın Amacı ve Önemi

Üniversiteye giriş sınavının geleceği belirlemedeki en önemli faktör olduğu ülkemizde ilköğretim ve üniversiteye yerleştirme puanına eklenen ortaöğretim başarı puanı da belirleyici bir faktör olmaktadır. Eğitim müfredatımızın sarmal bir yapıda olduğu düşünüldüğünde ilköğretimden başarılı bir şekilde mezun olan bir öğrencinin ortaöğretimde, ortaöğretimden başarı bir şekilde mezun olan bir öğrencinin de üniversite giriş sınavında ve yükseköğretimde başarılı olması beklenen bir durumdur. Buradan hareketle ortaokuldaki ilgili derslerin başarı puanlarının öğrencinin gelecekte gireceği yükseköğretime girişteki birinci aşama sınavı olan YGS'yi yordaması beklenir. Üniversiteye giriş puanına eklenerek yerleştirme puanını oluşturan ortaöğretim başarı puanı düşünüldüğünde ortaöğretimdeki ilgili derslerin başarı puanlarının YGS alt testi ham puanlarının ne ölçüde yordadığının bilinmesi eğitim sistemimizin geri bildirim ve bütünlüğü açısından önemlidir. Öğrencinin gelişimi bir bütün olarak ele alındığında, aynı öğrencinin ortaokul ve lise yılları boyunca ilgili derslerdeki başarı puanlarının boylamsal olarak izlenip bu başarının üniversiteye giriş sınavının birinci aşaması olan YGS alt testi ham puanlarını yordama gücüyle ilgili elde edilen bulgular eğitim sistemimize sağlayacağı geri bildirim açısından önemlidir. Yükseköğretime geçiş sınavı alt testlerinde sorulan soruların ölçtüğü bilgi ve beceriler, ortaokul ve lise müfredatı konuları içerisinde yer almaktadır.

Bu bilgilerden hareketle literatür incelendiğinde ortaokul başarı puanlarının YGS'yi yordamasıyla ilgili çok az çalışmaya rastlanması ve aynı öğrenciye ait ilgili derslerin başarı puanlarına ulaşılması, bu araştırmanın yapılmasına olanak sağlamıştır. Bu doğrultudayapılan bu araştırmada, ilköğretim ve ortaöğretimdeki ilgili derslerin yıl sonu başarı puanlarının YGS'nin alt testlerindeki ham puanları ne derecede yordadığı belirlenmeye çalışılmıştır. Hem ortaokul hem de lise ders notlarının alınmasının amacı sistemin sarmal yapısının da görülmek istenmesidir.

Bu bilgiler ışığında, araştırmanın problem cümlesi aşağıdaki gibi tanımlanmıştır.

Problem Cümlesi

İlköğretim ve ortaöğretimdeki ilgili derslerin yıl sonu başarı puanları, Yükseköğretime Geçiş Sınavı (YGS) alt testleri ham puanlarını ne düzeyde yordamaktadır?

Alt Problemler

Bu amaç doğrultusunda aşağıdaki alt problemlere yanıt aranmıştır:

1. Ortaokul 7-8. sınıflar Türkçe, ortaöğretim dil ve anlatım ile edebiyat dersleri yıl sonu başarı puanları YGS Türkçe alt testi ham puanlarını ne düzeyde yordamaktadır?
2. Ortaokul 7-8. sınıflar matematik ve ortaöğretim matematik dersleri yıl sonu başarı puanları YGS matematik alt testi ham puanlarını ne düzeyde yordamaktadır?
3. Ortaokul 7-8. sınıflar fen bilgisi ve ortaöğretim fen grubu dersleri yıl sonu başarı puanları YGS fen alt testi ham puanlarını ne düzeyde yordamaktadır?
4. Ortaokul 7-8. sınıflar ve ortaöğretim sosyal bilimler dersleri yıl sonu başarı puanları YGS sosyal bilimler alt testi ham puanlarını ne düzeyde yordamaktadır?

Sayıtlar

1. Farklı okul türündeki başarı puanlarının eşdeğer olduğu kabul edilmiştir. Çünkü çalışmada ele alınan okullar Anadolu lisesi, genel lise ve meslek lisesidir.
2. Farklı öğretmenler tarafından verilen aynı başarı puanının eşdeğer olduğu kabul edilmiştir. Çünkü veriler farklı okul türlerinden elde edilmiştir. Sonuç olarak bu okul türlerindeki öğrenci seviyelerine göre kazanımlar aynı olmasına rağmen öğrenci değerlendirme düzeylerinde değişimler olabilecektir.
3. Farklı ortamlarda yapılan sınavların öğrenci başarılarına etkisinin benzer olduğu kabul edilmiştir çünkü başarıyı etkileyen bir diğer faktör ortamdır. Verilerin toplandığı okul ortamlarındaki farklılığın başarıyı etkilemediği varsayılmıştır.
4. Türkçe okuduğunu anlama başarısının matematik, fen bilimleri ve sosyal bilimler alt testlerindeki başarıyı etkilemediği kabul edilmiştir.

Sınırlılıklar

1. Araştırma 2013 yılı YGS verileri ile sınırlıdır.
2. Araştırma Ankara ili Altındağ ilçesindeki üç lisesinin öğrenci verileri ile sınırlıdır.
3. İlköğretim başarı ölçütleri yedinci ve sekizinci sınıf Türkçe, matematik, fen ve teknoloji ile sosyal bilgiler dersleri yıl sonu ortalamalarıyla ve ortaöğretim başarı ölçütleri dokuzuncu, onuncu, on birinci ve on ikinci sınıf Türkçe, matematik, fizik, kimya ve biyoloji ile sosyal bilgiler dersleri yıl sonu ortalamalarıyla sınırlıdır.

Tanımlar

Yıl Sonu Başarı Puanı: Öğrencilerin o eğitim-öğretim yılına ait birinci dönem ve ikinci dönem başarı puanlarının aritmetik ortalamasıdır.

YGS Türkçe alt testi: YGS’de yer alan Türkçe dersine ait 40 soruyu ifade eder.

YGS Sosyal Bilgiler alt testi: YGS’de yer alan 15 adet tarih, 12 adet coğrafya, 8 adet felsefe ve 5 adet din kültürü ve ahlak bilgisi dersine ait soruları ifade eder.

YGS Matematik alt testi: YGS’de yer alan matematik dersine ait 40 soruyu ifade eder.

YGS Fen Bilimleri alt testi: YGS’de yer alan 14 adet fizik, 13 adet kimya, 13 adet biyoloji dersine ait soruları ifade eder.

İlgili Araştırmalar

Bu bölümde alan yazında yapılmış olan yurt içi ve yurt dışı çalışmalara yer verilmiştir. Taranan çalışmalar kronolojik olarak sıralanmıştır. Yurt içinde yapılan çalışmalarda Yıldırım (1972), Onay (1972), Akhun (1980), Oral (1985) ve Aşkar (1985) yılında yapmış olduğu çalışmalar incelenmiş ancak 1989 ve sonrası çalışmalar bu bölümde ele alınarak detaylı olarak bahsedilmiştir. Ayrıca yurt dışında yapılan yordama çalışmalarına konu olan sınavların içerikleriyle ilgili birtakım bilgiler verilmiştir. Amerika'da üniversite giriş sınavı olarak kullanılan en yaygın standartlaşmış iki sınav vardır. Bunlardan bir tanesi Amerikan College Test (ACT), diğeri ise Scholastic Assesment Test (SAT)'tır. Yurt dışında yapılan çalışmalarda bu iki sınav türüyle de ilgili pek çok araştırmaya rastlanmıştır. Alan yazın taraması sonucu bulunan çalışmalar aşağıdaki şekildedir:

Tavşancıl (1989) araştırmasında 1984-1985 öğretim yılında Ankara Üniversitesinin çeşitli bölümlerine kayıt yaptırmış öğrencilerin, mezun oldukları lise tür ve koluna göre akademik başarı karşılaştırmaları yapmıştır. Bu karşılaştırma bölümler düzeyinde; söz konusu her bir bilim alanında birden fazla fakülte olduğunda bölümler birleştirilerek fakülte; son olarak da bölüm ve/veya fakülteler birleştirilerek bilim alanları düzeyinde yapılmıştır. Çalışma toplam 944 öğrenciden oluşturmaktadır. Araştırmada kontrol edilen değişkenlerden biri lise not ortalamaları (LNO) diğer kontrol edilen değişken ise öğrenci yerleştirme sınavı puanları (ÖYSP) ve bağımsız değişken olarak kullanılan öğrencilerin mezun oldukları lise tür ve kollan, ayrıca geldikleri bölge, öğretim durumları ve yerleştirildikleri yükseköğretim kurumunu tercih sıralan ile ilgili bilgilerdir. Yapılan on beş karşılaştırmanın on birinde, LNO ile akademik başarı arasındaki korelasyon katsayısı manidar bulunmuş; ÖYSP ile akademik başarı arasında hesaplanan on beş korelasyon katsayısından bir tanesinin pozitif ve manidar, üç tanesinin negatif olduğu belirlenmiştir.

Kutlu (1989), yapmış olduğu çalışmada 1987-Eylül döneminde uygulanan Tıpta Uz manlık Eğitimi Giriş Sınavına katılan adayların, tıp fakültesi başarı ölçüleri ile, Tıpta Uzmanlık Eğitimi Giriş Sınavı puanları arasında nasıl bir ilişkinin (korelasyon) olduğunu bulmuştur. Araştırmada Ankara'da bulunan Hacettepe Üniversitesi Tıp Fakültesi ve Gazi Üniversitesi Tıp Fakültesini 1985, 1986 ve 1987 yıllarında bitirerek 1987-Eylül döneminde uygulanan Tıpta Uzmanlık Eğitimi Giriş Sınavı'na katılan adayların verileri kullanılmıştır. Araştırma sonucunda elde edilen bulgular;tıp fakültesi başarı ölçüleri ile tıp uzmanlık eğitimi giriş sınavı puanlarından hesaplanan korelasyonların, temel tıp bilimlerinde 0.07 ile 0.67 arasında, klinik tıp bilimlerinde ise -0.13 ile 0.55 arasında değişen değerler aldığını göstermektedir. Elde edilen bu korelasyonlardan başarı ölçüleriyle sınav puanları arasında anlamlı bir ilişkinin bulunduğu sonucuna ulaşılmıştır. Tıp fakültelerindeki başarı ölçüleri ile tıpta uzmanlık eğitimi giriş sınavı puanları arasında bulunmuş kanonik korelasyonlar 0.64 ile 0.76 arasında değişen değerler almıştır. Bu durum başarı ölçülerinde yansıyan zihinsel faktörlerle, sınav puanlarında yansıyan zihinsel faktörler bakımın dan önemli ölçüde benzerlikler bulunduğunu göstermektedir.

Gelbal (1989) yapmış olduğu çalışmada; ÖZ-DE-BİR'in hazırladığı 1988 Öğrenci Seçme Deneme Sınavı ile ÖSYM'nin hazırladığı ÖSS arasında öğrenci başarıları yönünden nasıl bir ilişki bulunduğunu, bu iki sınavdan hangisinin daha güvenilir olduğunu ve ÖYS puanlarını daha iyi yordadığını araştırmıştır. Sonuç olarak deneme sınavının sözel ve sayısal bölümleri ile ÖSS'nin sözel ve sayısal bölümlerinin ortalama güçlük dereceleri arasında anlamlı fark olmadığı, deneme sınavının sayısal bölümleri ile ÖSS'nin sayısal bölümleri arasındaki farkın anlamlı olmadığı, sözel bölümler arasında ise farkın anlamlı çıktığı sonucuna varmıştır. 0,05 anlamlılık düzeyinde deneme sınavı sözel ve sayısal bölümleri ile ÖSS'nin aynı bölümleri arasındaki ilişkiler anlamlı bulunmuştur. Her iki sınavın KR-21 güvenilirlikleri arasında anlamlı bir farkın olmadığı görülmüştür. ÖYS'deki alt testleri yordamada deneme sınavının alt bölümleri ile ÖSS'nin alt bölümleri arasında 0,05 anlamlılık düzeyinde, sadece sözel bölümlerin Türkçe alt testini yordamada anlamlı bir fark olduğu bulunmuştur.

Morgan (1990), SAT'ın yordama geçerliliğini belirlemek amacıyla 1978, 1981 ve 1985 yıllarında çeşitli üniversitelere yerleşen öğrencilerin SAT puanları ile üniversite birinci sınıf başarı düzeyleri arasındaki ilişkiyi incelemiş 1985 yılında yerleşen öğrencilerin, diğer yıllarda yerleşen öğrencilere

göre SAT puanları ile üniversite birinci sınıf başarı düzeyleri arasında güçlü bir ilişki olduğunu tespit etmiştir.

Tezbaşaran (1991) araştırmasında, Türkiye'deki Öğrenci Seçme ve Yerleştirme Sınavı Sistemi'nde 1987 yılında yapılan değişikliklerden sonra sınavın amacına hizmet etmesi yönünde beklenen değişmelerin ne ölçüde gerçekleştiğini bulmaya çalışmıştır. Bu amaçla 1983-1989 yılları arasında, ÖYS testlerine ait madde gücü ve güvenilirlik katsayıları, madde ayırıcılığı dağılımları ile testlerdeki mutlak başarı yüzdeleri ve yordama gücü, adayların tercihlerinde ortaya çıkan değişmeler, sınavı tekrarlama oranları incelenmiştir. ÖYS testlerine ait güvenilirlik katsayılarının 1983-1989 yılları arasında 0.86-0.96; 1987 öncesinde 0.86-0.93; 1987 ve sonrasında ise 0.91-0.96 arasında değiştiği; 1987 ve sonrasında, beklendiği gibi, ÖYS testlerinin güvenilirliklerinde yükselme olduğu görülmüştür. ÖYS testlerinin madde ayırıcılığı bakımından genel olarak seçme amaçlarına uygun olduğu, 1987 ve sonrasında, beklendiği gibi, maddelerin vasat ayırıcılık gücünde artış olduğu görülmüştür. Araştırmada kapsanan yükseköğretim programlarına 1983-1989 yıllarında yerleştirilen adayların, yerleşmelerine esas olan puan türünün hesaplanmasında büyük ağırlık, alan testteki mutlak başarı yüzdeleri ve bunların ortalamaları hesaplanmıştır. İncelemeler sonucunda, beklendiği gibi, 1987 ve sonrasında, öncesine göre yükseköğretim programlarına ÖYS testlerinde daha başarılı ol an adayların yerleştirilmeye başlandığı görülmüştür. 1987 yılında ÖYS'de kullanılan yordayıcılar ile üniversite birinci sınıf başarısının yordanmasıyla ilgili olarak 35 yükseköğretim programı 14 gruba ayrılarak incelenmiştir. Bu programlara yerleştirilen tüm adayların birinci yıl derslerinden aldıkları notlardan yıl sonu başarı ortalamaları hesaplanmıştır. Sonuçlar, 1987 ve sonrasında öncekilere göre, sayısal alanda güçlü adayların daha çok bu alanla, sözel alanda güçlü adayların da daha çok sözel alanla tutarlı tercihler yapmaya başladıklarını göstermiştir. 1983-1989 yıllarında yükseköğretim programlarına yerleştirilen tüm adayların, ilk yerleştikleri yılı izleyen yıllardaki sınavı tekrarlama oranları hesaplanmıştır. 1987 ve sonrasında, genel olarak, sınavı tekrarlama oranlarının beklendiği gibi düşmediği görülmüştür.

Myers & Pyles'in (1992) yürütmüş olduğu çalışmada ACT ile lise not ortalamasının üniversite başarısıyla ilişkisi araştırılmıştır. Yapılan çoklu regresyon sonucunda ACT puanları ile lise not ortalamasının üniversite başarısının %32'sini açıkladığı görülmüştür, lise başarısının üniversite başarısını açıklama yüzdesinin %22 olduğu, ACT puanlarının üniversite başarısını açıklama yüzdesinin ise %46 olduğu hesaplanmıştır. Araştırmanın sonunda lise başarısıyla ACT puanlarının beraber değerlendirildiğinde üniversite başarısını daha iyi açıkladığı sonucuna varılmıştır.

Karakaş (1998), araştırmasında eğitimde Ölçme Ve Değerlendirme Ana Bilim Dalındaki öğrencilerin ortaöğretim okullarındaki ders başarıları, ÖYS puanları ile bu ana bilim dalındaki akademik başarıları arasındaki ilişkileri ortaya koymayı amaçlamıştır. Araştırma Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme (EÖD) Ana Bilim Dalında 1993-1994, 1994-1995, 1995-1996 ve 1996-1997 öğretim yıllarında alınan 81 öğrencinin ortaöğretim okullarındaki ders başarıları, ÖYS puanları ve yükseköğretimdeki akademik başarıları ile sınırlandırılmıştır. Araştırmadan elde edilen bulgulara göre ortaöğretim okullarındaki matematik, Türk dili ve edebiyatı, tarih, inkılap tarihi, coğrafya, psikoloji, sosyoloji, felsefe ders başarıları ile ÖYS matematik, Türkçe, sosyal bilimler alt test puanları arasında 0.05 düzeyinde manidar bir ilişki bulunmamıştır. Ancak öğrencilerin ortaöğretim okullarındaki Türk dili ve edebiyatı ders başarıları ile yükseköğretimdeki Türk dili derslerine ilişkin başarıları arasında 0.05 düzeyinde manidar bir ilişki bulunmuştur. Ortaöğretim okul programlarının birbirinden farklı olması ve ders başarısını değerlendirmede, değerlendirmeciler arasındaki farklılıklar nedeniyle elde edilen başarı notlarının güvenilirliğinin düşük olmasının da korelasyonların düşük çıkmasına sebep olabileceği belirtilmiştir.

Üniversiteye giriş sınavlarıyla ilgili bir başka yordama çalışmasında Doğan (1999), Dershane Deneme Sınavları (DDS) ile Öğrenci Seçme Sınavı (ÖSS) ve Öğrenci Yerleştirme Sınavı (ÖYS) arasındaki ilişkiyi ve Dershane Deneme Sınavlarının yordama güçlerini araştırmıştır. DDS'nin ÖSS'yi yordama gücü, en düşük sözel ağırlıklı puanlar için olmak üzere 0,79; en yüksek sayısal ağırlıklı puanlar için olmak üzere 0,93 olarak bulunmuştur. DDS'nin ÖYS'yi yordama gücü en düşük

Türkçe - sosyal ağırlıklı puanlar için olmak üzere 0,72; en yüksek matematik ağırlıklı puanlar için olmak üzere 0,90 olarak bulunmuştur. DDS'nın ağırlıklı puanlarının ÖSS ve ÖYS'nin ağırlıklı puanlarını yordama güçlerinin yüksek olduğu sonucuna varılmıştır.

Erdoğan (1999), Ankara Üniversitesi Eğitim Bilimleri Fakültesinin dört ayrı bölümünden 1996-1997 öğretim yılı sonunda mezun olanların ÖSS ve ÖYS puanları ile üniversitedeki başarıları arasındaki ilişkilerin saptanmasını ve bu çerçevede mezunların ÖYS- Sosyal puanları ile Akademik başarıları arasındaki ilişkilerin değerlendirilmesini amaçlamıştır. Araştırma 1993-1994 öğretim yılında Ankara Üniversitesi Eğitim Bilimleri Fakültesini kazanmış ve 1996-1997 öğretim yılında mezun olmuş 242 kişi ile yapılmıştır. Araştırmada mezunların ÖSS ve ÖYS testlerinden aldıkları puanlar ile 1. sınıf Akademik Başarı Not Ortalaması ve 4 yıllık Ağırlıklı Akademik Başarı Not Ortalamaları (ABNO) arasındaki ilişkilere bakılmıştır. Sonuç olarak mezunların ÖSS-Sözel puanları ile 1. sınıf ABNO arasında, bölümlere göre saptanmış olan ilişkilerden (korelasyon(ardan) hiçbirisi manidar bulunmamıştır. Mezunların ÖSS-Sayısal puanları ile 1. sınıf ABNO arasındaki ilişkilere bölümlere göre bakıldığında bir bölüme ait olanı hariç (EPÖ bölümü mezunları için $r=0.27$ diğerleri manidar değildir. Mezunların ÖSS-Eşit Ağırlıklı puanları ile 1. sınıf ABNO arasındaki ilişkilere bölümlere göre bakıldığında, yine bir bölüme ait olanı hariç (EPÖ bölümü mezunları için ($r=0.27$)) diğerleri manidar değildir. Mezunların ÖSS-Sözel puanları ile 4 yıllık ABNO arasında bölümlere göre hesaplanan ilişkilerden hiçbirisi manidar değildir.

Tross, Harper, Osher & Kneidinger (2000), kolej öğrencilerinin üç kişilik özelliklerinin (başarı, bilinçlilik ve esneklik), SAT puanlarının ve lise akademik ortalamalarının üniversite birinci yıl akademik başarı ortalamasını yordama gücünü incelemişlerdir. Sonuç olarak "bilinçlilik" özelliğinin akademik varyansın 0,36'sını açıklayabildiği sonucuna varılmıştır.

Örnek (2002), çalışmasında Hacettepe Üniversitesi Sosyal Bilimler Enstitüsündeki yüksek lisans öğrencilerinin Lisansüstü Eğitime Giriş Sınavı (LES) puanları ile bir yıllık akademik ortalamaları arasındaki ilişkileri belirlemeyi ve bu doğrultuda yüksek lisans programlarına öğrenci seçiminde kullanılan LES ağırlıklı puan türlerinin uygunluğu hakkında bir ipucu elde edebilmeyi amaçlamıştır. Araştırmada Hacettepe Üniversitesi (HÜ) Sosyal Bilimler Enstitüsündeki yüksek lisans programlarına 1997-1998 öğretim yılı güz dönemi ile 1999-2000 öğretim yılı güz dönemi arasında LES sonuç belgesi ile kayıt yaptıran ve yüksek lisans programının bir yılını tamamlamış 507 öğrenciye ait veriler kullanılmıştır. Araştırma kapsamına alınan öğrencilerin LES puan türlerine ait puanlar ile yüksek lisanstaki akademik başarı puanları arasındaki ilişkiler Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılarak hesaplanmıştır. Bulunan korelasyon katsayılarının sıfırdan manidar derecede farklı olup olmadıkları, 0,05 düzeyinde, t-testi ile test edilmiştir. HÜ Sosyal Bilimler Enstitüsündeki dört fakülte ve Devlet Konservatuarı'ndaki yüksek lisans öğrencilerinin LES puan türleri ile akademik ortalamaları arasında manidar bir ilişki bulunmamıştır. Araştırmada ayrıca LES puan türleri ile akademik ortalamalar arasındaki ilişkilere fakülte ve devlet konservatuarındaki programların öğrenci alırken tercih ettikleri LES ağırlıklı puan türleri göz önünde bulundurularak cevap aranmıştır. Bunun sonucunda yalnızca Edebiyat Fakültesine LES-SÖZ puanı ile öğrenci alan programlardaki korelasyonlar, üç LES ağırlıklı puan türü için de 0,05 düzeyinde manidar bulunmuştur.

Kelecioğlu (2004), ÖSS'nin geçerliğini, ÖYS'yi ölçüt kabul ederek incelemiş ve sınav sistemindeki değişikliğin amacına uygun olup olmadığını tespit etmeyi amaçlamıştır. 1998 yılında ÖSS'ye giren adaylardan ÖYS'yi kazanan 38769 adayın ÖSS ve ÖYS puanları kullanılarak ÖSS puanının ÖYS alt test puanlarını ne derece yordadığı basit doğrusal regresyon tekniği ile test edilmiş ve ÖSS'nin sözel ve sayısal bölümleri ile ÖYS'nin bunlarla ilgili testleri arasında önemli ölçüde benzerlik bulunduğu görülmüştür. Sonuç olarak ÖYS'ye olan ihtiyacın tamamen ortadan kalktığına ilişkin yeterli kanıt bulunmadığını belirtmiştir.

Kan (2005), ÖSS'de soru çıkan ortaöğretim alan derslerindeki başarı ile ÖSS'den elde edilen puanlar arasındaki ilişkiyi ve bu ilişkiye dayalı sınıflama ve kararların geçerliğini incelemiştir. Sonuç olarak ÖSS sayısal puanlarını en iyi yordayan değişkenin fizik dersi akademik başarısı olduğu, diskriminant analiz modelinden elde edilen sonuçlar ile gerçek sonuçlar arasındaki doğru sınıflama yüzdesinin

toplamda %78,64; başarısız sınıflamada %81,48; başarılı sınıflama da ise %75,51 olduğu belirlenmiştir.

Olani (2009) çalışmasında birinci sınıf üniversite öğrencilerinin akademik başarılarının yordanması üzerinde durmuştur. Araştırma, Etiyopya'daki bir üniversitede eğitim gören bir grup öğrenci üzerinde yürütülmüştür. Üniversite akademik başarısını açıklamak için yapılan çoklu regresyon analizinde istatistiki olarak anlamlı çıkan tek katsayının lise not ortalaması olduğu bulunmuştur.

Sayın (2010), araştırmasında Gazi Üniversitesi Gazi Eğitim Fakültesi bölümlerine 2007, 2008 ve 2009 yıllarında ÖSS ile yerleştirilen öğrencilerin ÖSS-SÖZ2 ve ÖSSEA2 puanları ile üniversitedeki akademik başarı not ortalamaları (ABNO) arasındaki ilişkileri saptamaya çalışmıştır. Araştırmada ayrıca öğrencilerin ÖSS'nin alt testlerindeki Türkçe ve edebiyat test başarısı ve OBP değişkenleri de incelemeye alınmıştır. Ek olarak ÖSS'nin alt testleri olan Türkçe ve edebiyat testlerindeki soruların ortaöğretim Türk edebiyatı ile dil ve anlatım öğretim programlarıyla ne derece örtüştüğü araştırılmıştır. İlişkisel türden betimsel bir çalışma niteliği taşıyan bu araştırma, 2007, 2008 ve 2009 yıllarında Gazi Eğitim Fakültesinin bölümlerine ÖSS-SÖZ2 ve ÖSS-EA2 puanlarıyla yerleşen toplam 1940 öğrenci ile sınırlandırılmıştır. Ayrıca ortaöğretim kurumlarında görev yapan 75 Türk dili ve edebiyatı öğretmenine anket uygulanmıştır. Araştırmada değişkenler arasındaki ilişkilerin ortaya konması için Pearson Momentler Çarpımı Korelasyon Katsayısı ve soruların konulara göre dağılımında içerik analizi kullanılmıştır. Araştırmanın sonucunda ÖSS Türkçe testlerinde yer alan soruların %93'ünün 9. sınıf dil ve anlatım dersi konularını kapsadığı; edebiyat testinde yer alan soruların ise dört sınıf düzeyini kapsadığı bulunmuştur. Öğrencilerin ÖSS-SÖZ2, ÖSS-EA2 puanları ve ÖSS Türkçe testi doğru sayısı ile OBP arasında negatif yönde bir ilişki olduğu bulunmuştur. ÖSS edebiyat testi doğru sayısı ile OBP arasında istatistiki olarak anlamlı bir ilişki bulunmamıştır. Öğrencilerin ABNO'ları ile ÖSS Türkçe testi doğru sayısı, edebiyat testi doğru sayısı, ÖSS-SÖZ2 puanları arasında anlamlı bir ilişki bulunamamış; 2008 ÖSS-EA2 puanı ile ABNO arasında hesaplanan ilişki katsayısı -0,22 olarak saptanmıştır. ABNO ile OBP arasında pozitif yönde orta düzeyde ilişkiler hesaplanmış ($r=0,409$, $r=0,329$); üniversite başarısını açıklayan tek bağımsız değişkenin OBP olduğu saptanmıştır.

İncelenen araştırmalarda yordayıcı ile ölçüt arasındaki ilişkilere bakılarak yordayıcının ölçütü ne derece yordadığı bulunmaya çalışılmıştır. Çalışmaların birçoğunda ilköğretim akademik başarısının ortaöğretim geçiş sınavlarını yordama gücü veya ortaöğretim akademik başarısının üniversite giriş sınavı ile üniversite başarısını yordama gücüne bakılmıştır. Araştırma sonuçları incelendiğinde birçoğunda akademik ders başarıları ile giriş sınavları arasında doğrusal ilişkilerin olduğu ancak akademik başarının giriş sınavlarındaki başarıyı her zaman çok iyi yordamadığı anlaşılmaktadır. Bu çalışmada ilköğretim ve ortaöğretim akademik başarısının üniversite giriş sınavını yordaması üzerinde durulmuştur.

YÖNTEM

Araştırmanın Yöntemi

Bu araştırmada öğrencilerin YGS alt testleri ham puanları ile ilgili derslere ait ortaöğretim ve ilköğretim başarı puanları arasındaki ilişkilere bakılmıştır. Bu araştırma iki ya da daha fazla değişken arasındaki birlikte değişim varlığını ve derecesini belirlemeye çalışan ilişkiselsel türden betimsel bir çalışma niteliğindedir.

Çalışmanın Evren ve Örneklemi

Araştırmanın hedef evrenini 2012-2013 öğretim yılında Ankara ilinde YGS'ye giren öğrenciler, çalışma evrenini ise Altındağ ilçesinde 2013 yılında üç okulda YGS'ye giren 533 öğrenci oluşturmaktadır. Bu grubun demografik özellikleri Tablo 1'de verilmiştir.

Tablo 1. Çalışma Grubunu Oluşturan Öğrencilerin Cinsiyete ve Okullara Dağılımı

Okulun Adı	Öğrenci Sayısı		Toplam	Yüzde
	Kız	Erkek		
A okulu	107	69	176	33
B okulu	60	48	108	20
C okulu	157	92	248	47
Toplam	324	209	533	100

Çalışmanın verileri Ankara ili Altındağ ilçesinde bulunan üç farklı okuldaki 533 öğrenciden elde edilmiştir. Bu öğrencilerin okullara ve cinsiyete göre dağılımı Tablo 1’de verilmiştir.

Veri Toplama Yöntemleri

Araştırmada belirlenen sorulara yanıt bulmak amacıyla aynı öğrenciye ait ortaokul ve ortaöğretim yıl sonu başarı puanları ile YGS’deki doğru sayıları kullanılmıştır.

Verilerin İşlenmesi ve Çözümlemesi

Araştırma sorularını yanıtlamak üzere çoklu regresyon analizi kullanılmıştır. Baykul’a (2000) göre regresyon, iki değişken arasındaki fonksiyonel ilişkiyi belirten, bunlardan biri yardımıyla diğerinin değerinin kestirilmesine imkân sağlayan tekniktir. Büyüköztürk’e (2011) göre çoklu regresyon analizi, bağımlı değişkenler ile ilişkili olan iki ya da daha çok bağımsız değişkene (yordayıcı değişkenlere) dayalı olarak, bağımlı değişkenin tahmin edilmesine yönelik bir analiz türüdür. Çoklu regresyon analizi, yordayıcı değişkenler tarafından bağımlı değişkende açıklanan toplam varyansın yorumlanmasına, açıklanan varyansın istatistiksel olarak anlamlılığına ve yordayıcı değişkenlerle bağımlı değişken arasındaki ilişkinin yönüne ilişkin yorum yapma olanağı verir. Çoklu regresyonun en önemli avantajı, bağımsız değişkenlerin toplu etkisini göstermesi yanında, diğer değişkenler kontrol altında tutularak, her değişkenin etkisinin de tek tek görülebilmesidir (Akgül, 2005).

Çoklu regresyon analizinde yaygın olarak kullanılan üç yöntem standart, adımsal ve hiyerarşik yaklaşımlardır (Büyüköztürk, 2011). Bu çalışmada adımsal çoklu regresyon yöntemi kullanılmıştır. Adımsal yaklaşımda, her aşamada modele alınan her bir bağımsız değişkenin modelde denklemdaki önemlilik derecesi araştırılır. Modele önce bağımlı değişkenle en yüksek korelasyona sahip yani bağımlı değişkenin varyansına en büyük katkıyı verecek tek bir bağımsız değişken alınır. Daha sonra ikinci bağımsız değişken analize alınır. Bir değişken analize girdiğinde diğer bağımsız değişkenle modele hemen hemen aynı katkıyı sağlıyorsa, bu değişkenin yeni modelde yer almasına gerek kalmayarak modelden çıkarılır. Bu süreç, bağımlı değişkendeki değişkenlikle önemli bir fark görülmeğe kadar devam eder (Büyüköztürk, 2011).

Çoklu regresyon analizinde aşağıdaki kestirim denklemi kullanılmıştır (Pedhazur, 1997).

$$\hat{Y} = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

\hat{Y} : Kestirilen bağımlı değişken y’nin değeri

a : Sabit terim

b_i : Kısmi regresyon katsayısı

x_i : Bağımsız değişken

Çoklu regresyon analizinde, bağımsız değişkenlerin birlikte bağımlı değişkendeki değişimi açıklaması R^2 determinasyon katsayısı ile verilir. R^2 determinasyon katsayısı, bağımsız değişkenlerin tümü tarafından açıklanan bağımlı değişkenin değişim oranıdır. Çoklu regresyon analizinde eğimler,

diğer değişkenler sabit tutulduğunda, o bağımsız değişkendeki birim artışa karşılık bağımlı değişkendeki değişim miktarını gösterir ve kısmi eğim ya da kısmi regresyon katsayısı olarak da isimlendirilir (Mertler ve Vannatta, 2002).

Bu çalışmada araştırmaya başlamadan önce kullanılacak olan veride çoklu bağlantılılık problemi olup olmadığı incelenmiştir. Çoklu regresyon analizinde yordayıcı değişkenler arasında çoklu bağlantılılık olarak tanımlanan bir sorunla karşılaşılabilir. Araştırmada böyle bir sorun olup olmadığının belirlenmesi için tolerans değeri, varyans büyütme faktörü ve durum indeks değerleri incelenmiştir (Büyüköztürk, 2011). Elde edilen değerlerin sonucunda verinin çoklu regresyon analizine uygun bir veri olduğu anlaşılmış ve uygulamaya geçilmiştir. Çoklu regresyon analizinde bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkinin anlamlı olup olmadığı belirlenirken F testi kullanılmıştır. F testi sonucu anlamlı ise denkleme anlamlı katkısı olan bağımsız değişkenlerin belirlenmesi için bu bağımsız değişkenlerin kısmi regresyon katsayılarının sıfırdan farklı olup olmadığı test edilmiştir. Bu katsayıların anlamlılığı t-testi ile belirlenir. Bu araştırmada testlerde, anlamlılık düzeyi 0,05 olarak alınmıştır. Araştırma sonucunda elde edilen veriler yorumlanırken her bir dersle ilgili mesleki kıdemleri on ile on beş yıl arasında değişen ve farklı okullarda görev yapan üç öğretmenin görüşüne başvurulmuştur. Öğretmenlerin seçiminde ilk olarak mesleki kıdem dikkate alınmıştır. En az on yıllık tecrübeye sahip öğretmenler arasından gönüllülük esasına dayalı olarak seçim yapılmıştır. YGS kapsamı baz alınarak ortaokul Türkçe, matematik, fen bilgisi ve sosyal bilgiler öğretmenleri ile liselerde görev yapan Türk dili ve edebiyatı, matematik, fizik, kimya, biyoloji, inkılap tarihi ve coğrafya öğretmenlerinin görüşleri dikkate alınmıştır.

BULGULAR

Araştırmanın bu bölümünde toplanan veriler, amaçlar doğrultusunda analiz edilerek araştırmanın problemine cevap olacak bulgulara yer verilmiştir. Bulguların verilmesinde alt problemler tekrar yazılmıştır.

1. Ortaokul 7-8. sınıflar Türkçe, ortaöğretim dil ve anlatım ile edebiyat dersleri yıl sonu başarı puanları YGS Türkçe alt testi ham puanlarını ne düzeyde yordamaktadır?

Araştırmanın bu sorusuna cevap bulabilmek için öğrencilerin ilköğretim 7. ve 8. Sınıf Türkçe ile ortaöğretim 9.-12. sınıf dil anlatım ve edebiyat derslerindeki başarı puanlarının YGS Türkçe alt testi ham puanlarını yordama gücüne bakılmıştır. Bu amaçla çoklu regresyon analizi yapılmıştır. Çoklu regresyon analizinden önce değişkenler arasındaki korelasyonlar incelenmiştir. Değişkenler arasındaki korelasyonlar Tablo 2’de, VIF ve tolerans değerleri Tablo 3’te, regresyon analizi sonuçları ise Tablo 4’te verilmiştir.

Tablo 2. Ortaokul Türkçe, Ortaöğretim Dil ve Anlatım ile Edebiyat Dersleri Yıl Sonu Başarı Puanları YGS Türkçe Alt Testi Puanları Arasındaki Korelasyonlar

	TUR	7TUR	8TUR	9DIL	9EDB	10DIL	10EDB	11DIL	11EDB	12DIL	12EDB
TUR	1										
7TUR	0,63	1									
8TUR	0,67	0,77	1								
9DIL	0,55	0,60	0,64	1							
9EDB	0,44	0,48	0,52	0,69	1						
10DIL	0,37	0,33	0,41	0,57	0,57	1					
10EDB	0,61	0,59	0,63	0,67	0,66	0,51	1				
11DIL	0,70	0,66	0,71	0,68	0,59	0,52	0,76	1			
11EDB	0,64	0,59	0,65	0,67	0,64	0,54	0,86	0,82	1		
12DIL	0,68	0,67	0,72	0,63	0,51	0,37	0,73	0,78	0,74	1	
12EDB	0,66	0,64	0,67	0,58	0,53	0,36	0,72	0,78	0,76	0,82	1

TUR: Türkçe, DIL: Dil ve Anlatım, EDB: Edebiyat

Tablo 2 incelendiğinde en düşük korelasyonun 10. sınıf dil ve anlatım dersi ile 7. sınıf Türkçe dersi arasında olduğu (0,33), en yüksek korelasyonun ise 10. Sınıf edebiyat dersi ile 11. sınıf edebiyat dersi (0,86) arasında olduğu bulunmuştur.

Regresyon analizi öncesinde modele giren değişkenlerin korelasyon değerleri incelenmiş ve elde edilen bu korelasyon değerlerinden bazılarının 0,80 den büyük olduğu görülmüştür. Çoklu bağlantı olmaması için her bir korelasyon değerinin 0,80'den küçük olması gerektiğinden çoklu bağlantı sorununun olup olmadığını belirlemek için VIF ve tolerans değerlerine bakılmıştır (Büyüköztürk, 2011). Elde edilen sonuçlar Tablo 3'te verilmiştir.

Tablo 3. YGS Türkçe Alt Testini Yordamasına İlişkin Çoklu Bağlantı Kontrolü

Değişkenler	VIF Değerleri	Tolerans Değerleri
11DİL	3,28	0,31
8TUR	3,11	0,32
12DİL	3,84	0,26
7TUR	2,67	0,37
12EDB	3,57	0,28

Tablo 3 incelendiğinde VIF değerlerinin 10'dan düşük olduğu, tolerans değerlerinin 0,20 den yüksek olduğu bulunmuştur. Elde edilen bu korelasyon değerlerine göre çoklu bağlantı olmadığı görülmüştür (Büyüköztürk, 2011). Yapılan çoklu adımsal (stepwise) regresyon analizinde toplam beş model kurulmuştur. Bu modellerden beşincisi alınarak regresyon denklemi kurulmuştur. Bu modele ait regresyon analizi Tablo 4'te verilmiştir.

Tablo 4. YGS Türkçe Alt Testi Regresyon Analizi Sonuçları

Alt test	Beta	Beta std hata	Std beta	t	p	R	R ²	F	p	Kısmi korelasyon
Model 5						0,76	0,58	131,05	0,00	
Sabit	-14,59	1,71		-8,51	0,00					
11 DİL	0,14	0,03	0,23	4,82	0,00					0,21
8 TUR	0,12	0,03	0,18	3,49	0,00					0,16
12 DİL	0,09	0,03	0,15	2,53	0,01					0,11
7 TUR	0,09	0,03	0,14	2,78	0,00					0,13
12 EDB	0,07	0,03	0,14	2,41	0,01					0,11

Tablo 4 incelendiğinde model 5'e ait regresyon sonuçlarına göre, YGS Türkçe alt testini sırasıyla 11. sınıf dil ve anlatım, 8. sınıf Türkçe, 12. sınıf dil ve anlatım, 7. sınıf Türkçe ve 12. sınıf edebiyat değişkenlerinin önemli düzeyde yordadığı anlaşılmaktadır. Diğer değişkenler ise (9. sınıf edebiyat-dil ve anlatım, 10. sınıf edebiyat-dil ve anlatım ve 11. sınıf edebiyat) YGS Türkçe alt testini önemli düzeyde yordamamaktadır.

Yordayan değişkenler arasında en önemli yordayıcı değişkenin 11. sınıf dil ve anlatım yıl sonu başarı puanı değişkeni olduğu anlaşılmaktadır. 11. sınıf dil ve anlatım yıl sonu başarı puanı tek başına YGS Türkçe alt testi başarısının %48,5'ini açıklamaktadır. Bunu sırasıyla 8. sınıf Türkçe, 12. sınıf dil ve anlatım, 7. sınıf Türkçe ve 12. sınıf edebiyat yıl sonu başarı puanları izlemektedir. 8. sınıf Türkçe yıl sonu başarı puanı YGS Türkçe alt testinin açıklanma oranını %5,9 düzeyinde arttırmıştır. Bu beş değişkenin YGS Türkçe alt testinin %58'ini açıkladığı görülmektedir. 2013 YGS Türkçe alt testi sorularının konulara göre dağılımı incelendiğinde; paragrafta yardımcı düşünce (8 soru), paragrafta konu-ana düşünce (5 soru), paragraf anlatım teknikleri (4 soru), cümle yorumu (4 soru), paragrafta yapı (3 soru), karma dil bilgisi (3 soru), sözcük anlamı ve söz yorumu (3 soru), cümle

anlamı (karma) (2 soru), dil bilgisi (2 soru), yazım kuralları (2 soru), cümle anlamı (kavramlar) (1 soru), anlatım bozukluğu (1 soru), ses bilgisi (1 soru) ve noktalama işaretleri (1 soru) ile toplamda kırk sorudan oluşmaktadır

(Kaynak: http://mebk12.meb.gov.tr/meb_iys_dosyalar/44/01/340710/icerikler/2013ygsanalizi_1048368.html adresinden 01.07.2014 tarihinde alınmıştır).

Sonuçlarla ilgili uzman görüşüne başvurulduğunda, 2013 YGS Türkçe alt testinde paragraf sorularının bir önceki yıla oranla azaltıldığı, bu soruların yerine dil ve anlatım sorularının artırıldığı tespiti yapılmıştır. YGS Türkçe alt testi sorularının ağırlıklı olarak 9. ve 10. sınıf edebiyat ile dil ve anlatım derslerinin öğretim programından çıktığı ve içerik olarak 8. ve 7. sınıf Türkçe dil bilgisi konuları ile paralellik gösterdiği bilinmektedir. 12. sınıf düzeyinde öğrenciler YGS ile ilgili genel tekrarlar yapmakta aynı zamanda 12. sınıf dil ve anlatım dersinin konuları da genel tekrar niteliği göstermektedir. Bu yüzden YGS Türkçe alt testini yordamada 8., 7. sınıflar Türkçe ile 11. sınıf dil ve anlatım ve 12. sınıf edebiyat derslerinin yordamasının beklenen bir durum olduğu düşünülmektedir. Yordaması beklenen 9. sınıf edebiyat-dil ve anlatım ile 10. sınıf dil ve anlatım değişkenlerinin YGS Türkçe alt testini yordamama sebebinin eğitim müfredatımızın sarmal yapısı nedeniyle üst sınıf düzeylerinde de konuların tekrar edilmesi ve pekiştirilmiş olması düşünülmektedir. Ayrıca örneklem grubunun küçüklüğü de bir diğer etken olabilir.

Bu regresyon analizi sonucunda elde edilen regresyon denklemi aşağıda verilmiştir.

$$\text{YGS Türkçe} = -14,59 + 0,14 \text{ DIL11} + 0,12 \text{ TUR 8} + 0,09 \text{ DIL 12} + 0,09 \text{ TUR 7} + 0,07 \text{ EDB 12}$$

2. Ortaokul 7-8. sınıflar matematik ve ortaöğretim matematik dersleri yıl sonu başarı puanları YGS matematik alt testi ham puanlarını ne düzeyde yordamaktadır?

Araştırmanın bu sorusuna cevap bulabilmek için öğrencilerin ilköğretim 7. ve 8. sınıf matematik ve ortaöğretim 9.-12. matematik ve geometri derslerindeki başarı puanlarının YGS Matematik alt testi ham puanlarını yordama gücüne bakılmıştır. Bu amaçla çoklu regresyon analizi yapılmıştır. Çoklu regresyon analizinden önce değişkenler arasındaki korelasyonlar incelenmiştir. Değişkenler arasındaki korelasyonlar Tablo 5’te, VIF ve tolerans değerleri Tablo 6’da, regresyon analizi sonuçları ise Tablo 7’de verilmiştir.

Tablo 5. Ortaokul ve Ortaöğretim Matematik Dersleri Başarı Puanları YGS Matematik Alt Testi Puanları Arasındaki Korelasyonlar

	MAT	7MAT	8MAT	9GEO	9MAT	10SMAT	11SMAT	12SMAT	12SGEO
MAT	1								
7MAT	0,65	1							
8MAT	0,73	0,79	1						
9GEO	0,45	0,40	0,48	1					
9MAT	0,46	0,44	0,51	0,75	1				
10SMAT	0,51	0,41	0,48	0,58	0,68	1			
11SMAT	0,71	0,56	0,64	0,59	0,64	0,78	1		
12MAT	0,76	0,61	0,68	0,45	0,48	0,48	0,69	1	
12SGEO	0,60	0,44	0,53	0,29	0,27	0,35	0,49	0,60	1

MAT: Matematik, SMAT: Seçmeli Matematik, GEO: Geometri, SGEO: Seçmeli Geometri

Tablo 5 incelendiğinde en düşük korelasyonun 9. sınıf matematik dersi ile 12. sınıf seçmeli geometri dersi arasında olduğu (0,27), en yüksek korelasyonların ise 7. sınıf matematik dersi ile 8. sınıf matematik dersi arasında (0,79) olduğu bulunmuştur.

Regresyon analizi öncesinde modele giren değişkenlerin korelasyon değerleri incelenmiş ve elde edilen bu korelasyon değerlerinde çoklu bağlantı olmadığı yani her bir korelasyon değerinin 0,80'den küçük olduğu görülmüştür (Büyüköztürk, 2011). Aynı zamanda değişkenler arası çoklu bağlantı sorunu olup olmadığını belirlemek için VIF ve tolerans değerlerine bakılmıştır. Elde edilen sonuçlar Tablo 6'da verilmiştir.

Tablo 6. YGS Matematik Alt Testini Yordamasına İlişkin Çoklu Bağlantı Kontrolü

Değişkenler	VIF Değerleri	Tolerans Değerleri
12mat	2,63	0,38
8mat	3,58	0,28
12geo	2,14	0,47
10mat	1,64	0,61
7mat	2,88	0,35

Tablo 6 incelendiğinde VIF değerlerinin 10'dan düşük olduğu, tolerans değerlerinin ,20 den yüksek olduğu bulunmuştur. Elde edilen bu korelasyon değerlerine göre çoklu bağlantı olmadığı görülmüştür (Büyüköztürk, 2011).

Yapılan çoklu adımsal (stepwise) regresyon analizinde toplam beş model kurulmuştur. Bu modellerden beşincisi alınarak regresyon denklemi kurulmuştur. Bu modele ait regresyon analizi tablo 7'de verilmiştir.

Tablo 7. YGS Matematik Regresyon Analizi Sonuçları

Alt test	Beta	Beta std hata	Std beta	t	p	R	R ²	F	p	Kısmi korelasyon
Model 5						,84	0,71	230,06	0,00	
sabit	-29,08	1,80		-16,14	0,00					
12 mat	0,17	0,02	0,38	7,53	0,00					0,33
8 mat	0,16	0,04	0,21	4,37	0,00					0,20
11smat	0,15	0,02	0,24	6,64	0,00					0,29
12sgeo	0,11	0,02	0,15	4,69	0,00					0,12
7 mat	0,08	0,03	0,10	2,46	0,01					0,06

Tablo 7 incelendiğinde model 5'e ait regresyon analizi sonuçlarına göre, YGS matematik alt testini sırasıyla 12. sınıf matematik, 8. sınıf matematik, 11. sınıf seçmeli geometri, 12. sınıf seçmeli geometri ve 7. sınıf matematik değişkenlerinin önemli düzeyde yordadığı anlaşılmaktadır. Diğer değişkenler ise (9. sınıf matematik, 9. sınıf geometri, 10. sınıf seçmeli geometri ve 10. sınıf seçmeli matematik) YGS matematik alt testini önemli düzeyde yordamamaktadır.

Yordayan değişkenler arasında en önemli yordayıcı değişkenin 12. sınıf matematik yıl sonu başarı puanı değişkeni olduğu anlaşılmaktadır. 12. sınıf matematik yıl sonu başarı puanı tek başına YGS matematik alt testi başarı puanının %57,8'ini açıklamaktadır. Bunu sırasıyla 8. sınıf matematik, 12. sınıf geometri, 10. sınıf matematik ve 7. sınıf matematik yıl sonu başarı puanları izlemektedir. 8. sınıf yıl sonu matematik başarı puanı YGS matematik alt testi başarı puanının açıklanma oranını %8,3 oranında arttırmıştır. Bu beş değişkenin toplamda YGS matematik alt testinin %71'ini açıkladığı görülmektedir.

2013 YGS matematik alt testi sorularının konulara göre dağılımı incelendiğinde; denklem kurma problemleri (10 soru), denklem çözme (3 soru), temel kavramlar (2 soru), sayı basamakları (2 soru), bölen sayıları, OBEB OKEK (2 soru), oran orantı (2 soru), fonksiyonlar (2 soru), ondalıklı sayılar (1 soru), basit eşitsizlikler (1 soru), mutlak değer (1 soru), üslü ifadeler (1 soru), çarpımlara ayırma (1

soru), işlem (1 soru), olasılık (1 soru), mantık (1 soru) ve kombinasyon (1 soru) ile toplam 32 soru cebir ayrıca geometri alanından çokgenler (1 soru), dikdörtgen (1 soru), kare (1 soru), yamuk (1 soru), çemberde uzunluk (1 soru), katı cisimler (1 soru) ve analitik geometri (2 soru) olmak üzere toplamda sekiz soru sorulmuştur. 2013 YGS matematik alt testinde toplamda kırk soru sorulmuştur (MEB, 2004).

2013 YGS matematik alt test sorularının konulara göre dağılımı uzman görüşü alınarak incelendiğinde; 7. ve 8. sınıf matematik dersi konularından temel işlem becerilerine ait 6 soru sorulduğu görülmüştür. 8. sınıf matematik dersi konularından ise üslü ve köklü sayılar konularıyla ilişkili 5 soru, harfli ifadeler konusu ile ilişkili 3 soru ve eşitsizlikler konusu ile ilgili 1 soru sorulmuştur. Dokuzuncu sınıf matematik dersi konularından olan mutlak değer konusundan 1 soru ve fonksiyon ve denklem kurma problemleri konularından 9 soru sorulmuştur. Fonksiyon ve denklem kurma problemleri 9. sınıf konusu gibi gözüксе de öğrenciler fonksiyon konusunu özel tanımlı fonksiyonlar konusu 12. sınıf öğretim programındadır. Ayrıca öğrenciler 12. sınıf matematik dersi konularından olan limit, süreklilik, türev ve integral gibi konular ile fonksiyon konusunu pekiştirirler.

2013 YGS matematik alt testinde 10. sınıf matematik dersi konularından olan faktöriyel, olasılık, permütasyon ve kombinasyon konularından 3 soru sorulmuştur. YGS’de çıkan sorular konu bazında incelendiğinde yaklaşık olarak 9 soru 12. sınıf konuları ile ilgili sorulduğu görülmektedir. Bu nedenle YGS matematiğini en iyi şekilde 12. sınıf matematik başarısının yordadığı söylenebilir.

Ayrıca 8. sınıf konularından üslü sayılar, köklü sayılar, harfli ifadeler ve eşitsizlikler konularından 9 soru sorulmuştur. Bu nedenle 8. sınıf matematik yıl sonu başarı puanı öğrencilerin YGS puanını iyi bir şekilde yordamaktadır. 9. sınıf yıl sonu matematik başarı puanlarının YGS matematik alt testinde anlamlı bir yordayıcı olmamasına ilişkin olarak soruların tamamına yakını 9. sınıf konuları ile ilişkili gibi gözüксе de sadece 1 mutlak değer, 1 kümeler ve 1 EBOB, EKOK sorusu bulunmaktadır. Yani toplamda sadece 3 soru bu sınıf düzeyi ile doğrudan ilişkilidir. Bu nedenle 9. sınıfın yıl sonu matematik başarı puanının YGS matematik alt testini yordama gücü çoklu regresyon analizine göre istatistiksel olarak anlamlı çıkmamıştır.

2013 YGS matematik alt testi incelendiğinde 40 sorudan 8’inin geometri sorusu olduğu, bu sorulardan da dördünün yani geometri sorularının %50’sinin, 12. sınıf müfredatında olduğu; bu yüzden de 12. sınıf seçmeli geometri dersinin YGS matematik başarısını yordayan önemli bir değişken olduğu görülmektedir.

Bu regresyon analizi sonucunda elde edilen regresyon denklemi aşağıda verilmiştir.

$$YGS \text{ Mat} = -29,08 + 0,17 \text{ Mat } 12 + 0,16 \text{ Mat } 8 + 0,15 \text{ Seç. Mat } 11 + 0,11 \text{ Seç. Geo } 11 + 0,08 \text{ Mat } 7$$

3. Ortaokul 7-8. sınıflar fen bilgisi ve ortaöğretim fen grubu dersleri yıl sonu başarı puanları YGS fen alt testi ham puanlarını ne düzeyde yordamaktadır?

Araştırmanın bu sorusuna cevap bulabilmek için öğrencilerin ilköğretim 7. ve 8. sınıf fen bilgisi ve ortaöğretim 9.-12. fizik, kimya ve biyoloji derslerindeki başarı puanlarının YGS fen bilimleri alt testi ham puanlarını yordama gücüne bakılmıştır.

Bu amaçla çoklu regresyon analizi yapılmıştır. Çoklu regresyon analizinden önce değişkenler arasındaki korelasyonlar incelenmiştir. Değişkenler arasındaki korelasyonlar Tablo 8’de, VIF ve tolerans değerleri Tablo 9’da, regresyon analizi sonuçları ise Tablo 10’da verilmiştir.

Tablo 8. Ortaokul Fen Bilgisi ve Ortaöğretim Fen Grubu Dersleri Başarı Puanları YGS Fen Alt Testi Puanları Arasındaki Korelasyonlar

	7	8	9	9	9	10	10	10	11	11	11	12	12	12
FEN	FEN	FEN	BIY	FIZ	KIM	SBIY	SFIZ	SKIM	SBIY	SFIZ	SKIM	SBIY	SFIZ	SKIM
FEN	1													

12 skim	0,17	0,05	0,26	5,586	0,00	0,20
11 skim	0,17	0,04	0,27	2,463	0,00	0,25
12 sbiy	0,13	0,04	0,19	3,642	0,00	0,19
11 sfiz	0,09	0,04	0,14	2,293	0,03	0,13

Tablo 10 incelendiğinde model 4'e ait regresyon analizi sonuçlarına göre, YGS fen bilimleri alt testini 12. sınıf seçmeli kimya, 11. sınıf seçmeli kimya, 12. sınıf seçmeli biyoloji ve 11. sınıf seçmeli fizik değişkenlerinin önemli düzeyde yordadığı anlaşılmaktadır. Diğer değişkenler ise (7. sınıf fen bilgisi, 8. sınıf fen bilgisi, 9. Sınıf biyoloji, 9. sınıf fizik, 9. sınıf kimya, 10. sınıf seçmeli biyoloji, 10. sınıf seçmeli kimya, 11. sınıf seçmeli biyoloji ve 12. sınıf seçmeli fizik) YGS fen bilimleri testiniönemli düzeyde yordamamaktadır.

Yordayan değişkenler arasında en önemli yordayıcı değişkenin 12. sınıf seçmeli kimya değişkeni olduğu anlaşılmaktadır. 12. sınıf seçmeli kimya dersi yıl sonu başarı puanı tek başına YGS fen bilimleri alt testinin %39'unu açıklamaktadır. Bunu sırasıyla 11. sınıf seçmeli kimya, 12. sınıf seçmeli biyoloji ve 11. sınıf seçmeli fizik yıl sonu başarı puanları izlemektedir. 11. sınıf seçmeli kimya yıl sonu başarı puanı YGS fen bilimleri alt testinin açıklanmasını %8,1 oranında arttırmıştır. Bu dört değişkenin YGS fen bilimleri alt testinin %50'sini açıkladığı görülmektedir.

2013 YGS fen bilimleri alt testi sorularının konulara göre dağılımı incelendiğinde; fizik dersi sorularının elektrik (3 soru), ısı ve sıcaklık (2 soru), fiziğin doğası (1 soru), sıvıların kaldırma ilkesi (1 soru), basınç (1 soru), doğrusal hareket (1 soru), dinamik (1 soru), iş ve enerji (1 soru), mıknatıslık (1 soru), optik (1 soru) ve ses (1 soru) ile toplam on dört sorudan oluştuğu, kimya dersi sorularının bileşikler (4 soru), kimyasal değişimler (2 soru), çözünürlük (2 soru), madde ve özellikleri (1 soru), atomun yapısı (1 soru), bağlar (1 soru), periyodik sistem (1 soru) ve hayatımızda kimya (1 soru) ile toplam on üç sorudan oluştuğu, biyoloji dersi sorularının ekoloji (2 soru), hücre (2 soru), bakteri (2 soru), endokrin sistem (1 soru), bitki fizyolojisi (1 soru), hücre bölünmeleri (1 soru), canlıların temel bil. (1 soru), duyu organları (1 soru), kalıtım (1 soru) ve metabolizma (1 soru) ile toplam on üç sorudan oluşmaktadır. YGS fen bilimleri alt testi toplamda kırk sorudan oluşmaktadır (MEB, 2004).

YGS fen bilimleri alt testiyle ilgili uzman görüşüne başvurulduğunda, YGS fen bilimleri alt testindeki 13 kimya sorusundan altısının 12. sınıf kimya dersi müfredatından olduğu görülmüştür. Öğrencilerin konuları yeni görmüş olmaları ve soruların bilgi düzeyinde olması 12. sınıf kimya dersini YGS fen bilimleri alt testinin en yüksek yordayıcısı haline getirmiş olabileceği düşünülmektedir. Ayrıca 12. sınıf kimya dersinin yüksek yordama gücünün bir diğer sebebinin ise 12. sınıftaki öğrencilerin YGS'ye hazırlanırken hem okul derslerine hem de geçmiş dönem derslerine çalışmaları olabilir. YGS fen bilimleri alt testi biyoloji sorularıyla ilgili uzman görüşüne başvurulduğunda ise 2013 YGS fen bilimleri biyoloji testindeki 13 sorudan 5 tanesinin 12. sınıf biyoloji müfredatından olduğu bu yüzden YGS fen bilimleri alt testinin bir diğer önemli yordayıcısı olan 12. sınıf biyoloji dersi için, 12.sınıfta öğrencilerin kimya dersinde olduğu gibi biyoloji dersi için de YGS hazırlığı sırasında hem buldukları dönem hem de geçmiş dönem derslerine çalışmalarının önemli bir etken olduğu düşünülmektedir. Uzman görüşleri çerçevesinde YGS fen bilimleri alt testinde 14 fizik sorusunun olduğu ve soruların konu dağılımına göre incelendiğinde 11. sınıf seçmeli fizik dersi müfredatı ağırlıklı olduğu bu yüzden de yüksek bir yordayıcı olması beklenen bir durumdur.

Sonuç olarak ÖSS sayısal puanlarını en iyi yordayan değişkenin fizik dersi akademik başarısı olduğu sonucuna bulmuştur. Bu çalışmada da YGS fen bilimleri alt testini yordayan değişkenler arasında fizik dersi yer almaktadır. Genel anlamda YGS fen bilimleri alt testi incelendiğinde tüm alt testlerin (fizik, kimya ve biyoloji) yordayıcı değişkenler arasında olması beklenen bir durumdur.

Bu regresyon analizi sonucunda elde edilen regresyon denklemi aşağıda verilmiştir.

$$\text{YGS Fen Bilimleri} = -20,78 + 0,17 \text{ S.Kim 11} + 0,17 \text{ S. Kim 12} + 0,13 \text{ S. Biy 12} + 0,09 \text{ S. Fiz 11}$$

4. Ortaokul 7-8. sınıflar ve ortaöğretim sosyal bilimler dersleri yıl sonu başarı puanları YGS sosyal bilimler alt testi ham puanlarını ne düzeyde yordamaktadır?

Araştırmanın bu sorusuna cevap bulabilmek için öğrencilerin ilköğretim 7. Sınıf sosyal bilgiler, 8. sınıf T.C. inkılap tarihi ve Atatürkçülük ve ortaöğretim 9.-12. coğrafya ve tarih derslerindeki başarı puanlarının YGS sosyal bilimler alt testi ham puanlarını yordama gücüne bakılmıştır.

Bu amaçla çoklu regresyon analizi yapılmıştır. Çoklu regresyon analizinden önce değişkenler arasındaki korelasyonlar incelenmiştir. Değişkenler arasındaki korelasyonlar Tablo 11’de, VIF ve tolerans değerleri Tablo 12’de, regresyon analizi sonuçları ise Tablo 13’te verilmiştir.

Tablo 11. Ortaokul ve Ortaöğretim Sosyal Bilimler Dersleri Başarı Puanları YGS Sosyal Bilimler Alt Testi Puanları Arasındaki Korelasyonlar

	SOS	7SOS	8INK	9COG	9TAR	10COG	10TAR	11İNK	12SCOĞ
SOS	1								
7SOS	0,47	1							
8INK	0,58	0,77	1						
9COG	0,53	0,52	0,68	1					
9TAR	0,57	0,63	0,73	0,77	1				
10COG	0,56	0,63	0,71	0,77	0,77	1			
10TAR	0,57	0,62	0,68	0,70	0,76	0,77	1		
11İNK	0,57	0,66	0,77	0,72	0,78	0,72	0,85	1	
12SCOĞ	0,59	0,66	0,78	0,72	0,75	0,78	0,74	0,73	1

SOS: Sosyal Bilgiler, İNK: TC İnkılap Tarihi ve Atatürkçülük, COG: Coğrafya, TAR: Tarih, SCOĞ: Seçmeli Coğrafya

Tablo 11 incelendiğinde en düşük korelasyonun 7. sınıf sosyal bilgiler dersi ile 9. sınıf coğrafya dersi (0,52) arasında olduğu, en yüksek korelasyonun ise 10. sınıf tarih dersi ile 11. sınıf TC inkılap tarihi ve Atatürkçülük dersi (0,85) arasında olduğu bulunmuştur. Regresyon analizi öncesinde korelasyon değerleri incelenmiş ve elde edilen bu korelasyon değerlerinden yordama değerine giren değişkenlerin bazılarının 0,80’den büyük olduğu görülmüştür. Bu değişkenler arasında çoklu bağlantı sorunu olup olmadığını belirlemek için VIF ve tolerans değerlerine bakılmıştır. Elde edilen sonuçlar Tablo 12’de verilmiştir. Diğer korelasyon değerlerinin her birinin 0,80’den küçük olduğu görülmüştür.

Tablo 12. YGS Sosyal Bilimler Alt Testini Yordamasına İlişkin Çoklu Bağlantı Kontrolü

Değişkenler	VIF Değerleri	Tolerans Değerleri
12scoğ	4,59	0,22
11ink	4,13	0,24
10coğ	4,28	0,23

Tablo 12 incelendiğinde VIF değerlerinin 10’dan düşük olduğu, tolerans değerlerinin ,20 den yüksek olduğu bulunmuştur. Elde edilen bu korelasyon değerlerine göre çoklu bağlantı sorunu olmadığı görülmüştür (Büyükköztürk, 2011)

Tablo 13. YGS Sosyal Bilimler Regresyon Analizi Sonuçları

Alt test	Beta	Beta std hata	Std beta	t	p	R	R ²	F	p	Kısmi korelasyon
Model 3						,696	0,50	52,00	0,00	
sabit	-3,50	2,14		-1,60	0,10					
12scoğ	0,19	0,04	0,42	4,73	0,00					0,34
11ink	0,07	0,03	0,19	2,45	0,00					0,19
10coğ	0,09	0,04	0,17	2,10	0,00					0,16

Tablo 13 incelendiğinde model 3'e ait regresyon analizi sonuçlarına göre, YGS sosyal bilimler alt testini 12. sınıf seçmeli coğrafya, 11. sınıf TC inkılap tarihi ve Atatürkçülük ve 10. sınıf coğrafya değişkenlerinin önemli düzeyde yordadığı anlaşılmaktadır. Diğer değişkenler ise (7. sınıf sosyal bilgiler, 8. sınıf TC inkılap tarihi ve Atatürkçülük, 9. sınıf coğrafya, 9. sınıf tarih ve 10. sınıf tarih) önemli düzeyde yordamamaktadır.

Yordayan değişkenler arasında en önemli yordayıcı değişkenin 12. sınıf seçmeli coğrafya değişkeni olduğu anlaşılmaktadır. 12. sınıf seçmeli coğrafya dersi yıl sonu başarı puanı tek başına YGS sosyal bilimler alt testinin %44,3'ünü açıklamaktadır. Bunu sırasıyla 11. sınıf T.C. inkılap tarihi ve Atatürkçülük ve 10. sınıf coğrafya dersleri yıl sonu başarı puanları izlemektedir. 11. sınıf T.C. inkılap tarihi ve Atatürkçülük dersi yıl sonu başarı puanı YGS sosyal bilimler alt testinin açıklanmasında %2,8'lik bir artışa sebep olmuştur. Bu üç değişken YGS sosyal bilimler alt testinin %50'sini açıkladığı görülmektedir.

2013 YGS sosyal bilimler alt testi sorularının konulara göre dağılımı incelendiğinde; tarih dersi sorularının ilk çağ medeniyetleri (1 soru), İslam öncesi Türk tarihi (1 soru), İslam tarihi (1 soru), Türkiye Selçuklu Devleti (1 soru), yakın çağ'da Avrupa (1 soru), beylikten devlete osmanlı (1300-1453) (1 soru), en uzun yüzyıl (1800-1909) (1 soru), 20. yüzyılda osmanlı devleti (1 soru), kurtuluş savaşına hazırlık dönemi (2 soru), kurtuluş savaşı (1 soru), inkılaplar (1 soru), Atatürk ilkeleri (1 soru), Türkiye'nin dış politikası (1 soru) ve jeopolitik konum (1 soru) ile toplam on beş soru olduğu, coğrafya dersi sorularının coğrafya'nın bölümleri (1 soru), harita bilgisi (1 soru), rüzgarlar (1 soru), Türkiye'nin iklimi (1 soru), iç kuvvetler (1 soru), toprak tipleri (1 soru), dış kuvvetler (1 soru), doğal afetler (1 soru), nüfus (1 soru), ortak payda: bölge (1 soru), ulaşım yolları (1 soru) ve çevre ve insan (1 soru) ile toplam on iki sorudan oluştuğu, felsefe dersi sorularının felsefenin alanı (2 soru), bilgi felsefesi (1 soru), bilim felsefesi (1 soru), varlık felsefesi (1 soru), ahlak felsefesi (1 soru), siyaset felsefesi (1 soru) ve sanat felsefesi (1 soru) ile toplam sekiz soru olduğu, din kültürü ve ahlak bilgisi sorularının kuranı kerim ve akıl (1 soru), hazreti peygamberin görevleri (1 soru), ilahi kutsal dinlerin ortak mesajları (1 soru), musahiplik (yol kardeşliği) (1 soru) ve ayetlere inanma (1 soru) ile beş sorudan oluşmaktadır. YGS sosyal bilimler alt testi toplamda 40 sorudan oluşmaktadır (MEB, 2014).

YGS sosyal bilimler alt testinde çıkan soruların alanlara göre dağılımlarına bakıldığında tarih, coğrafya, felsefe ile din kültürü ve ahlak bilgisi olduğu görülmektedir. Sonuçlarla ilgili uzman görüşlerine başvurulduğunda, felsefe ile din kültürü ve ahlak bilgisi derslerinden gelen soruların bilgi düzeyi belirlemeden ziyade okuduğunu anlama ve yorum yapmaya dayalı olduğu görülmüştür. Yordayan değişkenler arasında öne çıkan en önemli değişkenlerin 12. sınıf seçmeli coğrafya, 11. sınıf T.C. inkılap tarihi ve Atatürkçülük ile 10. sınıf coğrafya derslerinin olduğu bulunmuştur. YGS coğrafya sorularının 9. ve 10. sınıf müfredatından sorulduğu ancak 12. sınıf konularının içerik açısından daha güncel konular bulunmaktadır. Ayrıca öğrenciler son sene YGS hazırlıklarını hızlandırdıkları ve öğrencilerin bilgi birikimindeki artışın etkisi olduğu düşünülmektedir. YGS sosyal bilimler alt testi tarih sorularının 9., 10. ve 11. sınıf müfredatlarından ağırlıklı olarak sorulduğu bu yüzden YGS sosyal bilimler alt testini yordamada 11. sınıf T.C. inkılap tarihi ve Atatürkçülük dersinin olmasının beklenen bir durum olduğu düşünülmektedir.

Bu regresyon analizi sonucunda elde edilen regresyon denklemi aşağıda verilmiştir.

$$\text{YGS Sosyal Bilimler} = -3,5 + 0,19 \text{ S. Coğ } 12 + 0,07 \text{ İnk } 11 + 0,09 \text{ Coğ } 11$$

SONUÇLAR ve TARTIŞMA

Bu araştırmada öğrencilerin YGS alt testi ham puanlarını yordamada 7.-12. sınıflar arasındaki dört temel dersten (Türkçe, Sosyal Bilimler, Matematik, Fen Bilimleri) elde edilen yıl sonu başarı puanlarının etkisi incelenmiştir.

Sonuç olarak;

YGS Türkçe alt testini en iyi yordayan değişkenin 11. sınıf dil ve anlatım yıl sonu başarı puanı değişkeni olduğu anlaşılmaktadır. 11. sınıf dil ve anlatım yıl sonu başarı puanı tek başına YGS Türkçe alt testi başarısının %48,5'ini açıklamaktadır. Bunu sırasıyla 8. sınıf Türkçe, 12. sınıf dil ve anlatım, 7. sınıf Türkçe ve 12. sınıf edebiyat yıl sonu başarı puanları izlemektedir. 8. sınıf Türkçe yıl sonu başarı puanı YGS Türkçe alt testinin açıklanma oranını %5,9 düzeyinde arttırmıştır. Bu beş değişkenin YGS Türkçe alt testinin %58'ini açıkladığı görülmektedir.

YGS Türkçe alt testi sonuçlarıyla ilgili uzman görüşüne başvurulduğunda, 2013 YGS Türkçe alt testinde paragraf sorularının bir önceki yıla oranla azaltıldığı yerine dil ve anlatım sorularının artırıldığı tespiti yapılmıştır. YGS Türkçe alt testi sorularının ağırlıklı olarak 9. ve 10. sınıf edebiyat ile dil ve anlatım derslerinin öğretim programından çıktığı ve içerik olarak 8. ve 7. sınıf Türkçe dilbilgisi konuları ile paralellik gösterdiği bilinmektedir. 12. sınıf düzeyinde öğrenciler YGS ile ilgili genel tekrarlar yapmakta aynı zamanda 12. sınıf dil ve anlatım dersinin konuları da genel tekrar niteliği göstermektedir. Bu yüzden YGS Türkçe alt testini yordamada 8., 7. sınıflar Türkçe ile 11. sınıf dil ve anlatım ve 12. sınıf edebiyat derslerinin yordamasının beklenen bir durum olduğu düşünülmektedir. Yordaması beklenen 9. sınıf edebiyat-dil ve anlatım ile 10. sınıf dil ve anlatım değişkenlerinin YGS Türkçe alt testini yordamama sebebinin eğitim müfredatımızın sarmal yapısı nedeniyle üst sınıf düzeylerinde de konuların tekrar edilmesi ve pekiştirilmiş olması düşünülmektedir. Ayrıca örneklem grubunun küçüklüğü de bir diğer etken olabilir.

YGS matematik alt testini en iyi yordayan değişkenin 12. sınıf matematik yıl sonu başarı puanı değişkeni olduğu anlaşılmaktadır. 12. sınıf matematik yıl sonu başarı puanı tek başına YGS matematik alt testi başarı puanının %57,8'ini açıklamaktadır. Bunu sırasıyla 8. sınıf matematik, 12. sınıf geometri, 10. sınıf matematik ve 7. sınıf matematik yıl sonu başarı puanları izlemektedir. 8. sınıf yıl sonu matematik başarı puanı YGS matematik alt testi başarı puanının açıklanma oranını %8,3 oranında arttırmıştır. Bu beş değişken toplamda YGS matematik alt testinin %71'ini açıkladığı görülmektedir.

2013 YGS matematik alt test sorularının konulara göre dağılımıyla ilgili matematik eğitimi alanındaki üç akademisyenden uzman görüşü alınarak incelenmiştir. Sonuç olarak 7. ve 8. sınıf matematik dersi konularından temel işlem becerilerine ait 6 soru sorulduğu görülmüştür. 8. sınıf matematik dersi konularından ise üslü ve köklü sayılar konularıyla ilişkili 5 soru, harfli ifadeler konusu ile ilişkili 3 soru ve eşitsizlikler konusu ile ilgili 1 soru sorulmuştur. Dokuzuncu sınıf matematik dersi konularından olan mutlak değer konusundan 1 soru ve fonksiyon ve denklem kurma problemleri konularından 9 soru sorulmuştur. Fonksiyon ve denklem kurma problemleri 9. sınıf konusu gibi gözüke de öğrenciler fonksiyon konusunu özel tanımlı fonksiyonlar konusu 12. sınıf öğretim programındadır. Ayrıca öğrenciler 12. sınıf matematik dersi konularından olan limit, süreklilik, türev ve integral gibi konular ile fonksiyon konusunu pekiştirirler.

2013 YGS matematik alt testinde 10. sınıf matematik dersi konularından olan faktöriyel, olasılık, permütasyon ve kombinasyon konularından 3 soru sorulmuştur. YGS'de çıkan sorular konu bazında incelendiğinde yaklaşık olarak 9 soru 12. sınıf konuları ile ilgili sorulmaktadır. Bu nedenle 12. sınıf matematik başarısı YGS matematiğini en iyi şekilde yordamaktadır denilebilir.

Ayrıca 8. sınıf konularından üslü sayılar, köklü sayılar, harfli ifadeler ve eşitsizlikler konularından 9 soru sorulmuştur. Bu nedenle 8. sınıf matematik yıl sonu başarı puanı öğrencilerin YGS puanını iyi bir şekilde yordamaktadır.

9. sınıf yıl sonu matematik başarı puanlarının YGS matematik alt testinde anlamlı bir yordayıcı olmamasına ilişkin olarak, soruların tamamına yakını 9. sınıf konuları ile ilişkili gibi gözüke de sadece 1 mutlak değer, 1 kümeler ve 1 EBOB, EKOK sorusu bulunmaktadır. Yani toplamda sadece 3 soru bu sınıf düzeyi ile doğrudan ilişkilidir. Bu nedenle 9. sınıfın yıl sonu matematik başarı puanının YGS matematik alt testini yordama gücü çoklu regresyon analizine göre istatistiksel olarak anlamlı çıkmamıştır.

2013 YGS matematik alt testi incelendiğinde 40 sorudan 8'inin geometri sorusu olduğu, bu sorulardan da dördünün yani geometri sorularının %50'sinin, 12. sınıf müfredatında olduğu bu

yüzden 12. sınıf seçmeli geometri dersinin YGS matematik başarısını yordayan önemli bir değişken olduğu görülmektedir.

YGS fen bilimleri alt testini en iyi yordayan değişkenin 12. sınıf seçmeli kimya değişkeni olduğu anlaşılmaktadır. 12. sınıf seçmeli kimya dersi yıl sonu başarı puanı tek başına YGS fen bilimleri alt testinin %39'unu açıklamaktadır. Bunu sırasıyla 11. sınıf seçmeli kimya, 12. sınıf seçmeli biyoloji ve 11. sınıf seçmeli fizik yıl sonu başarı puanları izlemektedir. 11. sınıf seçmeli kimya yıl sonu başarı puanı YGS fen bilimleri alt testinin açıklanmasını %8,1 oranında arttırmıştır. Bu dört değişkenin YGS fen bilimleri alt testinin %50'sini açıkladığı görülmektedir.

YGS fen bilimleri alt testiyle ilgili uzman görüşüne başvurulduğunda, YGS fen bilimleri alt testindeki 13 kimya sorusundan altısının 12. sınıf kimya dersi müfredatından olduğu görülmüştür. Öğrencilerin konuları yeni görmüş olmaları ve soruların bilgi düzeyinde olması 12. sınıf kimya dersini YGS fen bilimleri alt testinin en yüksek yordayıcısı haline getirmiş olabileceği düşünülmektedir. Ayrıca 12. sınıf kimya dersinin yüksek yordama gücünün bir diğer sebebinin ise 12. sınıftaki öğrencilerin YGS ye hazırlanırken hem okul derslerine hem de geçmiş dönem derslerine çalışmaları olabilir. YGS fen bilimleri alt testi biyoloji sorularıyla ilgili uzman görüşüne başvurulduğunda ise 2013 YGS fen bilimleri biyoloji testindeki 13 sorudan 5 tanesinin 12. sınıf biyoloji müfredatından olduğu bu yüzden YGS fen bilimleri alt testinin bir diğer önemli yordayıcısı olan 12. sınıf biyoloji dersi için, 12.sınıfta öğrencilerin kimya dersinde olduğu gibi biyoloji dersi için de YGS hazırlığı sırasında hem buldukları dönem hem de geçmiş dönem derslerine çalışmalarının önemli bir etken olduğu düşünülmektedir. Uzman görüşleri çerçevesinde YGS fen bilimleri alt testinde 14 fizik sorusunun olduğu ve soruların konu dağılımına göre incelendiğinde 11. sınıf seçmeli fizik dersi müfredatı ağırlıklı olduğu bu yüzden de yüksek bir yordayıcı olması beklenen bir durumdur. Kan (2005), yapmış olduğu çalışmada ÖSS'de soru çıkan ortaöğretim alan derslerindeki başarı ile ÖSS'den elde edilen puanlar arasındaki ilişkiyi ve bu ilişkiye dayalı sınıflama ve kararların geçerliliğini incelemiştir. Sonuç olarak, ÖSS sayısal puanlarını en iyi yordayan değişkenin fizik dersi akademik başarısı olduğu sonucuna bulmuştur. Bu çalışmada da YGS fen bilimleri alt testini yordayan değişkenler arasında fizik dersi yer almaktadır. Genel anlamda YGS fen bilimleri alt testi incelendiğinde tüm alt testlerin (fizik, kimya ve biyoloji) yordayıcı değişkenler arasında olması beklenen bir durumdur.

YGS sosyal bilimler alt testini en iyi yordayan değişkenin 12. sınıf seçmeli coğrafya değişkeni olduğu anlaşılmaktadır. 12. sınıf seçmeli coğrafya dersi yıl sonu başarı puanı tek başına YGS sosyal bilimler alt testinin %44,3'ünü açıklamaktadır. Bunu sırasıyla 11. sınıf T.C. inkılap tarihi ve Atatürkçülük ve 10. sınıf coğrafya dersleri yıl sonu başarı puanları izlemektedir. 11. sınıf T.C. inkılap tarihi ve Atatürkçülük dersi yıl sonu başarı puanı YGS sosyal bilimler alt testinin açıklanmasında %2,8'lik bir artışa sebep olmuştur. Bu üç değişken YGS sosyal bilimler alt testinin %50'sini açıkladığı görülmektedir.

YGS sosyal bilimler alt testinde çıkan soruların alanlara göre dağılımlarına bakıldığında tarih, coğrafya, felsefe ile din kültürü ve ahlak bilgisi olduğu görülmektedir. Sonuçlarla ilgili uzman görüşlerine başvurulduğunda, felsefe ile din kültürü ve ahlak bilgisi derslerinden gelen soruların bilgi düzeyi belirlemeden ziyade okuduğunu anlama ve yorum yapmaya dayalı olduğu görülmüştür. Yordayan değişkenler arasında öne çıkan en önemli değişkenlerin 12. sınıf seçmeli coğrafya, 11. sınıf T.C. inkılap tarihi ve Atatürkçülük ile 10. sınıf coğrafya derslerinin olduğu bulunmuştur. YGS coğrafya sorularının 9. ve 10. sınıf müfredatından sorulduğu ancak 12. sınıf konularının içerik açısından daha güncel konular bulunmaktadır. Ayrıca öğrenciler son sene YGS hazırlıklarını hızlandırdıkları ve öğrencilerin bilgi birikimlerindeki artışın etkisi olduğu düşünülmektedir. YGS sosyal bilimler alt testi tarih sorularının 9., 10. ve 11. sınıf müfredatlarından ağırlıklı olarak sorulduğu, bu yüzden YGS sosyal bilimler alt testini yordamada 11. sınıf T.C. inkılap tarihi ve Atatürkçülük dersinin olmasının beklenen bir durum olduğu düşünülmektedir.

Genel anlamda sonuçlar değerlendirildiğinde yordaması beklenen bazı sınıf düzeyindeki derslerin istenilen sonuçları vermediği görülmüştür. Yapılan sınavın içeriğinin geniş bir kapsama sahip olması ve kazanım yerine yetenek ölçüyor olmasında diğer etkenler olabilir.

Eğitim politikamız açısından üniversite kariyeri planlayan birçok adayın katıldığı YGS'nin her sınıf düzeyi sonunda yapılması ve aynen TEOG'da olduğu gibi sınavın içeriğindeki ilgili sınıf düzeyindeki konularla ilgili her bir sorunun kazanımlarla ilişkilendirilerek yapılmasının daha faydalı olacağı düşünülmektedir.

Öneriler

Araştırmacılara Dönük Öneriler

1. Bu çalışmada, aynı ilçedeki ve farklı okul türlerindeki üç okulun öğrencilerinin verileri kullanılmıştır. Yeni yapılacak çalışmalarda farklı ilçelerdeki benzer ve farklı okul türleri seçilebilir.
2. Araştırmanın kapsamı genişletilerek aynı öğrenci grubuna ait ilköğretim ve ortaöğretim yıl sonu başarı puanlarının Yükseköğretime Geçiş Sınavı (YGS) ve Lisans Yerleştirme Sınavı (LYS) alt testlerini yordama gücüne bakılabilir.
3. İlköğretim puanlarının YGS'yi yordamadaki etkisine bakılarak TEOG sonuçlarının YGS ve LYS'yi yordamasına yönelik çalışmalar yapılabilir.
4. Elde edilen sonuçlar Ankara ili Altındağ ilçesindeki okullar için bulunmuştur. Türkiye'ye genelleme yapabilmek için daha büyük gruplarla çalışmalar yapılabilir.

Uygulamada Çalışanlara Öneriler

1. YGS alt testlerinde sorulan soruların ağırlıklı olarak 9. ve 10. sınıf müfredatını kapsadığı düşünüldüğünde yapılan sınavın 12. sınıf sonunda değil de temel eğitimden ortaöğretime geçişte olduğu gibi her yılın sonunda yapılmasının daha etkili olacağı düşünülmektedir.
2. Yapılan çalışmada YGS fen bilimleri alt testi dışındaki bütün testlerde ilköğretim başarı puanlarının yordayıcılar arasında olduğu bulunmuştur. Dolayısıyla ortaöğretimden yükseköğretime geçişte hesaplanan puanlara ortaöğretim başarı puanları ile birlikte ilköğretim başarı puanlarının da eklenmesi önerilebilir.

KAYNAKÇA

- Aiken, L. R. (1971). *Psychological Testing and Assessment*. (2. Edition) Boston: Pearson Education Group.
- Akgül A. (2005). *Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri- SPSS Uygulamaları* 3. Baskı. Ankara: Emek Ofset Ltd.Şti.
- Akhun, İ. İ. (1980). Akademik Başarının Kestirilmesi: Çoklu Regresyon Yaklaşımının Uygulanmasına İlişkin Bir Araştırma. Ankara: *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Yayınları*. No: 88.
- Anastasi, A. (1988). *Psychological Testing*. (6. Edition). New York: Macmillan Publishing Company.
- Arıcı, H. (2001). *İstatistik: Yöntemler ve Uygulamalar*. Ankara: Meteksan Basımevi.
- Aşkar, P. (1985). Yükseköğretime Öğrenci Seçme ve Yerleştirme Sisteminin Geçerliliği. Yayımlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Aydın, A., Sarier, Y. ve Uysal, Ş. (2012). Sosyoekonomik ve Sosyokültürel Değişkenler Açısından PISA Matematik Sonuçlarının Karşılaştırılması. *Eğitim ve Bilim*, 37(164), Cilt 37, Sayı 164.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Büyüköztürk, Ş. (2011). *Sosyal Bilimler İçin Veri Analizi El Kitabı*. (11. Baskı). Ankara: Pegem Akademi Yayıncılık. s.100.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. (1. Edition). Orlando: Holt, Rinehart and Winston.
- Doğan, N. (1999). Dersane Deneme Sınavları İle ÖSS Ve ÖYS Arasındaki İlişki. Yayımlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Dünya Bankası Eğitim Raporu. (2005). Türkiye Eğitim Raporu Çalışması, Rapor No: 32450-TU. Web: http://siteresources.worldbank.org/INTTURKEY/Resources/3616161142415001082/ESS_Main_Report_V1.pdf adresinden 05.03.2013 tarihinde alınmıştır.
- Erdem, A. R. (2005). İlköğretimimizin Gelişimi ve Bugün Gelinek Nokta. Pamukkale Üniversitesi Bilim, Eğitim ve Düşünce Dergisi, 5(2), sayfa numarası yok. Web: www.universite-toplum.org/pdf/pdf.php?id=240 adresinden 03.05.2013 tarihinde alınmıştır.

- Erdođdu, Y. (1999). Öğrenci Seçme ve Yerleştirme Sınavında Kullanılan Testlerinin Yordama Geçerliğine İlişkin Bir Araştırma. Yayınlanmamış Doktora Tezi, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Gelbal, S. (1989). Öğrenci Seçme Deneme Sınavı ile Öğrenci Seçme Sınavı Testlerinin Öğrenci Başarıları Yönünden İlişkileri, Güvenirlikleri ve ÖYS'yi Yordama Güçleri. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Gürdal, O. (2000). Yaşam Boyu Öğrenme Etkinliği Enformasyon Okuryazarlığı. *Türk Kütüphaneciliği*, 14(2), 176-187.
- MEB (2014). http://mebk12.meb.gov.tr/meb_iys_dosyalar/44/01/340710/icerikler/2013_ygsanalizi_1048368.html adresinden 01.07.2014 tarihinde alınmıştır.
- Kan, A. (2005). ÖSS'ye Kaynaklık Eden Alan Derslerindeki Başarı ile ÖSS'den Elde Edilen Puanlar Arasındaki İlişkinin İncelenmesi. *Eğitim ve Bilim Dergisi*, 30(137), 38-44.
- Karakaş, D. (1998). Eğitimde Ölçme ve Değerlendirme Anabilim Dalındaki Öğrencilerin Yükseköğretime Giriş Puanları ile Bu Anabilim Dalındaki Başarı Arasındaki İlişkiler. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Karip, E. ve Köksal, K. (1996). Etkili Eğitim Sistemlerinin Geliştirilmesi. *Eğitim Yönetimi Dergisi*, 2(2), 245-247.
- Keleciođlu, H. (2004). Üniversiteye Girişte İki Aşamalı Sınavda Uygulanmış Olan Öğrenci Seçme Sınavı Puanları ile Öğrenci Yerleştirme Sınavı Puanları Arasındaki İlişkiler. *Eğitim ve Bilim Dergisi*, 29(133), 60-70.
- Kutlu, Ö. (1989). Tıpta Uzmanlık Eğitimi Giriş Sınavı (TUS) ile İlgili Bir Araştırma. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Mertler, C. A. ve Vannatta, R. A., (2002). Advanced and Multivariate Statistical Methods Practical Application and Interpretation. Pyrczak Publishing, Los Angeles. Akt: Verim, A. (2006). İlköğretim Düzeyindeki Bazı Başarı Ölçülerinin Ortaöğretim Kurumları Öğrenci Seçme Sınavını Yordama Gücü. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ölçme ve Değerlendirme Bilim Dalı, Ankara.
- Morgan, R. (1990). *Predictive Validity Within Categoriations of Collage Students 1978, 1981 and 1985*. Princeton, Nerw Jersey: Educational Testing Service.
- Myers, R. S. ve Pyles M. R. (1992). Relationships Among High School Grades, ACT Test Scores, and College Grades (ACT ile Lise Not Ortalamasının Üniversite Başarıyla İlişkisi), Paper Presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN.
- Olani, A. (2009). Predicting First Year University Students' Academic Success. *Electronic Journal of Research in Education al Psychology*.
- Onay, P. (1972). Öğrencilerin Lise, Üniversitelerarası Giriş İmtihanı ve Üniversite Başarıları Arasındaki İlişkiler. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Örnek, E. (2002). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü'ne Bağlı Yüksek Lisans Programlarındaki Lisansüstü Eğitimi Giriş Sınavı Puanları ile Akademik Başarıları Arasındaki İlişkiler. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Oral, T. (1985). Lise Başarı Ölçüleri ile ÖSYS Puanları Arasındaki Uyum. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction*, (Third Edition). United Kingdom: Wadsworth Thomson Learning. s.136.
- Saban, A. (2007). *Okul Teknoloji Planlaması ve Koordinasyonu*. Ankara: Pegem Akademi Yayıncılık.
- Sayın, A. (2010). Öğrenci Seçme ve Yerleştirme Sınavının Üniversite Başarı Not Ortalaması ve Ortaöğretim Başarı Puanı ile İlişkisi. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Sönmez, V. (2010). *Program Geliştirmede Öğretmen El Kitabı*. (16. Baskı). Ankara: Anı Yayıncılık.
- Tavşancıl, E. (1989). Lise Tür ve Kolunun Yükseköğretimdeki Akademik Başarıya Etkisi, Yayınlanmamış Doktora Tezi, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Tekin, H. (2007). *Eğitimde Ölçme ve Değerlendirme*. (18. Baskı). Ankara: Yargı Yayınevi.
- Tezbaşaran, A. (1991). Yükseköğretime Öğrenci Seçme ve Yerleştirme Sisteminde 1987 Yılında Yapılan Değişiklikler Üzerine Bir Araştırma, Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghtan Mifflin Company.
- Tross, S. A., Harper, J. P., Osher, L. W. ve Kneidinger, L. M. (2000). "Not Just The Usual Cast of Characteristics: Using Personality to Predict College Performance and Retention." *Journal of College Student Development*.

- Turgut, M. F. ve Baykul, Y. (2012). *Eğitimde Ölçme ve Değerlendirme*. (4. Baskı). Ankara: Pegem Akademi Yayıncılık.
- Yıldırım, C. (1972). Ortadoğu Teknik Üniversitesine Giriş ve Üniversitedeki Başarıyı Etkileyen Faktörler. Yayımlanmamış Doktora Tezi. Ortadoğu Teknik Üniversitesi, Ankara.
- Yıldırım, C. (1999). *Eğitimde Ölçme ve Değerlendirme*. (4. Baskı). Ankara: ÖSYM Yayınları.

EXTENDED ABSTRACT

Introduction

Human beings have been working to explore nature and increase the quality of live for ages. Thanks to this, each generation contributes to this effort by using former generation's accumulation of knowledge. Schools and education applied at those schools have been developed as a way of transiting this accumulation of knowledge systematically. Education is a system which supplies desired behaviors on individuals. (Turgut and Baykul, 2012). Although this aim seems as a sole purpose, it is clear that this phenomena has more dimensions than a single one. Those dimensions confront us not only as to develop individual but also as to develop social, political and economic functions of education. (Sönmez, 2010). In the early and mid-part of 21st century education, having an advanced level of reading and writing skills was accepted as sufficient in terms of education. (Erdem, 2005; Gürdal, 2000). Yet in the later 1900s, individuals who gained only those skills had difficulty in adopting increasing demand of the century and remained incapable of catching the era. (Karip and Köksal, 1996; Saban, 2007). Thus, countries had to change their education policies and philosophies. In order to convey this alteration, education philosophies had to be reorganized.

To choose student for a higher education level, many assessments have done at a national level. For instance, "Transition from Primary to Secondary Education System" - TEOG has been used lastly at transiting from primary education through secondary education. "Transition to Higher Education Examination" - YGS and "Undergraduate Placement Exam" - LYS have been used to transit from secondary education to higher education. One of the validity evidences of competitive examinations is the predictive power of those exams on students' future success. And the other evidence is exam relation level with students' former learnings. (Aiken, 1971). Regression analysis is a way to make prediction/guess about future in the light of information on hand. In terms of statistical way, regression is a function to make a guess about prudential prediction on unknown situations by using statistical techniques and benefitting from known. (Arıcı, 2001).

In this research, students' middle and high school success scores, were considered as YGS (The Transition to Higher Education Examination) score predictors. Defining how these variables predict YGS scores forms the fundamental problem of the research. 2013-2014 academic year raw scores of subtests at YGS are accepted as standard in this research.

Research Question

At which level, do year-end success scores of related courses at primary and secondary education predict raw scores of subtests at YGS (the transition to higher education examination)?

Sub Questions

In accordance with this purpose; subproblems below are questioned:

1. At which level, do middle school, 7th and 8th grades' Turkish and high school Language and Expression and Literature courses' year-end success scores predict raw scores of Turkish sub tests at YGS?

2. At which level, do middle school, 7th and 8th grades' math and high school math courses' year-end success scores predict raw scores of math sub tests at YGS?
3. At which level, do middle school, 7th and 8th grades' science and high school science group courses' year-end success scores predict raw scores of Science sub tests at YGS?
4. At which level, do middle school, 7th and 8th grades' and high school social sciences' year-end success scores predict raw scores of social sciences sub tests at YGS?

Method

In this research, relation between students' YGS sub tests raw scores and relevant courses' secondary and high school grade point averages were examined. This research has characteristics of correlational descriptive study that try to determine shared variance existence and degree between one or two variances.

Aimed population of the research were students who took YGS exam in 2013-2014 academic years in Ankara. Target population of the survey were from 533 students who took YGS exam in 2013 year in Altındağ district.

In order to get correct responses, same students' middle and high education year-end grade point averages and their YGS correct scores were used.

Multiple regression analysis was used to answer the research questions. According to Baykul (2000), regression defines the relationship between two variables and provides a technique to predict the values of one variable with the help of the values of the other variable. According to Büyüköztürk (2011), multiple regression analysis is a kind of analysis used to predict dependent variable from one or more independent variables (predictor variables). Multiple regression analysis allows predicting total variance explained by predictor variable(s), commenting on significance of explained variance statistically, and making interpretation on the direction of relation between the dependent variable and the predictor variable(s). The most important advantage of multiple regression is that the independent variables express their collective effect. In addition, each variable's effect can be seen one by one by keeping other variables under control. (Akgül, 2005).

Three methods used in multiple regression analysis are standard, stepwise and hierarchical approaches. (Büyüköztürk, 2011). Stepwise multiple regression approach was used in this research. In stepwise approach, significance degree of each independent variable at model, taken to model, at each step is investigated. At first, a single independent variable which has the highest correlation with dependent variable, that is, gives the biggest support to dependent variable's variance is taken. Afterwards, the second independent variable is taken into the analysis. Provided that a variable supplies pretty much the same contribution to the model with other independent variable when analyzed, this variable does not need to be taken part in the new model and it is excluded. This process continues till significant difference is seen at dependent variable (Büyüköztürk, 2011).

Results and Discussion

In this research, the effect of students' year-end grade point averages obtained from four basic courses - Turkish, social sciences, mathematics and physical sciences- from 7th through 12th grades were investigated in order to predict students' YGS subtests raw scores.

As a result;

It was understood that the best variable that predicts YGS Turkish subtest is 11th Grade's Language and Expression year-end grade point average. 11th grade's Language and Expression year-end grade point average explains about the 48,5% of variance in YGS Turkish subtest scores. It is followed by 8th grade's Turkish, 12th grade's Language and Expression, 7th grade's Turkish and 12th grade's

Literature year-end grade point averages in turn. 8th grade's Turkish year-end grade point averages increase explained variance of YGS Turkish subtest scores by the 5,9%. It is seen that those five variables explain about the 58% of variance in YGS Turkish subtest scores.

It was understood that the best variable that predicts YGS mathematics subtest is 12th grade's mathematics year-end grade point average. 12th grade's mathematics year-end grade point average explains about the 57,8% of variance in YGS Turkish subtest scores separately. It is followed by 8th grade's mathematics, 12th grade's geometry, 10th grade's mathematics and 7th grade's mathematics year-end grade point averages in turn. 8th grade's mathematics year-end grade point averages increase explained variance of YGS Turkish subtest scores by the 8,3%. It is seen that those five variables explain the 7% of YGS mathematics subtest scores.

It was understood that the best variable that predicts YGS physical sciences subtest is 12th grade's elective chemistry year-end grade point average. 12th grade's elective chemistry year-end grade point average explains the 39% of YGS physical sciences subtest scores separately. It is followed by 11th grade's elective chemistry, 12th grade's elective biology and 11th grade's elective physics year-end grade point averages in turn. 11th grade's elective chemistry year-end grade point averages increase variance explanation level of YGS Turkish subtest scores by the 8,1%. It is seen that those four variables explain the 50% of YGS physical sciences subtest scores.

It was understood that the the best variable that predicts YGS social sciences subtest is 12th grade's elective geography year-end grade point average. 12th grade's elective geography year-end grade point average explains about the 44,3% of YGS social sciences subtest scores separately. It is followed by 11th grade's republic of Turkey revolution history and kemalism and 10th grade's geography year-end grade point averages in turn. 11th grade's republic of Turkey revolution history and kemalism year-end grade point averages increase variance explanation level of YGS Turkish subtest scores by 2,8%.

When considered that asked questions at YGS subtests cover 9-10th Grades' curriculums predominantly, it's thought that applying that exam at each year instead of applying at the end of 12th Grade as done at transiting from primary education to secondary education will be more effective.

It's found in the research that primary school grade point averages are in between precursors at all tests except YGS Physical Sciences subtests. Thus, it's suggestible that secondary school grade point averages and along with primary school grade point averages are able to be added to calculated points at transiting from secondary school to higher education as well.

Sosyal Bilgiler Odaklı Başarı Yönelimleri Ölçeği (SOBYÖ): Geçerlik ve Güvenirlik Çalışması*

Social Studies Oriented Achievement Goal Scale (SOAGS): Validity and Reliability Study

Melehat GEZER**

İbrahim Fevzi ŞAHİN***

Öz

Bu araştırmanın amacı, öğrencilerin sosyal bilgiler başarı yönelimlerini geçerli ve güvenilir olarak ölçmeye olanak tanıyacak bir ölçme aracı geliştirmektir. Araştırma, 2014-2015 eğitim-öğretim yılı güz döneminde Diyarbakır ili merkez ilçelerinde öğrenim gören toplam 374 ortaokul öğrencisinden oluşan bir çalışma grubu üzerinde gerçekleştirilmiştir. Ölçeğin kapsam ve görünüş geçerliği için uzman görüşüne başvurulmuş, yapı geçerliği için Açıklayıcı Faktör Analizi (AFA) ve Doğrulamalı Faktör Analizi (DFA) uygulanmıştır. AFA sonucunda, toplam varyansın %50.82'sini açıklayan, 29 madde ve altı faktörden oluşan bir yapı elde edilmiştir. Bu doğrultuda birinci faktör içsel yaklaşma (İY), ikinci faktör mutlak yaklaşma (MY), üçüncü faktör bağıl yaklaşma (BY), dördüncü faktör mutlak kaçınma (MK), beşinci faktör bağıl kaçınma (BK) ve altıncı faktör içsel kaçınma (İK) olarak isimlendirilmiştir. DFA'dan elde edilen bulgular ve hesaplanan uyum indeksleri, Sosyal Bilgiler Odaklı Başarı Yönelimleri Ölçeği'ne ilişkin 29 madde ve altı faktörlü yapının doğrulandığını göstermiştir. Ölçeğin güvenirliliği iç tutarlılık yöntemiyle incelenmiş ve hesaplanan güvenirlilik katsayılarının kabul edilebilir sınırlar içerisinde yer aldığı saptanmıştır. Madde korelasyonu ile %27'lik alt-üst grup karşılaştırmaları sonucunda ölçekten çıkarılması gereken bir madde olmadığı kararına varılmıştır. Araştırmadan elde edilen sonuçlar, ölçeğin öğrencilerin sosyal bilgiler başarı yönelimlerini ölçmek amacıyla kullanılabilir geçerli ve güvenilir bir ölçme aracı olduğunu göstermektedir.

Anahtar Kelimeler: Başarı yönelimleri, hedef yönelimi, 3x2 başarı yönelimi modeli, sosyal bilgiler odaklı başarı yönelimleri, sosyal bilgiler odaklı başarı yönelimleri ölçeği.

Abstract

This study aims to develop a valid and reliable instrument for measuring students' social studies achievement goal. The research was conducted on a study group consisted of 374 middle school students studying in the central district of Diyarbakır in 2014-2015 school year fall semester. Expert opinion was consulted with regard to the scale's content and face validity. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were performed in order to measure the scale's construct validity. As a result of EFA, a 29-item and a six-factor structure model which explains 50.82% of the total variance was obtained. The emerging factors were called as a self-approach, task-approach, other-approach, task-avoidance, other-avoidance and self-avoidance respectively. The findings acquired CFA indicated that the 29-item and six-factor structure related to social studies oriented achievement goal scale have acceptable goodness of fit indices. The scale's reliability coefficients were calculated by means of internal consistency method. As a result of reliability analysis, it was determined that the reliability coefficients were within admissible limits. The finding of the item correlation and 27% of upper and lower group comparisons demonstrated that all of the items in the scale should remain. In light of these results, it could be argued that the scale is reliable and valid instrument and can be used in order to test students' social studies achievement goals.

* Bu çalışma, birinci yazarın doktora tezinden üretilmiştir.

**Arş. Gör. Dr. Melehat GEZER, Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, İlköğretim Bölümü, Diyarbakır/Türkiye, melehatgezer@gmail.com

***Prof. Dr. İbrahim Fevzi ŞAHİN, Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi, İlköğretim Bölümü, Erzurum/Türkiye, ifevzi@atauni.edu.tr

Keywords: Achievement goal, goal theory, 3x2 achievement goal model, social studies oriented achievement goal, social studies oriented achievement goal scale.

GİRİŞ

Motivasyon, insanların eylemlerinin öngörülebilir sonuçlarına dair beklentileri ile bu eylemleri gerçekleştirmeye dair öz yeterlikleri tarafından harekete geçirilen ve devamı sağlanan hedef güdümlü bir davranıştır (Bandura, 1977). Bir başka deyişle motivasyon; bir eylemi yapıp yapmama kararını etkileyen neden, nitelik ve durum olarak tanımlanabilir. Bireyi harekete geçiren motivasyon, bir ihtiyacın giderilmesi düşüncesinden doğmaktadır. Örneğin öğrenmeye yönelik duyulan ihtiyaç, bireyin motivasyon düzeyini etkilemektedir.

Son yıllarda yapılan araştırmalarda öğrenme ile ilişkili olarak motivasyon ve motivasyonel süreçlere çok fazla vurgu yapıldığı görülmektedir (Pintrich, 2000). Başarı motivasyonu teorileri bireyin algısının ve inançlarının, davranışlar üzerindeki etkilerini ön plana çıkartmıştır. Bu durum, motivasyon odağının içsel ihtiyaçlardan ve çevresel faktörlerden kişinin öznel dünyasına kaymasına neden olmuştur. Bununla birlikte başarı davranışı üzerinde öğrencilerin başarıya dair beklentilerini ve yapılacak göreve dahil olmakla ilgili algılanan değer duygusunun da etkili olduğu ortaya konmuştur. Bu modellere ayrıca hedefler ve yeteneklere dair algılar gibi bilişsel değişkenler de eklenmiştir. Kısacası motivasyon, bireyi bir davranışa yönlendiren şeyin ne olduğu, bireyin neyi amaçladığı/hedeflediği hakkında bilgi sunmaktadır. Amaçlar (hedefler), önemli motivasyon mekanizmalarından birini oluşturmaktadır (Bandura, 1977, 1986, 1991; Schunk ve Ertmer, 2000; Zimmerman, 2000). Hedef ve hedefe ulaşma konusunda gösterilen performans arasındaki algı, değişim için itici bir rol oynamaktadır. Hedef belirleme, sonuçlara ilişkin beklentiler ile birlikte işlev görmektedir. Örneğin insanlar amaçlarına ulaşmalarına yardım edeceğine inandıkları şekillerde davranırlar. İnsanlar hedefleri doğrultusunda çalıştıkça bu konudaki gelişimlerini fark eder ve motivasyonlarını sürdürürler.

Hedef Yönelimi Teorisi, daha önce ileri sürülen diğer teorilerin önemli gördüğü birçok değişkeni içermektedir. Bunun yanı sıra hedef yönelimi teorisi, motivasyon kavramını farklı bir bakış açısıyla ele almaktadır (Pintrich ve Schunk, 2002). Bu teori; hedefler, beklentiler, beceri kavramları, motivasyonla ilgili yönelimler, sosyal ve öz karşılaştırmalar ile başarı davranışları arasında önemli ilişkiler olduğunu varsaymaktadır (Ames, 1992a, 1992b; Anderman ve Wolters, 2006). Eğitim ve gelişim psikologları hedef yönelimi teorisini öğrencilerin başarı davranışlarını açıklamak ve öngörmek için geliştirmişlerdir. Bu teorideki temel yapı, başarı etkinliklerinde bireyin amacı ve odaklanmasına atıfta bulunan hedef yönelimi üzerine oturtulmuştur. Hedef yönelimi teorisi, daha çok davranışı teşvik eden ve yönlendiren hedeflerin nasıl saptandığıyla ve bunların özellikleriyle (örneğin; özgün olmak, zorluk ve yakınlık) ilgilenmektedir. Hedef yönelimi teorisi farklı hedef türlerinin başarıyı nasıl etkilediğine odaklanmaktadır (Anderman ve Wolters, 2006; Elliot, 2005; Pintrich, 2003; Pintrich ve Schunk, 2002). Bir diğer anlatımla hedef yönelimi; öğrencilerin, öğrenme görevlerine katılmalarının altında yatan gerekçelerle ilgilenmektedir (Anderman, Austin ve Johnson, 2002). Araştırmacılar tarafından hedef yönelimi kapsamında farklı yönelimler tanımlanmıştır (Elliot ve McGregor, 2001; Elliot ve Thrash, 2001). Örneğin Başarı yönelimleri; Carole Ames (1984), Carol Dweck (1986), Marty Maehr (Maehr ve Nicholls, 1980) ve John Nicholls'un (1984) çalışmaları sonucunda ileri sürülen yeni bir teoridir (Ames, 1992b; Elliot, 1997). Bu teori sosyal-bilişsel kurama dayalı bir bakış açısıyla öğrencilerin kendi performanslarını nasıl değerlendirdikleri üzerinde durmaktadır (Ames, 1992b). Başarı yönelimleri, bir yandan öğrencilerin başarıya ulaşmaya yönelik inançlarını ve performanslarını değerlendirmede esas aldıkları ölçütleri incelerken (Pintrich ve Schunck, 2002), bir yandan da öğrencilerin kendi yeterliğini nasıl yorumladığını, başarı ya da başarısızlık karşısındaki tepkilerini açıklamaktadır (Zweig ve Webster, 2004).

Başarı yönelimleri, öğrenme ve performans olmak üzere iki boyutlu bir yapıdan oluşmaktadır (Dweck ve Leggett, 1988). Öğrenme yönelimi, öğrencinin öğrenme birimini tam olarak anlaması ve öğrendiklerini geliştirme konusundaki istekliliğini ifade eder (Ames, 1992b). Öğrenme yönelimli öğrencilerin; bilişsel çaba gerektiren derin düşünme stratejilerini kullanma (Pintrich ve Garcia, 1991), derse karşı olumlu tutumlara sahip olma, başarının çabaya bağlı olduğuna inanma (Tuominen Soini, Aro ve Niemivirta, 2008), yeterlik düzeyini kendi normlarına göre değerlendirme (Jagacinski

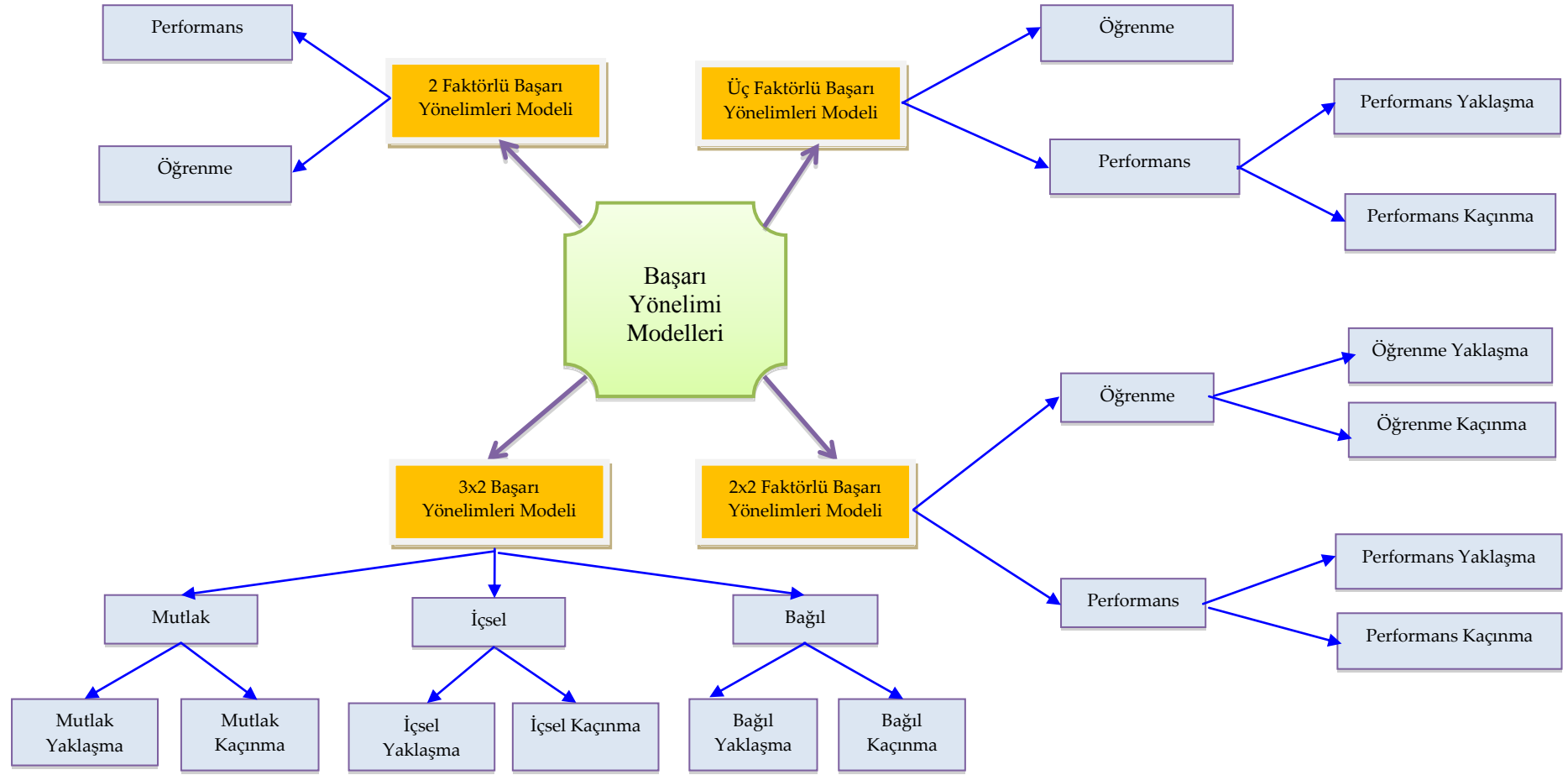
ve Nicholls, 1987; Jagacinski ve Strickland, 2000) ve göreve uygun etkinliklere katılmak için motive olma gibi davranışlara sahip oldukları belirlenmiştir (Schunk, 1989). Nitekim yapılan araştırmalarda da öğrenme yöneliminin birçok olumlu değişkenle pozitif yönde ilişkili olduğu tespit edilmiştir (Pajares ve Cheong, 2003). Bu değişkenler arasında öğrencinin yeterlilik algısı, bilişsel stratejileri kullanma becerisi, başarının çabaya bağlı olduğuna inanması ve güç bir problem durumu ile karşılaştığında gayretli olması sayılabilir (Nicholls, Patashnick ve Nolen, 1985). Performans yönelimi ise öğrencinin yeteneğe odaklanması (Ames ve Archer, 1988), sosyal karşılaştırmaya önem vermesi (Jagacinski ve Strickland, 2000), kendisini akranlarından daha başarılı, zeki ve yetenekli göstermek istemesi gibi özellikleri yansıtmaktadır (Ames, 1992b; Schunk ve Swartz, 1993). Performans yönelimli öğrencilerin olumsuz değerlendirilmelerden ve yeteneksiz görünmekten kaçınma gibi özelliklere sahip oldukları saptanmıştır (Ames, 1992b; Meece, Blumfeld and Hoyle, 1988; Nichols, Jones ve Hancock, 2003). Performans yönelimli öğrenciler için sosyal karşılaştırma bilgisi çok önemlidir. Çünkü bu öğrenciler başarıyı yorumlamak için sosyal normları kullanırlar (Jagacinski ve Strickland, 2000). Ancak sosyal karşılaştırmalarda zorluk çeken öğrencilerde düşük yetenek algısı oluşabileceğinden öğrencinin görev motivasyonu bu durumdan olumsuz etkilenabilir (Schunk, 1991) ve öğrenci başarısızlık yaşamaya başladığında performansında da düşüş gerçekleşebilir (Meece ve Holt, 1993).

Öğrenme ve performans boyutundan oluşan *iki faktörlü başarı yönelimleri* teorisi, akademik dünyada oldukça geniş bir kabul görmesine rağmen araştırmacıların bir kısmı; performans yöneliminin tümüyle olumsuz olmadığını, bazı durumlarda işlevsel olabileceğini ve öğrencileri olumlu davranışlara yöneltebileceğini savunmuşlardır (Elliot ve Harackiewicz 1996; Midgley, Kaplan ve Middleton, 2001). İki faktörlü başarı yönelimi modeli ile ilgili araştırmaların bir kısmında performans yönelimi ile benlik kavramı, akademik öz-yeterlik, akademik çalışmaya verilen değer ile çaba arasında pozitif ilişki bulunurken (Elliot ve Church, 1997; Pintrich ve Garcia, 1991); bazılarında ilişki olmadığı veya negatif ilişki olduğu saptanmıştır (Midgley vd., 2001). Middleton ve Midgley (1997), bu tutarsızlığın performans yöneliminin dikkate alınmayan bir unsurundan kaynaklanabileceğini ortaya atmıştır. Buna bağlı olarak birçok araştırmacı (Elliot ve Harackiewicz, 1996; Midgley, vd., 2001), performans yöneliminin iki alt boyuta ayrılabilceğini ileri süren bir model geliştirmiştir. Geliştirilen bu modelde, öğrenme boyutu aynı kalırken performans boyutu yaklaşma ve kaçınma olarak iki boyuta ayrılmıştır (Elliot ve Harackiewicz, 1996; Kaplan ve Midgley, 1999; Skaalvik, 1997). Performans yaklaşma ve performans kaçınmanın eklenmesiyle *üç faktörlü başarı yönelimleri modeli* elde edilmiştir. Bu modele göre performans yaklaşma yönelimini benimseyen öğrenciler; takdir edilmek, iyi bir öğrenci profili ortaya koyabilmek ve önemli biri olduğunu hissetmek amacıyla çalışırlar. Performans kaçınma yönelimli öğrenciler ise yeteneksiz görünmeme ve utanç duyulacak duruma düşmemek için çalışırlar (Elliot, 1999).

Üç faktörlü başarı yönelimleri modelinin ileri sürülmesinden sonra birçok araştırmacı, öğrenme yöneliminin de öğrenme yaklaşma ve öğrenme kaçınma şeklinde alt boyutlara ayrılabilceğini savunmuştur (Elliot, 1999; Elliot ve Covington, 2001; Pintrich, Conley ve Kempner, 2003). Bu ayırım, yeterlik kavramı üzerine inşa edilmiştir. Bu modele göre yeterlik, değer ve tanım olmak üzere iki bileşenli bir yapı şeklinde düzenlenmiştir. Değer bileşeni; pozitif yaklaşma ve negatif kaçınma yönelimleri olarak ifade edilmiştir. Tanım bileşeni; performans ve öğrenme olarak iki boyutlu bir yapıdan oluşmaktadır. Performans boyutu, yine üç faktörlü modelde olduğu gibi, performans yaklaşma ve performans kaçınma olarak yer almıştır. Öğrenme boyutunda ise mutlak ve içsel temelli hedefler tek bir başlık altında (öğrenme) düşünülmüştür. Çünkü mutlak ve içsel temelli hedefler, kişinin başkalarının performansından etkilenmeyen değerlendirme ölçütlerine sahip olması açısından benzerlik göstermektedir. Mutlak ve içsel hedefler, hem kaynaşık bir yapıda olması hem de diğer bireylerin performansından etkilenmeyen değerlendirme ölçütlerini içermeleri açısından bir arada ele alınmıştır (Elliot ve McGregor, 2001).

Öğrenme yaklaşma yönelimli öğrenciler; dersleri tam olarak öğrenmeyi ve öğrendikleri konularda uzmanlaşmayı amaç edinirler (Elliot ve McGregor, 2001). Öğrenme kaçınma yönelimli öğrenciler ise dersleri tam anlamıyla öğrenememekten korkarlar. Bu öğrenciler; öğrendiklerini unutmamak, konuları yanlış anlamamak ve hata yapmamak için çalışırlar (Akın, 2006; Elliot ve McGregor, 2001).

2x2 başarı yönelimleri modelinde, mutlak ve içsel temelli hedefler bir arada alınmıştır. Elliot, Murayama ve Pekrun (2011) ise mutlak ve içsel hedeflerin ayrı ayrı ele alınabileceğini belirtmiştir. Mutlak ve içsel hedeflerin ayrı ayrı ele alınabileceğine yönelik bu düşünceden hareketle *2x2 başarı yönelimleri modeli* revize edilmiştir. Yapılan revizyonla, *3x2 başarı yönelimleri* olarak adlandırılan yeni bir modele ulaşılmıştır. Bu modele göre günlük yaşamda mutlak ve içsel temelli öğrenme hedeflerinin birbirinden bağımsız olarak ele alınabileceğinin birçok örneği vardır. Örneğin çapraz bulmaca çözen iki bireyden biri daha önceki bulmaca çözüme deneyimlerini göz önünde bulundurmadan bulmacadaki bütün kelimelerin yerini bulmak için çaba gösterirken diğeri, bulmacadaki bütün kelimelerin yerini bulmaya odaklanmayıp bugünkü bulmacada çünkü bulmacaya göre daha fazla kelime bulmak için çaba sarf edebilir. Buna göre mutlak ve içsel temelli hedef düzenleme tekniklerinin farklılık gösterdiği söylenebilir (Elliot, vd., 2011). Bağlı temelli hedeflerde ise yeterlik ölçütü olarak, bireylerarası değerlendirmeler kullanılır. Bir başka deyişle bağlı hedeflerde bireyin diğerlerine göre başarılı veya başarısız olması yeterliğin ölçütünü oluşturur. Görüldüğü gibi *2x2 başarı yönelimlerinden* farklı olarak *3x2 başarı yönelimleri modelinde* yeterliğin tanım bileşeni mutlak, içsel ve bağlı olmak üzere üç boyutlu bir yapı olarak kavramsallaştırılmıştır. Ancak yeterliğin değer bileşenine ilişkin kavramsallaştırmada bir değişiklik yapılmıştır. Yani, tıpkı *2x2 başarı yönelimleri modelinde* olduğu gibi, *3x2 başarı yönelimleri modelinde* yeterliğin değer bileşeni yaklaşma ve kaçınma şeklinde iki boyutlu bir yapıya sahiptir. Anlaşılacağı üzere başarı yönelimleri teorisinin ortaya atılmasından bu yana konuyla ilgili dört farklı model ileri sürülmüştür. İleri sürülen bu modeller Şekil 1’de özetlenmiştir.



Şekil 1. Başarı Yönelimleri Modelleri (Elliot, Murayama ve Pekrun, 2011)

Araştırmanın Amacı

Başarı yönelimlerinin öğretim bağlamı üzerindeki etkisinin tespit edilebilmesinde konu alanının esas alınarak incelenmesinin daha faydalı olacağı yönünde görüşler belirtilmiştir (Anderman ve Anderman, 2000). Bu durum başarı yönelimleri teorisinin; matematik (Anderman ve Midgley, 1996; Middleton, Kaplan ve Midgley, 2004; Keys, Conley, Duncan ve Domina, 2012; Wolters, Yu ve Pintrich, 1996), fen (Pajares, Britner ve Valiante, 2000; Wolters, vd., 1996), İngilizce (Anderman ve Midgley, 1996; Wolters, vd., 1996) ve psikoloji (Harackiewicz, Barron, Karter, Lahte ve Elliot, 1997) gibi derslerde alan odaklı ele alınmasına kaynaklık etmiştir. Söz konusu çalışmalar, başarı yönelimlerinin uluslararası literatürde konu alanı esas alınarak incelendiğine örnek teşkil etmektedir. Buna karşın Türkçe literatürde başarı yönelimlerini yalnızca matematik konu alanı kapsamında ele alan bir araştırmaya (İlhan ve Çetin, 2014) rastlanmıştır. Ancak bu çalışmada başarı yönelimi 2x2 başarı yönelimi modeli ile ele alınmıştır. Bu araştırmada ise 3x2 başarı yönelimi modelinin sosyal bilgiler odaklı olarak incelenmesi amaçlanmıştır. Bu çalışmada 3x2 başarı yönelimleri modelinin ilk defa kullanılmasının, modelin Türk kültürüne uyum sağlayıp sağlamayacağını göstermesi açısından da önemli olduğu düşünülmektedir. Bununla birlikte bu araştırmada 3x2 başarı yönelimi modelinin sosyal bilgiler odaklı incelenmesinin Türkçe literatürdeki boşluğu dolduracak olması bakımından literatüre önemli katkısının olacağına inanılmaktadır.

YÖNTEM

Bu bölümde araştırma desenine, çalışma grubuna, veri toplama aracına ve ölçeğin geliştirilme sürecine ilişkin bilgiler sunulmuştur.

Çalışma Grubu

Araştırma, 2014-2015 eğitim-öğretim yılı güz döneminde Diyarbakır ili merkez ilçelerinde bulunan dört ortaokulda yürütülmüştür. Bu dört ortaokuldan altıncı, yedinci ve sekizinci sınıfı temsilen birer sınıf örneklem olarak seçilmiştir. Sosyal Bilgiler Odaklı Başarı Yönelimleri Ölçeği (SOBYÖ) toplamda 374 öğrencinin cevapları ile geliştirilmiştir. Araştırmaya katılan öğrencilerin okul, sınıf ve cinsiyete göre dağılımı Tablo 1’de gösterilmiştir.

Tablo 1. Ölçek Geliştirme Çalışmasının Yürütüldüğü Araştırma Grubuna Ait Demografik Bilgiler

Araştırma Grubu	Okul	Sınıf			Kadın		Erkek		Toplam
		6.sınıf	7.sınıf	8.sınıf	n	%	n	%	
SOBYÖ	1.Okul	26	30	35	176	47	198	53	374
	2.Okul	29	31	32					
	3.Okul	30	27	30					
	4.Okul	33	34	37					

Ölçek geliştirme çalışmaları kapsamında faktör analizinin yapılabilmesi için öncelikle örneklem büyüklüğünün dikkate alınması önerilmektedir. Örneklem büyüklüğü için araştırmacılar tarafından çeşitli öneriler sunulmuştur. Bu önerilerden bazıları şu şekilde sıralanabilir: Crowley ve Lee (1992) faktör analizi için 100 katılımcıyı yetersiz, 200’ü ortalama, 300’ü iyi, 500’ü çok iyi ve 1000 katılımcıyı mükemmel olarak belirtmiştir (Akt: Akbulut, 2010). Kline (2011) ise faktör analizinde 200 kişilik bir örneklem yeterli olarak kabul edilebileceğini ancak bu sayının faktör yapısının açık ve boyut sayısının az olması durumunda 100’e kadar indirilebileceğini ifade etmiştir. Bu ölçütler göz önünde bulundurulduğunda Tablo 1’deki ölçek geliştirme çalışmasının yürütüldüğü örneklem sayısının faktör analizinin yapılabilmesi için yeterli olduğu ifade edilebilir.

İşlem

SOBYÖ geliştirilirken Elliot, vd. (2011) tarafından geliştirilen 3x2 başarı yönelimleri modeli referans alınmıştır. Ölçeğin geliştirilmesi sürecinde, 3x2 modeli tanım bileşeninde yer alan görev temelli yeterliklerini kazanılmasına odaklanan *mutlak yaklaşma* yönelimi, görev temelli yeterliklerden kaçınmaya odaklanan *mutlak kaçınma* yönelimi, içsel temelli hedeflerin kazanımına ilişkin (öncekinden daha iyisini yapma) *içsel yaklaşma* yönelimi, içsel temelli hedeflerden kaçınmaya odaklanan (öncekinden daha kötüsünü yapmaktan kaçınma) *içsel kaçınma* yönelimi, diğer temelli yeterliklerin kazanımına yönelik (diğerlerinden daha iyisini yapma) *bağıl yaklaşma* yönelimi ve diğer temelli yetersizliklere odaklanan (diğerlerinden daha kötüsünü yapmaktan kaçınma) *bağıl kaçınma* yönelimi olmak üzere üç boyut temel alınmıştır. Literatürün taranmasının ardından, literatür doğrultusunda belirlenen yapı/yapıları tam olarak yansıtabilecek bir madde havuzu oluşturulması aşamasına geçilmiştir. Ölçek maddelerinin hazırlanması aşamasında söz konusu boyutlardan hareketle SOBYÖ maddelerinin yazımında, literatürdeki başarı yönelimleri ölçeklerinden (Akın, 2006; Akın ve Çetin, 2007; Elliot, vd., 2011; İlhan ve Çetin, 2014) yararlanılmıştır. Bu kapsamda ilk olarak Elliot vd., (2011) tarafından geliştirilen ölçekteki bazı maddeler kültürel ve dilsel farklılıklardan dolayı kısmen değiştirilerek Türkçeye çevrilmiş ve SOBYÖ'nün madde havuzuna dâhil edilmiştir. Örneğin Elliot vd., (2011) tarafından geliştirilen ölçekteki "To out perform other students on the exams in this class." maddesi, Türkçeye çevrilirken kısmen değiştirilmiş ve "Sosyal bilgiler dersinde sınıf arkadaşlarımdan daha iyi performans göstermeye çalışırım." biçiminde ifade edilmiştir. Ölçek maddelerinin hazırlanması aşamasında bir ölçme değerlendirme uzmanı ve iki eğitim programları ve öğretimi alanında çalışan uzman olmak üzere toplam üç uzmandan görüş alınmıştır. Uzman görüşleri, taslak ölçme formundaki maddelerin ifade edilmiş şekilleri ile ilgili herhangi bir değişiklik önerisinin olmadığını göstermiştir. Uzmanlardan alınan görüşler ve ölçek geliştirme sürecinde temel alınan boyutlardan yola çıkılarak madde havuzu oluşturulmuştur. Ayrıca uzmanlardan alınan görüşler doğrultusunda "Sosyal bilgiler dersine çalışmanın temel nedenlerinden biri sınıf arkadaşlarımdan başarılı bir öğrenci olduğumu düşünmesini sağlamaktır.", "Sosyal bilgiler dersinde başarılı olmamı sağlayacak fırsatlara önem veririm.", "Öğrendiğim sosyal bilgiler dersi konularını zamanla unutmaya düşüncesi beni endişelendirir.", "Sosyal bilgiler dersinde öğrenmem gereken konuların tümünü öğrenemeyeceğimden korkarım." maddeleri ölçme aracına eklenmiş ve madde havuzu ölçek geliştirme sürecinde temele alınan mutlak yaklaşma (MY), mutlak kaçınma (MK), içsel yaklaşma (İY), içsel kaçınma (İK), bağıl yaklaşma (BY) ve bağıl kaçınma (BK) boyutlarından yola çıkılarak oluşturulmuştur. MY boyutunu oluşturan 5 madde, MK boyutunu oluşturan 5 madde, İY boyutunu oluşturan 7 madde, İK boyutunu oluşturan 6 madde, BY boyutunu oluşturan 7 madde ve BK boyutunu oluşturan 7 madde olmak üzere toplam 37 maddelik bir taslak forma ulaşılmıştır. Ölçek maddeleri için beşli Likert tipi derecelendirme (Kesinlikle Katılıyorum 5, Katılıyorum 4, Kararsızım 3, Katılmıyorum 2 ve Kesinlikle Katılmıyorum 1) kullanılmıştır. Bunun ardından elde edilen madde havuzundaki maddelerden hangilerinin ölçülmesi amaçlanan yapıyı daha doğru ve iyi ölçtüğünün saptanması için iki ölçme değerlendirme uzmanı ve dört sosyal bilgiler eğitimi uzmanı olmak üzere toplam altı uzmanın görüşüne başvurulmuştur. Kapsam ve görünüş geçerliğinin sağlanması için uzmanlardan, taslak ölçek formunu, ölçekteki ifadelerin yazılmasında esas alınan mutlak, içsel ve diğer yaklaşma/kaçınma boyutları göz önünde bulundurarak değerlendirmeleri istenmiştir. Değerlendirme işleminde uzmanlardan ölçekte "kalmalı", "faydalı ama ölçekten çıkarılabilir" ve "ölçekten çıkarılmalı" şeklinde üç kategorili bir puanlama ölçeğini kullanmaları talep edilmiştir. Hangi maddelerin ölçme aracında yer alması gerektiğine uzmanlardan en az dördünün ölçekte kalmalı şeklinde görüş bildirmiş olması kuralına bağlı kalınarak karar verilmiştir. Ölçme aracının taslak formunda bulunan 37 maddeye ilişkin ölçekte kalmalı şeklinde görüş beyan eden uzman sayısının beşten çok olduğu tespit edilmiştir. Uzman görüşlerinin alınmasının ardından, ölçme aracının dilsel anlaşılabilirliğinin sağlanması için ölçme aracı üç Türk dili uzmanına sunulmuştur. Türk dili uzmanlarının ölçme aracını yazım kuralları ve noktalama işaretlerine göre incelemesinden sonra taslak ölçek maddeleri bu doğrultuda yeniden gözden geçirilmiştir. Ardından SOBYÖ'de bulunan maddelerin anlaşılabilirliğinin test edilmesi ve uygulama sürecine ilişkin bilgi edinmek amacıyla, 13 ortaokul öğrencisiyle (6 kız ve 7 erkek) ön uygulama yapılmıştır. Uygulama sonrası ölçek maddelerinin anlaşılabilirlik düzeyi hakkında söz konusu 13 öğrenciyle görüşülmüştür. Görüşme esnasında

öğrencilere ölçme aracının uygulama yönergesi hakkındaki düşünceleri sorulmuştur. Alınan öğrenci görüşleri doğrultusunda ölçme aracının uygulama yönergesi ile ölçek maddelerinde herhangi bir değişiklik yapılmasına ihtiyaç olmadığına karar verilmiştir.

Verilerin Toplanması

Uygulamaya geçilmeden önce ilgili kurumlardan gerekli izinler alınmış ve ölçme aracı öğrencilere sınıfta uygulanmıştır. Ölçeğin uygulama sürecine geçmeden öğrencilere çalışmanın amacı konusunda bilgi verilmiştir. Öğrencilere, toplanan verilerin sadece araştırma amacı doğrultusunda kullanılacağı, bunun dışında herhangi bir amaca hizmet edecek şekilde kullanılmasına izin verilmeyeceği ifade edilmiştir. Ayrıca öğrencilere ölçek uygulamasına katılımın gönüllülük ilkesine göre yürütüleceği hatırlatılmış, ölçek maddelerini nasıl kodlayacakları anlatılmış, ölçek maddelerinde doğru ya da yanlış cevap gibi bir şeyin olmadığı belirtilmiş ve maddeleri okuyup uygun buldukları seçeneği işaretlemeleri ifade edilmiştir. Yine öğrencilerin boş madde bırakmamaları, her madde için yalnızca bir seçeneği işaretlemeleri ve birbirlerine bakmadan ölçeği doldurmaları konularında öğrencilere uyarılarda bulunulmuştur. Veri toplama işleminin tamamlanmasından sonra ölçme aracının psikometrik özelliklerinin tespit edilmesine ilişkin istatistiksel analizlerin yapılmasına geçilmiştir.

Verilerin Analizi

Ölçeğin ön uygulamasının gerçekleştirildiği katılımcılardan toplanan veriler asıl uygulama dışında bırakılmıştır. Çalışma grubuna ölçme aracının uygulanmasının ardından ölçeğin psikometrik özelliklerinin ortaya konulması için istatistiksel analizler yapılmıştır. İlk olarak AFA ve DFA uygulanıp ölçeğin yapı geçerliği kontrol edilmiştir. Ölçeğin güvenilirliği iç tutarlılık yöntemiyle incelenmiştir. Ölçekteki maddelerin ayırt ediciliklerini saptamak için düzeltilmiş madde toplam korelasyonuna bakılmıştır.

BULGULAR

Açımlayıcı Faktör Analizi (AFA):

Çalışmada SOBYÖ'nün yapı geçerliğini ortaya koymak üzere ilk olarak AFA uygulanmıştır. Bu kapsamda yapılan ilk işlem verilerin faktörlenebilirliğinin test edilmesidir. Verilerin faktör analizine uygun olabilmesi için Kaiser-Mayer-Olkin (KMO) değerinin .60'dan yüksek ve Barlett testinin anlamlı olması gerekir. Bu çalışmada, KMO değeri .80 bulunmuş ve Barlett testinin istatistiksel olarak anlamlı olduğu ($\chi^2=2353.489$, $sd=406$) görülmüştür. Bu işlem sonrasında AFA'da temel bileşenler faktörleştirme tekniği ve direct oblimin döndürme tekniği kullanılmıştır. AFA sonuçları Tablo 3'te verilmiştir.

Tablo 3. SOBYÖ AFA İlk Analiz Bulguları

Madde No	MK	BK	BY	İK	İY	MY
SOBYÖ1	.289	-.002	-0.70	-.049	-.663	.249
SOBYÖ2	-.095	.082	.185	.064	-.092	.625
SOBYÖ3	.185	-.196	.126	.168	-.659	.218
SOBYÖ4	.238	-0.26	.015	-.007	-.707	.076
SOBYÖ5	.251	-.057	-.028	.186	-.577	.035
SOBYÖ6	.176	.088	.102	-.021	-.485	.238
SOBYÖ7	.366	-.154	-.076	.128	-.576	.268
SOBYÖ8	.148	-.055	-.041	.319	-.187	.314
SOBYÖ9	.469	.125	-.049	-.018	-.535	.038
SOBYÖ10	.460	-.038	-.050	.202	-.568	.164
SOBYÖ11	.609	.045	.055	.101	-.237	.115
SOBYÖ12	.714	.098	.021	.078	-.335	.143
SOBYÖ13	.676	.090	.158	.128	-.225	.093
SOBYÖ14	.070	.075	.633	-.010	-.122	.074
SOBYÖ15	.085	.063	.601	.031	-.029	.188
SOBYÖ16	-.033	-.048	.266	.033	.523	.179
SOBYÖ17	.282	.158	.642	.110	-.014	-.080
SOBYÖ18	-.001	.168	.721	-.081	.051	-.068
SOBYÖ19	-.146	.210	.670	.037	.152	.111
SOBYÖ20	.438	-.036	-.066	.049	-.339	.519
SOBYÖ21	.284	.296	.035	.186	-.056	.599
SOBYÖ22	.371	.388	.072	.267	-.122	.342
SOBYÖ23	.095	.288	.287	-.203	-.104	.138
SOBYÖ24	.444	.171	.089	.238	-.409	.259
SOBYÖ25	.466	.467	.183	-.179	-.211	.139
SOBYÖ26	-.042	.695	.181	.206	.075	.264
SOBYÖ27	.037	.731	.125	.123	-.007	.038
SOBYÖ28	.176	.703	.255	.219	-.025	.076
SOBYÖ29	.171	.610	.353	.212	.059	.095
SOBYÖ30	-.280	.282	.532	.110	.174	.265
SOBYÖ31	.110	.341	.641	.173	-.032	.075
SOBYÖ32	.268	-.004	.094	.572	-.035	.239
SOBYÖ33	.158	.402	.131	.628	-.126	.270
SOBYÖ34	.103	.424	.176	.611	-.044	.160
SOBYÖ35	-.153	.590	.166	.468	-.047	.011
SOBYÖ36	.186	.372	.216	.044	-.485	.000
SOBYÖ37	.010	.317	.071	.663	-.087	-.109

AFA’da ulaşılan bulgular yorumlanırken maddenin yer alması beklenen boyutta .30’un üzerinde bir faktör yüküne sahip olması esas alınmıştır (Büyüköztürk, 2010). İlk yapılan AFA sonuçları, 2, 8, 13, 16, 23 ve 36 numaralı maddelerin bu ölçütü karşılamadığını göstermiştir. Bu sebeple söz konusu altı madde ölçek dışında bırakılarak analiz tekrarlanmıştır. Tekrarlanan AFA sonucunda, 30 ve 31 numaralı maddelerin kuramsal olarak örtüşmediği faktörlere kaydığı ve yer alması öngörülen boyutta kabul edilebilir bir faktör yüküne (.30) sahip olmadığı tespit edilmiştir. Buna bağlı olarak 30 ve 31 numaralı maddeler de ölçme aracından çıkarılmış ve bu işlem sonrasında öngörülen kuramsal çerçeve ile uyumlu altı faktörlü bir yapıya ulaşılmıştır. Toplam varyansın %50.82’sini açıklayan bu yapıda ortaya çıkan altı faktör sırasıyla; içsel yaklaşma (İY), mutlak yaklaşma (MY), bağıl yaklaşma (BY), mutlak kaçınma (MK), bağıl kaçınma (BK) ve içsel kaçınma (İK) şeklinde isimlendirilmiştir. Bu yapıda maddelerin bir kısmının (m6, m7, m8, m21, m25, m26), birden fazla boyutta faktör yüküne sahip olduğu belirlenmiştir. Maddenin hangi boyutta yer alacağına karar verilmesinde ilgili boyut ile kuramsal açıdan örtüşmesi esas alınmıştır. Bu boyutların her biri için açıklanan varyans oranı ile maddelerin faktör yükleri Tablo 4’te gösterilmiştir.

Tablo 4. SOBYÖ'den Çıkarılan Maddelerin Ardından Tekrarlanan AFA Sonuçları

Madde No	MK	BK	BY	İK	İY	MY
SOBYÖ1	.098	.107	-.091	-.064	.718	.237
SOBYÖ2	.202	-.131	.145	.054	.718	.078
SOBYÖ3	.066	.020	.033	-.038	.711	.313
SOBYÖ4	.145	-.056	-.038	.098	.594	.185
SOBYÖ5	-.118	.127	.076	.065	.477	.435
SOBYÖ6	.535	.013	-.032	-.089	.635	.096
SOBYÖ7	.204	.244	-0.15	-.105	.539	.430
SOBYÖ8	.293	.034	-.019	.064	.595	.373
SOBYÖ9	.254	.047	.109	.036	.239	.745
SOBYÖ10	.205	.165	.049	.031	.354	.774
SOBYÖ11	-.075	.065	.653	.052	.108	.158
SOBYÖ12	.116	.069	.662	.033	-.004	.205
SOBYÖ13	.067	.262	.661	.056	.099	.081
SOBYÖ14	-.099	.214	.723	-.031	-.012	-.063
SOBYÖ15	.060	.087	.669	.188	-.216	-.157
SOBYÖ16	.665	-.019	-.052	-.044	.306	.284
SOBYÖ17	.552	.308	.033	.210	.059	.291
SOBYÖ18	.342	.342	.051	.371	.109	.353
SOBYÖ19	.518	.129	.120	.155	.351	.386
SOBYÖ20	.325	.537	.186	-.157	.188	.363
SOBYÖ21	.057	.604	.105	.445	-.156	.130
SOBYÖ22	-.040	.728	.053	.295	-.034	.072
SOBYÖ23	.072	.785	.228	.286	.028	.120
SOBYÖ24	.124	.689	.325	.264	-.035	.016
SOBYÖ25	.447	-.013	.073	.441	.116	.029
SOBYÖ26	.435	.351	.080	.649	.111	.048
SOBYÖ27	.207	.370	.107	.676	.063	.044
SOBYÖ28	-.063	.409	.119	.695	-.078	.098
SOBYÖ29	-.006	.178	.051	.732	.036	.101
Açıklanan Varyans %	17.05	12.99	7.58	4.96	4.57	3.65

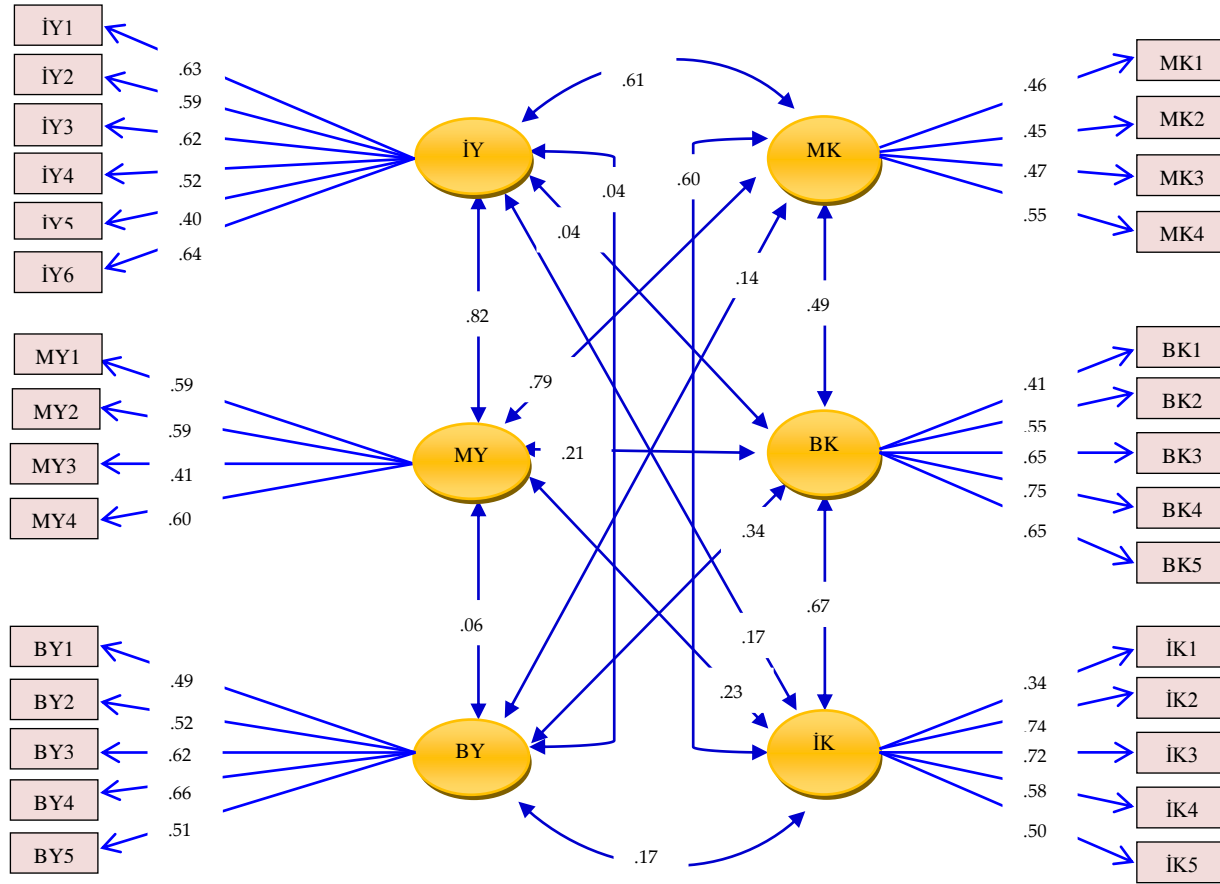
Doğrulayıcı Faktör Analizi

SOBYÖ'nün yapı geçerliğinin test edilmesi kapsamında uygulanan ikinci işlem AFA'da ulaşılan 29 madde ve altı faktörlü yapı üzerinden DFA'nın uygulanması olmuştur. DFA'da test edilen modelin yeterliliğini ortaya koymak için incelenen uyum indeksleri, bu indekslere ait mükemmel ve kabul edilebilir ölçütler ile birlikte Tablo 5'te verilmiştir. Uyum indeksleri için Tablo 5'te verilen ölçütler, DFA'da test edilen altı faktörlü modelin uyumlu olduğunu yansıtmaktadır.

Tablo 5. SOBYÖ için Uyum İndekslerine İlişkin Mükemmel ve Kabul Edilebilir Uyum Ölçütleri ile DFA'dan Elde Edilen Uyum İndeksi Değerleri

İncelenen Uyum İndeksleri	Mükemmel Uyum Ölçütleri	Kabul Edilebilir Uyum Ölçütleri	DFA'dan Elde Edilen Uyum İndeksleri	Sonuç
¹ χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 5$	1.53	Mükemmel Uyum
² CFI	$.95 \leq CFI \leq 1.00$	$.90 \leq CFI \leq .95$.91	Kabul Edilebilir Uyum
² NNFI	$.95 \leq TLI \leq 1.00$	$.90 \leq TLI \leq .95$.90	Kabul Edilebilir Uyum
² IFI	$.95 \leq IFI \leq 1.00$	$.90 \leq IFI \leq .95$.91	Kabul Edilebilir Uyum
³ RMSEA	$.00 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$.58	Kabul Edilebilir Uyum
³ SRMR	$.00 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$.68	Kabul Edilebilir Uyum
⁴ PNFI	$.95 \leq PNFI \leq 1.00$	$.50 \leq PNFI \leq .95$.76	Kabul Edilebilir Uyum
⁴ PGFI	$.95 \leq PGFI \leq 1.00$	$.50 \leq PGFI \leq .95$.72	Kabul Edilebilir Uyum

¹(Kline, 2011), ²(Bentler, 1980; Bentler ve Bonett, 1980; Marsh, Hau, Artelt, Baumert ve Peschar, 2006), ³(Hu ve Bentler, 1999), ⁴(Meyers, Gamst ve Guarino, 2006)



Şekil 2. SOBYÖ'ye İlişkin Ölçüm Modeli

Şekil 2’de görüldüğü gibi altı boyutlu modele ilişkin faktör yükleri; İY alt ölçeği için .40 ile .64 arasında, MY alt boyutu için .41 ile .60 arasında, BY alt boyutu için .49 ile .66 arasında, MK alt boyutu için .45 ile .55 arasında, BK alt boyutu için .41 ile .75 arasında ve İK alt boyutu için .34 ile .74 arasında değişmektedir. DFA sonucu elde edilen altı faktörlü modele ilişkin t-testi değerleri ise Tablo 6’da sunulmuştur. t-değeri için 1.96’dan büyük değerler .05 düzeyinde; 2.58’i üzerindeki değerler ise .01 düzeyinde anlamlıdır (Kline, 2011). Buna göre Tablo 6’daki t-değerlerinin tamamı .01 düzeyinde anlamlı olduğu anlaşılmaktadır. Bu bulgu Byrne (2010), tarafından da ifade edildiği üzere ölçüm modelinden çıkarılması gereken herhangi bir madde olmadığını yansıtmaktadır.

Tablo 6. SOBYÖ için DFA’dan Elde Edilen t-testi Değerleri

Madde No	t-değeri	Madde No	t-değeri	Madde No	t-değeri	Madde No	t-değeri
İY1	12.12*	MY3	7.28*	MK2	7.77*	İK1	6.12*
İY2	11.28*	MY4	11.16*	MK3	8.22*	İK2	14.82*
İY3	11.88*	BY1	8.44*	MK4	9.75*	İK3	14.32*
İY4	9.65*	BY2	9.12*	BK1	7.37*	İK4	11.05*
İY5	7.20*	BY3	11.02*	BK2	10.26*	İK5	9.23*
İY6	12.42*	BY4	11.73*	BK3	12.54*		
MY1	10.82*	BY5	8.90*	BK4	14.95*		
MY2	10.93*	MK1	8.04*	BK5	12.53*		

* $p < .01$

Güvenirlilik ve Madde Analizi

SOBYÖ’nün güvenirliğini hesaplamak için Cronbach Alpha iç tutarlılık katsayısından yararlanılmıştır. Ölçekteki maddelerin ayırt ediciliklerini saptamak amacıyla düzeltilmiş madde toplam korelasyonu kullanılmıştır. Madde analizi kapsamında düzeltilmiş madde toplam korelasyonu ile gerçekleştirilen seçim işleminin güvenirliği hakkında bilgi vermesi açısından %27’lik alt-üst grup karşılaştırmalarına başvurulmuştur. Güvenirlik katsayısı için dikkate alınması önerilen ölçüt .70 olmakla birlikte (Tezbaşaran, 1997) madde sayısının az olması durumunda .50 ve üzerindeki değerler kabul edilebilir olarak yorumlanmaktadır (Nunnally, 1978). Madde ayırt ediciliği için ise .30’un üzerindeki madde toplam korelasyonu değerlerinin ölçüt olarak alınabileceği ifade edilmektedir (Erkuş, 2012). Son olarak seçilen maddelerin güvenirliğinin tespiti için gerçekleştirilen %27’lik alt-üst grup karşılaştırmalarında hesaplanan t-değerlerinin anlamlı olması gerekmektedir. Araştırmada her bir boyut için hesaplanan güvenirlik katsayıları, ölçek maddelerine ilişkin madde toplam korelasyonları ile birlikte Tablo 7’de sunulmuştur. Tablo 7’ye bakıldığında ölçekteki maddelerin tümünün yeterli ayırt edicilikte olduğu görülmektedir. İK alt ölçeğinde yer alan bir nolu maddeye ait madde toplam korelasyonu değeri .30’un altında olmasına rağmen %27’lik alt-üst grup karşılaştırmaları sonucunda rapor edilen t-değeri bu maddenin ölçekte kalması şeklinde alınacak bir kararın güvenilir olacağı anlamına gelmektedir. Bu sebeple söz konusu maddenin de ölçekte kalması uygun görülmüştür. Sonuç olarak gerek hesaplanan güvenirlik katsayıları gerekse de madde analizi sonuçları dikkate alınması önerilen ölçütleri karşılamaktadır.

Tablo 7. SOBYÖ Madde Analizi Sonuçları

Alt Boyut	Madde No	Madde Çıkarıldığında Ölçek Alfası	Düzeltilmiş Madde Toplam Korelasyonu	Gruplar	Ortalama	Standart Sapma	t	
İçsel Yaklaşma N=374 Cronbach Alpha=.73	İY1	.666	.546	Üst _{%27} Alt _{%27}	4.94 3.50	.242 .991	13.05**	
	İY2	.678	.507	Üst _{%27} Alt _{%27}	4.95 3.48	.218 1.065	12.44**	
				Üst _{%27} Alt _{%27}	4.90 2.97	.300 .958		
	İY3	.655	.566	Üst _{%27} Alt _{%27}	4.94 3.87	.242 1.027	9.35**	
				Üst _{%27} Alt _{%27}	4.74 3.04	.443 1.304		
	t testi için sd=166 **p<.001	İY5	.745	.324	Üst _{%27} Alt _{%27}	4.95 3.60	.218 .898	13.48**
İY6		.687	.482	Üst _{%27} Alt _{%27}	5.00 3.93	.000 .941		
Mutlak Yaklaşma N=374 Cronbach Alpha=.63	MY1	.577	.386	Üst _{%27} Alt _{%27}	5.00 3.93	.000 .941	11.44**	
	MY2	.566	.403	Üst _{%27} Alt _{%27}	5.00 4.08	.000 .904	10.35*	
				Üst _{%27} Alt _{%27}	5.00 3.53	.000 1.046		
	t testi için sd=230 **p<.001	MY3	.602	.362	Üst _{%27} Alt _{%27}	5.00 3.76	.000 1.073	14.22**
		MY4	.486	.500	Üst _{%27} Alt _{%27}	5.00 3.76	.000 1.073	
	Bağıl Yaklaşma N=374 Cronbach Alpha=.71	BY1	.673	.446	Üst _{%27} Alt _{%27}	4.53 2.15	.747 1.342	15.30**
BY2		.668	.457	Üst _{%27} Alt _{%27}	4.53 2.00	.944 1.231	16.04**	
				Üst _{%27} Alt _{%27}	4.78 2.41	.508 1.429		
t testi için sd=192 **p<.001		BY3	.658	.485	Üst _{%27} Alt _{%27}	4.40 1.76	.858 .970	20.24**
		BY4	.640	.526	Üst _{%27} Alt _{%27}	4.47 1.88	.962 1.100	
Mutlak Kaçınma N=374 Cronbach Alpha=.55		MK1	.463	.348	Üst _{%27} Alt _{%27}	5.00 3.59	.000 1.275	11.83**
	MK2	.458	.352	Üst _{%27} Alt _{%27}	5.00 3.54	.000 1.434	10.84*	
				Üst _{%27} Alt _{%27}	5.00 3.45	.000 1.464		
	t testi için sd=223 **p<.001	MK3	.531	.269	Üst _{%27} Alt _{%27}	5.00 3.48	.000 1.358	11.25**
		MK4	.447	.365	Üst _{%27} Alt _{%27}	5.00 3.48	.000 1.358	
	Bağıl Kaçınma N=374 Cronbach Alpha=.73	BK1	.735	.337	Üst _{%27} Alt _{%27}	4.79 3.59	.448 1.520	7.36**
BK2		.708	.436	Üst _{%27} Alt _{%27}	4.80 2.36	.459 1.546	14.57*	
				Üst _{%27} Alt _{%27}	4.85 1.86	.407 1.151		
t testi için sd=206 **p<.001		BK3	.666	.538	Üst _{%27} Alt _{%27}	4.88 2.11	.353 1.181	23.96**
		BK4	.625	.635	Üst _{%27} Alt _{%27}	4.89 2.63	.316 1.437	
İçsel Kaçınma N=374 Cronbach Alpha=.70		İK1	.733	.269	Üst _{%27} Alt _{%27}	4.65 2.37	.696 1.664	12.22**
	İK2	.608	.553	Üst _{%27} Alt _{%27}	4.70 2.38	.499 1.298	16.06*	
				Üst _{%27} Alt _{%27}	4.77 2.44	.442 1.340		
	t testi için sd=197 **p<.001	İK3	.608	.554	Üst _{%27} Alt _{%27}	4.77 2.10	.504 1.225	15.76**
		İK4	.641	.464	Üst _{%27} Alt _{%27}	4.74 2.34	.518 1.416	
	İK5	.637	.472	Üst _{%27} Alt _{%27}	4.74 2.34	.518 1.416	15.33*	
Üst _{%27} Alt _{%27}				4.74 2.34	.518 1.416			

SOBYÖ'nün Puanlanması

SOBYÖ'de 29 madde yer almaktadır (Ek-1). Ölçekte 5'li likert tipi bir derecelendirme kullanılmıştır. Ölçek, İY, MY, BY, MK, BK ve İK olmak üzere altı boyutlu bir yapıya sahiptir. SOBYÖ'den toplam bir puan hesaplanmamaktadır. Ölçekte sadece alt boyutlardan alınan puanlar üzerinden işlem yapılabilmektedir.

TARTIŞMA VE SONUÇ

Bu araştırmada, ortaokul öğrencilerinin sosyal bilgiler dersi başarı yönelimlerinin ölçülmesine hizmet edecek bir ölçme aracının geliştirilmesi amaçlanmıştır. SOBYÖ'nün geliştirilmesi sürecinde, 3x2 başarı yönelimleri modeline ilişkin mutlak yaklaşma yönelimi, mutlak kaçınma yönelimi, içsel yaklaşma yönelimi, içsel kaçınma yönelimi, bağıl yaklaşma yönelimi ve bağıl kaçınma yönelimi olmak üzere üç boyut temel alınmış (Elliot, vd., 2011) ve söz konusu boyutlardan hareketle SOBYÖ maddelerinin yazımında, literatürdeki başarı yönelimleri ölçeklerinden yararlanılmıştır (Akın, 2006; Akın ve Çetin, 2007; İlhan ve Çetin, 2014). MY boyutunu oluşturan 5, MK boyutunu oluşturan 5 madde, İY boyutunu oluşturan 7 madde, İK boyutunu oluşturan 6 madde, BY boyutunu oluşturan 7 madde ve BK boyutunu oluşturan 7 madde yazılmasıyla toplam 37 maddelik bir taslak ölçek formu elde edilmiştir. Ölçeğin kapsam ve görünüş geçerliğinin test edilmesi amacıyla uzman görüşüne başvurulmuş ve taslak form asıl uygulama öncesinde öğrencilere uygulanmıştır. Ölçme aracında yer alan 37 madde için uzmanlardan ve öğrencilerden alınan dönütler, ölçek maddelerinde herhangi bir değişikliğe ihtiyaç duyulmadığını ortaya koymuştur. Taslak ölçek formu, beşli likert tipi bir derecelendirme (Kesinlikle Katılıyorum 5; Kesinlikle Katılmıyorum 1) kullanılarak öğrencilere uygulanmıştır.

SOBYÖ'nün yapı geçerliğinin test edilmesi amacıyla AFA ve DFA yapılmıştır. AFA sonuçları incelendiğinde AFA'nın ilk sonuçlarına göre madde 2, madde 8, madde 13, madde 16, madde 23 ve madde 36'nın kuramsal olarak yer alması gereken boyutta yeterli faktör yüküne sahip olmadığı görülmüştür. Bu doğrultuda söz konusu altı madde ölçekten çıkarılmıştır. Bu altı maddenin çıkarılmasının ardından yapılan ikinci AFA sonucunda; madde 30 ile 31'in kuramsal olarak desteklenmediği başka bir faktöre kaydığı belirlenmiş ve bu iki madde ölçme aracından çıkarılmıştır. Sonuç olarak 29 maddelik toplam varyansın %50.82'sini açıklayan ve kuramsal yapıyla birebir uyum gösteren 6 faktörlü bir ölçek elde edilmiştir. Faktörlerde yer alan maddelerin kapsamı ve ilgili literatür doğrultusunda birinci faktör içsel yaklaşma (İY), ikinci faktör mutlak yaklaşma (MY), üçüncü faktör bağıl yaklaşma (BY), dördüncü faktör mutlak kaçınma (MK), beşinci faktör bağıl kaçınma (BK) ve altıncı faktör içsel kaçınma (İK) olarak isimlendirilmiştir. AFA sonucunda ulaşılan 29 madde ve altı faktörlü yapının doğrulanmasına ek kanıt elde etmek için DFA yapılmıştır. DFA sonucunda ortaya çıkan bulgular, uyum indekslerine ilişkin mükemmel ve kabul edilebilir uyum ölçütlerine göre, altı faktörlü modelin uyumlu olduğunu göstermektedir. Ayrıca DFA sonucunda elde edilen *t* değerlerinin tamamının .01 düzeyinde anlamlı olduğu belirlenmiştir. *t* değerlerinin anlamlı olmaması; söz konusu *t* değerine sahip maddelerin model dışında bırakılması gerektiği ya da çalışmadaki örneklem sayısının faktör analizinin gerçekleştirilmesi için yeterli olmadığı anlamına gelmektedir (Byrne, 2010). Bu bağlamda, ölçek maddelerinin faktör yükünün .30 alt sınırının üzerinde olduğunu gösteren AFA sonuçları (Büyüköztürk, 2010) ile ölçek maddelerinin *t* değerlerinin anlamlı olduğunu ve kabul edilebilir sınırlar içerisinde yer alan uyum indekslerine sahip olduğunu gösteren DFA sonuçları, SOBYÖ'nün yapı geçerliğinin sağlandığına işaret etmektedir.

Ölçek güvenilirliği, Cronbach Alpha iç tutarlılık yöntemiyle incelenmiştir. Ölçeğe ilişkin hesaplanan güvenilirlik katsayıları ise şu şekildedir: İY alt ölçeği için .73, MY alt ölçeği için .63, BY alt ölçeği için .71, MK alt ölçeği için .55, BK alt ölçeği için .73, İK alt ölçeği için ise .70. Güvenirlik katsayısı olarak .70 ve üzerinin sınır kabul edilmesi (Tezbaşaran, 1997), az sayıda maddeye sahip (10 ya da daha az) ölçekler için bu sayının .60 (Sipahi, Yurtkoru ve Çinko, 2010; Şeker ve Gençdoğan, 2006) veya .50'nin üzeri olarak ölçüt alınabileceği (Nunnally, 1978) göz önüne alındığında SOBYÖ'yü oluşturan alt ölçeklerin yeterli düzeyde güvenilir olduğu söylenebilir.

SOBYÖ'de kalması uygun olan maddelerin seçimine yönelik olarak madde korelasyonuna ve gerçekleştirilen seçim işleminin güvenilirliği için ise %27'lik alt-üst grup karşılaştırmalarına başvurulmuştur. Analiz sonuçlarına göre madde toplam korelasyonları; İY alt ölçeği için .32 ile .56 arasında, MY alt ölçeği için .36 ile .50 arasında, BY alt ölçeği için .43 ile .52 arasında, MK alt ölçeği için .26 ile .36 arasında, BK alt ölçeği için .33 ile .63 arasında ve İK alt ölçeği için ise .26 ile .55 arasında değişmektedir. Madde toplam korelasyonu hakkında karar verirken, .30 ölçütünün karşılanıp karşılanmadığına bakılmaktadır (Büyüköztürk, 2010). Buna göre MK alt ölçeğindeki üçüncü madde ve İK alt ölçeğindeki birinci madde dışındaki bütün maddelerin bu kriteri sağladığı söylenebilir. Bununla birlikte %27'lik alt ve üst grup karşılaştırmalarındaki *t* değerlerinin madde korelasyonu .30'un altında yer alan bu iki madde için de anlamlı bulunmuştur. Bu sonuç, ölçekten çıkarılması gereken bir madde olmadığını göstermiştir. Sonuç olarak çalışmada, SOBYÖ'nün psikometrik özelliklerinin incelenmesi için yapılan istatistiksel analizler ışığında ortaya konulan bulgular, ölçeğin ortaokul öğrencilerinin sosyal bilgiler dersi başarı yönelimlerinin belirlenmesinde kullanılabilecek geçerli ve güvenilir bir ölçme aracı olduğunu göstermektedir. Ayrıca bu araştırma ile 3x2 başarı yönelimleri modelinin Türk literatürüne kazandırılması sağlanmıştır.

ARAŞTIRMANIN SINIRLILIKLARI ve İLERİ ARAŞTIRMALARA YÖNELİK ÖNERİLER

Bu çalışmada SOBYÖ'nün geliştirilmesi amaçlanmıştır. Türkçe literatür incelendiğinde 3x2 başarı yönelimlerinin kullanıldığı ya da sosyal bilgiler odaklı başarı yönelimlerinin ele alındığı bir çalışmaya rastlanmamıştır. Bu yönüyle araştırmanın özgün olduğu ve literatüre katkı sunacağı ifade edilebilir. Diğer yandan bu araştırmanın bir takım sınırlılıkları bulunmaktadır. Bu sınırlılıkların aşılabilmesi için yeni araştırmaların yapılması ihtiyacı belirlemektedir. İlk olarak SOBYÖ'nün yapı geçerliği kapsamında sadece AFA ve DFA yapılmıştır. SOBYÖ'yü kullanacak araştırmacıların, ölçme aracının yapı geçerliğine yönelik yakınsak ve iraksak geçerliği yapmalarının ölçeğin geçerliğine katkı sağlayacağı düşünülmektedir. SOBYÖ'nün ölçüt bağımlı geçerlik çalışmasının yapılmamış olması, araştırmanın ikinci sınırlılığını oluşturmaktadır. Bu sınırlılığın ortadan kaldırılması için ileri araştırmalarda öğrencilerin SOBYÖ'de yer alan alt ölçek puanları ile sosyal bilgiler dersi başarıları arasındaki ilişkiye bakılabilir. Araştırmanın üçüncü sınırlılığı ise bu çalışmada, 3x2 başarı yönelimleri ölçeğinin sosyal bilgiler odaklı olarak geliştirilmesi yani sosyal bilgiler konu alanı ile sınırlandırılmış olmasıdır. Bu sınırlılığın başarı yönelimlerinin farklı konu alanlarına göre incelenmesiyle aşılabileceği düşünülmektedir. Son olarak başarı yönelimlerine ilişkin literatür incelendiğinde başarı yönelimleri; akademik başarı (Akın, 2006; İzci ve Koç, 2012; Tapola ve Niemivirta, 2008; Pekrun, Elliot ve Maier, 2009) biliş ötesi farkındalık (Akın, 2006), öğrenme yaklaşımları (Camdemir, 2013), sınıf iklimi (Anderman 1999; Midgley, Kaplan, Middleton, Maehr ve Urdan 1998; Ryan ve Patrick 2001), sınıf değerlendirme atmosferi (Buldur, 2014; Brookhart ve Bronowicz, 2003; Brookhart ve DeVoge, 1999; Wang, 2004) gibi çok sayıda faktörle ilişkilidir. Bu kapsamda sosyal bilgiler odaklı başarı yönelimlerinin sözü geçen faktörlerle olan ilişkisinin ortaya konulması amacıyla ileri çalışmalarda ele alınması önerilmektedir. Böylelikle sosyal bilgiler odaklı başarı yönelimleri ölçeğinin ölçme gücünün daha da arttırılacağı düşünülmektedir.

KAYNAKÇA

- Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları*. İstanbul: İdeal Kültür Yayıncılık.
- Akın, A. ve Çetin, B. (2007). Başarı Yönelimleri Ölçeği, Geçerlik ve Güvenirlik Çalışması. *Eğitim Araştırmaları*, 26, 1-12.
- Akın, A. (2006). 2x2 başarı yönelimleri ölçeği: geçerlik ve güvenilirlik çalışması. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 12, 1-13.
- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: A cognitive-motivational analysis. In C. Ames & R. Ames (Eds.), *Research on motivation in education*, (3, pp. 177-207). New York, NY: Academic Press.
- Ames, C. (1992a). Achievement goals and the classroom motivational climate. In D.H. Schunk & J. Meece (Eds.), *Student perceptions in the classroom* (pp. 327-348). Hillsdale, NJ: Erlbaum.
- Ames, C. (1992b). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271.

- Ames, C., and Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260-267.
- Anderman L. (1999). Classroom goal orientation, school belonging, and social goals as predictors of students' positive and negative affect following the transition to middle school. *Journal of Research and Development in Education*. 32, 89-103.
- Anderman, E.M., and Midgley, C. (1996). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle level schools. *Contemporary Educational Psychology*, 22, 269-298.
- Anderman, E.M., and Wolters, C.A. (2006). Goals, values, and affect: Influences on student motivation. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 369-389). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Anderman, E.M., Austin, C.C., and Johnson, D.M. (2002). The development of goal orientation. In A. Wigfield & J.S. Eccles (Eds.), *Development of achievement motivation* (pp. 197-220). San Diego: Academic Press.
- Anderman, L.H., and Anderman, E.M. (2000). Considering contexts in educational psychology: introduction to the special issue. *Educational Psychologist*, 35(2), 67-68.
- Bandura, A. (1977) Self-efficacy: toward a unifying theory of behavioural change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1986). *The social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1991). Social cognitive theory of moral thought and action. In Kurtines W.M. & Gewirtz, J.L. (Eds.), *Handbook of moral behavior and development* (pp. 45-103). Hillsdale, NJ: Erlbaum.
- Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419-456.
- Bentler, P.M., and Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Brookhart, S.M., and Bronowicz, D.L. (2003). I don't like writing. It makes my fingers hurt: Students talk about their classroom assessments. *Assessment in education: Principles, Policy & Practice*, 10(2), 221-242.
- Brookhart, S.M., and DeVoge, J.G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, 12, 409- 425.
- Buldur, S. (2014). Öğrencilerin sınıf-içi değerlendirme ortamına ilişkin algıları ile başarı amaç oryantasyonları arasındaki ilişkinin incelenmesi: Cinsiyet perspektifi. *Eğitim ve Bilim*, 39(176), 213-225.
- Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi Yayınları
- Byrne, B.M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*. New York, NY: Taylor and Francis Group.
- Canidemir, A. (2013). *Ortaöğretim öğrencilerinin öğrenme yaklaşımları ve başarı amaç yönelimlerinin akademik başarı ile ilişkisinin incelenmesi*. Yayınlanmamış yüksek lisans tezi. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040 –1048.
- Dweck, C.S., and Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Elliot, A., and McGregor, H. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-509.
- Elliot, A.J. (1997). Integrating "classic" and "contemporary" approaches to achievement motivation: A hierarchical model of approach and avoidance achievement motivation. In P. Pintrich & M. Maehr (Eds.), *Advances in motivation and achievement*, (pp. 143-179). Greenwich, CT: JAI Press.
- Elliot, A.J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34, 169-189.
- Elliot, A.J. (2005). A conceptual history of the achievement goal construct. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford Press
- Elliot, A.J., and Church, M.A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 218-232.
- Elliot, A.J., and Covington, M.V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, 13(2), 73-92.
- Elliot, A.J., and Harackiewicz, J.M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70, 461-470.
- Elliot, A.J., and Thrash, T.M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review*, 13, 139-156.

- Elliot, A.J., Murayama, K., and Pekrun, R. (2011). A 3 x 2 achievement goal model. *Journal of Educational Psychology, 103*, 632-648.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi Yayınları.
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications.
- Harackiewicz, J.M., Barron, K.E., Karter, S.M., Lahte, A.T., and Elliot, A.J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology, 73*(6), 1284-1295.
- Hu, L.T., and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55.
- İlhan, M. ve Çetin, B. (2014). Matematik başarı yönelimleri ölçeğinin geliştirilmesi (MBYÖ): geçerlik ve güvenilirlik çalışması. *International Online Journal of Educational Sciences, 6*(3), 682-705.
- İzci, E. ve Koç, S. (2012). Pedagojik formasyon eğitimi alan öğrencilerin başarı yönelim düzeylerinin incelenmesi. *Dicle Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 4*(8), 31-43.
- Jagacinski, C., and Nicholls, J. (1987). Competence and affect in task involvement and ego involvement: The impact of social comparison information. *Journal of Educational Psychology, 79*, 107-114.
- Jagacinski, C.M., and Strickland, O.J. (2000). Task and ego orientation the role of goal orientations in anticipated affective reactions to achievement outcomes. *Learning and Individual Differences, 12*, 189-208.
- Kaplan, A., and Midgley, C. (1999). The relationship between perceptions of the classroom goal structure and early adolescents' affect in school: The mediating role of coping strategies. *Learning and Individual Differences, 11*, 187-212.
- Keys, T.D., Conley, A.M., Duncan, G.J., and Domina, T. (2012). The role of goal orientations for adolescent mathematics achievement. *Contemporary Educational Psychology, 37*(1), 47-54.
- Kline, R.B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Maehr, M.L., & Nicholls, J.G. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in cross-cultural psychology* (3, pp. 221-267). New York, NY: Academic Press.
- Marsh, H.W., Hau, K.T., Artelt, C., Baumert, J., and Peschar, J.L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*(4), 311-360.
- Meece, J.L., and Holt, K. (1993). A pattern analysis of students' achievement goals. *Journal of Educational Psychology, 85*, 582-590.
- Meece, J.L., Blumfeld, P.C., and Hoyle, R.K. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*(4), 514-523.
- Meyers, L.S., Gamst, G., and Guarino, A. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.
- Middleton, M.J., and Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology, 89*, 710-718.
- Middleton, M.J., Kaplan, A., and Midgley, C. (2004). The change in middle school students' achievement goals in mathematics over time. *Social Psychology of Education, 7*, 289-311.
- Midgley, C., Kaplan, A., and Middleton, M.J. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology, 93*, 77-86.
- Midgley, C., Kaplan, A., Middleton, M., Maehr, M.L., Urdan, T. et al. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology, 23*, 113-131.
- Nicholls, J.G., Patashnick, M., and Nolen, S.B. (1985). Adolescents' theories of education. *Journal of Educational Psychology, 77*, 683-692.
- Nichols, W.D., Jones, J., and Hancock, D. (2003). Teachers' influence on goal orientation: Exploring the relationship between eighth graders' goal orientation, their emotional development, their perceptions of learning, and their teachers' instructional strategies. *Reading Psychology, 24*(1), 57-87.
- Nunnally, J.C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.
- Pajares, F., and Cheong, Y.F. (2003). Achievement goal orientations in writing: A developmental perspective. *International Journal of Educational Research, 39*, 437-455.
- Pajares, F., Britner, S.L., and Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology, 25*(4), 406-422.
- Pekrun, R., Elliot, A.J., and Maier, M.A. (2009). Achievement goals and achievement emotions: testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115-135.
- Pintrich, P.R. (2000). An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research. *Contemporary Educational Psychology, 25*, 92-104.

- Pintrich, P.R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95, 667-686.
- Pintrich, P.R., and Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In M.L. Maehr, & P.R. Pintrich (Eds.), *Advances in motivation and achievement* (pp. 371-402). Greenwich: JAI Press.
- Pintrich, P.R., and Schunk, D.H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Pintrich, P.R., Conley, A.M., and Kempler, T.M. (2003). Current issues in achievement goal theory and research. *International Journal of Educational Research*, 39, 319-337.
- Ryan, A.M., and Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38, 437-460.
- Schunk, D.H. (1989). Self-efficacy and cognitive achievement: Implications for students with learning problems. *Journal of Learning Disabilities*, 22, 14-22.
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26, 207-231.
- Schunk, D.H., and Ertmer, P.A. (2000). Self-regulation and academic learning: self-efficacy enhancing interventions. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 631-649). San Diego: Academic Press.
- Schunk, D.H., and Swartz, C.W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18, 337-354.
- Şeker, H., ve Gençdoğan, B. (2006). *Psikolojide ve eğitimde ölçme aracı geliştirme*. Ankara: Nobel Yayın Dağıtım.
- Sipahi, B., Yurtkoru, S. ve Çinko, M. (2010). *Sosyal bilimlerde SPSS'le veri analizi*. İstanbul: Beta Basım Yayım Dağıtım
- Skaalvik, E.M. (1997). Self-enhancing and self-defeating ego orientation: relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology*, 89, 71-81.
- Tapola, A., and Niemivirta, M. (2008). The role of achievement goal orientations in students' perceptions of and preferences for classroom environment. *British Journal of Educational Psychology*, 78, 291-312.
- Tezbaşaran, A. (1997). *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği.
- Tuominen-Soini, H., Salmela-Aro, K., and Niemivirta, M. (2008). Achievement goal orientations and subjective well-being: A person-centred analysis. *Learning and Instruction*, 18, 251-266.
- Wang, X. (2004). *Chinese EFL students' perceptions of classroom assessment environment and their goal orientations in the college English course*. Unpublished master's thesis. Queen's University, Kingston, Ontario, Canada.
- Wolters, C.A., Yu, S.L., and Pintrich, P.R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences*, 8, 211-238.
- Zimmerman, B.J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P.R., Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.
- Zweig, D., and Webster, J. (2004). Validation of a multidimensional measure of goal orientation. *Canadian Journal of Behavioral Science*, 36(3), 232-243.

EXTENDED ABSTRACT

Introduction

A review of the literature revealed that an instrument to measure students' social studies achievement goal does not exist in literature. Also 3x2 achievement goal model does not exist in Turkish culture. Identifying students' social studies achievement goal may help to determine their purpose for engaging in achievement behaviour. For this reason, it is of great significance to develop a Turkish instrument to determine students' achievement goal regarding social studies. In this context, the present study aims to develop a valid and reliable instrument to measure social studies oriented of achievement goal.

Method

The participants were 374 middle school students studying in the central district of Diyarbakır in 2014-2015 school year fall semester. Expert opinion was consulted with regard to the scale's content

and face validity. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were performed in order to measure the scale's construct validity. The reliability of the Social Studies Oriented Achievement Goal (SOAGS) was tested through internal consistency coefficient. If the items should remain in the scale or not, was determined regarding the item total correlation and a comparison between the top and bottom 27% groups. The validity and reliability analyses were carried out with SPSS 20.0 and LISREL 8.54.

Results and Discussion

According to EFA results, a six-factor structure which accounted 50.82% of the total variance was acquired. Taking into consideration the items' content and theoretical structure, the primary factor was named self-approach, the second factor was named task-approach, the third factor was named other-approach, the fourth factor was named task-avoidance, the fifth factor was named other-avoidance, and finally the sixth factor was named self-avoidance. In order to determine whether the 29 items and six-factor structure obtained as a result of the EFA gives adequate goodness of fit indices, and to obtain further support for construct validity, CFA was applied. The CFA findings revealed that the scale has adequate goodness of fit indices [$\chi^2/sd=1.53$, CFI=.91, NNFI=.90, IFI=.91, RMSEA=.58, SRMR=.68, PNFI=.76, and PGFI=.72]. The reliability analysis indicated that the internal consistency coefficients were .73 for the self-approach sub-scale, .63 for the task-approach sub-scale, .71 for the other-approach sub-scale, .55 for the task-avoidance sub-scale, .73 for the other-avoidance sub-scale, and finally .70 for the self-avoidance sub-scale respectively. Generally, it is accepted that the instruments which have .70 or higher internal consistency coefficient is reliable (Tezbaşaran, 1997). However, when the number of items in the scale is less, the reliability coefficient is accepted as reliable when it is .60 (Sipahi, et al., 2010; Şeker and Gençdoğan, 2006) or above .50 (Nunnally, 1978). Accordingly this scale has reliable values in terms of internal consistency. In order to detect discriminatory and predictive power of the items, corrected-item total correlations were computed and both top and bottom 27% of groups were compared. Pearson Product Moment Correlation was used to determine corrected-item total correlation. In order to compare both top and bottom 27% of groups, independent-samples t-test was performed. The item analysis reported that the corrected item total correlations ranged from .32 and .56 for the self-approach, .36 and .50 for the task-approach, .43 and .52 for the other approach, .26 and .58 for the task-avoidance, .33 and .63 for the other-avoidance, and .26 and .55 for the self-avoidance. In addition to the differences between the top and bottom 27% groups were significant for all the items included in the scale. Considering that the item total correlations are above .30 (Büyüköztürk, 2010; Erkuş, 2012; Field, 2009), and the significantly *t*-values related to the top and bottom 27% group differences can be regarded as evidence for the item' discrimination. Thus, it can be said that all of the items in the scale were discriminatory.

In this study, the findings from statistical analyses of psychometric characteristics of the SOAGS revealed that the scale can be used as a valid and reliable instrument to measure students' social studies achievement goal. As a result of the present study which aims to contribute to the Turkish literature with the SOAGS, it can be argued that an instrument which has satisfactory psychometric properties to measure students' social studies achievement goal has been developed. A review of the literature revealed that an instrument to measure students' Social studies achievement goal does not exist in literature and 3x2 achievement goal model has not been developed in Turkish literature. In this respect, the research can be expressed as an original research and will contribute to the literature. On the other hand, this study has some limitations. In order to overcome these limitations, it is proposed making in the new research. Firstly, in order to measure the scale's construct validity only Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were performed. When SOAG is used in further research, the construct validity of the scale tested in convergent and divergent validity studies are expected to contribute to the validity of the scale. The second limitation of the research is not to study about criterion-related validity. On the basis of this, it is proposed that in the future research the relationship between the students of grades in social studies courses and subscales in the SOAGS scores can be examined. The third limitation of the research in this study is that it was aimed to develop the domain oriented achievement goals scale of 3x2 and research was limited to social studies. In order to overcome this limitation, it is suggested to develop achievement

goals' scale for different courses. Finally, in reviewing the literature related to achievement goals suggests that students' achievement goal have an influence on their success, metacognitive awareness, learning approach, classroom climate and classroom asesment environment. In this context, further studies could be conducted on the correlation between the SOAGS and academic success as well as their metacognitive awareness, learning approach, classroom climate and classroom asesment environment. It is believed that the studies suggested above will also make great contributions to the literature which the SOAGS measure what is intended.

Tam Öğrenme Yaklaşımının Öğrencilerin Akademik Başarıları Üzerindeki Etkisi: Bir Meta-Analiz Çalışması

The Effect of Mastery Learning Model on Students Academic Achievement: A Meta-Analysis Study

Tarık BAŞAR*

İlkay AŞKIN TEKKOL**

Selahattin GELBAL***

Öz

Bu araştırmada tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi araştırılmıştır. Araştırmada ayrıca tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin alan ve öğrenim düzeyine göre farklılaşıp farklılaşmadığı incelenmiştir. Araştırma meta-analiz çalışması niteliğinde olup araştırmaya toplam 24 çalışma (tez, makale ve bildiriler) dahil edilmiştir. Araştırma sonuçlarına göre çalışmaların tamamı pozitif (+) etki büyüklüğüne sahiptir. Meta-analize dahil edilen çalışmaların %50'sinin muazzam düzeyde, %29,17'sinin geniş düzeyde, %12,5'inin orta düzeyde ve %8,33'ünün ise çok geniş düzeyde etki büyüklüğüne sahip olduğu belirlenmiştir. Araştırmaya dahil edilen çalışmaların ortalama etki büyüklüğü, rastgele etkiler modeline göre 1,546 olarak hesaplanmıştır. Bu sonuca göre tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde oldukça yüksek düzeyde ve olumlu yönde bir etkiye sahip olduğu ortaya konmuştur. Öğrencilerin akademik başarılarının, dersin uygulandığı alanlar bakımından anlamlı bir farklılık göstermezken öğrenim düzeyi bakımından anlamlı bir farklılık gösterdiği sonucuna ulaşılmıştır. Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde, lise ve ilköğretim düzeyinde muazzam düzeyde etkili olduğu belirlenmiştir.

Anahtar Kelimeler: tam öğrenme, meta-analiz, akademik başarı, yöntem, öğrenciler

Abstract

In this research the effect of mastery learning model on academic achievement of students is investigated. Whether the effect of mastery learning on academic achievement changes according to the field and education level or not is also investigated. This research is a meta-analysis study and consists of 24 research (dissertations, articles and oral presentations). According to the results of the study, all of the studies have positive (+) effect size. It was found that 50% of the studies that were included to the meta-analysis had huge effect size, 29,17% of them showed large, 12,5% of them showed middle and 8,33% of them showed very large effect size. According to random effect model, average effect size of the studies that were included to this research is calculated as 1,546. According to this finding, mastery learning approach has a very high and positive effect on academic achievement of students. While the academic achievement of the students doesn't change in terms of the fields that the lessons applied; a significant difference is found in terms of education level. Mastery learning approach has a huge effect on academic achievement for primary school and high school level.

Keywords: mastery learning, meta-analysis, academic achievement, method, students

*Dr., Ahi Evran Üniversitesi, Eğitim Fakültesi, Kırşehir-Türkiye, e-posta: tarikbasar84@hotmail.com

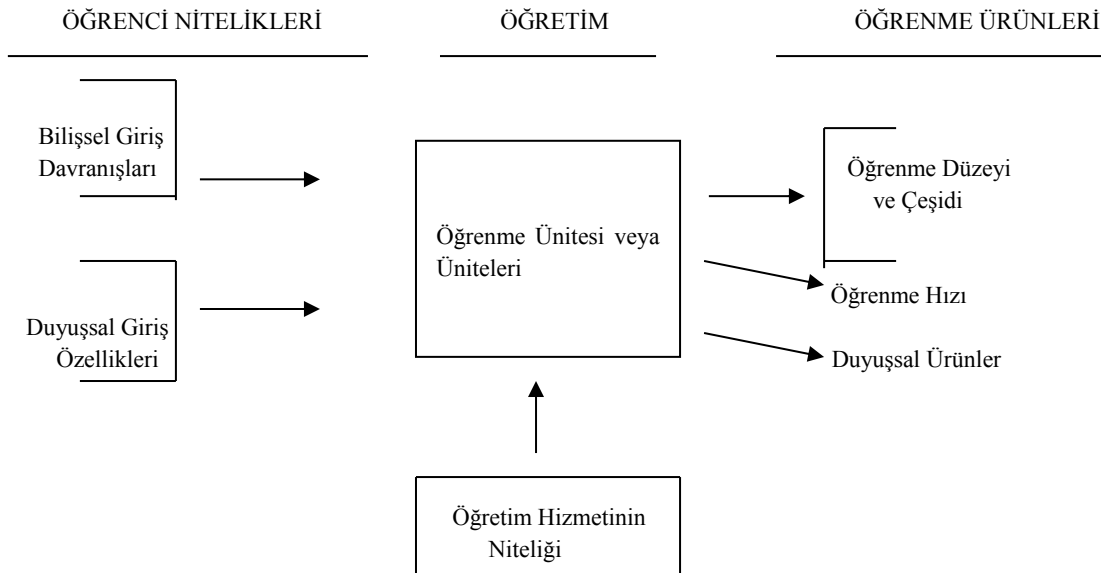
**Dr., Kastamonu Üniversitesi, Eğitim Fakültesi, Kastamonu-Türkiye, e-posta: ilkayaskin@hotmail.com

***Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, e-posta: sgelbal@gmail.com

GİRİŞ

Bireyler arasındaki öğrenme seviyesi farklılıklarının temel nedenleri incelendiğinde okullardaki öğretme-öğrenme durumlarının ve çevresel etmenlerin, doğuştan getirilen özelliklerden daha etkili olduğu görülmektedir (Senemoğlu, 2013). Okuldaki öğrenmeyi etkileyen faktörlerden bazıları, zekâ, öğretmen özellikleri, ailenin sosyo-ekonomik durumu gibi, öğretme-öğrenme süreciyle doğrudan değiştirilemeyecek faktörler iken bazıları ise, ön-öğrenmeler, ilgi, tutum, başarı algısı, öğretim hizmetinin niteliği gibi, öğretme-öğrenme süreci yoluyla değiştirilebilen faktörlerdir (Senemoğlu, 2013). Öğretme-öğrenme süreci yoluyla değişim gösteren bu özellikler tüm öğrencilerin, öğretilmek istenen davranışları öğrenebileceğini vurgulayan tam öğrenme modelinin temelini oluşturmaktadır.

Tam öğrenme, öğrencilerin tümünün okulların öğretmeyi amaçladığı davranışları öğrenebileceğini ifade eden bir yaklaşımdır. Tam öğrenme yaklaşımında, öğrencilere planlı bir öğretim hizmeti sağlanması ve öğrenme gücünü yaşayan öğrencilere yardım edilmesi söz konusudur. Ayrıca öğrencilerin belirlenen öğrenme hedeflerine ulaşabilmeleri amacıyla onlara ihtiyaç duydukları zaman verilerek onlar için de anlamlı olacak “tam öğrenme” ölçütü belirlenmekte ve hemen hemen tüm öğrencilerin yüksek düzeyde öğrenmeler gerçekleştirmesi sağlanmaktadır (Bloom, 1976). Tam öğrenme yaklaşımında, yer alan başlıca değişkenler Şekil 1’de gösterilmiştir (Bloom, 1976).



Şekil 1. Tam Öğrenme Yaklaşımında Yer Alan Değişkenler

Bloom (1976), bilişsel giriş davranışları ve duyuşsal giriş özelliklerinin üzerinde önemle durmuştur. Öğrenci niteliklerinde yer alan bilişsel giriş davranışları, öğrenme ünitesi/ünitelerinin öğrenilebilmesi için gerekli kabul edilen ön öğrenmeleri ifade etmektedir. Bloom ve arkadaşlarının (1976) yaptığı araştırmalar sonucunda öğrencilerin bilişsel giriş davranışları ile daha sonraki öğrenme ünitelerindeki başarıları arasında güçlü bir ilişki olduğu ortaya konmuştur. Duyuşsal giriş özellikleri ise öğrencilerin öğretme-öğrenme sürecine girerken sahip oldukları ilgilerini, tutumlarını ve başarılı olacıklarına olan inançlarını ve kendilerine güvenlerini ifade etmektedir (Erden ve Akman, 1996).

Öğretim hizmetinin niteliği değişkeninin, öğretimle ilgili en önemli değişken olduğu düşünülmektedir. Öğretim hizmetinin niteliği değişkeninin kapsamı içine öğrencilerin neyi nasıl öğreneceklerini bildiren uyarıcılar, öğrencilerin öğrenme sürecine katılma olanakları ve öğrenilenlerin pekiştirilmesi için kullanılan uyarıcılar ile öğrenci ihtiyaçları arasındaki uygunluk derecesi girmektedir (Bloom, 1976).

Bilişsel giriş davranışları, duyuşsal giriş özellikleri ve öğretim hizmetinin niteliği değişkenleri öğrenme ürünlerini belirlemektedir. Öğrenmenin çeşidi ve düzeyi, öğrenme hızı ve öğrencinin öğrenme ünitesi

ve kendi kişiliği ile ilgili olarak geliştirdiği duyuşsal özellikler de bu etkenlerin belirlediği ürünlerdir. Öğrencinin ilgili giriş özellikleriyle öğretim hizmetinin niteliğinin olumlu olması halinde öğrenciler, bütün bu öğrenme ürünlerine yüksek düzeylere erişecek ve bu ürünler bakımından öğrenciler arası farklılaşma da en aza inecektir. Tam tersi bir durumda ise yani öğrenciler arasında giriş özellikleri bakımından büyük farklar bulunması ve öğretim hizmetinin niteliğinin de tüm öğrenciler için “en iyi” olmaması halinde ise öğrenme ürünlerinde de büyük farklar görülecektir. Bu değişkenlerin birinin ya da birkaçının istenilen düzeyin altına düşüp düşmemesi; öğrenmenin çeşit ve düzeyini, öğrenme sırasında karşılaşılabilecek güçlükleri, belli düzeyde öğrenme için gerekecek zaman ve çabayı, öğrencinin öğrenmeye, öğrenme sürecine ve ilgili yönleriyle kendi kişiliğine karşı duygularını belirleyecektir (Bloom, 1976).

Ülkemizde tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirlemek amacıyla birçok araştırma yapılmıştır (Arslan, 1996; Balta, 2007; Çelik ve Şengül, 2005; Demirci ve Yıldırım, 1996; Hevedanlı, Oral ve Akbayın, 2005; İşeri, 2004; Karaca, 2007; Kuzu, Kuzu ve Özdemir, 2005; Kurtuldu ve Bakıoğlu, 2012; Öner, 2005; Özder, 2000; Özer, 2012; Savcı ve Kırkıç, 2012; Sever, 1993; Yıldırım ve Afreşa, 1994; Yıldırım ve Aydın, 1994; Yıldırım ve Eğinlioğlu, 1987; Yıldırım ve Kırkıç, 1996; Yıldırım ve Nwabueze, 1991; Yıldırım ve Tuğal, 2004; Yıldırım ve Sayar, 1987; Yurteri, 2005; Yüksel, 2010; Zengin, 2005). Bu araştırmaların tamamı deneysel çalışmalardan oluşmakta ve tam öğrenmeye dayalı öğretme-öğrenme süreci sonunda öğrencilerin başarılarında anlamlı düzeyde artış olduğu görülmektedir. Araştırmaların farklı alanlarda, farklı öğrenim düzeyindeki öğrenciler üzerinde yapılması ve sonuçta elde edilen verilerin de birbirinden farklı olması bu araştırmalara yönelik bir meta-analizin yapılması ihtiyacını doğurmaktadır. Bu sayede Türkiye’de yapılmış olan ve tam öğrenme yaklaşımının öğrencilerin başarıları üzerindeki etkilerini ortaya koyan araştırmaların tümü incelenerek tam öğrenme yaklaşımının akademik başarı üzerindeki etkisine yönelik bütüncül bir yorum yapmanın mümkün olacağı düşünülmektedir.

Araştırmanın Amacı

Bu çalışmada, ülkemizde tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirlemeye yönelik yapılmış olan benzer çalışmaların bulgularının meta-analiz yöntemiyle bir araya getirilerek bütünleştirilmesi ve yorumlanması amaçlanmıştır. Ayrıca çalışmada, tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin alanlara ve öğrencilerin öğrenim düzeyine göre farklılaşp farklılaşmadığı da incelenmiştir.

Bu çalışma tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirlemeye yönelik yapılmış olan benzer çalışmaların bulgularının bütünleştirilerek bir sentezin yapılmasına imkân verdiği için önemli görülmektedir. Bir konu üzerinde yapılmış benzer çalışmaların bulgularının bir araya getirilerek bunların bir sentezinin yapılması, tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi hakkında anlamlı yorumlar yapılmasına fırsat sağlayacaktır. Böylece öğrencilerin akademik başarıları üzerindeki etkisi hakkında daha geniş ve kapsamlı bir görüşe ulaşılması sağlanarak gerek bu konuda yapılacak uygulamalara gerekse bundan sonra yapılacak araştırmalara katkı sağlayacağı düşünülmektedir.

Türkiye’de tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini inceleyen bir meta-analiz çalışmasına rastlanmaması nedeniyle bu çalışmada tam öğrenme yaklaşımı ele alınmıştır. Bu konudaki yurt dışındaki alan yazın incelendiğinde ise Kulik, Kulik ve Bangert-Drowns’un 1990 yılında yaptıkları ve tam öğrenmenin akademik başarı üzerindeki etkisini inceledikleri meta-analiz çalışmasının yer aldığı görülmektedir. Araştırmada aşağıdaki sorulara cevap aranmıştır:

- Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi nedir?
- Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi dersin uygulandığı alana göre anlamlı bir farklılık göstermekte midir?

- Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi öğrencilerin öğrenim düzeyine göre anlamlı bir farklılık göstermekte midir?

YÖNTEM

Araştırma Yöntemi

Çalışmada tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirlemeye yönelik yapılmış olan benzer çalışmaların bulgularının bir araya getirilerek bütünleştirilmesini sağlayan meta-analiz yöntemi kullanılmıştır. Glass'a (1976) göre meta-analiz, bulguların kaynaştırılması amacıyla, geniş çapta analizlerin sonuçlarının toplanmasının istatistiksel bir analizidir. Meta-analiz yönteminin temel amacı, deneysel araştırmada ihtiyaç duyulan alan yazın taramasına aynı yöntemsel kesinliği sağlamaktır (DeCoster, 2004).

Lipsey ve Wilson'a (2000) göre meta-analiz yöntemi ile farklı çalışmaların sonuçlarını özetleyen nicel verilere ulaşılarak sonuçların birleştirilebilmesiyle ortak bir yargıya ulaşılabılır (Akt. Sağlam ve Yüksel, 2007). Abramson ve Abramson'a (2001) göre ise bu şekilde elde edilecek sonuçlar, hem benzer bulgulara sahip bireysel araştırmaların geçerliliğini artırabilir hem de farklı bulgulara sahip bireysel araştırmaların da nedenlerinin ortaya koyulması ile yeni bilgiler elde edilmesini sağlayabilir (Akt. Sağlam ve Yüksel, 2007). Meta-analiz aynı zamanda, bireysel araştırmalarda gözlenen sonuçların bulgularının şansa bağlı olup olmadığı hakkında da ipuçları sağlar. Ayrıca örneklem büyüklüğünün artırılması, çalışmanın istatistiksel olarak anlamlılık düzeyini de artırabilir (Sağlam ve Yüksel, 2007).

Meta-analiz yönteminin en önemli üstünlüklerinden biri araştırmanın açık ve sistemli bir şekilde özetlenmesi ile yazarın sayıltıları, süreci, bulguları ve sonucu belirgin olarak değerlendirmesine katkı sağlamasıdır. Bunun yanı sıra meta-analiz, araştırma bulgularının kullanışlı bir şekilde özetlenerek araştırmalardaki bulguların daha farklı ve daha incelikli olarak sunulmasına olanak sağlar (Lipsey ve Wilson 2001).

Meta Analiz Yönteminin Aşamaları

Bir meta analiz çalışması aşağıdaki basamaklardan oluşmaktadır:

Bir araştırma probleminin kesin ve açık olarak ifade edilmesi: Bir meta analizde birinci adım olan araştırma probleminin belirlenmesi demek; açıklayıcı ve cevaplayıcı değişkenlerin kullanılacağı kuramsal yapılara karar vermek demektir. Ayrıca araştırmacı incelemek istediği konuyla ilgili evren büyüklüğünü de belirlemelidir. Yani araştırmacı, evreninin sınırlarını çizmelidir. Araştırmacı, belirlediği ölçütlere göre çalışmaya dahil edilecek veya edilmeyecek araştırmalara karar verebilmelidir. Bu aşamadaki en önemli nokta, incelenen araştırmalardan iyi bir senteze ulaşmayı sağlayacak uygun evreni belirlemektir. Araştırmacı, evrenini belirlerken evrenin sınırlarını, hem araştırma konusuyla ilgili tüm çalışmaları içerecek şekilde hem de bir senteze ulaşılmasını sağlayacak şekilde belirlemelidir (DeCoster, 2004).

Alan yazın araştırması: Araştırma konusunun belirlenmesinin ardından meta-analizin uygulanabilirliği açısından alan yazının taranması oldukça önemlidir (Dinçer, 2014). Eğer araştırmacı, yapılacak meta-analiz çalışmasından gerçekçi sonuçlara ulaşmak istiyorsa yapacağı analiz çalışmasında yayımlanmamış çalışmalara da yer vermelidir. Çünkü yayımlanmış çalışmaların çoğunda, anlamsız bulunan bulgulardan çok, anlamlı bulunan bulgular tercih edilmektedir. Yani araştırma konusuyla istenildiği gibi anlamlı bir farklılık çıkan bulguların yer aldığı çalışmalar yayınlanırken araştırma konusuyla istenildiği gibi anlamlı bir farklılık çıkmayan bulguların yer aldığı çalışmalar yayımlanmamaktadır (DeCoster, 2004).

Çalışmaların kodlanması: Meta-analiz araştırmalarında araştırmacı, öncelikle ulaştığı çalışmaların bir örneklemini oluşturmalı ve daha sonra da bu çalışmaları özelliklerine ve hesaplanan etki büyüklüğüne göre kodlamalıdır (DeCoster, 2004). Ayrıca kodlamaların eksik veri içermemesine dikkat edilmesi gerekmektedir. Yanlış bir kodlama tüm basamakların yeniden yapılmasına yol açabilir (Dinçer, 2014).

Çalışmalardan elde edilen bulguların farklı istatistiksel yöntemlerle birleştirilmesi: Meta-analiz çalışmasına dahil edilen araştırmalarda kullanılan istatistiksel yöntemler (F testi, t-testi vb.) öncelikle ortak ölçülere dönüştürülmelidir. Etki büyüklüğü yöntemi ise ortak ölçülerin uygulanmasının zor olduğu durumlarda kullanılmalıdır. Etki Büyüklüğü İndeksi, Pearson Momentler Çarpımı Korelasyon Katsayısı ya da Fisher's z Dönüştürme Değeri bu yöntemlerin uygulanabilmesi için kullanılabilir. Bu işlem bittikten sonra ise bulguların birleştirilmesi için "birleştirilmiş testler" uygulanır. Bunun için de genellikle Winer, Stouffer ve Fisher birleştirilmiş testleri tercih edilmektedir (Sağlam ve Yüksel, 2007:9).

DeneySEL çalışmalarda etki büyüklüğü, faktörün olduğu durumda ortaya çıkan sonucun, faktörün olmadığı durumda ortaya çıkan sonuca göre etki gücünü belirtmektedir. Etki büyüklüğü deney-kontrol araştırmalarında X etkisinin kontrol grubuna göre deneysel grupta standart sapmaya bağlı olarak ne kadar büyük, etkili ve güçlü olduğunu göstermektedir (Şahin, 1999).

Thalheimer ve Cook (2002), etki büyüklüğü değerleri ile ilgili aşağıdaki sınıflandırmayı yapmıştır:

$-0,15 \leq$ Etki büyüklüğü değeri $< 0,15$ önemsiz

$0,15 \leq$ Etki büyüklüğü değeri $< 0,40$ küçük

$0,40 \leq$ Etki büyüklüğü değeri $< 0,75$ orta

$0,75 \leq$ Etki büyüklüğü değeri $< 1,10$ geniş

$1,10 \leq$ Etki büyüklüğü değeri $< 1,45$ çok geniş

$1,45 \leq$ Etki büyüklüğü değeri muazzam

Meta analiz raporu yazma: Bir meta analiz raporu olabildiğince eksiksiz ve anlaşılır bir şekilde yazılmalıdır. Raporun giriş kısmında, yapılan analizin konusu tanımlanmalı ve bu konu geniş bir psikolojik bağlam içerisine yerleştirilmelidir (DeCoster, 2004).

İstatistiksel Model Seçimi

Meta-analizde, sonuçların birleştirilmesi amacıyla bazı istatistiksel yöntemler kullanılmaktadır. Bu yöntemlerde ilk aşamada, çalışmadaki sonuçların hangi modele daha uygun olduğu tespit edilmektedir. Bu modeller, "sabit etki modeli (fixed effect model)" ve "rastgele etki modeli (random effects model)"dir. Sabit etki modelinin varsayımı, araştırmaya dahil edilen çalışmaların tamamının aynı etkiyi tahmin etmesidir. Bu varsayımın doğruluğu altında, kişisel olarak yapılmış çalışmaların sonuçlarının varyansının tersi ile en küçük varyanslı ağırlıklı ortalamasının bulunması gerekmektedir. Sabit etki modellerinde çalışma sonuçları arasındaki varyansın, birbirleri ile ilişkili görülen verilerden kaynaklandığı kabul edilir. Sabit etki modelinin varsayımları karşılanmadığında rastgele etkiler modeli kullanılmaktadır. Rastgele etkiler modelinde, çalışmaların kendi içlerindeki varyans ve çalışmalar arası varyans dikkate alınarak bir değerlendirme yapılır (Sutton ve diğerleri, 2000: Akt.:Özdemirli, 2011).

Verilerin Toplanması

Çalışmada öncelikle Yüksek Öğretim Kuruluna (YÖK) ait tez veri tabanından "tam öğrenme yaklaşımı", "tam öğrenme modeli" ve "tam öğrenme", "mastery learning" anahtar kavramları kullanılarak tarama yapılmış ve toplam 25 yüksek lisans ile doktora tezine ulaşılmıştır. Bu tezler incelendiğinde 6 tanesinde tam öğrenme yaklaşımının akademik başarı üzerindeki etkisinin incelenmediği görülerek bu çalışmalar, araştırmanın kapsamı dışında tutulmuştur. Kalan 19 tezden erişime açık olan 5 teze YÖK'e ait veri tabanından ulaşılmıştır. Erişime açık olmayan tezlerden 2 tanesine üniversite kütüphanesinden ve Milli Kütüphane'den ulaşılmıştır. Bunun yanı sıra yine erişime açık olmayan tezler için e-posta adreslerine ulaşılabilen yazarlara e-posta gönderilmiş ancak bu yazarlardan e-postalara bir dönüş olmamıştır. Bu durumda erişime açık olmayan ve erişilemeyen 12 tezin makalelerine ulaşmak üzere tarama yapılmıştır ve bu tezlerden 6 tanesinin makalesine, 1

tanisinin ise kongre bildirisine ulaşılmıştır. Sonuç olarak toplamda 7 tez, 6 tez makalesi ve 1 kongre bildirisine ulaşılmıştır.

Makale tarama sürecinde de benzer aşamalar takip edilerek “Google Akademik”ten ve “Hacettepe Üniversitesi Veri Tabanı”ndan aynı anahtar sözcükler aranmış ve yapılan tarama sonucunda istenen niteliklere sahip 5 makaleye ulaşılmıştır. Ulaşılamayan, ancak konuyla ilgili olduğu düşünülen 5 makaleye ise üniversite kütüphanelerinden ve Milli Kütüphane’den ulaşılmıştır. Kütüphanelerden ulaşılan makalelerle birlikte toplam 10 makale araştırmaya dahil edilmiştir.

Sonuç olarak bu araştırma, tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin incelendiği 16 makale, 7 tez ve 1 tane bildiri olmak üzere toplam 24 deneysel çalışmayı içermektedir.

Dahil Edilme Ölçütleri

Meta-analiz çalışmasına dahil edilen araştırmaların seçilmesinde aşağıdaki ölçütler kullanılmıştır:

- Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin incelendiği bir araştırma olması,
- Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin incelendiği araştırmanın Türkiye’de yayımlanan yüksek lisans, doktora tezi, makale ya da bildiri olması,
- Standartlaştırılmış etki büyüklüğüne ulaşabilmek için araştırmalarda kontrol ve deney gruplarının yer alması,
- Etki büyüklüğünün hesaplanabilmesi için araştırmalarda yeterli bilginin (örneklem büyüklüğü, ortalama, standart sapma) yer alması

Çalışmaların Kodlanması

Meta analiz çalışmalarında birbirinden farklı özelliklere sahip araştırmalar arasında karşılaştırmalar yapabilmek için bir kodlama yöntemine gereksinim vardır. Kodlama yapılırken çalışmaların sahip oldukları karakteristikler belirlenir. Bu çalışma kapsamında kodlanacak çalışma karakteristikleri şunlardır:

- Çalışmanın yazarı,
- Çalışmanın yılı,
- Çalışmanın yapıldığı öğrenim düzeyi,
- Çalışmanın yapıldığı ders,
- Çalışmanın yapıldığı alan,
- Çalışmanın yayın türü

Bağımlı-Bağımsız Değişkenler

Meta-analiz kapsamında tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkililiği ile ilgili olarak hesaplanan etki büyüklükleri, bu çalışmanın bağımlı değişkenlerini oluşturmaktadır. Çalışmanın yapıldığı alan ve öğrencilerin öğrenim düzeyi ise çalışmanın bağımsız değişkenlerini oluşturmaktadır.

Verilerin Analizi

Çalışmada her bir araştırmaya ait etki büyüklükleri ve birleştirilmiş etki büyüklüğü “Comprehensive Meta-Analysis” programı kullanılarak hesaplanmıştır. Bu hesaplama ise her bir araştırmanın deney ve kontrol gruplarına ait ortalama, standart sapma ve örneklem büyüklüğü değerlerinden yararlanılarak

yapılmıştır. Çalışmada, etki büyüklüğü hesaplamasında ise “Hedge’s g” kullanılmıştır. Çünkü Üstün ve Eryılmaz’a (2014) göre diğer etki büyüklüğü indeksleri, yani hem Cohen d hem de Glass Δ popülasyonun etki büyüklüğünün kestirilmesinde, özellikle küçük örneklerde hafif bir yanlılığa sahiptir.

Çalışmada istatistiksel model olarak ise rasgele etkiler modeli tercih edilmiştir. Bu modelin tercih edilmesinin nedeni ise yapılan heterojenlik testi sonucunda dağılımın homojen olmadığı belirlenmesidir. Ayrıca çalışmada istatistiksel analizlerin anlamlılık düzeyi .05 olarak belirlenmiştir. Çalışmada betimleyici istatistiklere ilişkin bulguların elde edilmesinde ise yüzde ve frekans değerleri kullanılmıştır.

BULGULAR

Tam Öğrenmeye İlişkin Yapılan Çalışmalara Ait Betimleyici İstatistikler

Çalışma karakteristiklerine ilişkin betimleyici istatistikler (çalışmanın yılı, çalışmanın yapıldığı öğrenim düzeyi, çalışmanın yapıldığı ders, çalışmanın yapıldığı alan ve çalışmanın yayın türü) aşağıdaki tablolarda (Tablo 1; Tablo 2; Tablo 3; Tablo 4 ve Tablo 5) verilmiştir:

Tablo 1. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Yıllara Ait Frekans ve Yüzde Değerleri

Çalışma Yılı	Frekans	Yüzde Değeri
1987	2	8,33
1991	1	4,17
1993	1	4,17
1994	1	4,17
1996	3	12,5
2000	1	4,17
2004	2	8,33
2005	7	29,17
2007	2	8,33
2010	1	4,17
2012	3	12,5

Tablo 1’de meta-analize dahil edilen çalışmaların yapıldıkları yıllara göre dağılımı incelendiğinde en fazla çalışmanın 7 çalışma (%29,17) ile 2005 yılında gerçekleştiği görülmektedir.

Tablo 2. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Öğrenim Düzeylerine Ait Frekans ve Yüzde Değerleri

Çalışmanın Yapıldığı Öğrenim Düzeyi	Frekans	Yüzde Değeri
İlköğretim	14	58,33
Lise	8	33,33
Üniversite	2	8,33

Tablo 2’de meta-analize dahil edilen çalışmaların yapıldıkları öğrenim düzeylerine ilişkin dağılım incelendiğinde çalışmaların büyük çoğunluğunun 14 çalışma (%58,33) ile ilköğretim düzeyinde gerçekleştiği görülmektedir. Tablo 2 incelendiğinde lise düzeyinde 8 (%33,33) ve üniversite düzeyinde de 2 (%8,33) çalışmanın yapıldığı görülmektedir.

Tablo 3. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Derslere Ait Frekans ve Yüzde Değerleri

Çalışmanın yapıldığı ders	Frekans	Yüzde Değeri
Matematik	8	33,33
Fen bilgisi	4	16,67
İngilizce	3	12,5
Biyoloji	1	4,17
Sosyal Bilgiler	1	4,17
Türkçe	1	4,17
Kimya	1	4,17
Fizik	1	4,17
Coğrafya	1	4,17
Müzik	1	4,17
Temel bilgi teknolojisi kullanımı	1	4,17
Öğretimde Planlama ve Değerlendirme	1	4,17

Tablo 3’de meta-analize dahil edilen çalışmaların yapıldıkları derslere ilişkin dağılım incelendiğinde en fazla çalışmanın 8 çalışma (%33,33) ile matematik dersinde gerçekleştiği görülmektedir. Tablo 3 incelendiğinde fen bilgisi dersine ait 4 (%16,67); İngilizce dersine ait 3 (12,50) ve diğer derslere ilişkin ise birer çalışma (%4,17) yapıldığı görülmektedir.

Tablo 4. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Alanlara Ait Frekans ve Yüzde Değerleri

Çalışmanın yapıldığı alan	Frekans	Yüzde Değeri
Sayısal	15	62,5
Sözel	3	12,5
Yabancı Dil	3	12,5
Diğer alanlar	3	12,5

Tablo 4’te meta-analize dahil edilen çalışmaların yapıldıkları alanlara ilişkin dağılım incelendiğinde en fazla çalışmanın 15 çalışma (%62,5) ile sayısal dersinde gerçekleştiği görülmektedir. Tablo 4 incelendiğinde sözel, yabancı dil ve diğer alanlarda 3’er çalışmanın (%12,50) yapıldığı görülmektedir.

Tablo 5. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Yayın Türlerine Ait Frekans ve Yüzde Değerleri

Çalışmanın yapıldığı öğrenim düzeyi	Frekans	Yüzde Değeri
Yüksek lisans tezi	5	20,83
Doktora tezi	2	8,33
Makale	16	66,67
Bildiri	1	4,17

Tablo 5’te meta-analize dahil edilen çalışmaların yayın türlerine ilişkin dağılım incelendiğinde en fazla çalışmanın, 16 çalışma ile makale (%66,67) türünde olduğu görülmektedir. Tablo 5 incelendiğinde, 5 çalışmanın (%20,83) yüksek lisans tezi, 2 çalışmanın (%8,33) doktora tezi ve 1 çalışmanın da (%4,17) bildiri türünde olduğu görülmektedir.

Meta-Analizden Elde Edilen Bulgular

Çalışmada tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirleyebilmek için meta-analize dahil edilen çalışmaların etki büyüklükleri hesaplanmıştır. Meta-analize dahil edilen 24 çalışmaya ait etki büyüklükleri Tablo 6’da verilmiştir.

Tablo 6. Meta-analize Dahil Edilen Çalışmaların Etki Büyüklükleri

Çalışma No	Çalışmanın Yazarı	Çalışmanın Yılı	Etki Büyüklüğü	Standart Hata	Varyans
1	Arslan	1996	1,485	0,237	0,056
2	Balta	2007	2,164	0,336	0,113
3	Çelik ve Şengül	2005	0,669	0,244	0,059
4	Demirci ve Yıldırım	1996	2,272	0,401	0,161
5	Hevedanlı, Oral ve Akbayın	2005	0,954	0,265	0,070
6	İşeri	2004	1,216	0,287	0,083
7	Karaca	2007	2,910	0,439	0,192
8	Kazu, Kazu ve Özdemir	2005	0,943	0,230	0,053
9	Kurtuldu ve Bakıoğlu	2012	2,911	0,629	0,395
10	Öner	2005	0,554	0,204	0,042
11	Özder	2000	1,540	0,328	0,108
12	Özer	2012	1,359	0,263	0,069
13	Savcı ve Kırkıç	2012	1,008	0,398	0,158
14	Sever	1993	1,817	0,304	0,092
15	Yıldırım ve Afreşa	1994	3,004	0,357	0,128
16	Yıldırım ve Aydın	2005	0,825	0,165	0,027
17	Yıldırım ve Eğinlioğlu	1987	1,026	0,281	0,079
18	Yıldırım ve Kırkıç	1996	3,084	0,536	0,287
19	Yıldırım ve Nwabueze	1991	1,001	0,273	0,075
20	Yıldırım ve Tuğal	2004	2,250	0,369	0,136
21	Yıldırım ve Sayar	1987	1,745	0,350	0,122
22	Yurteri	2005	2,591	0,360	0,129
23	Yüksel	2010	0,717	0,290	0,084
24	Zengin	2005	1,029	0,249	0,062

Tablo 6 incelendiğinde en büyük etki büyüklüğüne 18 no'lu çalışmanın (Yıldırım ve Kırkıç, 1996), en küçük etki büyüklüğüne ise 10 no'lu çalışmanın (Öner, 2005) sahip olduğu görülmektedir. Meta-analize dahil edilen çalışmaların etki büyüklüğü yönüne ait frekans ve yüzde değerleri ise Tablo 7'de verilmiştir.

Tablo 7. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Etki Büyüklükleri Yönüne Ait Frekans ve Yüzde Değerleri

Etki büyüklüğü yönü	Frekans	Yüzde (%)
0 (sıfır)	-	-
+ (pozitif)	24	100
- (negatif)	-	-

Tablo 7 incelendiğinde, çalışmaların tamamının + (pozitif) etki büyüklüğüne sahip olduğu görülmektedir. Bu bulgu da tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde pozitif yani olumlu bir etkiye sahip olduğunu göstermektedir.

Meta-analize dahil edilen çalışmaların etki büyüklükleri, Thalheimer ve Cook'un (2002) sınıflandırmasına göre sınıflandırıldığında oluşan frekans ve yüzde değerleri Tablo 8'de verilmiştir.

Tablo 8. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Etki Büyüklükleri, Thalheimer ve Cook'a (2002) Göre Sınıflandırıldığında Oluşan Frekans ve Yüzde Değerleri

Etki büyüklüğü düzeyi	Frekans	Yüzde (%)
Önemsiz	-	-
Küçük	-	-
Orta	3	12,5
Geniş	7	29,17
Çok geniş	2	8,33
Muazzam	12	50

Tablo 8 incelendiğinde meta-analize dahil edilen çalışmaların yarısının (%50) muazzam düzeyde, %29,17'sinin geniş düzeyde, %12,5'inin orta düzeyde ve %8,33'ünün ise çok geniş düzeyde etki büyüklüklerine sahip oldukları görülmektedir.

Meta-analize dahil edilen çalışmaların etki modeline göre ortalama etki büyüklüğü, güven aralıkları ve heterojen dağılım değeri Tablo 9'da verilmiştir.

Tablo 9. Etki Modeline Göre Ortalama Etki Büyüklüğü, Güven Aralıkları ve Heterojen Dağılım Değeri

Model	N	Serbestlik Derecesi	Hedges's g	Z değeri	%95 Güven Aralığı		Q-değeri	p
					Alt Sınır	Üst Sınır		
Sabit Etkiler Modeli	24	23	1,297	22,192	1,182	1,411	131,970	0,000
Rastgele Etkiler Modeli	24		1,546	10,709	1,263	1,829		

Tablo 9'da, sabit etkiler modeline göre etki büyüklüklerinin %95 güven aralığının üst sınırının 1,411 ve alt sınırının 1,182 olduğu görülmektedir. Yine Tablo 9'dan ortalama etki büyüklüğü değerinin ise 1,297 olduğu görülmektedir. Ancak bu meta-analiz çalışmasına sabit etkiler modelinin uygun olup olmadığını belirlemek için yapılan heterojenlik testi sonucu incelendiğinde, dağılımın homojen olmadığı sonucuna ulaşılmaktadır. Çünkü heterojenlik testi sonucunda elde edilen Q istatistiksel değeri (Q=131,970) %95 önem düzeyinde ki kare tablosundan elde edilen değeri ($X^2_{(0,95)}=35,172$) aşmaktadır. Bu yüzden de çalışmada sabit etkiler modeli yerine rastgele etkiler modeli tercih edilmiştir.

Tablo 9'da rastgele etkiler modeline göre ise etki büyüklüklerinin %95 güven aralığının üst sınırının 1,829, alt sınırının 1,263 ve ortalama etki büyüklüğü değerinin de 1,546 olduğu görülmektedir. Ortalama etki büyüklüğü değerinin (+) değerde olması, yani pozitif yönde olması, tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisinin olumlu yönde olduğunu göstermektedir. Analizden elde edilen ortalama etki büyüklüğü değeri, Thalheimer ve Cook'a (2002) göre ise muazzam düzeydedir. Bir başka deyişle ortalama etki büyüklüğü değeri Thalheimer ve Cook'un (2002) sınıflandırmasına göre yorumlandığında, tam öğrenme yaklaşımının, öğrencilerin akademik başarılarını artırma açısından oldukça yüksek düzeyde bir etkiye sahip olduğu söylenebilir.

Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde oldukça yüksek düzeyde (Hedge's $g=1,546$) bir etkiye sahip olduğunun geçerli bir bulgu olup olmadığını belirlemek amaçlanmıştır. Bunun için meta-analiz çalışmasının gücünü belirlemede kullanılan "Classic Fail-Safe N" analizinden yararlanılmıştır. "Classic Fail-Safe N" analizi yapılan meta-analiz çalışmasında yayın yanlılığı olup olmadığına ilişkin bilgi vermektedir. Long'a (2001) göre yayın yanlılığı meta-analiz bulgularının geçerliliğini etkileyen önemli bir sorundur. Yayın yanlılığı, meta-analize dahil edilen

çalışmaların genellikle yayınlanmış çalışmalardan oluşması ve yalnızca anlamlı sonuç veren çalışmaların yayınlanıyor olması varsayımına dayanmaktadır. “Classic Fail-Safe N” analizi ile meta-analiz sonuçlarını anlamsız hale getirecek kaç çalışmaya ihtiyaç duyulduğu hesaplanarak meta-analiz çalışmasında yayın yanlılığı olup olmadığı belirlenmektedir. Meta-analiz sonuçlarını anlamsız hale getirecek çalışma sayısı ne kadar fazla olursa meta-analiz sonuçları da o kadar güvenilir olmaktadır (Long, 2001). “Classic Fail-Safe N” analizi sonucunda elde edilen veriler, Tablo 10’da verilmiştir.

Tablo 10. Yayın Yanlılığına İlişkin Yapılan Classic Fail-Safe N Analizi Sonuçları

Z-değeri	23,83382
p-değeri	0,000
Alfa değeri	0,050
Alfa değeri için Z-değeri	1,95996
N	24
p>alpha sonucu için gerekli eksik çalışma sayısı	3525

Tablo 10’da Classic Fail-Safe N analizinden elde edilen verilere göre yapılan meta-analiz çalışmasının sonucunu geçersiz kılmak için tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde hiçbir etkisinin olmadığı (nötr) ya da olumsuz yönde etkisinin olduğu (negatif) bulgularına ulaşılmış 3525 çalışmaya daha ihtiyaç olduğu görülmektedir ($p<.05$). Meta-analiz çalışmasının 24 çalışma ile yapıldığı düşünüldüğünde 3525 sayısının oldukça yüksek bir sayı olduğu söylenebilir. Bu sayının yüksek olması ise meta-analiz sonuçlarının güvenilir olduğunu yani tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerinde oldukça yüksek düzeyde bir etkiye sahip olduğu bulgusunun geçerli bir bulgu olduğunu göstermektedir.

Tam Öğrenme Yaklaşımının Öğrencilerin Akademik Başarıları Üzerindeki Etkisinin Alanlara Göre Karşılaştırılması

Tam öğrenme yaklaşımının uygulandığı derslerin toplam etki büyüklüğü üzerindeki etkisini belirlemek amacıyla dersler 4 farklı alana ayrılmıştır. “Fen Bilgisi, Fizik, Kimya, Biyoloji, Matematik” dersleri “sayısal alan”; “Türkçe, Sosyal Bilgiler, Coğrafya” dersleri “sözel alan” ve “İngilizce” dersi ise “yabancı dil” alanı içerisinde yer almıştır. “Müzik, Öğretimde Planlama ve Değerlendirme ve Temel Bilgi Teknolojisi Kullanımı” dersleri ise belirlenen alanlara uymadığı ve her biri farklı alanda birer ders olduğu için “diğer alanlar” başlığı altında toplanmıştır. Tablo 11’de her bir alana ait etki büyüklükleri verilmiştir.

Tablo 11. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Alanlara Göre Etki Büyüklükleri

Model	N	Hedges's g	%95 Güven Aralığı		Heterojenlik testi	
			Alt Sınır	Üst Sınır	Q-değeri	p
Rastgele Etkiler Modeli						
Sayısal	15	1,404	1,072	1,737		
Sözel	3	2,388	1,730	3,046		
Yabancı Dil	3	1,644	0,922	2,365		
Diğer Alanlar	3	1,322	0,446	2,199		
Ara Toplam					7,223	0,065

Tablo 11’de rastgele etkiler modeline göre tam öğrenme yaklaşımının uygulandığı sayısal alanda ortalama etki büyüklüğünün 1,404; sözel alanda 2,388; yabancı dil alanında 1,644 ve diğer alanlarda ise 1,322 olduğu görülmektedir. Tam öğrenme yaklaşımının uygulandığı alanlara göre çalışmalardan elde edilen ortalama etki büyüklükleri arasında anlamlı bir farklılık olup olmadığını belirlemek için yapılan heterojenlik testi sonucuna göre akademik başarı açısından çalışmaların ortalama etki büyüklükleri arasında anlamlı bir fark bulunmamaktadır ($Q=7,223$; $p>.05$). Yani çalışmaların ortalama etki büyüklükleri arasında homojenlik vardır ve farklı alanlarda uygulanan tam öğrenme yaklaşımı

sonucunda öğrencilerin ulaştıkları akademik başarı düzeyleri arasında anlamlı bir farklılık bulunmamaktadır.

Tam Öğrenme Yaklaşımının Öğrencilerin Akademik Başarıları Üzerindeki Etkisinin Öğrenim Düzeyine Göre Karşılaştırılması

Öğrencilerin öğrenim düzeylerinin, tam öğrenme yaklaşımının toplam etki büyüklüğü üzerindeki etkisini belirlemek amacıyla öğrencilerin öğrenim düzeyleri, “ilköğretim”, “lise” ve “üniversite” olmak üzere 3 farklı gruba ayrılmıştır. Tablo 12’de her bir öğrenim düzeyine ait etki büyüklükleri verilmiştir.

Tablo 12. Tam Öğrenmeye İlişkin Yapılan Çalışmaların Öğrenim Düzeylerine Göre Etki Büyüklükleri

Model	N	Hedges's g	%95 Güven Aralığı		Heterojenlik testi	
			Alt Sınır	Üst Sınır	Q-değeri	p
Rastgele Etkiler Modeli						
İlköğretim	14	1,502	1,141	1,863		
Lise	8	1,863	1,287	2,440		
Üniversite	2	0,856	0,502	1,210		
Ara Toplam					10,912	0,004

Tablo 12’de rastgele etkiler modeline göre tam öğrenme yaklaşımının uygulandığı “ilköğretim” düzeyinde etki büyüklüğünün 1,502; “lise düzeyinde” 1,863 ve “üniversite” düzeyinde 0,856 olduğu görülmektedir. Tam öğrenme yaklaşımının uygulandığı öğrenim düzeylerine göre çalışmalardan elde edilen ortalama etki büyüklükleri arasında anlamlı bir farklılık olup olmadığını belirlemek için yapılan heterojenlik testi sonucuna göre akademik başarı açısından çalışmaların ortalama etki büyüklükleri arasında anlamlı bir fark bulunmamaktadır (Q=10,912; p<.05). Yani tam öğrenme yaklaşımının öğrencilerin akademik başarısı üzerindeki etkisi, öğrencilerin öğrenim düzeylerine göre anlamlı bir farklılık göstermektedir.

Tablo 12’de tam öğrenme yaklaşımının öğrencilerin akademik başarısı üzerinde en çok “lise” düzeyinde (Hedge’s g=1,863) etkili olduğu görülmektedir. Daha sonra ise “ilköğretim” düzeyi (Hedge’s g=1,502) gelmektedir. “Üniversite” düzeyinde (Hedge’s g=0,856) ise tam öğrenme modelinin diğer düzeylere göre daha az etkili olduğu görülmektedir. Etki büyüklükleri, Thalheimer ve Cook’un (2002) sınıflandırmasına göre yorumlandığında tam öğrenme yaklaşımının, “ilköğretim” ve “lise” düzeyinde, öğrencilerin akademik başarıları üzerinde muazzam düzeyde; “üniversite” düzeyinde ise geniş düzeyde etkili olduğu söylenebilir.

SONUÇLAR ve TARTIŞMA

Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini belirlemek amacıyla yapılan meta-analiz çalışmasında toplam 24 çalışmanın verilerinden yararlanılmıştır. Bu 24 çalışmanın çoğunluğu makale (%66,67) türündedir. Ayrıca çalışmalar içerisinde 5 yüksek lisans tezi (%20,83), 2 doktora tezi (%8,33) ve 1 de (%4,17) bildiri yer almaktadır. Meta-analiz çalışmasına dahil edilen bu 24 çalışma yapıldıkları yıllara göre incelendiğinde en fazla çalışmanın (%29,17) 2005 yılında yapıldığı belirlenmiştir. Tam öğrenme konusundaki çalışmaların 2005 yılında yoğunlaşmasının nedeni, 2005 yılında uygulanmaya başlanan ilköğretim programları üzerinde tam öğrenme yaklaşımının etkililiğinin incelenmesi olarak düşünülebilir. Çalışmalar yapıldıkları öğrenim düzeylerine göre incelendiğinde ise en fazla çalışmanın ilköğretim düzeyinde (%58,33), en az çalışmanın ise üniversite düzeyinde (%8,33) gerçekleştiği görülmüştür. Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisini araştıran bu 24 çalışma en fazla matematik dersinde (%33,33) ve yine aynı şekilde en fazla sayısal alanda (%62,5) gerçekleştirilmiştir. Meta-analize dahil edilen çalışmaların ağırlıklı olarak (%33,33) matematik alanında yapılmasının, matematik dersinin yüksek düzeyde aşamalılık ilişkisi göstermesinden kaynaklandığı düşünülebilir (Özçelik, 1981). Tam öğrenme yaklaşımının yüksek aşamalılık ilişkisi gösteren derslerde daha farklı sonuçlar gösterdiği bilinmektedir (Senemoğlu, 1987).

Meta-analize dahil edilen çalışmaların tamamı pozitif etki büyüklüğü yönüne sahiptir. Bu da tam öğrenme yaklaşımının öğrencilerin akademik başarılarını artırdığını yani olumlu yönde etkilediğini

göstermektedir. Meta-analize dahil edilen çalışmalar için ortalama etki büyüklüğü 1,546 olarak hesaplanmıştır. Bulunan etki büyüklüğü değeri, Thalheimer ve Cook'un (2002) sınıflandırmasına göre muazzam düzeyde bir etkiye sahiptir. Bu değere göre tam öğrenme yaklaşımı öğrencilerin akademik başarılarını artırmada oldukça etkilidir. Bir başka deyişle tam öğrenme yaklaşımının uygulandığı sınıflarda öğrencilerin akademik başarıları önemli ölçüde artmaktadır. Meta-analize dahil edilen çalışmalara ait etki büyüklükleri; Thalheimer ve Cook'a (2002) göre sınıflandırıldığında, çalışmaların yarısı (%50) muazzam düzeyde, %29,17'si geniş düzeyde, %12,5'i orta düzeyde ve %8,33'ü ise çok geniş düzeyde yer almaktadır. Bu yüzden de meta-analize dahil edilen çalışmaların büyük çoğunluğunun öğrencilerin akademik başarılarını artırma açısından yüksek düzeyde etkili oldukları söylenebilir. Bu sonuç, Kulik, Kulik ve Bangert-Drowns'un (1990) tam öğrenmenin öğrencilerin akademik başarılarını olumlu yönde etkilediğini ortaya koydukları meta-analiz çalışması ile tutarlık göstermektedir. Tam öğrenme, yeterli zaman ve ilgi ile tüm öğrencilerin öğrenebileceğini ifade eden ve buna yönelik uygulamaları ortaya koyan bir yaklaşımdır (Bloom, 1976). Buna göre etkili bir şekilde uygulandığında tam öğrenme yaklaşımının öğrencilerin öğrenme düzeyini yükselterek akademik başarılarını arttırması, beklenen bir durum olarak kabul edilebilir. Alan yazında da tam öğrenme yaklaşımının akademik başarıyı olumlu olarak etkilediğini, öğrenme düzeyini arttırdığını ortaya koyan araştırma bulguları olduğu görülmektedir (Senemoğlu, 1988; Miles, 2010; Block ve Burns, 1976; Mevarech, 2015).

Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi, yaklaşımın uygulandığı ders alanlarına göre anlamlı bir farklılık göstermemektedir. Tam öğrenme yaklaşımının sayısal, sözel ya da yabancı dil alanında uygulanması öğrencilerin akademik başarılarını anlamlı bir şekilde etkilememektedir. Yani tam öğrenme yaklaşımının uygulandığı alanlar arasında öğrencilerin akademik başarıları açısından bir farklılık olmadığı belirlenmiştir.

Tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkisi, öğrencilerin öğrenim düzeylerine göre anlamlı bir farklılık göstermektedir. Thalheimer ve Cook'un (2002) sınıflandırmasına göre tam öğrenme yaklaşımı; ilköğretim ve lise düzeyinde, öğrencilerin akademik başarıları üzerinde muazzam düzeyde, üniversite düzeyinde ise geniş düzeyde etkilidir. Tam öğrenme yaklaşımı, en fazla lise düzeyinde (Hedge's $g=1,863$) yapılan uygulamalarda öğrencilerin akademik başarıları üzerinde etkili olmaktadır. Tam öğrenme yaklaşımının lise düzeyinden sonra en fazla etkili olduğu öğrenim düzeyi ise ilköğretimdir (Hedge's $g=1,502$). Tam öğrenme yaklaşımı, üniversite düzeyinde (Hedge's $g=0,856$) ise öğrencilerin akademik başarılarını artırma açısından diğer alanlara göre daha az etkilidir. Bu durum, tam öğrenme yaklaşımının lise ve ilköğretim düzeyine üniversite düzeyinden daha uygun olduğu anlamına gelebilir. Ancak meta-analize dahil edilen çalışmaların uygulandıkları öğrenim düzeylerine göre dağılımı incelendiğinde ilköğretim ve lise düzeyinde 22 çalışma yer alırken üniversite düzeyinde ise 2 çalışmanın yer aldığı görülmektedir. Bu yüzden de tam öğrenme yaklaşımının öğrencilerin akademik başarıları üzerindeki etkililiğinin üniversitelerde daha fazla araştırılması gerektiği söylenebilir.

Yapılan meta-analiz çalışması, tam öğrenme yaklaşımının öğrencilerin sadece akademik başarıları üzerindeki etkisini belirlemeye yöneliktir. Tam öğrenme yaklaşımının tutum, kalıcılık vb. değişkenler üzerindeki etkilerine yönelik de meta-analiz çalışmaları yapılabilir.

KAYNAKÇA

- Arslan, M. (1996). *Altı çizili materyalle çalışma ve tam öğrenme yönteminin öğrenme düzeyine, hatırlamaya ve akademik benlik kavramına etkisi*. Yayımlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Balta, Y.A.N (2007). *İlköğretim okullarında uygulanan sınavlarda tam öğrenmenin (bloom taksonominin) kullanılmasının önemi*. Yayımlanmamış Yüksek Lisans Tezi, Yeditepe Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Block, J. H. ve Burns, R. B. (1976). Mastery learning. *Review of Research in Education*. 4, 3-49.
- Bloom, B. S. (1976). *Human characteristics and school learning*; Özçelik, D. A. (Çeviren), (2012). *İnsan nitelikleri ve okulda öğrenme*. Ankara: Pegem Akademi.

- Çelik, N. G. ve Şengül, S. (2005). Tam öğrenme yönteminin ilköğretim 6. sınıf matematik öğrencilerinin akademik başarıları ile kalıcılık düzeylerine etkisi. *Uludağ Eğitim Fakültesi Dergisi*, XVIII(1), 107-122.
- Decoster, J. (2004). *Meta-analysis notes*. Retrieved from <http://www.stat-help.com/notes.html>.
- Demirci, G. Y. ve Yıldırım, G. (1996). The effects of mastery learning method of instruction and a particular conceptual change strategy on achievement and misconception levels of eighth grade science students. *Boğaziçi university journal: Educational Sciences*, XVI, 113-141.
- Dinçer, S. (2014). *Eğitim bilimlerinde uygulamalı meta-analiz*. Ankara: Pegem Akademi.
- Erden, M. ve Akman, Y. (1996). *Eğitim psikolojisi*. Ankara: Arkadaş Yayınevi.
- Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Hevedanlı, M., Oral, B. ve Akbayın, H. (2005). Biyoloji öğretiminde işbirlikli öğrenme ve tam öğrenme yöntemleriyle geleneksel öğretim yöntemlerinin öğrenci başarısına etkisi. *Milli Eğitim Dergisi*, 166, 234-246.
- İşeri, H. (2004). *İlköğretim matematik öğretiminde tam öğrenme modeli*. Yayımlanmamış Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, İzmir.
- Karaca, A. (2007). *Sosyal bilgiler öğretiminde tam öğrenme modelinin uygulanabilirliği*. Yayımlanmamış Yüksek Lisans Tezi, Fırat Üniversitesi, Sosyal Bilimler Enstitüsü, Elazığ.
- Kazu, İ. Y., Kazu, H. ve Özdemir, O. (2005). The effects of mastery learning model on the success of the students who attended "usage of basic information technologies" course. *Educational Technology & Society*, 8(4), 233-243.
- Kulik, C. C. L., Kulik, J. A. ve Drowns, R. L. B (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research Summer*, 60(2), 265-299.
- Kurtuldu, M. K. ve Bakıoğlu, Ç. (2012). Tam öğrenme modeline dayalı müzik öğretiminde öğrenci başarılarının karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43, 330-339.
- Lipsey, M ve Wilson, D. (2001). *Practical meta-analysis*. Londra: Sage Publication.
- Long, J. (2001). *An Introduction to and generalization of the fail-safe n*. Annual meeting of the southwest educational research association. New Orleans, LA.
- Mevarech, Z. R. (2015). The effects of cooperative mastery learning strategies on mathematics achievement. *The Journal of Educational Research*, 78 (6), 372-377.
- Miles, K. (2010). *Mastery learning and academic achievement*. Unpublished Doctoral Thesis. Minnesota: Walden University, Middle School Education.
- Öner, M. (2005). *Tam öğrenme destekli çoklu zeka kuramı uygulamalarının fen bilgisi dersindeki erişimi, tutum ve kalıcılığa etkisi*. Yayımlanmamış Yüksek Lisans Tezi, Dicle Üniversitesi, Sosyal Bilimler Enstitüsü, Diyarbakır.
- Özçelik, D. A. (1981). *Okullarda ölçme ve değerlendirme*. Ankara: ÜSYM Eğitim Yayınları.
- Özder, H. (2000). Tam öğrenmeye dayalı işbirlikli öğrenme modelinin etkililiği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 19, 114 – 121.
- Özdemirli, G. (2011). *İşbirlikli öğrenme yönteminin öğrencinin matematik başarısı ve matematiğe ilişkin tutumu üzerindeki etkililiği: Bir meta-analiz çalışması*. Yayımlanmamış Yüksek Lisans Tezi, Çukurova Üniversitesi, Sosyal Bilimler Enstitüsü, Adana.
- Özer, Ö. (2012). *İlköğretim beşinci sınıf öğrencilerinin tam öğrenme modeline göre matematik dersindeki erişim düzeyleri ile öğrenme stratejileri arasındaki ilişki*: 202-215. IV. Uluslararası Türkiye eğitim araştırmaları kongresi, İstanbul.
- Sağlam, M ve Yüksel İ. (2007). Program değerlendirmede meta-analiz ve meta değerlendirme yöntemleri. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 18, 175-187.
- Savcı, H. ve Kırkıç, K. A. (2012). Tam öğrenme yönteminin kavram haritalarıyla birlikte kullanıldığında kimya öğrencilerinin başarı düzeylerine etkisi. *İstanbul Sabahattin Zaim Üniversitesi Sosyal Bilimler Dergisi*, 1, 194-226.
- Senemoğlu, N. (1987). Tam öğrenme-avantajları ve sınırlılıkları. *Eğitim ve Bilim*, 12 (66), 28-34.
- Senemoğlu, N. (1988). Öğrenme düzeyini yükseltme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 3, 105-115.
- Sever, S. (1993). *Türkçe öğretiminde uygulanan tam öğrenme kuramı ilkelerinin, öğrencilerin okuduğunu anlama ve yazılı anlatım becerilerindeki erişime etkisi*. Yayımlanmamış Doktora Tezi, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Şahin, F. (1999). *Meta analiz in tıpta kullanımı ve bir uygulama*. Yayımlanmamış Doktora Tezi, Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
- Senemoğlu, N. (2013). *Gelişim, öğrenme ve öğretim*. Ankara: Yargı Yayınevi.
- Thalheimer, W. ve Cook, S. (2002). *How to calculate effect size from published research: a simplified spreadsheet*. Erişim Ağustos 14 2014, http://www.work-learning.com/white_papers/effect_sizes/Effect_Sizes_Spreadsheet.xls.

- Üstün, U. ve Eryılmaz, A. (2014). Etkili araştırma sentezleri yapabilmek için bir araştırma yöntemi: meta-analiz. *Eğitim ve Bilim Dergisi*, 39(174), 1-32.
- Yıldıran, G. ve Eğinlioğlu, U. (1987). The effects of mastery learning and improved materials on English achievement levels of Turkish students. *Boğaziçi University Journal: Educational Sciences*, XII, 73-85.
- Yıldıran, G. ve Sayar, Y. (1987). Tam öğrenme yöntemiyle birlikte öğrencilerin bilişsel giriş davranışlarına sahip olmalarının başarı üzerindeki etkileri. *Psikoloji Dergisi*, 6(21), 67-72.
- Yıldıran, G. ve Nwabueze, B. (1991). The effects of mastery learning and improved teaching on mathematics achievement of seventh grade Turkish students. *Boğaziçi University Journal: Educational Sciences*, XIV, 111-123.
- Yıldıran, G. ve Afreşa, O. (1994). Tam öğrenme ve geleneksel öğretim yöntemlerinin beşinci sınıf öğrencilerinin fen bilgisi dersi başarı ve hatırlama düzeyine etkileri. *Boğaziçi University Journal: Educational Sciences*, XV, 37-48.
- Yıldıran, G. ve Kırkıç, A. K. (1996). The effects of mastery learning method of instruction and laboratory experiments on achievement levels and science misconception scores of secondary school Turkish students. *Boğaziçi University Journal: Educational Sciences*, XVI, 95-112.
- Yıldıran, G. ve Tuğal, İ. (2004). The Effects of grouping and mastery learning method of instruction on english achievement levels of lycée one students in kuleli military high school. *Boğaziçi University Journal of Education*, 21(1), 1-33.
- Yıldıran, G. ve Aydın, E. (2005). The effects of mastery learning and cooperative, competitive and individualistic learning environment organizations on achievement and attitudes in mathematics. *Journal of Research in Mathematical Education*, 9(1), 69-96.
- Yüksel, S. (2010). Öğrencilerin ölçme ve değerlendirme konularındaki başarılarında tam öğrenme yaklaşımı ve aktif öğretim yöntemlerinin etki derecelerinin incelenmesi. *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, 26, 1-11.
- Yurteri, E. (2005). *Coğrafya öğretiminde tam öğrenme modelinin öğrenci başarısına etkisi*. Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Zengin, N. (2005). *Tam öğrenme ilkeleri doğrultusunda farklı öğretim yöntemleri ile işlenen matematik dersinin ilköğretim 7. sınıf öğrencilerinin matematik başarı düzeylerine etkisi*. Yayımlanmamış Yüksek Lisans Tezi, Marmara Üniversitesi, Eğitim Bilimleri Enstitüsü, İstanbul.

EXTENDED ABSTRACT

Introduction

Mastery learning approach is based on the assumption that almost all students can learn all the new behaviors schools aim to teach. In addition, mastery learning can provide a high level of learning with enough time and patience (Bloom, 1976). This study is considered as important because it presents synthesis of studies about the effects of mastery learning approach on students' academic achievement in terms of finding comprehensive ideas.

The reason of conducting a meta-analysis on mastery learning is that there is limited research focusing on the effects of mastery learning on academic achievement of students in Turkey. Even, when the international literature of this topic is examined, a few studies are found. For example, one of these studies conducted by Kulik, Kulik and Bangert-Drowns (1990).

The present study aims to find answers to the following questions:

- What is the effect of mastery learning approach on academic achievement of students?
- Is the effect of mastery learning model on academic achievement of students statistically significant in terms of subject matter?
- Is the effect of mastery learning model on academic achievement of students statistically significant in terms of education level?

Method

In order to decide mastery learning model's effect on academic achievement of students, studies focusing on mastery learning approach in the literature was investigated. In this research, the findings of the studies focusing on mastery learning approach is combined by using meta-analysis method to decide mastery learning model's effect on academic achievement of students.

According to Glass (1976), meta-analysis is a statistical analysis that is used for merging the findings. In the light of this idea, this meta-analysis study contains 16 articles, 7 thesis and 1 oral presentation. In total, 24 experimental studies are included in the study.

"Comprehensive Meta-Analysis" program is used to calculate effect sizes of each study and the combined effect size. In this research, "Hedge's g " is used to calculate effect size. Percentage and frequency values are given to show descriptive statistics.

Results and Discussion

The aim of the present study was to determine the effect of mastery learning approach on academic achievement of students. For this purpose, when the studies were investigated in terms of the years they are conducted, it was found that most of them were done in 2005 (29, 17%) and they were conducted at primary school level (58, 33%). Moreover, while few research was carried out at university level (8, 33%), the researchers did mostly quantitative studies (62, 5%), and they mostly focused on math (33, 33%).

All the studies were integrated with meta-analysis had positive (+) effect size direction. This demonstrated that mastery learning increase academic achievement of the students, and it influenced on success positively. This finding was consistent with the result of Kulik, Kulik and Bangert-Drowns' meta-analysis study (1990) in terms of positive effect of mastery learning on academic achievement of students.

When the effect sizes of the studies were compared according to the classification of Thalheimer and Cook (2002), half of the studies (50%) were considered huge, 29, 17% were large, 12, 5% were medium and 8, 33% of them were at small level. Therefore, most of the studies were integrated with meta-analysis could be taken into consideration as effective in terms of increasing the academic achievement of the students.

Average effect size of the studies were calculated as 1,546. This effect size value has huge according to the classification of Thalheimer and Cook (2002). According to this value, mastery learning approach was considerably effective in terms of increasing the academic achievement of students. In other words, mastery learning increased statistically significant academic achievement of students in the classes that it was applied.

There was no significant difference on the effect of mastery learning on academic achievement of students in terms of subject matters. Application of mastery learning in quantitative, verbal or foreign language fields did not affect the academic achievement of students in a meaningful way. It can be said that there was no significant difference in terms of students' academic achievement between the fields.

Another finding of the present study was that there is a significant difference on the effect of mastery learning on academic achievement in terms of students' education level. According to the classification of Thalheimer and Cook (2002) mastery learning approach has a huge effect on academic success of students at the primary and high school level, and it has a large effect at the university level.

The application of mastery learning approach was maximum effective on academic achievement of students at high school level (hedge's $g=1,863$). Mastery learning approach had medium effective at primary school (hedge's $g=1,502$). It was less effective at university level in terms of the academic achievement (hedge's $g=0,856$). This might mean mastery learning was more appropriate for high

school and primary school level. However, when the studies were investigated in terms of education level, it was found that 22 studies were conducted at primary and high school level, and only 2 studies were conducted at university level. Because of this finding, it is suggested that mastery learning's effect on academic achievement at university level might be new research area in future studies. Moreover, the present meta-analysis study was conducted for determining the effect of mastery learning approach on students' academic achievement. New meta-analysis studies might be done on the effect of mastery learning on the variables as attitude, and permanence.

Eğitimde Ölçme ve Değerlendirme Dersinde Başarıyı Etkileyen Faktörlerin Çoklu Göstergeler Çoklu Nedenler Modeliyle İncelenmesi*

The Investigation of the Factors Affecting Achievement in Measurement and Evaluation Course with Multiple Indicators Multiple Causes Model

Kemal KURŞUN ** Derya ÇOBANOĞLU AKTAN ***

Öz

Bu çalışmanın amacı, öğretmen adaylarının ölçme ve değerlendirme dersindeki başarılarına etki eden faktörlerin çoklu göstergeler çoklu nedenler modeli kullanılarak incelenmesi ve bu faktörlerin başarı üzerindeki etki derecelerinin ortaya çıkarılmasıdır. Çalışmanın verilerini, ölçme ve değerlendirme dersi almış 443 öğretmen adayının, bölümleri, ebeveyn eğitim durumu ve aylık gelirleri, matematik başarı ve kaygı düzeyleri, genel akademik başarı notları, anadolu öğretmen lisesinden mezun olma durumları, eğitimde ölçme ve değerlendirme dersine karşı tutumları, ders çalışmaya ayırdıkları zaman ile ölçme ve değerlendirme ders başarılarıyla ilgili sorulara verdikleri cevaplar oluşturmuştur. Veriler ile ölçme modelinin doğrulanmasından sonra yapısal modelde bağımsız gösterge değişkenler ile gizil değişkenler arasındaki ilişkiler çoklu göstergeler ve çoklu nedenler modeli (ÇGÇN) ile incelenmiştir. Araştırma sonuçlarına göre öğretmen adaylarının eğitimde ölçme ve değerlendirme ders başarılarını, derse karşı tutumları, genel akademik başarıları ve anadolu öğretmen lisesinden mezun olma durumları yordamaktadır. Sosyoekonomik düzeyin, öğretmen adaylarının genel akademik, matematik ve ölçme değerlendirme ders başarılarına manidar bir etkisi yoktur. Duyuşsal özelliklerden matematik kaygısının matematik başarısını negatif yönde, genel akademik başarısını ise pozitif yönde manidar olarak yordadığı sonucuna ulaşılmıştır.

Anahtar Kelimeler: yapısal eşitlik modeli, çoklu göstergeler ve çoklu nedenler (ÇGÇN), eğitimde ölçme ve değerlendirme ders başarısı.

Abstract

The aim of this study is to investigate the factors affecting teacher candidates' measurement and evaluation course achievement with multiple indicators multiple causes modeling and to explore the effects of this factors on the achievement. The data in the study was obtained with a measurement instrument prepared for this purpose. Questions about departments, parent education level and income, mathematics achievement level and anxiety, academic grade, type of graduated high school, attitude toward measurement and evaluation course, achievement of measurement and evaluation course and the time allocated to the study were asked to teacher candidates. After analyzing fit indexes of the measurement model, the relations between the independent indicator variables and latent variables were investigated by multiple indicators and multiple causes model. According to the results general academic achievement grade, attitudes toward measurement and evaluation course and type of high graduated school have predicted the achievement level of teacher candidates' in measurement and evaluation course. Socioeconomic level does not have significant impact on general academic achievement grade, mathematics achievement level and achievement of assessment and evaluation course of teacher candidates. It has been concluded that mathematics anxiety variable from affective characteristics significantly has predicted mathematics achievement negatively, and general academic achievement positively.

* Bu makale, birinci yazar tarafından ikinci yazar danışmanlığında hazırlanan yüksek lisans tez çalışmasından türetilmiştir.

** Kemal KURŞUN, kursun_kemal@hotmail.com

*** Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN, Hacettepe Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü, Ankara- TÜRKİYE, coderya@gmail.com

Keywords: structural equation modeling, multiple indicators and multiple causes (MIMIC), measurement and evaluation course achievement.

GİRİŞ

Eğitim ve öğretim faaliyetlerinin başarılı olması için öğretmenlerin belirli alanlarda uzmanlaşma, mesleki bilgi birikimi ve yeterliklere sahip olması gerekmektedir. Bu yeterliliklerden birisi de ölçme ve değerlendirmedir. Ölçme ve değerlendirme eğitimi, öğretmen adaylarına lisans eğitimlerinde ölçme ve değerlendirme dersleri kapsamında verilmekte ve bu becerileri kazanmaları için öğretmen adaylarının dersi başarılı bir biçimde tamamlamaları önem kazanmaktadır.

Yapılan araştırmalar öğrencilerin derslerde başarılı olmasını önemli ölçüde etkileyen bilişsel, duyuşsal, toplumsal ve fizyolojik çok sayıda faktör olduğunu ve bu faktörlere bağlı olarak öğrenci başarısının arttığını ya da azaldığını göstermektedir (Kalender, 2010; Poyraz, 2012; Yenilmez ve Duman, 2008; Yavuz, 2009). Bu kapsamda ölçme ve değerlendirme dersinde başarılı olması gereken öğretmen adaylarının başarı düzeylerini etkileyen faktörlerin tespit edilmesi de önemli bir problem sahası olarak ortaya çıkmaktadır.

Eğitimde ölçme ve değerlendirme alanında yapılan araştırmaların, çoğunlukla öğretmenlerin ölçme ve değerlendirmeye yönelik yeterlilikleri konusunda ve derse karşı tutumla ilgili çalışmalar olduğu belirlenmiştir (Alkharusi, Kazem ve Al Musawai, 201; Erdoğan, 2010; Süral, 2014; Yaşar, 2014). Öte yandan genel akademik başarıyla ilgili çalışmalar incelendiğinde, öğrencilerin bölümü, cinsiyeti, sosyoekonomik düzeyi, öz-yeterlik, kaygı ve tutumları gibi değişkenlerin genel akademik başarı üzerinde etkisinin olduğu (Poyraz, 2012; Kart, 2012; Süral, 2014) görülmüştür. Matematik başarısında ise, anne-baba eğitim düzeyi, aylık gelir durumu, okul türü ve öğretmen cinsiyeti değişkenleri ile genel akademik başarı, ders kaygısı ve derse karşı tutumun etkili olduğu belirlenmiştir (Erden ve Akgül, 2010; Kalender, 2010; Peker ve Şentürk, 2012; Reçber, 2011; Yenilmez ve Duman, 2008). Ayrıca ders çalışmaya ayrılan zamanın ders başarısını olumlu yönde etkilediği (Özer ve Anıl, 2011; Savaş, Taş ve Duru, 2010) ve yüksek sosyoekonomik düzeyin öğrencilerin fen ve matematik dersi ile genel akademik başarısında etkisinin olduğu görülmüştür (Kalender, 2010; Savaş, Taş ve Duru, 2010; Yavuz, 2009).

Eğitimde ölçme ve değerlendirme özeline bakıldığında öğrencilerin derse yönelik tutumları, bölüm, matematiğe karşı tutum ve matematik dersini alıp almamanın etkilediği (Süral, 2014; Yaşar, 2014), öğrencilerin ölçme ve değerlendirme ders başarılarını baba eğitim durumu ve cinsiyetin yordadığı, tutum ve öz yeterlik algıları gibi duyuşsal özelliklerin manidar olarak yordamadığı görülmüştür (Kart, 2012).

Araştırmanın Amacı

Bu çalışmanın amacı, öğretmen adaylarının eğitimde ölçme ve değerlendirme ders başarısı ile bazı demografik ve duyuşsal değişkenler arasındaki ilişkilerin ortaya çıkarılmasıdır. Çalışmada ele alınan değişkenler ve kurulan model yukarıda bahsedilen literatür ışığında önceki çalışmaların sonuçlarına göre seçilmiş ve kurulmuştur. Yüksek Öğretim Kurumu (YÖK) ders kur tanımına göre güvenirlilik, korelasyon, madde ve test istatistikleri ve benzeri konuları öğrenmede matematik bilgisi gerekli olduğu için çalışma kapsamında öğrencilerin matematik başarıları ve matematik kaygıları ele alınmıştır. Ayrıca Anadolu öğretmen lisesi mezunları, lise eğitimlerinde ölçme ve değerlendirme dersi almaktadırlar. Bu yüzden, çalışmada ölçme ve değerlendirme ders başarılarını etkileyen değişkenler arasına Anadolu öğretmen lisesi mezunu olup olmama durumu dahil edilmiştir. İlgili araştırmalar incelendiğinde matematik kaygısı, matematik başarısı ve genel akademik başarının ölçme ve değerlendirme ders başarısı üzerinde, matematik kaygısı ve başarısının ölçme ve değerlendirme dersine karşı tutumla ilişkisi olup olmadığını inceleyen bir çalışma olmadığı görülmüştür. Bu araştırmada diğer çalışmalardan farklı olarak ölçme ve değerlendirme dersinin matematik alanı ile ilişkisi göz önüne alınarak matematik kaygı ve başarı faktörlerinin, ölçme ve

değerlendirme ders başarısı ve derse karşı tutum faktörleriyle ilişkisi olup olmadığı ortaya çıkarılmaya çalışılmıştır.

Araştırmada, ölçme ve değerlendirme dersi almış olan öğretmen adaylarının sosyoekonomik düzeyleri, anadolu öğretmen lisesi mezunu olma durumları, eğitimde ölçme ve değerlendirme dersine ayrılan zaman ve derse yönelik tutumları, matematik kaygısı, matematik başarı düzeyi ve genel akademik başarıları ile ölçme ve değerlendirme dersi başarıları arasındaki ilişkilerin incelenerek, bu değişkenlerin öğrencilerin ölçme ve değerlendirme dersi başarısıyla ve birbirleriyle olan ilişki düzeyi belirlenmeye çalışılmıştır. Bu ilişkileri ortaya çıkarmada kullanılan çoklu göstergeler çoklu nedenler modeli ile tahmin edicilerin değişkenleri ne düzeyde açıkladığının yanısıra, tahmin etmede en önemli unsurun hangi değişkenler olduğu da tespit edilmiştir.

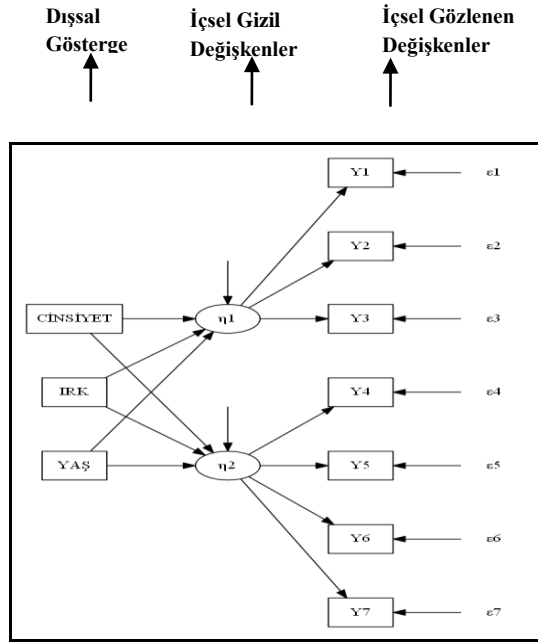
Diğer eğitim bilimleri derslerine kıyasla daha fazla sayısal ağırlıklı bir ders olduğu öngörülen ölçme ve değerlendirme dersinin matematik alanı ile ilişkisi göz önüne alındığında matematik kaygı ve başarısının, ölçme ve değerlendirme dersi başarısı ve derse karşı tutum faktörleriyle ilişkisi olup olmadığının araştırılarak öğretmen adaylarının başarısına etki eden faktörlerin belirlenmesinin, onların yeterlilikleri ve eksiklikleri hakkında öngörü sağlanarak başarının artırılması ve geliştirilmesi hususunda katkısı olacağı düşünülmektedir.

Yapılan alanyazın taramasında eğitimde ölçme ve değerlendirme (EÖD) ders başarısını konu alan Kart'ın (2012) yapmış olduğu çoklu regresyon analizi ile yapılan araştırma dışında yurt içi ve yurt dışı çalışmalarında, özellikle yapısal eşitlik modellemesi geliştirilerek model üzerinde karşılıklı etkilerin ve ilişkilerin analiz edildiği bir çalışmaya rastlanılmamıştır. Bu özelliği ile araştırmanın alana katkı sağlayacağı öngörülmektedir. Çalışmanın yöntemi hakkında detaylara geçmeden önce, aşağıdaki bölümde araştırmada kullanılan çoklu göstergeler ve çoklu nedenler modelinden kısaca bahsedilecektir.

Çoklu Göstergeler ve Çoklu Nedenler Modeli (ÇGÇN) (Multiple Indicators and Multiple Causes Model, MIMIC)

Yapısal eşitlik modelleri (YEM), aralarında ilişki olan birden fazla bağımlı ve bağımsız değişken arasındaki ilişkiyi matematiksel eşitlik yardımıyla açıklar (Çokluk, Şekercioğlu ve Büyüköztürk, 2014). Regresyon analizlerine benzemekle birlikte, yapılara ölçme hatalarını ve hatalar arasındaki ilişkileri de dahil eder (Yılmaz ve Çelik, 2013). Ayrıca gizil yapılar arası ilişkileri analiz eden yapısal eşitlik modeli bu özelliği ile, karşılaştırılan iki veya daha çok nedensel modelin korelasyon matrisinin uygunluğunu test etmek amacıyla kullanılan, regresyon modellerinin devamı niteliğindeki gözlenen değişkenler arası ilişkileri analiz eden yol analizinden de farklıdır (Wang ve Wang, 2012).

Çoklu göstergeler ve çoklu nedenler modeli yapısal eşitlik modelinin alt unsuru olarak tanımlanabilir. YEM analizlerinde olduğu gibi ölçme modeli ve yapısal modelden oluşmaktadır. Ölçme modeli gizil değişkenleri gözlenen değişkenlere bağlar, gözlenen değişkenlerin gizil değişken tarafından nasıl tanımlandığını gösterir. Yapısal model ise gizil değişkenler ile bağımsız gösterge değişkenler arasındaki ilişkiyi belirler (Brown, 2015; Baldemir, Özkoç ve İşçi, 2009; Kline, 2015).



Şekil 1. Çoklu Göstergeler ve Çoklu Nedenler (ÇGÇN) Modeli

Çoklu göstergeler ve çoklu nedenler (mimic) model yaklaşımı birçok neden ve gösterge değişkeni içeren gizil değişkenlerin istatistiksel teorisine dayanmaktadır. Özellikle sosyal araştırmalarda gözlenemeyen değişkenlerin incelenmesinde kullanılan, çok sayıda gösterge değişkeninin içsel gizil değişkeni temsil ettiği ve çok sayıda nedensel etkilerin de (tahmin edici gözlenenlerin) içsel gizil değişkenleri etkilediği bir yöntemdir (Brown, 2015; Baldemir, Özkoç ve İşçi, 2009; Kline, 2015). Şekil 1'de çoklu göstergeler ve çoklu nedenler modeline bir örnek sunulmuştur. Kovaryans yapı analizlerinde çoklu göstergeler ve çoklu nedenler modeli, ölçme modelinin bir gizil değişken yerine, bir ya da bir kaç gizil değişkenden oluştuğu, çok sayıda içsel göstergenin içsel gizil değişkeni ölçmek için kullanıldığı ve çoklu nedensel etkilerin gösterildiği, bütün gizil değişkenlerin bir ölçme modeli altında toplandığı bir yöntemdir. Gizil değişkenler, dışsal gizil değişkenleri mükemmel temsil ettiği varsayılan tahmin edici gösterge değişkenler (cinsiyet, ırk gibi değişkenleri ölçmede ölçme hatası olmadığı kabul edilir) tarafından etkilenir. Dışsal gizil değişkenlerin, bağımsız gösterge değişkenler olarak tanımlandığı bu modellerde gizil değişkenler ile demografik değişkenler arasındaki nedensel ilişkiler araştırılır. Şekil 1'de görülen cinsiyet, ırk ve yaş değişkenleri demografik değişkenlerdir ve çoklu gösterge çoklu neden (ÇGÇN) modelinde iki gizil değişkeni etkileyen bağımsız gösterge (gözlenen) değişkenlerdir.

YÖNTEM

Çalışma Grubu

Araştırmanın çalışma grubunu, 2015-2016 yılında orta Anadolu'da bölgesindeki bir üniversitede öğrenim gören 443 öğretmen adayı oluşturmaktadır. Çalışmanın uygulanması için gerekli olan etik kurul izni alınmıştır. Araştırmaya katılan 443 öğrenciden sınıf öğretmenliği programındaki öğrenci sayısı 69 (%15.6), felsefe bölümünde 85 (%19.2), ilköğretim matematik bölümünde 35 (%7.9), fen bilgisi öğretmenliği programında 28 (%6.3), İngiliz dili bölümünde 53 (%12), bilgisayar ve öğretim teknolojileri programında 84 (%19), psikolojik danışma ve rehberlik programında 51 (%11.5), okul öncesi öğretmenliğinde ise 38 (%8.6) dir.

Veri Toplama Araçları

Araştırmadaki veriler, üç bölümden oluşan ölçme aracı ile elde edilmiştir. Ölçme aracının ilk bölümünü, öğretmen adaylarının eğitimde ölçme ve değerlendirme dersine ait notları, genel

akademik başarı not ortalamaları, matematik başarı durumları, anadolu öğretmen lisesi mezunu olup olmadıkları, bölümleri, haftalık ders çalışmaya ayırdıkları zaman, ebeveyn öğrenim ve gelir düzeyleri ile ilgili demografik bilgilerini içeren 9 madde oluşturmaktadır. İkinci bölümde Fennema ve Sherman tarafından 1976 yılında geliştirilen ve Tağ (2000) tarafından Türkçeye uyarlanan Matematik ve test kaygısı olmak üzere 2 alt faktörden oluşan ölçme aracındaki matematik kaygısını ölçen 9 maddelik bölüm kullanılmıştır ($\alpha=0.84$). Ölçme aracının üçüncü bölümünde Aktaş ve Alıcı (2012) tarafından geliştirilen "Eğitimde Ölçme ve Değerlendirme Dersine Yönelik Tutum Ölçeği" kullanılmıştır. Ölçek, öğretmen adaylarının ölçme değerlendirme dersine yönelik olan tutumlarını bilişsel, duyuşsal ve davranışsal anlamda ölçmek amacıyla 20 maddeden oluşmuştur. Üç faktörden oluşan ölçeğin alt boyutları sırasıyla bilişsel ($\alpha=0,911$), davranışsal ($\alpha=0,898$) ve duyuşsal ($\alpha=0,845$) tutum olarak adlandırılmıştır. Ölçeğin tümüne ait güvenirlilik katsayısı $\alpha=0.944$ bulunmuştur.

Verilerin Analizi

Bu çalışmada verilerin analizi için SPSS 15.0 ve MPLUS 6.0 programları kullanılmıştır. Öncelikle, matematik başarı düzeylerine, genel akademik başarı notları, eğitimde ölçme ve değerlendirme dersi başarı notları ve derse karşı tutumları ile matematik kaygı düzeylerine ilişkin betimsel istatistikler hesaplanmıştır. Verilerin yapılacak analizler için uygun olup olmadığı incelenmek üzere, verilerin doğrulanması ile kayıp değer ve uç değer analizleri yapılmıştır. Ayrıca veriler düzenlenirken hatalı veri girişi olup olmadığı, verilerin genel dağılımı ile ilgili gerekli varsayımlar kontrol edilmiştir. Varsayımların incelenmesi sonucu çoklu normal dağılım varsayımı dışında bir soruna rastlanılmadığı görülmüş ve verilerin analize uygun olduğu sonucuna varılmıştır. Çoklu normallik sağlanmadığı için yapısal eşitlik modeli analizinde standart hatalardan etkilenmeyen en çok olabilirlik tahmin edicisi metodu (maximum likelihood estimation with robust standard errors, MLR) kullanılmıştır (Wang ve Wang, 2012).

Araştırma problemini cevaplamak için, öğrencilerin başarılarını etkileyen faktörlere ilişkin çoklu göstergeler çoklu nedenler modeli analizi için öncelikle ölçme modeli testi gerçekleştirilmiştir. Bu çalışmada, ölçme modeli için sosyo ekonomik düzey (SED), KAYGI ve TUTUM (TUT1, TUT2, TUT3) gizil değişkenlerine ait gözlenen değişkenler ile aralarındaki standartlaştırılmış ve standartlaştırılmamış katsayılar elde edilmiştir. Araştırmada ölçme başarısı (ÖLÇME), genel akademik başarı (AKADEMİK) ve matematik başarısı (MATEMATİK) değişkenleri, sırasıyla OLCB, GAB ve MATB değişkenleri tarafından temsil edilen tek gözlenen değişkenli gizil değişkenlerdir. Ancak bunlarda ölçme hatası olmadığı düşünülerek işlem yapılması yanlı tahminler elde edilmesine neden olabilir. Bu sebeple modelde güvenirlilik katsayısı, $r=0,80$ olarak kabul edilmiş ve bu doğrultuda bu gizil değişkenler için hata varyansları da analize dahil edilerek gizil değişkenlere ait faktör yükleri elde edilmiştir.

Oluşturulan ölçme modelinin uyum sonuçları kontrol edilerek MPLUS 6'da yapılan analiz sonuçlarına göre modelin daha iyi uyum göstermesi için bazı modifikasyonlara gidilmiş ve nihai model oluşturulmuştur. Bu kapsamda, T11 "Sınıfta EÖD dersi ile ilgili konuları tartışmaktan hoşlanırım." ile T10 "EÖD dersindeki öğrendiğim bilgileri çevremdeki insanlara anlatmaktan zevk alırım." maddeleri arasında, T20 "EÖD dersinde işlenen konular bana karmaşık gelir." ile T17 "EÖD dersi benim için stres kaynağıdır." maddeleri arasında, T19 "EÖD dersine çalışmak beni mutsuz eder." ile T18 "EÖD dersine girmek benim için zaman kaybıdır." maddeleri arasında ve K9 "Matematik beni huzursuz eder ve kafamı karıştırır." ile K8 "Zor matematik problemleri ile uğraştığımı düşündüğüm zaman kendimi çaresiz hissedirim." maddeleri arasında modifikasyona gidilerek model uyum değerleri iyileştirilmiştir. Ayrıca, matematik kaygısı ile ilgili ölçme aracındaki maddelerin faktör yükleri ve içeriğinin incelenmesi sonucu K1 "Matematik beni hiç korkutmaz.", K2 "Matematik beni gerginleştirir." ve K3 "Matematik problemlerini çözebilmek konusunda genelde hiç endişelenmem." maddelerinin birbirine olan benzerliği dikkate alınarak madde birleştirme (item parcel) uygulaması ile MPLUS6 programında aritmetik ortalamaları alınmak suretiyle $(K1+K2+K3)/3=KAY1$ değişkeni olarak tek maddede birleştirilmiştir.

Ölçme modellerinin test edilmesinden sonra, oluşturulan yapısal modelde,

- a) TUTUM gizil değişkenine, KAYGI ve SED gizil değişkenlerinin etkisi,
b) MATEMATİK gizil değişkenine, TUTUM, KAYGI ve SED gizil değişkenlerinin etkisi,
c)AKADEMİK başarı gizil değişkenine, TUTUM, KAYGI, MATEMATİK ve SED gizil değişkenlerinin etkisi,
d) ÖLÇME gizil değişkenine ise, AKADEMİK, TUTUM, KAYGI, MATEMATİK ve SED gizil değişkenlerinin etkisi olup olmadığına dair yapısal ilişkiler incelenmiştir.
e) Ayrıca ZAMAN gösterge değişkeninin TUTUM ve ÖLÇME gizil değişkenlerine etkisi ile AOL gösterge değişkeninin TUTUM, ÖLÇME, MATEMATİK ve AKADEMİK gizil değişkenlerine etkisi olup olmadığı çoklu göstergeler ve çoklu nedenler (ÇGÇN) modeli kapsamına uygun olarak oluşturulan yapısal model altında incelenmiştir.

Araştırmada ölçme modeli ve yapısal modellerin test edilmesi aşamasında güçlendirilmiş (robust) en çok olabilirlik (Maximum Likelihood Robust, MLR) yöntemi kullanılmıştır. MLR, kategorik ve sürekli verilerden oluşan modellerde kullanabilen, kayıp veriler olduğunda analiz yapabilen standart hatalara dayanıklı bir yöntemdir (Wang ve Wang, 2012).

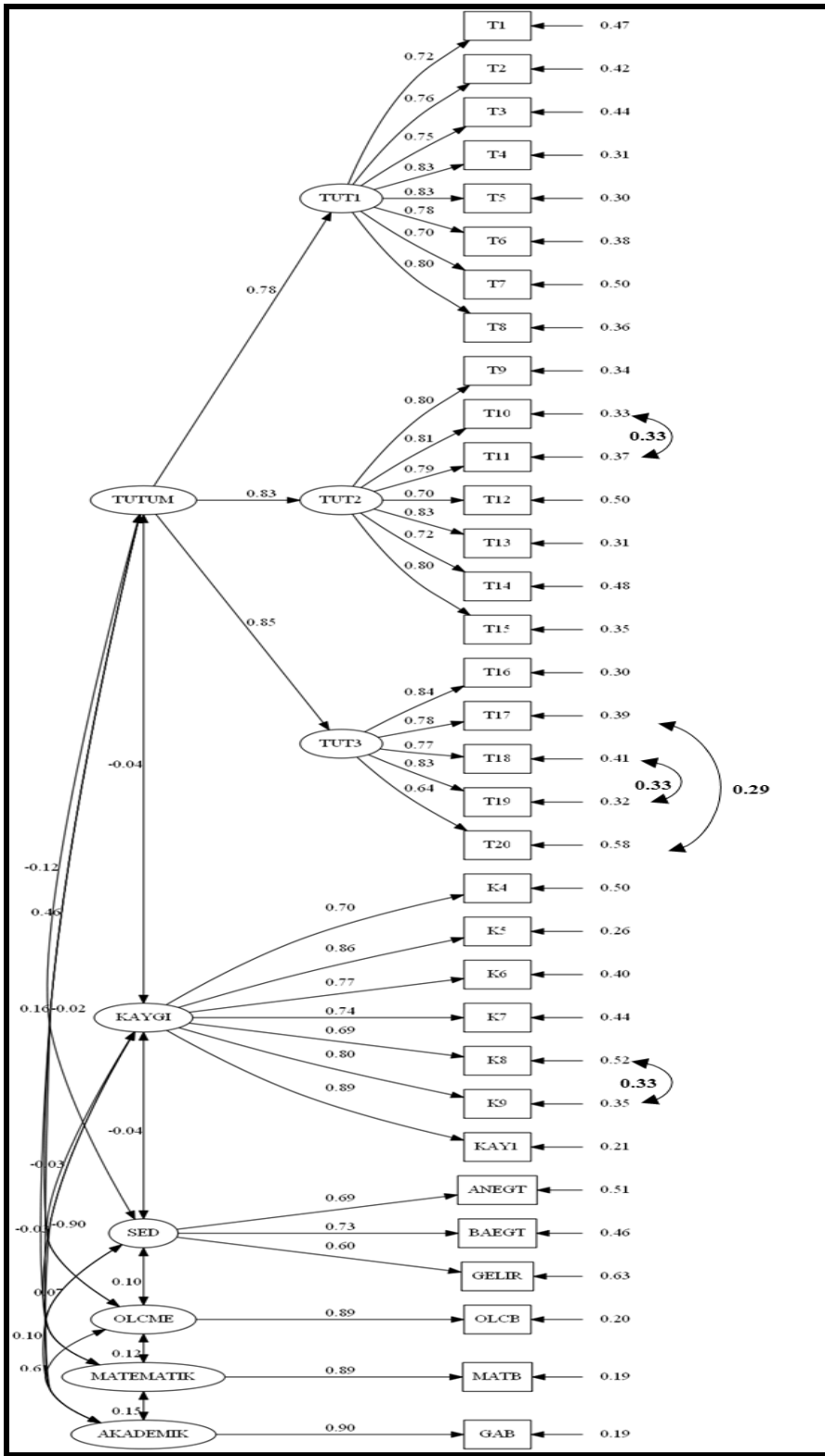
BULGULAR

Katılımcıların genel akademik başarı notları, eğitimde ölçme ve değerlendirme dersi başarı notları ve derse karşı tutumları ile matematik kaygı düzeylerine ilişkin betimsel istatistikler incelendiğinde aşağıdaki bulgular elde edilmiştir.

Tablo 1. Başarı, Tutum, ve Kaygı Değişkenlerine ait Betimsel İstatistikler

	<i>Min.</i>	<i>Max</i>	<i>Ort.</i>	<i>S.D</i>
2014-2015 Yılı Akademik Başarı Notu	1.68	3.95	3.097	.39401
Ölçme ve Değerlendirme Dersi Başarı Notu	37	98	82.19	11.797
Bilişsel Tutum	8.00	40.00	29.970	6.1459
Davranışsal Tutum	7.00	35.00	18.760	6.0459
Duyuşsal Tutum	5.00	25.00	18.186	4.5638
TUTUM	20.00	100.00	66.917	14.385
KAYGI	9.00	45.00	21.972	8.250

Tablo 1'deki bulgulara göre katılımcıların kaygı faktörü için toplam puanlarının 9 ile 45 arasında değişmekte olduğu ve 45 puan üzerinden 21.97 puan ortalaması ile matematik kaygı düzeylerinin düşük seviyelerde olduğu görülmektedir. Eğitimde ölçme ve değerlendirme dersine karşı tutumla ilgili bilişsel tutum alt boyutunun 40 puan üzerinden 29.98 ortalama puanı ve duyuşsal tutumun 25 puan üzerinden 18.18'lik puan ortalamasıyla yüksek, davranışsal tutumun 35 puan üzerinden 18.76 puan ortalamasıyla nispeten daha düşük olduğu görülmektedir. Tutum faktörüne ait toplam tutum puanları 20 ile 100 arasında değişmekte ve 100 puan üzerinden 66.91'lik puan ortalaması ile öğretmen adaylarının eğitimde ölçme ve değerlendirme dersine yönelik tutumlarının olumlu düzeyde olduğu ifade edilebilir. Genel akademik başarı notları en düşük 1.68 ile en yüksek 3.95 puanlar arasında değişmektedir. Araştırmaya katılan öğretmen adaylarına ait genel akademik başarı not ortalamalarının 4 üzerinden 3.07 olduğu görülmektedir. Bu nedenle öğretmen adaylarının akademik başarılarının iyi seviyede olduğu ifade edilebilir. Eğitimde ölçme ve değerlendirme dersi başarı notları, en düşük 37 ile en yüksek 98 puanları arasında değişmekte ve ortalamasının 100 üzerinden 82,19 olduğu ve bu nedenle genel olarak ölçme ve değerlendirme dersinde başarılı oldukları ifade edilebilir.



Şekil 2. Ölçme Modeli

Ölçme ve yapısal modele ilişkin bulgular

Şekil 2’de görülen ölçme modeline ilişkin uyum iyiliği indeksleri değerleri Tablo 2’de sunulmaktadır. Bu değerler incelendiğinde ki-kare değerinin anlamlı çıktığı görülmektedir.

Tablo 2. Ölçme Modeli Uyum İndeksi Sonuçları

	χ^2	$\chi^2/s.d.$	RMSEA	NNFI/TLI	CFI	SRMR
Ölçme Modeli	957.352 (p=0,00)	957.352/476 =2.01	0.048	0.934	0.940	0.052

Ki-kare değerinin genelde anlamlı çıktığını görürüz, çünkü bu değer örneklem büyüklüğüne oldukça duyarlıdır (Şimşek, 2007). Ki kare ve serbestlik derecesi oranının ($\chi^2/s.d.$) 2,01 olması ve RMSEA değerinin 0,048 olması iyi uyumu göstermektedir. NNFI değerinin 0,934 olması ve CFI değerinin 0,940 olması kabul edilebilir uyumu, SRMR değerinin 0,052 olması da modelin iyi uyum gösterdiğini belirtmektedir. Bu sonuçlara göre ölçme modelinin doğrulandığı söylenebilir. Ölçme modeline ait değişkenlerin faktör yükleri ve hata varyanslarına ilişkin değerlerin gösterimi Şekil 2’de, analize ilişkin standartlaştırılmış yükler, p değerleri, açıklanan varyans ve hata varyansları ise Tablo 3’de verilmiştir.

Tablo 3. Ölçme Modeline Ait Analiz Sonuçları

Gizil Değişken	Gözlenen Değişken	Standartlaştırılmış Yükler	P-değeri	Toplam Varyans	Hata Varyansları
BİLİŞSEL ÖZELLİKLER (TUT1)	T1	0.724	0.00	0.525	0.475
	T2	0.757	0.00	0.573	0.427
	T3	0.748	0.00	0.559	0.441
	T4	0.829	0.00	0.687	0.313
	T5	0.832	0.00	0.692	0.308
	T6	0.784	0.00	0.614	0.386
	T7	0.704	0.00	0.496	0.504
	T8	0.797	0.00	0.635	0.365
DAVRANISSAL ÖZELLİKLER (TUT2)	T9	0.808	0.00	0.654	0.346
	T10	0.814	0.00	0.662	0.338
	T11	0.793	0.00	0.628	0.372
	T12	0.702	0.00	0.492	0.508
	T13	0.830	0.00	0.689	0.311
	T14	0.720	0.00	0.519	0.481
	T15	0.804	0.00	0.646	0.354
DUYUŞSAL ÖZELLİKLER (TUT3)	T16	0.837	0.00	0.701	0.299
	T17	0.777	0.00	0.604	0.396
	T18	0.767	0.00	0.588	0.412
	T19	0.826	0.00	0.682	0.318
	T20	0.643	0.00	0.413	0.587
TUTUM	TUT1	0.777	0.00	0.604	0.396
	TUT2	0.832	0.00	0.693	0.307
	TUT3	0.851	0.00	0.724	0.276
KAYGI	K4	0.704	0.00	0.496	0.504
	K5	0.858	0.00	0.737	0.263
	K6	0.771	0.00	0.595	0.405
	K7	0.747	0.00	0.558	0.442
	K8	0.692	0.00	0.479	0.521
	K9	0.804	0.00	0.646	0.354
	KAY1	0.889	0.00	0.791	0.209
SED	ANEGT	0.697	0.00	0.486	0.516
	BAEGT	0.731	0.00	0.534	0.466
	GELIR	0.603	0.00	0.364	0.636
ÖLÇME	OLCB	0.894	0.00	0.80	0.20
AKADEMİK	GAB	0.898	0.00	0.806	0.194
MATEMATİK	MATB	0.897	0.00	0.804	0.196

Şekil 2’de görülen ölçme modelinde TUTUM’a ait TUT1 gizil değişkeni olan bilişsel tutumun ilk göstergesi ele alındığında T1 "EÖD dersi önem verdiğim bir derstir." değişkeni korelasyon katsayısı 0.72’dir. Bu değer, TUT1 gizil değişkenine ait T1 gözlenen değişkeninin faktör yükünü göstermektedir. Kline’a (2005) göre, faktör yük değeri, maddelerin faktörlerle olan ilişkisini açıklar. Bilişsel tutum boyutu en çok 0.83 yük katsayısı ile T5 "EÖD dersini almanın, öğrenci başarısını

belirlemede yapılacak çalışmaların önemini kavramak için gerekli olduğuna inanırım." gözlenen değişkeni tarafından açıklanmaktadır. TUT2 gizil değişkeni olan davranışsal tutum boyutunda en yüksek faktör yüküne sahip değişkenin ise 0.83 ile T13 "Arkadaşlarımla EÖD konularını konuşmayı severim." olduğu saptanmıştır. Aynı şekilde, duyuşsal tutum boyutu (TUT3) ise en çok T16 "EÖD dersine çalışmak bana sıkıcı gelir." tarafından açıklanmaktadır. İkinci düzeyde ise TUTUM gizil değişkeninin en çok 0.85 ile duyuşsal tutum boyutu olan TUT3 tarafından açıklandığı görülmüştür.

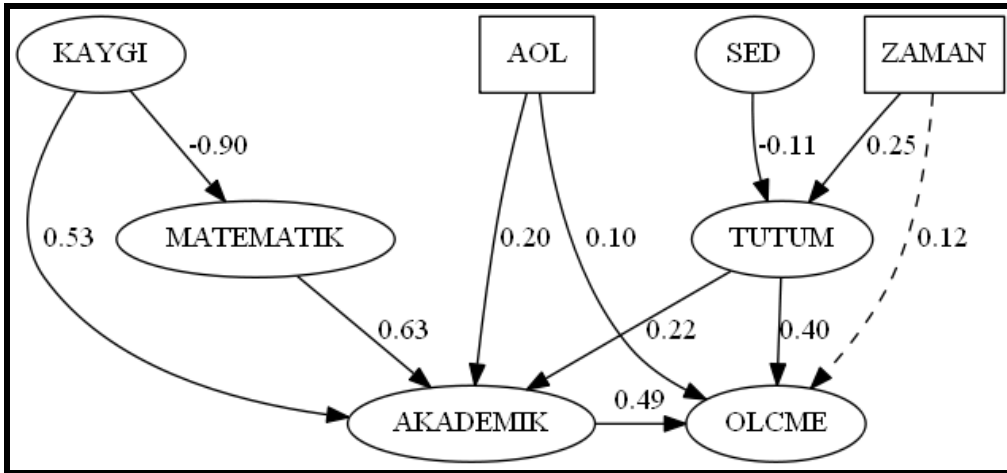
Tablo 3'ten de görüldüğü gibi KAYGI gizil değişkeninin 0.889 ile KAY1 değişkeni tarafından açıklandığı görülmüştür. KAY1 değişkeni K1 "Matematik beni hiç korkutmaz.", K2 "Matematik beni gerginleştirir." ve K3 "Matematik problemlerini çözebilmek konusunda genelde hiç endişelenmem." değişkenlerinin içerik benzerlikleri olduğunun değerlendirilmesi sonucu madde birleştirme (item parcel) uygulaması ile oluşturulmuştur. SED değişkeninin en çok 0.73 faktör yük katsayısı ile baba eğitim değişkeni (BAEGT) tarafından açıklandığı görülmektedir.

Yapısal modele ilişkin uyum indeksleri değerlendirildiğinde ki-kare değerinin anlamlı çıktığı görülmektedir. Anlamlı çıkmasının sorun teşkil etmesi nedeniyle bakılan diğer bir ölçüt olan ki kare ve serbestlik derecesi oranının ($\chi^2/s.d.$) 1,95 olması iyi uyumu göstermektedir. RMSEA değerinin 0.046, NNFI değerinin 0.931 ve SRMR değerinin 0.05 olması da iyi uyumu göstermektedir. CFI değerinin 0.938 olması kabul edilebilir uyumu belirtmektedir.

Tablo 4. Yapısal Model Uyum İndeksi Sonuçları

	χ^2	$\chi^2/s.d.$	RMSEA	NNFI/TLI	CFI	SRMR
Yapısal Model	1045.977 (p=0,00)	1045.977/536 =1,95	0.046	0.931	0.938	0.05

Yapısal model analizi sonucunda arasında ilişki olup olmadığı görülmek istenen gizil değişkenler arası etkiler ile çoklu göstergeler ve çoklu nedenler modeli olarak tanımlanan gösterge değişkenler (AOL, ZAMAN) ile gizil değişkenler arasındaki etkilerden anlamlı ($p<0,05$) olanlar Şekil 3'te verilmiştir. Bu bulgulara göre ölçme modelinin doğrulandığı ve yapısal modelin iyi uyum sağladığı görülmektedir.



Şekil 3. Yapısal Model

Yapısal modele ilişkin gizil değişkenlere ait açıklanan varyanslar (R^2) Tablo 5'te verilmiş olup gizil değişkenlere ait açıklanan varyans yüzdelерinin hepsinin anlamlı ($p<0,05$) olduğu görülmektedir. Tek faktörlü ölçeklerde açıklanan varyansın %30 ve daha fazla olması yeterli görülebilir. Çok faktörlü ölçeklerde ise açıklanan varyansın daha yüksek olması beklenir. Sosyal bilimlerde açıklanan varyansın %40 ile %60 arasında olması yeterli olarak kabul edilir (Çokluk vd., 2014). Elde edilen bu bulguların ardından gizil değişkenler arasındaki ve gösterge değişkenler ile gizil değişkenler arasındaki yapısal ilişkileri daha detaylı olarak aşağıda incelenmiştir.

Tablo 5. Yapısal Model Gizil Değişkenlere Ait Açıklanan Varyanslar

<i>Gizil Değişken</i>	<i>Tahmin</i>	<i>Std. Hata</i>	<i>Tahmin/ Std.Hata</i>	<i>P Değeri</i>
TUT1 Bilişsel Tutum	0.598	0.052	11.49	0.00
TUT2 Davranışsal Tutum	0.641	0.056	11.55	0.00
TUT3 Duyuşsal Tutum	0.780	0.061	12.85	0.00
TUTUM	0.079	0.028	2.86	0.004
ÖLÇME BAŞARI	0.534	0.066	8.08	0.00
AKADEMİK BAŞARI	0.161	0.051	3.15	0.002
MATEMATİK BAŞARI	0.821	0.044	18.60	0.00

Analiz sonuçlarına göre yapısal modelin tutumla ilgili kısmında, eğitimde ölçme ve değerlendirme dersine karşı tutum ile matematik kaygısı arasında -0.051'lik ($p; 0.381 > 0,05$) ve anadolu öğretmen lisesi mezunu olmaları arasında -0,048'lik istatistiksel olarak anlamlı olmayan ($p=0,351$) bir ilişki olduğu görülmektedir. Öte yandan tutum ile sosyoekonomik düzey arasında -0.113'lük ve derse ayrılan zaman arasında 0.246'lık istatistiksel olarak anlamlı bir ilişki bulunmuştur. Bu değer derse zaman ayıranların tutum puanlarının derse zaman ayırmayanlara oranla daha yüksek olduğunu göstermektedir.

$$\text{TUTUM} = 0.246 * \text{ZAMAN} - 0.113 * \text{SED} - 0.051 * \text{KAYGI} - 0.048 * \text{AOL}$$

Bu sonuçlara göre sosyo ekeonomik düzey, kaygı, Anadolu öğretmen lisesi mezunu olma durumu ve zaman değişkenlerinin tutumun yüzde sekizini açıkladığı, açıklanan varyansın çok düşük olduğu, eğitimde ölçme ve değerlendirme dersine karşı tutum gizil değişkenine en çok açıklayan değişkenin 0.246'lık etkiyle derse ayrılan zaman olduğu görülmektedir.

Modelin matematik başarısı ile ilgili kısmının sonuçları incelendiğinde, matematik başarısının sosyoekonomik düzey ile arasında 0.034'lük ($p=0.391 > 0,05$); EÖD dersine karşı tutum ile arasında -0.047'lük ($p=0,189$); anadolu öğretmen lisesi mezunu olmaları ile arasında 0,033'lük ($p=0,294$) düzeyde istatistiksel olarak anlamlı olmayan bir ilişki olduğu görülmektedir. Matematik kaygısı ile matematik başarısı arasında -0.904'lük negatif ve istatistiksel olarak anlamlı ($p=0,00$) bir ilişki bulunmuştur. Bu değer kaygıdaki artışın başarıda azalışa veya azalışın başarıda artışla sonuçlandığını göstermektedir.

$$\text{Matematik Başarısı} = 0.034 * \text{SED} - 0.904 * \text{KAYGI} + 0.033 * \text{AOL} - 0.047 * \text{TUTUM}$$

Açıklanan varyansa bakıldığında, sosyoekonomik düzey (SED), matematik kaygısı (KAYGI), Anadolu öğretmen lisesi mezunu olup olmama (AOL) ve eğitimde ölçme ve değerlendirme dersine karşı tutum (TUTUM) değişkenlerinin, matematik başarısının %82'sini açıkladığı görülmektedir. Matematik başarısına anlamlı etkiyi kaygı değişkeninin yaptığı görülmüştür.

Yapısal modelin genel akademik başarı gizil değişkenine ilişkin analiz sonuçlarına göre genel akademik başarının sosyoekonomik düzey ile arasında 0.086'lük pozitif yönde, zayıf ve istatistiksel olarak anlamlı olmayan ($p; 0.159 > 0,05$) bir ilişki mevcuttur. Matematik kaygısı ile akademik başarı arasında 0.533'lük ($p=0,005$) tutum ile arasında 0.218'lük ($p=0,001$) matematik başarısı ile arasında 0,629'lük düzeyde ve katılımcıların anadolu öğretmen lisesi mezunu olmaları ile arasında 0,204'lük ($p=0,00$) istatistiksel olarak anlamlı bir ilişki olduğu görülmektedir.

$$\text{Genel Akademik Başarı} = 0.086 * \text{SED} + 0.533 * \text{KAYGI} + 0.204 * \text{AOL} + 0.218 * \text{TUTUM} + 0.629 * \text{MATEMATİK}$$

Bu sonuçlara göre sed, kaygı, aol, tutum ve matematik değişkenlerinin genel akademik başarının %16.1'ini açıkladığı görülmektedir.

Yapısal modelde eğitimde ölçme ve değerlendirme ders başarısı gizil değişkeni ile ilgili ilişkiler incelendiğinde, sosyoekonomik düzey ile arasında 0.083'lük zayıf ve istatistiksel olarak anlamlı

olmayan ($p; 0.136 > 0,05$), matematik kaygısı ile arasında 0.222'lik düzeyde pozitif ancak anlamlı olmayan ($p=0,182$), matematik başarısı ile arasında ise, 0,247'lik düzeyde ve anlamlı olmayan ($p=0,149$) bir ilişki bulunmuştur. Ölçme ve değerlendirme dersine karşı tutum ile ders başarısı arasında 0.404'lük düzeyde ve istatistiksel olarak anlamlı ($p=0,00$) bir ilişki bulunmuştur. Genel akademik başarı notu ile ölçme ders başarı notu arasında da 0,489'lük düzeyde yüksek, pozitif yönde ve anlamlı ($p=0,00$) bir ilişki bulunmuştur. Katılımcıların Anadolu öğretmen lisesi mezunu olmaları ile ölçme başarısı arasında 0,101'lik pozitif yönde ve anlamlı ($p=0,007$) bir ilişki olduğu görülmektedir. Anadolu öğretmen lisesi mezunu olanların diğer liselere göre ölçme dersinde düşük düzeyde olsa da daha başarılı oldukları sonucuna ulaşılmıştır. Ölçme ve değerlendirme dersine ayrılan zaman ölçme ve değerlendirme ders başarısına doğrudan anlamlı etkisi olmadığı, ancak aşağıdaki tabloda görüldüğü üzere pozitif yönde ve anlamlı 0,119'lük dolaylı etkisi olduğu görülmektedir.

Tablo 6. Ölçme ve Değerlendirme Dersine Ayrılan Zaman ile Ders Başarısı Arasındaki Dolaylı İlişkiler

<i>Yapısal ilişkiler</i>	<i>Std. Yükler</i>	<i>P-değeri</i>
ZAMAN Tutum dolaylı ÖLÇME	0.099	0.00
ZAMAN Tutum Akademik dolaylı ÖLÇME	0.026	0.007
ZAMAN Tutum Matematik dolaylı ÖLÇME	- 0.003	0.353
ZAMAN Tutum Matematik Akademik dolaylı ÖLÇME	- 0.004	0.223
TOPLAM ZAMAN dolaylı ÖLÇME	0.119	0.00

$$\text{Ölçme Başarısı} = 0.083 \cdot \text{SED} + 0.222 \cdot \text{KAYGI} + 0.119 \cdot \text{ZAMAN} + \mathbf{0.101} \cdot \text{AOL} + \mathbf{0.404} \cdot \text{TUTUM} + \mathbf{0.489} \cdot \text{AKADEMIK} + 0.247 \cdot \text{MATEMATİK}$$

Bu sonuçlara göre sosyoekonomik düzey, matematik kaygısı, ölçme ve değerlendirme dersine ayrılan zaman, anadolu öğretmen lisesi mezunu olma, ölçme ve değerlendirme dersine karşı tutum, genel akademik başarı ve matematik başarısı değişkenlerinin ölçme ve değerlendirme ders başarısının %53.4'ünü açıkladığı görülmektedir. Ölçme ve değerlendirme ders başarısını en çok açıklayan değişkenin 0.489'lük etkiyle genel akademik başarı olduğu görülmektedir.

Bu çalışmada öğretmen adaylarının eğitimde ölçme ve değerlendirme ders başarısı ile bazı demografik ve duyuşsal değişkenler arasındaki ilişkiler çoklu göstergeler çoklu nedenler modeli ile incelenmiştir. Analiz sonuçlarına göre aşağıdaki sonuçlara ulaşılmıştır.

Eğitimde ölçme ve değerlendirme dersine karşı tutumu yordayan değişkenleri belirlemek için kurulan model iki aşamalı olarak test edilmiş ve elde edilen bulgular sonucunda modelin verilerle iyi uyum gösterdiği bulunmuştur. Ölçme modeline alınan değişkenlerden, tutumu belirleyen en yüksek faktör yüküne sahip yapının duyuşsal özellikler olduğu görülmüştür. Bu çalışmanın bulgularına göre öğretmen adaylarının eğitimde ölçme ve değerlendirmeye yönelik tutumlarının olumlu düzeyde olduğu söylenebilir. Bu sonuç, Erdoğan'ın (2010) genel anlamda öğretmenlerin ölçme ve değerlendirmeye yönelik tutumlarının olumlu yönde olduğu bulgusuyla desteklenirken, Yaşar'ın (2014), öğretmen adaylarının eğitimde ölçme ve değerlendirme dersine yönelik tutum düzeyleri ile tutumlarının bazı değişkenler açısından incelenmesi amacıyla yaptığı çalışmadaki öğretmen adaylarının eğitimde ölçme ve değerlendirmeye yönelik tutumlarının çok düşük düzeyde olduğu bulgusuyla desteklenmemektedir. Yapısal model incelendiğinde ise, tutumu açıklayan en yüksek faktörün ölçme ve değerlendirme dersi için çalışmaya ayrılan zaman olduğu görülmüştür. Kesiktaş'ın (2006) da belirttiği gibi verimli ders çalışma alışkanlıkları edinmiş öğrenciler, okula ve derse karşı olumlu tutum göstermektedir. Sosyoekonomik düzeyin ise negatif yönde derse karşı tutuma etki ettiği görülmüştür. Sosyoekonomik düzey gizil değişkenine ait en yüksek faktör yüküne sahip olan değişkenin baba eğitim durumu olduğu sonucuna ulaşılmıştır. Ele alınan matematik kaygısı ve anadolu öğretmen lisesi mezunu olup olmama durumunun ölçme değerlendirme dersine karşı tutuma etkisinin negatif yönde ve manidar olmadığı sonucuna ulaşılmıştır.

Yapısal modelde matematik başarısını yordayan değişkenler incelendiğinde, başarıyı açıklayan faktörün matematik kaygı değişkeni olduğu, matematik kaygısı arttıkça başarının düştüğü, azaldıkça başarının arttığı sonucuna ulaşılmıştır. Bu sonuç Erden ve Akgül'ün (2010) yapılan regresyon analizi sonucundaki matematik kaygısının matematik başarısının anlamlı yordayıcısı olduğu ve Peker ve Şentürk'ün (2012) matematik kaygıları ile matematik notları arasında negatif yönlü ve orta düzeyde anlamlı bir ilişkinin olduğu bulgusuyla desteklenmektedir. Sayısal ağırlıklı bir ders olması nedeniyle tutum puanı ile matematik başarısı arasında bir ilişki beklense de anlamlı bir ilişki olmadığı eğitimde ölçme ve değerlendirme dersine karşı tutumun matematik başarısı üzerinde herhangi bir etkisi olmadığı görülmüştür. Bu sonuç, Süral'in (2014) sayısal içerikli bölümlerde öğrenim gören öğrencilerin ölçme ve değerlendirme dersine yönelik daha yüksek bir tutum gösterdikleri ve öğrenim hayatında aldığı matematik dersi, matematik dersine karşı edindiği tutum, lisans eğitimindeki ölçme değerlendirme dersine olan bakış açısına yön verdiği bulgularıyla desteklenmemektedir. Ayrıca sosyoekonomik düzey ve anadolu öğretmen lisesi mezunu olma durumu ile matematik başarısı arasında istatistiksel olarak anlamlı olmayan bir ilişki bulunmuştur. Bu sonuç Yavuz'un (2009) anne ve baba eğitiminin doğrudan ayrıca baba eğitiminin aylık gelir artışı ve öğrencinin dershaneye gitme süresi aracılığı ile öğrencinin matematik-fen puanını pozitif olarak etkilediği ve Kalender'in (2010) sosyoekonomik durumun matematik başarısı üzerinde güçlü bir etkiye sahip olduğu bulgusuyla ve öğrencilerin matematik başarıları ile devam ettikleri okul türleri arasında anlamlı ilişki olduğu sonucuyla desteklenmemektedir.

Yapısal modelde genel akademik başarıyı yordayan değişkenler incelendiğinde, akademik başarıyı açıklayan en yüksek faktörün matematik başarısı olduğu, yine matematik kaygısı değişkeninin de genel akademik başarıyı pozitif yönde etkilediği görülmüştür. Yani matematik kaygısı yüksek düzeyde olan öğretmen adaylarının genel akademik başarısı da yüksektir. Bu durum oldukça ilginçtir. Matematik kaygısı düşük olan öğretmen adaylarının matematik başarı notları yüksek olmasına rağmen genel akademik başarılarının da matematik başarısıyla benzer eğilim göstermesi beklense de tam tersi bir durum ortaya çıkmıştır. Bu sonuçlar, Peker ve Şentürk'ün (2012) genel notları ile matematik kaygıları arasında negatif yönlü ve orta düzeyde anlamlı bir ilişkinin olduğu bulgusuyla desteklenmemekte, matematik notları ile genel notları arasında pozitif yönlü ve yüksek düzeyde anlamlı ilişki olması bulgusuyla desteklenmektedir. Öğretmen adaylarının anadolu öğretmen lisesi mezunu olmasının, diğer okullardan mezun olan öğretmen adaylarına göre genel akademik başarıyı pozitif yönde etkilediği sonucuna varılmıştır. Genel akademik başarıyı etkileyen diğer faktör ise eğitimde ölçme ve değerlendirme dersine karşı tutumdur. Tutum puanı yüksek öğretmen adaylarının genel akademik başarısı notlarının yüksek, tutum puanı düşük olanların akademik başarı notları da düşüktür. Bu sonuç, Süral'in (2014) genel başarı not ortalamaları yüksek olan öğrencilerin derse karşı gösterdiği önem ve bilişsel yeterlik boyutlarında daha yüksek bir tutum sergiledikleri, düşük öğrencilerin ise olumsuz yaklaşım alt boyutunda daha yüksek bir ortalamaya sahip oldukları bulgusuyla desteklenmektedir. Ele alınan demografik değişkenlerden sosyoekonomik düzeyin genel akademik başarıya, manidar bir etkisi olmadığı sonucuna ulaşılmıştır. Taningco (2006), ekonomik durum ile anne ve babanın eğitim seviyesinin öğrencilerin akademik başarısının artmasında önemli bir etken olduğunu belirtse de bu durum araştırma sonuçları ile desteklenmemektedir.

Yapısal modelin eğitimde ölçme ve değerlendirme ders başarısını yordayan değişkenleri incelendiğinde, ölçme ve değerlendirme ders başarısını açıklayan en yüksek faktörün genel akademik başarı değişkeni olduğu sonucuna ulaşılmıştır. Genel akademik başarısı yüksek düzeyde olan öğretmen adaylarının eğitimde ölçme ve değerlendirme ders başarısı da yüksektir. Ölçme ve değerlendirme ders başarısını etkileyen diğer faktör ise eğitimde ölçme ve değerlendirme dersine karşı tutumdur. Bu sonuç, Kart'ın (2012) öğretmen adaylarının ölçme ve değerlendirme dersi başarılarını duyuşsal özelliklerin manidar olarak yordamadığı bulgusuyla örtüşmemektedir. Öğretmen adaylarının anadolu öğretmen lisesi mezunu olmasının da, diğer okullardan mezun olan öğretmen adaylarına göre ölçme ve değerlendirme ders başarısını pozitif yönde etkilediği sonucuna varılmıştır. Ele alınan demografik değişkenlerden sosyoekonomik düzeyin ölçme ve değerlendirme ders başarısına manidar bir etkisi olmadığı sonucuna ulaşılmıştır. Sayısal ağırlıklı bir ders olarak

değerlendirilen eğitimde ölçme ve değerlendirme dersinin matematik başarısı ve matematik kaygı faktörleri ile ilişkisi olması beklense de bu faktörlerin ölçme ve değerlendirme ders başarısına manidar bir etkisi olmadığı sonucuna varılmıştır. Eğitimde ölçme ve değerlendirme dersine ayrılan zamanın ölçme ve değerlendirme ders başarısına istatistiksel olarak anlamlı doğrudan bir etkisi olmadığı görülsede pozitif yönde dolaylı etkisi olduğu sonucuna ulaşılmıştır. Bu sonuç Özer ve Anıl (2011) ile Savaş, Taş ve Duru'nun (2010) ders çalışma süresiyle ders (fen ve matematik) başarısı arasındaki ilişkinin de anlamlı olduğu sonucu bulgusuyla desteklenmektedir.

Bu sonuçlara göre araştırmacı ve eğitimciler için öneriler şunlardır. Öğretmen adaylarının eğitimde ölçme ve değerlendirme dersine yönelik tutum düzeylerinin orta seviyelerde olmasının nedenleri tespit edilerek özellikle davranışsal ve duyuşsal boyutta düşük veya olumsuz olan tutumlarının olumlu yönde artmasını sağlayacak çözümler (derse hazır gelmeleri için ödev, sunum, proje uygulama görevi vb.) araştırılmalıdır.

Öğrencinin derslerine çalışmak için ayırdığı zaman değişkeninin başarıya ve derse karşı tutumlarına etkisi olduğu görülmektedir. Derse hazırlanma kapsamında yapılan çalışmalarda, derste hazır olma düzeyinin artırılarak ders içi başarıyı olumlu etkileyen ve derse karşı olumlu tutum gösterilmesini sağlayan farkındalık eğitimleri verilebilir. Bu kapsamda, öğrencilere, ders çalışmaya zaman ayırarak derse hazırlıklı gelmelerinin derse karşı tutumlarına ve başarılarına olumlu etkisinin olacağı da vurgulanmalıdır.

Duyuşsal özelliklerden biri olan matematik kaygısının yapılan araştırmada matematik başarısına en büyük etki eden faktör olduğu sonucundan hareketle, öğrencilere matematiği sevdirmeye yönelik gerekli etkinlikler ve matematiğe karşı özgüven düzeylerini arttırmak için gerekli çalışmaları planlayıp, uygulamaları konusunda gerekli adımlar atılabilir.

Bu araştırma, orta Anadolu bölgesindeki bir üniversitede ölçme ve değerlendirme dersi almış öğrenciler üzerinde yapılmıştır. Farklı üniversite ya da üniversiteler arası karşılaştırma amaçlı ya da benzer nitelikte heterojen gruplarla çalışma yapılabilir.

Bu araştırmada öğrenci görüşlerine dayalı başarıyı etkilediği öngörülen bazı değişkenler kullanılmıştır. Modele alınmayan diğer değişkenlerin de (matematiğe karşı tutum, ölçme ve değerlendirme dersi kaygı düzeyi, cinsiyet, özyeterlilik, özgüven, ilgi düzeyi vb.) kullanılması ve bölümlere göre sayısal, sözel ve eşit ağırlıklı alanlarda eğitim görenler arasında farklılık olup olmadığını tespit etmek üzere araştırma yapılabilir.

KAYNAKÇA

- Aktaş, M. ve Alıcı, D. (2012). Eğitimde ölçme ve değerlendirme dersine yönelik tutum ölçeğinin (EÖD-TÖ) geliştirilmesi. *Journal of Qafqaz University, Philology and Pedagogy*, 33, 66-73.
- Alkharusi, H., Kazem, A.M. & Al Musawai, A. (2011). Knowledge, skills and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pasific Journal of Teacher Education*, 39(2), 113-123.
- Baldemir, E., Özkoç, H. ve İşçi, Ö. (2009). MIMIC model ve yolsuzluk üzerine Türkiye uygulaması. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24(2), 49-63.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2014). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. (3.baskı) Ankara: Pegem Akademi Yayınları.
- Erden, M. ve Akgül, S. (2010). İlköğretim öğrencilerinin matematik kaygısının ve öğretmen sosyal desteğinin matematik başarılarını yordama gücü. *Eğitimde Kuram ve Uygulama*, 6(1), 3-16.
- Erdoğan, M. Y. (2010). *Öğretmenlerin ölçme ve değerlendirmeye yönelik tutumlarının bazı değişkenler açısından İncelenmesi*. International Conference on New Trends in Education and Their Implications. Antalya, 11-13 November.
- Kalender, Ö. M. (2010). *The roles of affective, socioeconomic status and school factors on mathematics achievement: a structural equation modeling study*. Yayımlanmamış Doktora Tezi. Ortadoğu Teknik Üniversitesi, Ankara.

- Kart, A. (2012). *Demografik ve duyuşsal deęişkenlerin ölçme ve deęerlendirme dersi başarısını ve genel akademik başarıyı yordama gücü*. Yayınlanmamış Yüksek Lisans Tezi. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kesiktaş, A. D. (2006). Ders çalışma becerileri ve özel gereksinimli öğrenciler. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi* 2006, 7(1), 37-48.
- Kline, R. B. (2005). *Principles and practice of structural equation modelling*. (2nd Edition). New York: Guilford Publications, Inc.,
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. (4th edition). Guilford publications.
- 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.
- Peker, M. ve Şentürk, B. (2012). İlköğretim beşinci sınıf öğrencilerinin matematik kaygılarının bazı deęişkenler açısından incelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 34, 21-32.
- Poyraz, S. (2012). *İlköğretim ikinci kademe öğrencilerinde sınav kaygısı, matematik kaygısı, genel başarı ve matematik başarıları arasındaki ilişkilerin incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Üniversitesi, İstanbul.
- Reçber, Ş. (2011). *An investigation of the relationship among the seventh grade students' mathematics self efficacy, mathematics anxiety, attitudes towards mathematics and mathematics achievement regarding gender and school type*. Yayınlanmamış Yüksek Lisans Tezi. Ortadoęu Teknik Üniversitesi, Ankara.
- Savaş, E., Taş, S. ve Duru, A. (2010). Factors affecting students' achievement in mathematics. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 11(1), 113-132.
- Süral, S. (2014). Eğitim fakültesi öğrencileri ile pedagojik formasyon alan öğrencilerin ölçme deęerlendirme dersine yönelik tutumlarının incelenmesi. *Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi*, 1(3), 63-75.
- Şimşek, O. F. (2007). *Yapısal eşitlik modellemesine giriş (Temel ilkeler ve LISREL uygulamaları)*. Ankara: Ekinoks.
- Taę, Ş. (2000). *Reciprocal relationship between attitudes toward mathematics and achievement in mathematics*. Unpublished Master' Thesis. Middle East Technical University, Ankara.
- Taningco, M. T. V. (2006). *Assessing the effects of parental decisions about school type and involvement on early elementary education*. [Çevrim-içi: http://www.rand.org/content/dam/rand/pubs/rgs_dissertations/2006/RAND_RGSD205.pdf, Erişim tarihi: 25 Kasım 2015.]
- Wang, J. & Wang, X. (2012). *Structural equation modeling applications using mplus*. United Kingdom: Higher Education Press.
- Yaşar, M. (2014). Eğitimde ölçme deęerlendirme dersine yönelik tutum ölçeğinin geliştirilmesi. *Eğitim Bilimleri Araştırma Dergisi*. 4(1), 259-279.
- Yavuz, M. (2009). Ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavında öğrencilerin matematik- fen (mf) puanlarını etkiledięi düşünölen bazı faktörlerin yapısal eşitlik modeli ile incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri / Educational Sciences: Theory & Practice*, 9(3), 1557-1572.
- Yenilmez, K. ve Duman, A. (2008). İlköğretimde matematik başarısını etkileyen faktörlere ilişkin öğrenci görüşleri. *Sosyal Bilimler Dergisi*, 19, 251-268.
- Yılmaz, V. ve Çelik, H. E. (2013). *Lisrel 9.1 ile yapısal eşitlik modellemesi: Temel kavramlar, uygulamalar, programlama*. (Yenilenmiş 2. Baskı). Ankara: Anı Yayıncılık.

EXTENDED ABSTRACT

Introduction

Teacher candidates take educational measurement and evaluation (EME) course in their undergraduate education and it is important for them to successfully complete the course in order to be competent in their future occupation. Therefore, it is important to identify the variables that are affecting teacher candidates' achievement levels at educational measurement and evaluation course. When the studies conducted in the field of educational measurement and evaluation are examined it was found that they are mostly concerned with the competencies of teachers for measurement and evaluation and their attitudes towards the course. On the other hand, when studies related to general academic achievement are examined, it was found that variables such as department, gender, socioeconomic level, self-efficacy, anxiety and attitudes of the students have an effect on overall

academic achievement (Poyraz, 2012; Kart, 2012; Süral, 2014). Parental education level, monthly income, school type and gender variables as well as general academic achievement level, math-anxiety and attitude toward the course were found as effective variables in mathematics education (Yenilmez and Duman, 2008, Erden and Akgül, 2010; Rescher, 2011, Peker and Senturk, 2012). In addition, it has been observed that the time allocated for studying affects the course achievement level positively and high socioeconomic level has an effect on students' general academic achievement in science and mathematics lessons (Anıl ve Özer, 2011; Kalender, 2010; Savaş, Taş and Duru, 2010; Yavuz, 2009; Duru, 2010).

The aim of this study is to investigate the relationship among the prospective teachers' achievement levels in educational measurement and evaluation (EME) course and some demographic and affective variables. In the study, the relations among the socioeconomic status, being graduate of Anatolian teacher education high school, study time, attitudes towards the EME course, mathematics anxiety, mathematics achievement level, general academic achievement and the measurement and evaluation course achievement level were investigated. This study is guided by the following research problem:

What is the level of relationship among the teacher candidates' achievement levels of the EME course, socioeconomic status, being graduate of Anatolian teacher education high school, attitude towards the course, study time, general academic achievement, mathematics achievement level and math anxiety?

Method

The study group consists of 443 prospective teachers who are students at the university in middle Anatolia in 2015 fall and 2016 semesters. The data in the study were obtained by measurement instrument which has three parts. The first part of the instrument consists of 9 items that include the demographic information about parental education and income levels, whether the students graduated from Anatolian teacher education high school, departments, study time, EME course grade, general academic achievement grade, mathematics achievement level. The second part of the instrument included, 9-item scale to measure math anxiety and the third part included "Attitudes Toward Educational Measurement and Evaluation scale" developed by Aktaş and Buyer (2012).

In this study, descriptive statistics related to mathematics achievement levels, general academic achievement grades, achievement grades in educational measurement and evaluation course, attitudes towards the course and mathematics anxiety levels were calculated. In order to answer the research problem, multiple indicators multiple causes (MIMIC) method was used. MIMIC models can be described as a sub-component of the structural equation model. MIMIC model is consisted of measurement model and structural model as in SEM analysis. The measurement model links latent variables to observed variables, and shows how the observed variables are defined by the latent variable. The structural model determines the relationship among latent variables and independent indicator variables (Baldemir, Özkoç and İşçi, 2009). In the study, the robust maximum likelihood (MLR) method was used to test the measurement model and structural models. MLR is a standard error-tolerant method that can be used in models consisting of categorical and continuous data, and with missing data (Wang and Wang, 2012).

In this study as a first step of the MIMIC analysis, measurement model was analyzed. After that, the structural model was constructed with the following specifications,

- a. Effect of ANXIETY and SES latent variables to ATTITUDE latent variable,
- b. Effect of ATTITUDE, ANXIETY and SES latent variables to MATHEMATICS latent variable,
- c. Effect of ATTITUDE, ANXIETY, SES and MATHEMATICS latent variables to ACADEMIC latent variable,

- d. Effect of ATTITUDE, ANXIETY, SES, MATHEMATICS and ACADEMIC latent variables to MEASUREMENT latent variable,
- e. Structural relations of whether TIME indicator variable has an effect on ATTITUDE and MEASUREMENT the latent variables and whether AOL indicator variable has an effect on ATTITUDE, MEASUREMENT, MATHEMATICS and ACADEMIC are examined by MIMIC models.

Conclusion and Discussion

According to the findings, the measurement and the structural models showed good fit statistics. Following these findings, the structural relationships between latent variables and indicator variables are examined in more detail below.

According to the results, when the structural model is examined, it is seen that the study time indicator variable explains the attitude latent variable more than other variables. Mathematics anxiety and being Anatolian teacher high school graduate status have a negative effect on the attitude toward EME course and the effects were not significant statistically.

When the variables that predict mathematic achievement are examined, it is found that mathematics anxiety explained most of the variance in mathematics achievement. In other words, as the mathematical anxiety increases, the achievement decreases. Moreover, attitude toward EME course did not have a significant effect on mathematics achievement. In addition, there was a statistically insignificant effect of socioeconomic status and being Anatolian teacher education high school on mathematics achievement.

When the variables that predict general academic achievement were examined, it was seen that mathematics was the most important factor explaining the academic achievement and the mathematics anxiety variable also affected the overall academic achievement positively. In other words, the general academic achievement of teacher candidates increases with mathematics anxiety. This is quite interesting. Despite the fact that the mathematics achievement grades of teacher candidates with low mathematics anxiety are high, it is expected that general academic achievements will show a similar tendency with mathematics achievement. The prospective teachers who are graduated from Anatolian teacher education high school had higher general academic grades compared to the prospective teachers who graduated from other schools. Another factor affecting general academic achievement is attitude towards EME course. Candidates with high attitude scores have high academic achievement grades, and those with low attitude scores have low academic achievement grades. Among the demographic variables socioeconomic status did not have a significant effect on general academic achievement.

Finally, it is observed that the highest factor explaining the achievement level in EME course is the general academic achievement variable. Teacher candidates with a high level of academic achievement in general have higher success in EME course. Another factor affecting the achievement level in the EME course is the attitude towards measurement and evaluation course. Being graduate of Anatolian teacher education high school also had positive effect on achievement of EME course compared to being graduate of other school types. Socioeconomic status did not have a significant effect on the achievement level for the EME course. It is also found that mathematics achievement and mathematics anxiety factors did not have a significant influence on the achievement in EME course. Although study time did not have statistically significant direct effect on the achievement in EME course, it had an indirect positive effect.

Computer Adaptive Multistage Testing: Practical Issues, Challenges and Principles*

Bireye Uyarlanmış Çok Aşamalı Testler: Pratik Konular, Zorluklar ve Prensipler

Halil Ibrahim SARI** Hasibe YAHSI SARI *** Anne Corinne HUGGINS MANLEY****

Abstract

The purpose of many test in the educational and psychological measurement is to measure test takers' latent trait scores from responses given to a set of items. Over the years, this has been done by traditional methods (paper and pencil tests). However, compared to other test administration models (e.g., adaptive testing), traditional methods are extensively criticized in terms of producing low measurement accuracy and long test length. Adaptive testing has been proposed to overcome these problems. There are two popular adaptive testing approaches. These are computerized adaptive testing (CAT) and computer adaptive multistage testing (ca-MST). The former is a well-known approach that has been predominantly used in this field. We believe that researchers and practitioners are fairly familiar with many aspects of CAT because it has more than a hundred years of history. However, the same thing is not true for the latter one. Since ca-MST is relatively new, many researchers are not familiar with features of it. The purpose of this study is to closely examine the characteristics of ca-MST, including its working principle, the adaptation procedure called the routing method, test assembly, and scoring, and provide an overview to researchers, with the aim of drawing researchers' attention to ca-MST and encouraging them to contribute to the research in this area. The books, software and future work for ca-MST are also discussed.

Keywords: adaptive testing, computerized adaptive testing, computer adaptive multistage testing

Öz

Genel olarak eğitimdeki testlerin amacı testteki sorulara verilen cevaplarla testi alan bireylerin yetenek seviyelerini ölçmektir. Bu işlem yıllarca geleneksel yöntem olarak bilinen kağıt-kalem formundaki testlerle yapıldı. Ancak geleneksel yöntemler diğer test yöntemlerine (bireye uyarlanmış testler) göre yüksek ölçme hatası barındırmaları ve test uzunluğu gibi problemler nedeniyle çokça eleştirilmektedir. Bu problemlerin üstesinden gelebilmek için bireye uyarlanmış testler tasarlanmıştır. Günümüzde kullanılan en yaygın bireye uyarlanmış iki tip test bulunmaktadır: 1) bilgisayar ortamında madde bazında bireye uyarlanmış testler ve 2) bilgisayar ortamında modül bazında bireye uyarlanmış çok aşamalı testler. Madde bazında bireye uyarlanmış testler yüzyılı aşkın bir geçmise sahip olup bugüne kadar üzerinde çokça çalışma yapılmıştır. Bu yüzden eğitimde ve psikolojide ölçme alanı dışındaki araştırmacılar tarafından bile birçok yönü itibarıyla bilinmektedir. Fakat bireye uyarlanmış çok aşamalı testler, madde bazında bireye uyarlanan testlere göre çok daha yeni bir çalışma alanı. Bu sebeble de çok aşamalı testlerin birçok araştırmacı tarafından yeterince bilinmemektedir. Bu çalışmanın amacı bireye uyarlanmış çok aşamalı testlerin tüm özelliklerini, diğer testlerden farklılıklarını, avantajları ve dezavantajlarını araştırmacılarla paylaşmak, aynı zamanda araştırmacıların bu alana olan ilgilerini arttırmak ve bu alanın gelişmesine katkı sağlamalarına teşvik etmektir. Çalışmada ayrıca bu alanda yazılan kitaplar, kullanılan bilgisayar yazılımları ve alanla ilgili gelecekte yapılabilecek çalışmalar tartışılmıştır.

Anahtar Kelimeler: madde ve modül bazında bireye uyarlanmış testler, ölçme, psikometri

*This study is derived from a part of Ph.D. dissertation by Halil Ibrahim SARI entitled Content Control in Adaptive Tests: Computerized Adaptive Testing vs. Computerized Multistage Testing, submitted to University of Florida Graduate School under the supervision of Professor Anne Corinne HUGGINS-MANLEY.

**Post-Doctoral Associate, University of Florida, Department of Family, Youth and Community Services, Gainesville-Florida, United States of America, e-posta:hisari@ufl.edu

***Mathematics Teacher, Ministry of National Education, Yozgat-Turkey, e-posta:hsbyahsi@gmail.com

****Assistant Professor, University of Florida, Research and Evaluation Methodology Program, Gainesville-Florida, United States of America, e-posta:amanley@coe.ufl.edu

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Cilt 7, Sayı 2, Kış 2016, 388-406.

Journal of Measurement and Evaluation in Education and Psychology, Vol. 7, Issue 2, Winter 2016, 388-406.

Received: 27.09.2016

Accepted: 27.11.2016

DOI: 10.21031/epod.280183

INTRODUCTION TO TEST ADMINISTRATION MODELS

There are numerous test administration models used to measure student achievement in the educational realm. Each model has advantages and disadvantages in terms of test validity, score reliability, test fairness, cost, practical issues, test administration, and schedule. The most widely used traditional model today is paper and pencil test. In this testing approach, the exam is administered on paper, the same set of items is given to all examinees, and item order cannot change during the test (e.g., the American College Testing). Since all examinees receive the same set of items, it is relatively easy to construct the forms because they do not always necessitate creating an item pool, which requires additional time, effort, and money. In addition, easier test administration, flexible item format (e.g., open-ended items), and item review by the examinees are primary advantages associated with paper and pencil tests (Becker & Bergstrom, 2013). However, one big criticism of paper and pencil tests is their vulnerability to security breaches (i.e., cheating). This is because all questions are exposed to all test takers, which is a serious threat for test validity and score reliability (Thompson, 2008). It is possible to see some examples of paper and pencil tests that overcome this challenge by creating multiple linear forms (e.g., the Scholastic Aptitude Test-SAT). The drawback to this testing model is delayed scoring and late reporting. This requires test takers to wait for their test scores to become available, which can present a problem if application deadlines are approaching. Long test length and low measurement efficiency can be counted as other major disadvantages of paper and pencil tests (Yan, von Davier, & Lewis, 2014).

The increased role of computers in educational and psychological measurement has led testing companies (e.g., Educational Testing Service [ETS], Pearson, KAPLAN, the College Board, and American College Testing [ACT]) and practitioners to explore alternative ways to deal with the deficiencies of linear tests. One solution is adaptive testing. The most widely known adaptive testing method is computerized adaptive testing (CAT) (Weiss, 1973). CAT has a long history in the field of educational measurement. In fact, the first attempt of computerized adaptive testing was intelligence tests created by Alfred Binet in 1905 (Wainer et al., 2000). It was also used during World War I for army recruitment purposes in the USA (DuBois, 1970). Throughout the past 100 years, much research has been conducted on CAT, including new item selection methods (e.g., Barrada, Olea, Ponsoda, & Abad, 2008; Chang & Ying, 1996), stopping rules (e.g., Choi, Grady, & Dodd, 2010), and exposure control methods (e.g., Leung, Chang, & Hau, 2002; van der Linden & Chang, 2005).

The working principle of CAT is as follows: First, computer algorithms randomly administer an item (typically an item of medium difficulty) to an examinee. After her response to the first item, the computer estimates her latent score and selects the next item from the pool that best matches with her current trait level. This basically means that when she gets an item correct, the computer selects a harder question; if she gets it wrong, the computer selects an easier question. This process continues until the stopping rule is satisfied. A flowchart in Figure 1 visually summarizes this. Even though it is successful in how accurate it measures ability and how secure it is—in contrast to linear testing—CAT does have its own disadvantages. First, it requires a large item pool which incurs a high cost to testing companies. Second, it requires complicated software and fast computers to be available in test centers. The third drawback is that CAT generally does not allow examinees to review the items that have already been answered or to skip any item during the test. Test takers usually have to respond to all items, and cannot go back to previous items.

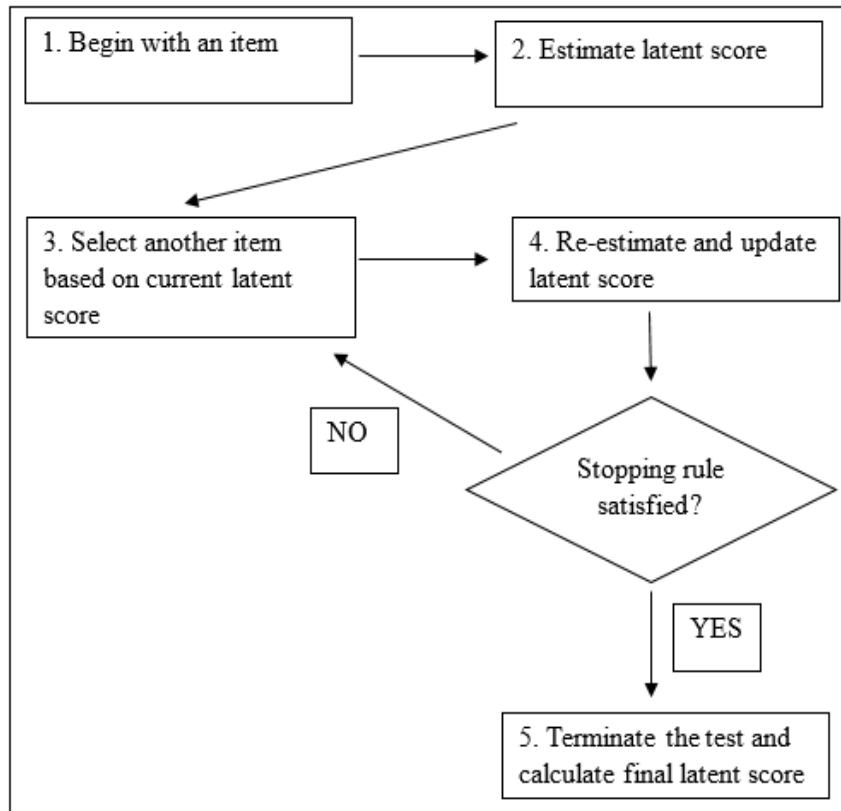


Figure 1. The Flowchart of Computerized Adaptive Testing

To deal with some of the latter's deficiencies, multistage testing (MST) has been proposed as an alternative test administration method. Over the years, it has been called by different names, such as two-stage testing (Kim & Plake, 1993), computerized mastery testing (Lewis & Sheehan, 1990), computer adaptive sequential testing (Luecht, 2000), and bundled multistage testing (Luecht, 2003). Although multistage testing is relatively new compared to linear tests and CAT, the MST idea is not new (Yan et al., 2014). Early MST designs date back to the 1950s (see Angoff & Huddleston, 1958). The first versions of MSTs were in paper-and-pencil format, and there was no adaptation from one point to another (see Cronbach & Glaser, 1965; Lord, 1971 & 1974; Weiss, 1973). Even though these initial attempts were amateur compared to today, they were invaluable in terms of growing the field of alternative test administration.

The early 1970s were the most critical years for MST advancements. That is because Fredrick Lord and David Weiss created the basis of the first item response theory (IRT)-based MST applications. When the research on CAT eclipsed MST, an adaptive version of multistage testing, which is what this study is concerned with, was proposed for use (Keng, 2008; Mead, 2006). It is called computer adaptive multistage testing (Yan et al., 2014), or ca-MST, and it has gained in popularity recently, with more than one hundred journal articles published in the past twenty years alone. In addition, the number of operational examples has increased recently—tests using the ca-MST format include the Massachusetts Adult Proficiency Test, Graduate Record Examination or GRE, Law School Admission Council or LSAC, and Certified Public Accountants or CPA Examination. After the GRE switched formats from CAT to ca-MST in 2011, interest in ca-MST increased exponentially. But despite this growing interest in ca-MST, we believe that it has not yet received the recognition that it deserves.

Purpose of the Study

Due to its long history, CAT is a well-known test administration model. However, many features of ca-MST are still unknown by many researchers, especially outside the U.S. So this paper briefly reviews the ca-MST literature, compiling recent developments for researchers newly interested in ca-MST.

The purpose of this study is to closely examine the characteristics of ca-MST, including its working principle, the adaptation procedure called the routing method, test assembly, and scoring, and provide an overview to researchers and practitioners, with the aim of drawing researchers' attention to ca-MST and encouraging them to contribute to the research in this area. To that end, this study first explains ca-MST in detail, summarizes the recent developments in the literature of the area of ca-MST, and then discusses future work in ca-MST research.

STRUCTURE AND WORKING PRINCIPLE OF CA-MULTISTAGE TESTING

The ca-MST terminology includes some special terms not used in other testing procedures. These include module, stage, panel, routing, path, and test assembly. Ca-MST is made up of different panels (e.g., a group of test forms), and those panels are, in turn, composed of different stages (e.g., division of a test). The stages themselves are made up of pre-constructed item sets, called modules, at different difficulty levels (Luecht & Sireci, 2011). This means that at each stage some of the modules are easier and some of them are harder. In other words, some modules are more appropriate for low-ability test takers, while some are more appropriate for high-ability test takers.

There is almost always one module in stage one, called the routing module, which is used to establish the test taker's proficiency level. The test taker moves to the next module of the test based on her performance on the routing module. The number of stages, the number of the modules in each stage, and the number of the items in each module can vary from test to test.

In general, the working principle of a ca-MST is as follows. After assigning a test taker to a panel, unlike individual items in CAT, ca-MST starts with a routing module (e.g., a set of five or ten items). After the routing module, the first stage of ca-MST, the computer calculates the test taker's latent performance. Then, based on her current performance, the computer selects one of the pre-constructed modules in the second stage, and routes the test taker to the appropriate module. For example, if her performance on the routing module is high, she receives a harder module in the second stage; otherwise an easier module is selected. After she completes the second stage, again, the computer calculates her performance, and routes her to the most appropriate module in the third stage. This process continues until the test taker completes all stages. This is the main distinction between CAT and ca-MST: there is an item level adaptation in CAT, in contrast to the module level adaptation in ca-MST. This feature brings the advantages of item review, item skip, higher control over test content, strict adherence to the target content distributions, and consistent item order.

Figure 2 shows an example of the simplest possible ca-MST structure, with two stages. There is one module in stage one and two modules in stage two; this structure is called the 1-2 panel design. As shown in this figure, there are two possible pathways that a test taker might draw (e.g., Routing-Easy and Routing-Hard).

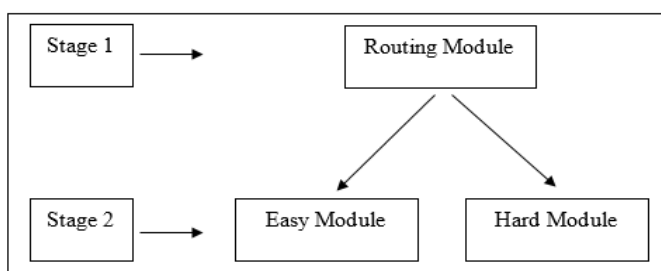


Figure 2. An Example of 1-2 Ca-MST Panel Design

Figure 3 shows a more complex ca-MST design, with three stages; there is one module in stage one and three modules in stages 2 and 3. Accordingly, this structure is called the 1-3-3 panel design. There are seven possible paths in this type of ca-MST design. These are: Routing-Easy-Easy, Routing-Easy-Medium, Routing-Medium-Easy, Routing-Medium-Medium, Routing-Medium-Hard, Routing-Hard-Medium, Routing-Hard-Hard. As can be understood from the figure, the pathways from a module to another module in the next stage that is not adjacent to the current module are ignored. For example, if a student receives easy module in stage 2, the student is not permitted to receive the hard module in stage 3, even if she performed very well in stage 2. The strategy of disallowing extreme jumps among the module is very common in many ca-MST designs (Luecht, Brumfield, & Breithaupt, 2006), and prevents aberrant item responses and inconsistent response patterns (Davis & Dodd, 2003; Yan et al., 2014). However, this is something that a test developer needs to decide prior to the test administration.

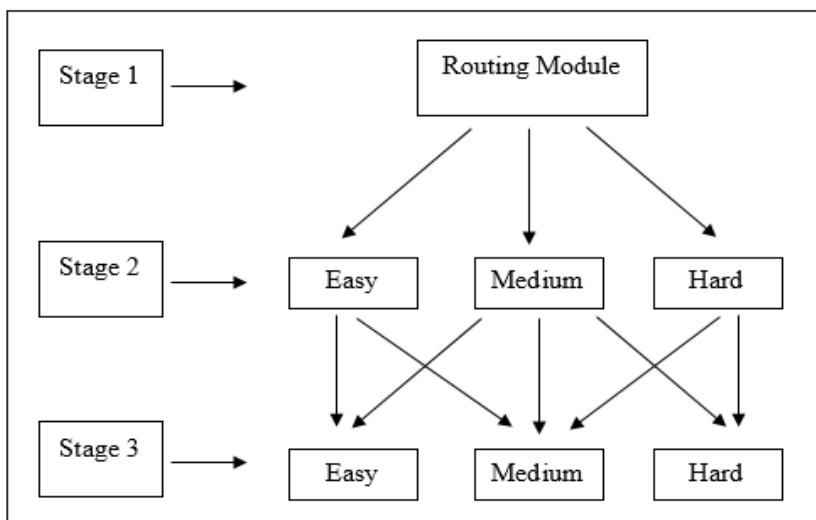


Figure 3. An Example of 1-3-3 Ca-MST Panel Design

Both Figures 2 and 3 display an example of one panel only, whereas a computer adaptive multistage panel is referred to as a collection of modules (Luecht & Nungester, 1998). In order to control panel exposure rate (and thereby module and item), ca-MST designs consist of multiple panels that are similar to each other, and each test taker is randomly assigned to one of them. Figure 4 shows an example of multiple panels. The panel construction procedure is described in the following sections.

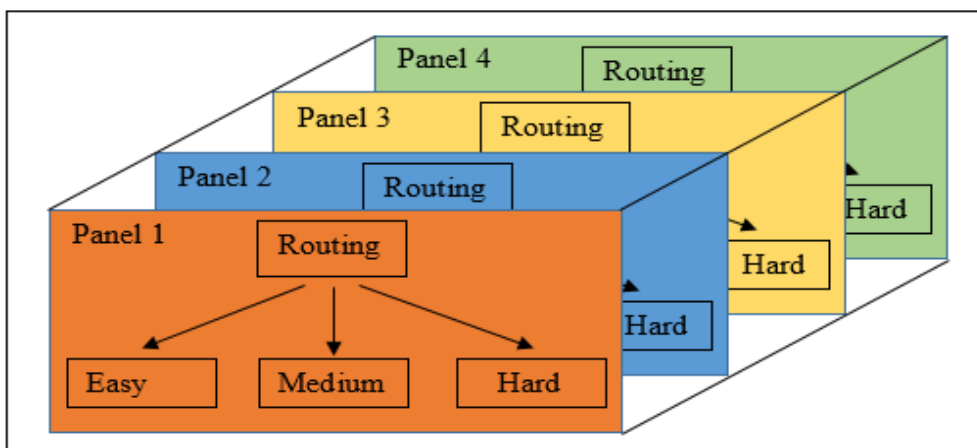


Figure 4. Illustration of Multiple Panels in Ca-MST

Like all other test administration models, ca-MST has advantages and disadvantages. Perhaps the most attractive advantage of ca-MST over CAT is that ca-MST is more flexible in terms of item review and item skipping. Ca-MST allows examinees to go back to the previous items within each module, and to skip any item as well. However, examinees are not allowed to go back to the previous stage(s), and review items in the previous module(s). Compared to linear tests, the ability to go back to a limited number of previous items is a drawback, but compared to traditional CAT, it is a remarkable feature. There is a well-known motto that the first answer is always correct, and students should always trust their initial responses (van der Linden, Jeon, & Ferrara, 2011). However, a research by Bridgeman (2012), an ETS practitioner, showed that this common belief is actually a superstition, finding an increase in student abilities when examinees had a chance to review answers (Bridgeman, 2012).

In terms of test length, ca-MST typically falls somewhere between linear tests and CAT (Hendrickson, 2007). Ca-MST is a fixed test, which means that test length is determined by the test developer; however, in CAT, the test length can be fixed or varied. In terms of test design control (e.g., content area, answer key balance, and word count), ca-MST allows more flexibility than CAT but less than linear tests. In terms of measuring ability, ca-MST is much more accurate than linear tests (Armstrong, Jones, Koppel, & Pashley, 2004; Patsula, 1999) but about as accurate as or slightly less accurate than CAT (Armstrong et al., 2004; Kim & Plake, 1993; Patsula, 1999).

However, some of CAT's drawbacks, such as the high cost of creating an item bank, are still a concern in ca-MST. Also, while CAT can stop at any point (if the stopping rule is satisfied), reducing the test time, due to the module level adaptation, ca-MST continues until all stages are completed (Zheng, Nozawa, Gao, & Chang, 2012).

By taking into account the advantages and disadvantages of ca-MST, it can be said that ca-MST is a highly promising test administration model because it combines the advantages of linear tests and CAT. Ca-MST shares some characteristics with CAT, such as test design and structure, routing method, ability estimation, content control and test assembly, and exposure control. These components are summarized in later sections.

BUILDING A COMPUTER ADAPTIVE MULTISTAGE TEST

Prior to administering a ca-MST, the test designer must determine several things. These include the number of panels, stages, modules, items, and total number of test items, as well as automated test assembly (ATA), content control, routing method, and interim and final ability estimation method. These steps are summarized in the following sections.

Number of Panels

As stated before, parallel test forms called panels comprise the ca-MST, and the assignment of test takers to the panels occurs randomly (Luecht, 2003). Having multiple panels helps to reduce panel, module, and item exposure rate, and prevents items from being overused. This is critical for test security; otherwise, test cheating and item sharing problems will arise (Yan et al., 2014). Depending on the importance of the exam (e.g., high stake or low stake), the number of panels changes, but in both operational examples and simulation studies it usually varies from one to forty (Yan et al., 2014). The preferred exposure rate in CAT generally ranges from 0.20 to .35. In order to achieve the similar exposure rate in the area of ca-MST—for example, if the desired panel exposure rate is .20 (as in the CAT version of the GRE)—one needs to create five ca-MST panels (the exposure rate is $1/r$, where r is the number of panels) (Eignor, Stocking, Way, & Steffen, 1993). Setting a low panel exposure rate means having more panels and more items.

We know that having a lower number of retired items is desired, because researchers and practitioners do not want to throw many items away after each administration. This is because item-writing is not easy work, and requires professionalism and money. A qualified item has to meet

certain criteria, such as good discrimination and content appropriateness. Thus items have to be created by content professionals. Furthermore, the estimated cost for a single qualified item ranges between \$1,500 and \$2,500 (Rudner, 2009). So if we want to have a bank size of 500, it will cost between \$750,000 and \$1,250,000. In fact, this amount will increase, as the bank needs maintenance. However, in reality, only the items in the routing module are received by all test takers that are assigned to that panel, and they are then retired for use. Yet, due to the adaptation feature of ca-MST, the following modules and thereby items are received by fewer test takers. Thus, the desired panel exposure rate is the maximum exposure rate for a panel. Since some of the items in different panels do not reach the pre-specified exposure rate, they can be used in later administrations. In other words, the number of panels should be decided by taking into account a combination of desired test security and cost. It is important to note that a test designer can have overlapping panels, which means that some of the items can be placed in more than one panel. Also, so some of the items can be saved for future administrations of the test. The panel construction procedure is handled in the following sections.

Number of Stages and Modules

As stated before, the modules in each stage differ in their average difficulty levels, but the number of items and the proportion of content specifications are the same across the modules in a stage (i.e., they are not necessarily the same in modules at different stages). There are several examples of ca-MST design configurations in the literature such as 1-2 (e.g., Wang, Fluegge, & Luecht, 2012), 1-3 (e.g., Schnipke & Reese 1999; Wang, Fluegge, & Luecht, 2012), 1-2-2 (e.g., Zenisky, 2004), 1-3-3 (e.g., Keng & Dodd, 2009; Luecht et al., 2006; Zenisky 2004), 1-2-3 (e.g., Armstrong & Roussos 2005; Zenisky 2004), 1-3-2 (e.g., Zenisky, 2004), 1-1-2-3 (e.g., Belov & Armstrong, 2008; Weissman, Belov, & Armstrong, 2007), 1-5-5-5-5 (e.g., Davey & Lee 2011), 1-1-2-3-3-4 (e.g., Armstrong et al. 2004), 5-5-5-5-5-5 (e.g., Crotts, Zenisky, & Sireci 2012). The two-stage design is the simplest and most widely used in both operational applications (e.g., the revised version of GRE) and simulation studies. This is because there is only one adaptation point in this configuration, but this same quality also has the disadvantage of introducing a higher likelihood of a routing error (Yan et al., 2014). Some have suggested using an additional module called a “recovery module” when necessary, but it was not found very interesting (see Schnipke & Reese, 1999). Thus, as previous studies showed, although the test complexity increases, adding more stages and modules into each stage and/or allowing more branching increases test outcomes (Luecht & Nungester 1998; Luecht, Nungester, & Hadadi 1996). This is likely due to having more adaptation points. Having more adaptation points makes ca-MST similar to CAT. In ca-MST, the number of adaptation points is associated with the number of stages (e.g., one minus number of stages), whereas in CAT, it is associated with the number of items (e.g., one minus number of items). For example, in a 1-3-3 ca-MST panel design, there are two adaptation points regardless of the test length. In a 20-item CAT, there are 19 adaptation points.

Some researchers believe that having multiple modules at the last stage is better for gaining accuracy in ability estimations, because the estimated abilities become closer to the test taker’s true abilities at the end of the test (Luecht & Nungester 1998). Armstrong et al. (2004) and Patsula and Hambleton’s (1999) huge simulation studies showed that having more than four stages does not produce meaningful gains in test outcomes, and two or three stages with two or three modules at each stage are sufficient for a successful ca-MST administration (Armstrong, et al., 2004; Patsula & Hambleton, 1999; Yan et al., 2014). The characteristics of modules are handled in the following sections.

Number of Items

Ca-MST is a fixed-length test, so it is necessary to decide the total test length for a test taker. The total test length varies from ca-MST to ca-MST (e.g., from 10 to 90). Then it is necessary to decide the number of items in each module. Based on the literature, tests with 20 or fewer items are usually

considered short tests (Thompson & Weiss, 2011), while tests with 60 or more questions are generally considered long tests (Zheng et al, 2012).

After deciding the total test length, it is possible to assign a different number of items to the modules at different stages. The research on number of items can be summarized as follows: a) adding more items to tests increases the reliability of test scores and thereby measurement accuracy (Crocker & Algina, 1986); b) varying the number of items in modules at different stages does not affect test outcomes (Patsula & Hambleton, 1999); c) having shorter modules is better because it allows for more adaptation points; and d) increasing the number of items in the routing module has a positive effect on test outcomes (Kim & Plake, 1993). For this last item, Kim and Plake (1993) claimed having more items decreases the likelihood of misrouting in the following stages. However, as discussed before, having more items in the modules at the last stage can also have a positive effect on test outcomes. So, it can be concluded that when we earn from one side, lose from another side (Zheng et al, 2012).

Routing Method

The routing method in ca-MST is analogous to the item selection method in CAT. Since there is a module level adaptation in ca-MST, the routing method is used to decide the subsequent modules that an examinee will receive (Luetch, 2003). The efficiency of the routing method affects the pathway a student draws during the test, so misrouting impacts the test outcomes, and therefore the usefulness of ca-MST (Lord, 1980). Thus, it is a tremendously important component in ca-MST. Routing methods are generally sorted into two categories, dynamic rules and static rules (Yan et al., 2014). The most widely known dynamic method of maximum Fisher's information (MFI) (Lord 1980; Thissen & Mislevy, 2000) is an information-based method, and uses item response theory for module assignment. MFI is analogous to the maximum information item selection method in CAT, where a computer algorithm calculates the examinee's ability level based on the previously administered module(s), and then selects the module that maximizes information at his/her current ability estimate (Weissman et al., 2007). One can refer to Lord (1980) and/or Weissman et al. (2007) for more technical details.

Both two static methods define cut points (e.g., routing points or upper and lower bounds) for latent traits when routing examinees to the modules, but they differ in defining cut points. The first static rule of the module selection method is the approximate maximum information method (AMI) (Luecht, 2000) which is mainly used in criterion-referenced test administration (Zenisky, 2004). In the AMI routing method, a computer algorithm calculates the cumulative test information function (TIF) based on the previously administered modules (see the next section for discussion on information), and the TIFs of the current alternative modules (easy, medium, or hard). Then it adds each alternative module's TIF to the current cumulative TIF separately, and defines the intersection points of TIFs as the cut points (e.g., if the two intersection points of three TIFs are -1 and 1, the cut points are -1 and 1). Finally, the computer routes examinees to the alternative module that provides the highest information for the provisional latent trait of an examinee. For example, if $\theta \leq -1$, the examinee is routed to the easy module, if $-1 \leq \theta \leq 1$, the examinee is routed to the medium module, and if $\theta \geq 1$, the examinee is routed to the hard module. By nature, AMI requires an ability estimation method after each module, which brings additional complexity to the test administration. One can refer to Luecht (2000) and/or Zenisky (2004) for more technical details.

The second static rule of the module selection method is the defined population interval (DPI) or number-correct (NC), which is mainly used in norm-referenced test administration (Zenisky, 2004). This method is currently used in the revised version of GRE (Yan et al., 2014). The main goal in this method is to route a specific proportion of people to the modules to ensure an equal or nearly equal number (or proportion) of people draws each possible pathway (e.g., 33% of examinees routed to easy module, 34% of examinees routed to medium module, and 33% of examinees routed to hard module). For example, in a 1-3-3 panel structure, first the θ values corresponding to the 33rd and 67th

percentiles of cumulative θ distribution are calculated (e.g., if the latent scores are normally distributed, θ -scores for the 33rd and 67th percentiles are -0.44 and 0.44, respectively). These are actually defined as the cut points in DPI method. Then these cut points are transformed to corresponding estimated true scores. Finally, examinees are routed to the one of the alternative modules based on the number of correct responses they got on the current module (e.g., people who got six or fewer items correct out of ten items are routed to the easy module, people who got seven or eight items correct are routed to the medium module, and people who got nine or ten items correct are routed to the hard module) (see Zenisky et al., 2010 for more details). Even if essentially or strictly parallel panels are built, it is always a good idea to check corresponding cut scores separately for each panel.

Compared to the MFI and AMI methods, the DPI is fairly straightforward to implement, but requires to an assumption on the distribution of theta scores to specify cut points prior to the test administration. Misrouting is more likely to occur when there is a huge discrepancy between the actual and assumed distributions. However, the number-correct method is a very understandable strategy by test takers (Hendrickson, 2007). Previous research has showed that in terms of routing decisions, DPI or NC is very practical and sufficient for a ca-MST design (Weissman et al., 2007). However, the information-based routing method is independent of theta distribution and may produce better outcomes than the static rules if there is discrepancy between the actual and assumed distributions. Yet, the choice of routing method is mainly determined by the purpose and consequences of the test (Zenisky et al., 2010).

Automated Test Assembly and Content Control

As explained, ca-MST is designed in such a way that test takers receive pre-constructed modules based on their performance on the previous module (Armstrong & Roussos, 2005). This means that each subsequent module has to match the current ability of a test taker. Thus, items must be carefully grouped in the modules. The most critical consideration here is to group items in modules based on the target information functions. Some have used the trial-error method and manually assembled modules (see Davis & Dodd, 2003; Reese & Schnipke, 1999), but it is quite difficult to satisfy all constraints (e.g., number of items, controlling content area) and to create parallel panels with manually assembled tests. Thus, it is always best to use a better strategy because automated test assembly (ATA) is a must in ca-MST (Luecht, 2000).

As Luecht describes, “The automated test assembly involves the use of mathematical optimization procedures to select items from an item bank for one or more test forms, subject to multiple constraints related to the content and other qualitative features” (2003, p.7). The optimum solution of ATA procedure ensures that items in modules and panels meet the desired constraints such as difficulty level, content control, word count, item and test overlap, and item format. As stated before, one has to decide ca-MST panel structure, total test length for a test taker, and total number of panels prior to ca-MST administration. Next, based on the ca-MST structure, the test designer has to determine the number of items in each module, and to pull a group of items from the given item bank that meet all desired constraints.

As shown by Luecht (1998), the automated test assembly provides a solution for maximizing the IRT information function at a fixed theta point. Let's denote θ_0 as the fixed theta point and suppose we want a total of 24 items in the test. We first define a binary decision variable, x_i , ($x_i=0$ means item i is not selected from the item bank, $x_i=1$ means item i is selected from the item bank). The information function we want maximize is;

$$I(\theta_0) = \sum_{i=1}^N I(\theta_0, \xi_i) x_i \quad (1)$$

where ξ_i represents the item parameters of item i (e.g., discrimination- a , difficulty- b , guessing- c parameters). Let's say we have two content areas (e.g., C_1 and C_2), and want to select an equal number of items from each content area. The automated test assembly is modeled to maximize

subject to

$$\sum_{i=1}^N I(\theta_0, \xi_i)x_i \tag{2}$$

$$\sum_{i \in C_1} x_i \geq 12 \tag{3}$$

$$\sum_{i \in C_2} x_i \geq 12 \tag{4}$$

$$\sum_{i=1}^N x_i \geq 24 \tag{5}$$

$$x_i \in (0,1), i = 1, \dots, N \tag{6}$$

which puts constraints on C_1 , C_2 , the total test length, and the range of decision variables, respectively. For illustration purposes, we provided an example of information functions for a 1-3 ca-MST panel design in Figure 5. As shown in this figure, the information functions for routing and the medium module peaks around $\theta=0$, the information functions for the easy and hard modules peak around $\theta=-1$ and $\theta=1$, respectively.

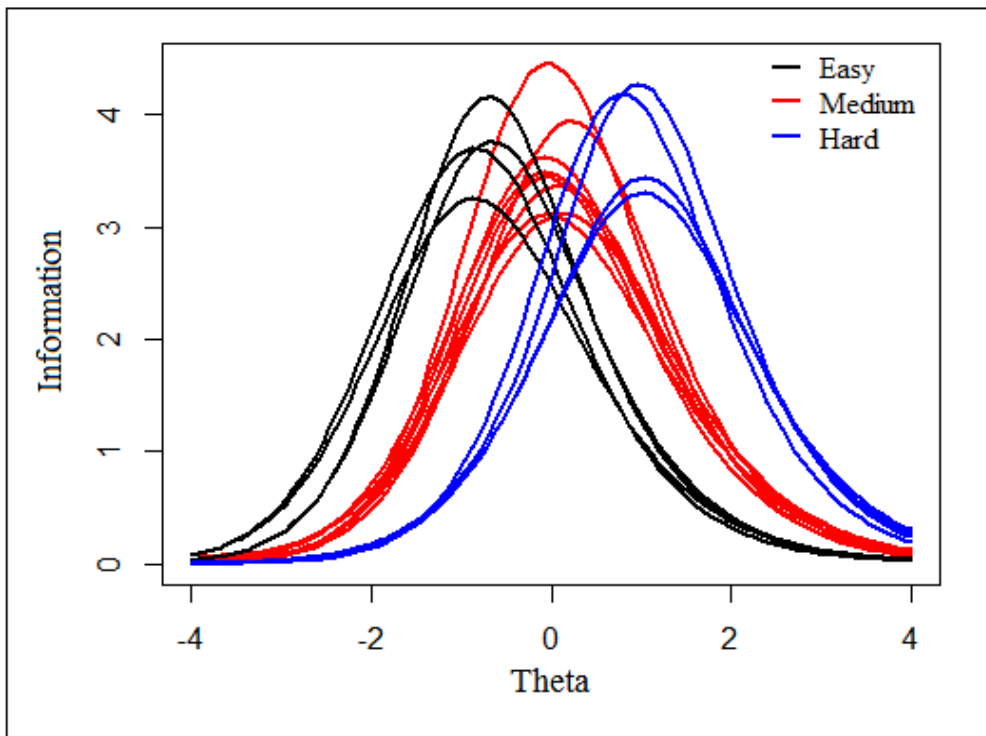


Figure 5. Test information functions across the modules at different difficulty levels

Adding content constraints in the equation is very important for meeting content requirements in the test. This is because for a successful and valid ca-MST, it is necessary to control content balancing so that each test taker takes has an equal and/or a similar number or percentage of items from each content area. It is well known that even if score precision is very high, as would be the case through proper adaptive algorithms in ca-MST, this does not necessarily ensure that the test has valid uses (Crocker & Algina, 1986). In other words, high score precision might not represent the intended construct if the content balance is not ensured. Furthermore, if all students are not tested on all aspects of the construct of interest, the test fairness is jeopardized.

For these reasons, any test form adapted to an examinee has to cover all related and required sub-content areas. Due to how it's assembled, ca-MST allows greater control over test design and content. (i.e., ca-MST designers can determine item and content order within modules). Since items

in the modules are placed by the test developer prior to the administration, ca-MST allows for strict adherence to content specification, no matter how complex it is (Yan et al., 2014). However, in CAT, content misspecifications are more likely to occur (see Leung, Chang, & Hau, 2003).

It is possible to add other constraints to the equations such as word count, especially if the test is a timed test. This is because students who receive too many items with high word counts will have a disadvantage even if the items are easy (Veldkamp & van der Linden, 2002).

One important issue when determining the structure of a ca-MST design, pulling items, and solving ATA problems is that it might not always be possible to extract a group of items from a large bank that are at the different difficulty levels. For example, if a researcher wants to build a 1-5 ca-MST panel design, and wants to allow more branching, she needs to have five modules in stage two (very easy, easy, medium, hard, and very hard). In such cases, the item bank must have groups of items with difficulty levels varying from very easy to very hard. In fact, the number of items needed for each module will increase for multiple panels. Otherwise, it will become impossible to find a proper solution for the desired structure. In short, the best solution of ATA for a desired ca-MST design can be found with a psychometrically rich item bank, because the quality of the item bank directly impacts especially complex ATA problems. It is also important to note that if item and/or test overlap is desired, a fewer number of items can be pulled from the item bank and placed in the modules. This can increase the likelihood of finding proper solution for an ATA problem.

There are some integer programming software programs used to solve ATA problems in ca-MST studies, such as IBM CPLEX (ILOG, Inc., 2006), CASTISEL (Luecht, 1998), LPSolve IDE¹, and IpSolve R package (Berkelaar, 2015). IBM Cplex is a commercial software, but a demo version allowing a limited number of constraints can be downloaded from www.ibm.com. The others are all non-commercial programs. Besides these, Microsoft Excel also provides a binary programming engine called Excel Solver Add-In. One can refer to Cor, Alves and Gierl (2009), and Diao and van der Linden (2011) for the informative works showing solving ATA problems in Excel and IpSolve R package, respectively.

After solving the ATA, the test designer has to place items into the modules and these modules into the panels. Two approaches are used to assign modules into the panels, bottom-up and top-down (Luecht & Nungester, 1998). In the bottom-up approach, all modules are built so as to meet module level specifications such as content and target difficulty. It is possible to think of each module as a mini test (Keng, 2008). Since modules constructed with this strategy are strictly parallel, the modules are exchangeable across the panels. In the bottom-down approach, the modules are built based on test level specifications. Since the modules are dependent and not parallel, they are not exchangeable across the panels.

Scoring and Ability Estimation Methods

The purpose of any test is to measure test takers' latent trait scores from responses given to a set of items (Crocker & Algina, 1986). At this point, ability estimation methods are used to calculate trait scores that represent student success. Even though classical test theory methods can be used for scoring, they are criticized by psychometric theoreticians and practitioners for producing test- and population-dependent outcomes and for focusing on true scores (Lord, 1980). Thus, we focus on IRT based scoring methods.

Like the routing method and automated test assembly, the ability estimation method is a key component in ca-MST administration. As discussed, if number-correct is the interested routing method, then ability estimation is not required when navigating examinees during the test. However, it is still required to estimate final ability level at the end of the test. If the ability cutoff-based or information-based routing methods are the interested routing methods, then it is necessary to use the

¹ <http://web.mit.edu/lpsolve/doc/>

ability estimation method both when navigating examinees and calculating final ability estimates at the end of the test.

There are two main ability estimation method groups that can be used to estimate both interim and final ability estimates: Bayesian methods and non-Bayesian methods. The most commonly used non-Bayesian method is maximum likelihood estimation (MLE) (Hambleton & Swaminathan, 1985). MLE begins with some a priori value (e.g., starting theta value) for the ability of the examinee and calculates the likelihood function with a given response pattern (e.g., 0101100111) for an ability level. The likelihood function estimates the probability of having that response vector and finds the value of theta that most likely results in that observed pattern. For example, if a person got many items wrong (e.g., response vector of 000010001), the likelihood function will tell us that this response pattern belongs to someone that has a very low latent trait. Then MLE finds the point that maximizes the likelihood of an examinee's item response pattern, and goes to the corresponding score on the latent trait scale. This score is the estimate of latent trait for a test taker. The first advantage of this method is that MLE is mathematically easier procedure compared to the other methods. Another advantage is that since item parameters are known in advance in ca-MST, estimations with MLE are unbiased compared to linear tests, in which parameters of items are unknown (Wang & Vispoel, 1998). One big disadvantage associated with MLE is that it does not provide an estimate for examinees that get all items right (i.e., perfect scores) or wrong (i.e., zero scores). This causes the likelihood function to infinitely increase and it becomes impossible to find the highest value on the likelihood function. This might be a serious problem for interim theta estimates in a ca-MST administration if there are few items in modules (Keller, 2000). One can refer to Hambleton and Swaminathan (1985) and/or Lord (1980) for technical details.

Two Bayesian methods that are commonly used are maximum a posteriori (Lord, 1986) and expected a posteriori (Hambleton & Swaminathan, 1985). Maximum a posteriori (MAP) is similar to MLE, but MAP specifies a prior distribution, multiplies this prior distribution by the likelihood function, and then, does the same thing with the MLE. Most often, the prior distribution is chosen from a normal distribution. One big advantage of MAP is that it provides estimates for perfect and zero scores, outperforming MLE (Wang & Vispoel, 1998). One can refer to Hambleton and Swaminathan (1985) and/or Keller (2000) for technical details.

Another Bayesian method is expected a posteriori (EAP). EAP also specifies a prior distribution, but unlike the highest point on the likelihood function found in MAP, it finds the mean of the posterior distribution which represents the estimate of the latent trait. Unlike MLE and MAP, it is a non-iterative procedure, so it is easier to implement. Also, it does not assume normal prior distribution as MAP, and EAP outperforms MLE and MAP, meaning that it produces a lower standard error and bias than other ability estimation methods. One possible disadvantage with EAP is that if inappropriate prior distribution is specified, this affects the accuracy of outcomes. One can refer to Wainer and Thissen (1987) and/or Hambleton and Swaminathan (1985) for technical details.

One important issue for scoring and ability estimation is that when scoring the theta values, it is recommended to count all items an examinee received during the test (except pre-tested or seeded items) (Weissman et al., 2007). Then the estimated theta scores can be transformed to the desired rating scale (e.g., from 0 to 100). This is because the theta scores theoretically ranges from $-\infty$ to $+\infty$, but typically in practice ranges from -3 to 3 (Baker, 1992).

ON THE FLY COMPUTER ADAPTIVE MULTISTAGE TESTING

The main property of computer adaptive multistage testing is that selects the most appropriate module to an examinee and everyone works his/her own pace. As discussed before, the modules are pre-constructed "in house" and cannot change during the test. Due to this feature, the ca-MST aligns with linear tests and a test developer knows the prospective modules a test taker will receive during the test. However, in CAT, a test developer does not know which items a test taker will receive from

a large item bank because CAT is a “on the fly” test. This feature of CAT has become an inspiration for ca-MST users, and on the fly multistage testing (OMST) has been proposed (Han & Guo, 2013).

Instead of administering pre-assembled modules in ca-MST, the OMST shapes item modules for theta interim values during the test. Due to this variation from ca-MST, some refer to on-the-fly MST as MST-shaping (MST-S), and ca-MST as MST-routing (MST-R) (see Han & Guo, 2014; Zheng & Chang, 2015). MST-S still holds other features of ca-MST such as item skipping and review, and it produces comparable results with ca-MST (Han & Guo, 2014). In summary, while ca-MST is somewhere between linear tests and CAT, MST-S is between ca-MST and CAT.

BOOKS AND SOFTWARE FOR COMPUTER ADAPTIVE MULTISTAGE TESTING

The first attempt to put together the work conducted on computer adaptive multistage testing was done by the Journal of Applied Measurement in Education, and a special volume (i.e., Volume 3) was organized in 2006. Then a special chapter (i.e., Chapter 18), “Multistage Testing: Issues, Designs, and Research,” put together by April Zenisky, Ronald Hambleton, and Richard Luecht (2010) was published in the book of Elements of Adaptive Testing, edited by Wim J. van der Linden and Cees A. W. Glass (2010). Next, three ETS researchers, Duanli Yan, Alina A. von Davier, and Charles Lewis published the first ca-MST book, “Computerized Multistage Testing: Theory and Applications” (2014). To the best of our knowledge, this is the first and only book written for computer adaptive multistage testing.

The field of computer adaptive multistage testing is not very rich in terms of available computer programs or software yet. Two software programs can be used by researchers for their simulation studies: a) MSTGen (Han, 2013) and b) R (R Development Core Team, 2009-2016). The former, written by Kyung T. Han in 2013, is a Windows-based program, and is fairly user-friendly. It is available at no cost and can be downloaded from the author’s website.² The current version of MSTGen supports both MST-R and MST-S. MSTGen supports three routing methods (maximum fisher information, matching-*b* value, and random module selection), as well as three theta scoring methods (MLE, MAP and EAP). MSTGen also supports creating and analyzing multiple panels. More information can be found in the user manual.³

The second program, R, is the most widely used open source program today and can be downloaded online.⁴ R consists of user-created packages which include pre-written statistical commands. However, there is no special R package written for computer adaptive multistage testing yet. Thus, researchers have to write their own commands to run simulations for testing their ca-MST designs. Han and Kosinski (2014) provided an example R code that analyzes a ca-MST panel design found in Yan et. al. (2014) (see pages from 417 to 419 in Chapter 26). However, it is important to note that this code may not always serve researchers’ intended study purposes. This is because this R code is limited to the maximum fisher information routing method, and is written for a special case where there must be equal number of items in all modules, which may not be always desired. This code also does not have a mechanism that prevents extreme jumps among the modules, which might cause aberrant response patterns.

Earlier this year, David Magis, the author of the computer adaptive testing R package called catR (Magis & Raiche, 2012) announced in the International Meeting of Psychometric Society Meeting in Asheville, N.C. that he is currently writing a ca-MST package named mstR. He plans to make it available for R users in fall 2017 (D. Magis, personal communication, July 11, 2016). MstR is going to support a variety of routing methods and ability scoring methods, and will not allow extreme jumps among the modules if user desires. The authors of this paper are also writing an R routine for

² <http://www.umass.edu/remf/software/simcata/mstgen/>

³ http://www.umass.edu/remf/software/simcata/mstgen/MSTGen_Manual.pdf

⁴ <https://cran.r-project.org/mirrors.html>

ca-MST analyses, and hope to release it in early 2017. The current version of the R routine can be requested from the first author of this paper.

DISCUSSION AND FUTURE RESEARCH ON CA-MST

Historically, student success has been mostly measured by linear test administration methods. However, these conventional methods are more accurate for measuring the ability of students at the medium ability level and less so for students at extreme ability points (Weiss & Kingsbury, 1984). They also produce longer tests (Segall, 2005). With modern advances in technology and measurement theory, adaptive testing has been proposed to cope with the problems posed by linear tests. So far, computerized adaptive testing has been the most popular and commonly preferred adaptive testing. During the last century, much research has been conducted on CAT. We strongly believe that even people outside of the educational and psychological measurement field are fairly familiar with CAT. Unfortunately, the same thing is not true for computer adaptive multistage testing.

This study does not argue against linear tests and/or computerized adaptive testing. Each test administration method has advantages and disadvantages when compared to the others. We believe that some superior features of linear tests, such as high test developer control on content balancing and item review, are vital. Similarly, some superior features of CAT, such as high measurement precision for theta estimates and shorter test lengths are also extremely important. To be honest, it is not possible to have all these features in the same test administration model.

This study argues that the computer adaptive multistage testing (ca-MST) “strikes a balance among adaptability, practicality, measurement accuracy, and control over test forms” (Zenisky, Hambleton, & Luetch, 2010, p.369) and combines all practical advantages of other test administration models. Despite these qualities, we believe that ca-MST has not been given enough attention and consideration, especially by researchers and practitioners in Turkey. We aim to arouse interest in ca-MST and encourage Turkish researchers to contribute this highly promising field.

Furthermore, we know that operational applications of adaptive testing are now being used in many European countries, the U.S., and Canada. Unfortunately, despite a large number of dedicated researchers and rapid advancements in technology, an operational example of adaptive testing has never been used in Turkey. Testing companies in the U.S. such as Educational Testing Service lead in adaptive testing. This was seen when the GRE switched from CAT to the ca-MST format, and the interest in ca-MST noticeably increased after 2011. The research reports released and distributed by ETS inspire researchers to find new research questions. We strongly believe that the Measuring, Selection and Placement Center (i.e. abbreviated as OSYM in Turkish) can have the same impact on the researchers in Turkey by releasing an operational example of adaptive testing and distributing research reports associated with it.

Ca-MST has fewer number of routing methods. Whereas some item selection methods used in CAT such as Kullback-Leibler (Chang & Ying, 1996), maximum likelihood weighted information (Veerkamp & Berger, 1997), the maximum posterior weighted information (van der Linden, 1998) can be easily adopted and modified, and then used as a routing method in the ca-MST environment. Furthermore, some special and popular topics such as differential item functioning, item parameter drift, item copying, and cognitive diagnostic models should be investigated in the ca-MST environment. Also, a computer software comparison study across MSTGen and R should be conducted to compare the effectiveness, usefulness, and accuracy of these two software programs. And last but not least, another comparison study looking at IBM Cplex, CASTISEL, LPSolve, and IpSolve R package should be conducted to explore the effect of integer programming software when creating ca-MST panel designs. Researchers can always contact the authors with questions and for help regarding ca-MST design, software, coding, ATA, etc.

REFERENCES

- Angoff, W. H., & Huddleston, E. M. (1958). *The multi-level experiment: a study of a two-level test system for the College Board Scholastic Aptitude Test (SR-58-21)*. Princeton, New Jersey: Educational Testing Service.
- Armstrong, R. D. & Roussos, L. (2005). *A method to determine targets for multi-stage adaptive tests*. (Research Report 02-07). Newtown, PA: Law School Admissions Council.
- Armstrong, R. D., Jones, D. H., Koppel, N. B., & Pashley, P. J. (2004). Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement*, 28, 147- 164.
- Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2008). Incorporating randomness to the Fisher information for improving item exposure control in CATS. *British Journal of Mathematical and Statistical Psychology*, 61, 493-513.
- Becker, K. A., & Bergstrom, B. A. (2013). Test administration models. *Practical Assessment, Research & Evaluation*, 18(14), 7.
- Belov, D. I., & Armstrong, R. D. (2005). Monte Carlo test assembly for item pool analysis and extension. *Applied Psychological Measurement*, 29, 239-261.
- Berkelaar, M. (2015). Package 'lpSolve'.
- Bridgeman, B. (2012). A Simple Answer to a Simple Question on Changing Answers. *Journal of Educational Measurement*, 49, 467-468.
- Chang, H.H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20, 213–229.
- Choi, S. W., Grady, M. W., & Dodd, B. G. (2010). A New Stopping Rule for Computerized Adaptive Testing. *Educational and Psychological Measurement*, 70(6), 1–17.
- Cor, K., Alves, C., & Gierl, M. (2009). Three Applications of Automated Test Assembly within a User-Friendly Modeling Environment. *Practical Assessment, Research & Evaluation*, 14(14).
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Cronbach, L. J., & Glaser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Crotts, K. M., Zenisky, A. L., & Sireci, S. G. (2012, April). *Estimating measurement precision in reduced-length multistage-adaptive testing*. Paper presented at the meeting of the National Council on Measurement in Education, Vancouver, BC, Canada.
- Davey, T., & Y.H. Lee. (2011). Potential impact of context effects on the scoring and equating of the multistage GRE Revised General Test. (GRE Board Research Report 08-01). Princeton, NJ: Educational Testing Service.
- Davis, L. L., & Dodd, B. G. (2003). Item Exposure Constraints for Testlets in the Verbal Reasoning Section of the MCAT. *Applied Psychological Measurement*, 27, 335-356.
- Diao, Q., & van der Linden, W. J. (2011). Automated test assembly using lp_solve version 5.5 in R. *Applied Psychological Measurement*, DOI: 0146621610392211.
- Dubois, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon
- Eignor, D. R., Stocking, M. L., Way, W. D., & Steffen, M. (1993). Case studies in computer adaptive test design through simulation. *ETS Research Report Series*, 1993(2).
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications* (Vol. 7). Springer Science & Business Media.
- Han, K. T. (2013). " MSTGen": Simulated Data Generator for Multistage Testing. *Applied Psychological Measurement*, 37, 666-668.
- Han, K. T., & Kosinski, M. (2014). Software Tools for Multistage Testing Simulations. In *Computerized Multistage Testing: Theory and Applications* (pp. 411-420). Chapman and Hall/CRC.
- Han, K.T., & Guo, F. (2013). *An Approach to Assembling Optimal Multistage Testing Modules on the Fly* (Report No. RR-13-01). Reston, Virginia: Graduate Management Admission Council. Retrieved from GMAC website: <http://www.gmac.com/market-intelligence-and-research/research-library/validity-and-testing/research-reports-validity-related/module-assembly-on-the-fly.aspx>
- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, 26(2), 44-52.
- ILOG. (2006). ILOG CPLEX 10.0 [User's manual]. Paris, France: ILOG SA.
- Keller, L. A. (2000). Ability estimation procedures in computerized adaptive testing. USA: American Institute of Certified Public Accountants-AICPA Research Consortium-Examination Teams.

- Keng, L. & Dodd, B.G. (2009, April). *A comparison of the performance of testlet based computer adaptive tests and multistage tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Keng, L. (2008). *A comparison of the performance of testlet-based computer adaptive tests and multistage tests* (Order No. 3315089).
- Kim, H., & Plake, B. S. (1993, April). *Monte Carlo simulation comparison of two-stage testing and computerized adaptive testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Atlanta, GA.
- Leung, C.K., Chang, H.H., & Hau, K.T. (2002). Item selection in computerized adaptive testing: Improving the a-Stratified design with the Sympon-Hetter algorithm. *Applied Psychological Measurement, 26*(4), 376–392.
- Leung, C.K., Chang, H.H., & Hau, K.T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. *Journal of Technology, Learning, and Assessment, 2*(5).
- Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement, 14*(4), 367-386.
- Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika, 36*, 227–242.
- Lord, F. M. (1974). Practical methods for redesigning a homogeneous test, also for designing a multilevel test. Educational Testing Service RB-74–30.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement, 23*, 157–162.
- Luecht, R. M. & Sireci, S. G. (2011). A review of models for computer-based testing. Research (Report No: 2011-12). New York: The College Board. Retrieved from website: <http://research.collegeboard.org/publications/content/2012/05/review-models-computer-based-testing>
- Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement, 22*, 224-236.
- Luecht, R. M. (2000, April). *Implementing the computer-adaptive sequential testing (CAST) framework to mass produce high quality computer-adaptive and mastery tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Luecht, R. M. (2003, April). *Exposure control using adaptive multi-stage item bundles*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Luecht, R. M., & Nungester, R. J. (1998). Some Practical Examples of Computer- Adaptive Sequential Testing. *Journal of Educational Measurement, 35*, 229-249.
- Luecht, R. M., Brumfield T., & Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. *Applied Measurement in Education, 19*, 189–202.
- Luecht, R. M., Nungester, R.J., & Hadadi, A. (1996, April). *Heuristic-based CAT: Balancing item information, content and exposure*. Paper presented at the annual meeting of the National Council of Measurement in Education, New York.
- Magis, D., & Raïche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software, 48*(8), 1-31.
- Mead, A. D. (2006). An introduction to multistage testing. *Applied Measurement in Education, 19*, 185-187.
- Patsula, L. N. & Hambleton, R.K. (1999, April). *A comparative study of ability estimates from computer adaptive testing and multi-stage testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec.
- Patsula, L. N. (1999). *A comparison of computerized adaptive testing and multistage testing* (Order No. 9950199). Available from ProQuest Dissertations & Theses Global. (304514969)
- R Development Core Team. (2013). *R: A language and environment for statistical computing, reference index* (Version 2.2.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Rudner, L. M. (2009). Implementing the graduate management admission test computerized adaptive test. In *Elements of adaptive testing* (pp. 151-165). Springer New York.
- Schnipke, D. L., & Reese, L. M. (1999). A Comparison [of] Testlet-Based Test Designs for Computerized Adaptive Testing. Law School Admission Council Computerized Testing Report. LSAC Research Report Series.
- Segall, D. O. (2005). Computerized adaptive testing. *Encyclopedia of social measurement, 1*, 429-438.
- Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2nd ed., pp. 101-133). Hillsdale, NJ: Lawrence Erlbaum.

- Thompson, N. A. (2008). A Proposed Framework of Test Administration Methods. *Journal of Applied Testing Technology*, 9(5), 1-17.
- Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, 16(1), 1-9.
- van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63, 201-216. doi: 10.1007/BF02294775
- van Der Linden, W. J., & Chang, H. H. (2003). Implementing content constraints in alpha-stratified adaptive testing using a shadow test approach. *Applied Psychological Measurement*, 27, 107-120.
- van der Linden, W.J., Jeon, M., & Ferrara, S. (2011). A paradox in the study of the benefits of test-item review. *Journal of Educational Measurement*, 48, 380-398.
- Veerkamp, W. J. J., & Berger, M. P. F. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22, 203-226. doi: 10.3102/10769986022002203
- Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67, 575-588.
- Wainer, H., & Thissen, D. (1987). Estimating ability with the wrong model. *Journal of Educational and Behavioral Statistics*, 12, 339-368.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislavy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge.
- Wang, T., & Vispoel, W. P. (1998). Properties of ability estimation methods in computerized adaptive testing. *Journal of Educational Measurement*, 35, 109-135.
- Wang, X., Fluegge, L., & Luecht, R.M. (2012, April). *A large-scale comparative study of the accuracy and efficiency of ca-MST panel design configurations*. Paper presented at the meeting of the National Council on Measurement in Education, Vancouver, BC, Canada.
- Weiss, D. J. (1973). *The stratified adaptive computerized ability test* (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Weiss, D. J., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.
- Weissman, A., Belov, D., & Armstrong, R. (2007). *Information-based versus number-correct routing in multistage classification tests*. (LSAC Research Report No:07-05). Newtown, PA: Law School Admissions Council.
- Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2014). *Computerized multistage testing: Theory and applications*. CRC Press.
- Zenisky, A. L. (2004). *Evaluating the effects of several multi-stage testing design variables on selected psychometric outcomes for certification and licensure assessment* (Order No. 3136800).
- Zenisky, A., Hambleton, R. K., & Luecht, R. M. (2010). Multistage testing: Issues, designs, and research. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 355-372). New York: Springer.
- Zheng, Y., & Chang, H. H. (2015). On-the-fly assembled multistage adaptive testing. *Applied Psychological Measurement*, 39(2), 104-118.
- Zheng, Y., Nozawa, Y., Gao, X., & Chang, H. H. (2012). *Multistage Adaptive Testing for a Large-Scale Classification Test: Design, Heuristic Assembly, and Comparison with Other Testing Modes*. *ACT Research Report Series*, 2012 (6). ACT.

GENİŞ ÖZET

Öğrenci yeteneğini veya başarısını ölçmek birçok testin öncelikli amacıdır. Bu amaç doğrultusunda en çok başvurulan yöntem klasik yöntem diye adlandırılan kağıt-kalem formatındaki testlerdir. Bu yöntemin en büyük avantajlarından biri test formlarının hazırlanmasında testi organize eden kişiye test içeriğini oluşturmada büyük kolaylık sağlamasıdır. Bunun nedeni testi hazırlayan kişinin test içeriğini, testteki maddelerin sırasını, soru sayısını istediği gibi belirleyebilmesidir. Ayrıca bu yöntem soru bankası oluşturmayı gerektirmediği için daha az maliyetlidir. Ancak bu yöntem öğrenci başarısını ölçmede yüksek yanlışlık (hata) ürettiği ve test uzunluğu nedeniyle oldukça eleştirilmektedir. Bu sebeple bilgisayar ortamında uygulanan bireye uyarlanmış testler geliştirilmiştir.

Günümüzde en yaygın kullanılan ve bilinen bireye uyarlanmış test; madde bazında bireye uyarlanmış testlerdir. Bu test yönteminin genel çalışma prensibi şu şekildedir; bilgisayar bireye başlangıç için bir soru verir (genellikle orta zorluk derecesinde), bireyin bu maddeye verdiği cevap sonrasında, bilgisayar bireyin yetenek seviyesini hesaplar, hesaplanan yeni yetenek seviyesine göre

başka bir soru verir ve yetenek seviyesini günceller. Bu işlem test sonuna kadar devam eder. Testi sona erdiren mekanizma testi organize eden kişi tarafından belirlenir ve a) süre, b) önceden belirlenen standart hata (örneğin kişinin yetenek seviyesi 0.3 hata oranıyla hesaplandığında testi durdur) veya c) önceden belirlenen test uzunluğu olabilir (örneğin her bir birey 30. sorusunu aldığı anda testi durdur). Madde bazında bireye uyarlanan testin en büyük avantajı yetenek seviyesini minimum hata ile hesaplayabilmesi ve soru sayısını %50 oranında azaltmasıdır. Kısacası daha az soru ile daha iyi bir ölçmeyi gerçekleştirmesidir. Ancak bu yöntemin de kendine özgü dezavantajları bulunmaktadır. Belki de en ciddi olanı yüksek maliyet gerektirmesidir çünkü bu test yöntemi soru bankası oluşturmayı zorunlu kılmaktadır. Bu yöntemin diğer dezavantajı bireylere önceki sorulara gerip dönüp cevaplarını değiştirme veya gözden geçirme şansı tanımaması ve her bir sorunun cevaplanmasını zorunlu kılmasıdır.

Bilgisayar ortamında uygulanan diğer metod ise bireye uyarlanmış çok aşamalı testlerdir. Bireye uyarlanmış çok aşamalı testler isminden anlaşılacağı üzere çeşitli bölümlerden oluşur ve her bir bölümde modül adı verilen farklı zorluk derecesinde soru kümeleri bulunmaktadır. Bu bağlamda her bir modülü mini bir test olarak düşünmek mümkündür. İlk aşamada genellikle tek modül bulunur ve yönlendirme modülü (routing module) olarak adlandırılır. Fakat diğer aşamalarda farklı zorluk derecelerinde birden fazla modül bulunmaktadır. Bireye uyarlanmış çok aşamalı testlerin çalışma prensibi şu şekildedir; birey önce yönlendirme modülünü alır ve bu modülde göstermiş olduğu performansa bağlı olarak ikinci aşamada düşük, orta veya yüksek zorluk seviyesindeki modülü alır ve ikinci aşamayı tamamlar. Bu işlem birey tüm aşamaları bitirinceye kadar devam eder. Anlaşıldığı üzere madde düzeyindeki bireyselleşmeden ziyade, bireyselleşme modül düzeyinde gerçekleşmektedir. Testi hazırlayan kişi testteki aşama sayısını kendisi belirleyebilir ve her aşamaya istediği kadar modüle yerleştirebilir. Şekil 2 ve 3 çok aşamalı testlerin yapısını gösteren birer örnektir. Bu şekillerde sadece bir panel gösterilmektedir. Halbuki çok aşamalı testler birbirine paralel olan çok sayıda panelden oluşmaktadır ve bireyler herhangi bir panele rastgele atanır. Bu panellerin bireylere maruz bırakılma veya kullanım oranını belirli seviyede tutma ve test güvenliğini sağlama açısından oldukça önemlidir. Bireye uyarlanan çok aşamalı testlerin madde düzeyinde bireye uyarlanan testlere göre en büyük avantajlarından biri bireylere her bir modül içerisindeki önceki sorulara geri dönüp cevapları gözden geçirmesine imkan vermesidir. Ancak bireylerin bir önceki aşamadaki modülde aldıkları sorulara geri dönmelerine izin verilmez. Diğer bir avantajı ise testi hazırlayan kişiye test içeriği üzerinde daha fazla imkan vermesidir. Bunun nedeni modül içerisindeki sorular testi hazırlayan tarafından önceden belirlenir, istenilen sayıda ve sırada soru yerleştirilebilir. Bireye uyarlanan çok aşamalı testler bireylerin yetenek seviyelerini ölçmedeki hata derecesi açısından madde bazında bireye uyarlanan testlerden kötü, kağıt kalem formundaki testlerden iyidir. Fakat madde bazında uyarlanmanın olduğu testlerde karşılaşılan yüksek maliyet gibi sorunlar çok aşamalı testlerde de görülmektedir. Bu nedenlerden dolayı bireye uyarlanan çok aşamalı testlerin diğer iki testin avantajlarını ve dezavantajlarını içermektedir. Çok aşamalı bir test yönlendirme metodu denilen adaptasyonu gerçekleştiren mekanizmayı kullanmayı gerektirir. Yönlendirme metodu bireylerin bir sonraki aşamada hangi modülü alacağını belirler ve bireylerin yetenek seviyelerinin hesaplanmasında çok önemli rol oynar. Bu nedenle çok aşamalı testlerin en önemli unsurlarından biridir. Çalışma en yaygın kullanılan üç farklı yönlendirme metodunu detaylı şekilde açıklamaktadır. Bunlar doğru sayısına göre, hesaplanan yetenek seviyesine göre ve test bilgisine (test information) göre yönlendirme metodlarıdır. Çok aşamalı testlerdeki önemli noktalardan biri de “automated test assembly” (ATA) olarak adlandırılan madde havuzundaki modülleri oluşturmaya yarayan test toplama/bir araya getirme metodudur. ATA farklı zorluk derecelerindeki maddeleri bir araya getirmeye yarar. Bunu yaparken farklı her bir modül içerisinde o teste ait farklı alt konulara (denklemler, fonksiyonlar, toplama/çıkarma vs.) ait soruları bir araya getirmeye dikkat edilmelidir. Çalışmada panellerin ve modüllerin nasıl oluşturulduğunu detaylıca anlatmaktadır. Diğer bir önemli unsur ise yetenek seviyesini hesaplama metodudur. En yaygın kullanılan metodlar “en çok olabilirlik (MLE), sonsal maksimum kestirim (MAP) ve sonsal beklenti kestirimi (EAP) metodlarıdır. Her bir metodun sınırlılıkları ve üstün yönleri çalışmada detaylıca anlatılmıştır. Bireye uyarlanan çok aşamalı testler üzerine yazılmış kitaplar ve tasarlanmış bilgisayar

yazılımları çalışmada tanıtılmıştır. Çalışmada ayrıca geçmişten günümüze yapılan çalışmalar özetlenmiş ve gelecekte yapılabilecek çalışmalar tartışılmıştır.

Madde bazında bireye uyarlanan testlerin ilk adımları 1905 yılında atılmış olup yüzyılı aşkın bir geçmişi bulunmaktadır. Günümüze gelene kadar üzerinde çokça çalışma yapılmış ve ciddi ilerleme sağlanmıştır. Bu yüzden eğitimde ve psikolojide ölçme alanı dışındaki araştırmacılar tarafından bile birçok yönü itibariyle bilinmektedir. Fakat modül bazında bireye uyarlanan veya bireye uyarlanan çok aşamalı testler diğer test yöntemlerine göre çok daha yeni olduğu için birçok araştırmacı tarafından bilinmemekte veya kısmen bilinmektedir. Bu çalışmanın amacı bireye uyarlanan çok aşamalı testleri tüm yönleriyle incelemek, temel prensiplerini anlatmak, gelinen noktayı ve yapılan çalışmaları araştırmacılar için özetlemektir. Türkiye’de bu alanda yapılan çalışmalar çok kısıtlı olduğu için hedefimiz Türkiye’deki araştırmacıların dikkatini bu alana çekmek ve katkıda bulunmalarına teşvik etmektir. Kuşkusuz bireye uyarlanan çok aşamalı testler avantajları itibariyle bireye uyarlanan madde bazındaki testlerin yerini almaya aday. Bu nedenle de çok aşamalı testlere olan ilgi gelecekte daha da artacaktır. Özellikle 2011 yılında Graduate Record Examination (GRE) madde bazında bireyselleştirmeden modül bazında bireyselleştirmeye geçtikten sonra, çok aşamalı testlere olan ilgi gözle görünür şekilde artmıştır. Bunu yayınlanan akademik araştırma ve yapılan simülasyon çalışmalarının sayısından rahatça görmekteyiz. Benzer ilginin Türkiye’de de artması için yapılabilecek çalışmalara değinilmiştir.

Öğretmen – Öğrenci İlişki Ölçeği'nin (ÖÖİÖ) Uyarlanması*

Adaptation of Teacher – Student Relationship Scale (STRS)

Selen DEMİRTAŞ ZORBAZ**

Arif ÖZER***

Dilek GENÇTANIRIM-KURT****

Tuncay ERGENE*****

Öz

Öğretmen – öğrenci ilişkisi, çocuğun okulla tanıştığı ilk andan itibaren sosyal ve duygusal gelişimini belirleyen bir unsurdur. Söz konusu ilişkinin niteliği birçok önleyici ve müdahale edici çalışmaya yol gösterebilecektir. Bu çalışmanın amacı Pianta (2001) tarafından Amerika'da geliştirilen ve İtalya, Hollanda, Yunanistan, Norveç ve Çin'de de uyarlama çalışması yapılan Öğretmen – Öğrenci İlişki Ölçeğinin Türkiye uyarlamasını gerçekleştirmektir. Araştırma, 46 öğretmenin 560 öğrenci için doldurdıkları ölçekten elde edilen veriler üzerinde yürütülmüştür. Ölçeğin geçerlik çalışması kapsamında doğrulayıcı ve açıklayıcı faktör analizi yapılmış ve ölçeğin iki faktör ve 23 maddeden oluştuğu belirlenmiştir. Güvenirlilik çalışmaları kapsamında Cronbach Alfa katsayıları yakınlık alt boyutu için .83; çatışma alt boyutu için .86 olduğu tespit edilmiştir. Yapılan bu çalışmalar ile ölçeğin, Türkiye'de kullanımı için gerekli geçerlik ve güvenilirlik kanıtları sağlanmış ve ölçek farklı çalışmalar için kullanılmak üzere alana kazandırılmıştır.

Anahtar Kelimeler: öğretmen – öğrenci ilişkisi, ilkökul, çatışma, yakınlık, STRS

Abstract

Teacher – student relationship is a critical factor which play a role on child's social and emotional development from the beginning of the school life. Assessing this relationship's quality can lead the preventive and intervention studies. From this point of view the aim of the study is to do the adaptation of the Student-Teacher Relationship Scale which developed by Pianta (2001) and adapted to other cultures such as Italy, Holland, Greek, Norway and China. 46 teachers were asked to fill the scale for 560 students. STRS's construct validity was assessed with confirmatory and exploratory factor analyses and it has seen that Turkish version of STRS has two dimension and 23 items. Cronbach's Alpha coefficient was .83 for closeness and Cronbach's Alpha coefficient was .86 for conflict. Results showed that STRS is a valid scale to be used in Turkey and can be used in different studies.

Keywords: teacher – student relationship, elementary school, conflict, closeness, STRS

GİRİŞ

Çocuklar doğumlarından itibaren yetişkinlerle etkileşim halinde olmaktadırlar. Çocuklara birincil bakım veren anne, baba, diğer aile bireyleri, bakıcı, dadı veya bakımevleri gibi kurumlardaki görevliler gibi yetişkinlerle kurulan ilişki farklı kuramcılarının (Bronfenbrenner, 1986; Bowlby; 2012) açıkladığı üzere çocukluk döneminde çocukların gelişimini belirleyicidir. Beş altı yaşlardan itibaren çocukların ilişki kurduğu önemli yetişkinlerin başında anne babalardan sonra öğretmenler gelmektedir. Çocuklar öğretmenlerle genellikle okul öncesi ya da ilkökul döneminde

* Bu makale, TÜBİTAK tarafından desteklenen Prof. Dr. Tuncay Ergene'nin yürütücülüğünü yaptığı "İlkokul Birinci Sınıf Öğrencilerinin Okula Uyumu" başlıklı ve 115K448 numaralı SOBAG projesinin bir kısmından ve Selen Demirtaş Zorbaz'ın Doktora tez çalışmasının bir kısmından oluşturulmuştur.

** Arş. Gör.Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, selendemirtas@hacettepe.edu.tr

***Doç. Dr., Hacettepe Üniversitesi, Ankara-Türkiye, e-posta: arifozer@hacettepe.edu.tr

****Yrd. Doç. Dr., Ahi Evran Üniversitesi, Kırşehir-Türkiye, e-posta: digenc@gmail.com

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

Prof. Dr. Tuncay Ergene, Boston Children's Hospital'da bulunan ve Principle Investigator (PI): Kerim M. Munir, MD, MPH, DSC.'nin olduğu NIH, ICORTHA, Fogarty International Mental Health and Developmental Disabilities (MH/DD) Research Training Program (D43TW05807) tarafından desteklenmektedir.

karşılaşmaktadırlar. İlkokul birinci sınıf öğrencilerinin ailelerinden ayrılarak yeni bir çevreye girdiği bir dönemde olduğu düşünüldüğünde, öğrencilerin yakın çevresinden ayrılıp okula başlaması ve ailesinin dışında bir yetişkinle uzun süre vakit geçirmesi öğrenciler için bu dönemi daha kritik hale getirmektedir (Demirkaya, 2013). Pomeroy (1999)'a göre öğrenci öğretmen ilişkisi okullardaki iletişim ağında anahtar rol oynayan bir unsurdur. Öğretmen öğrenci ilişkisi özellikle okul ortamında, “hem öğretmenin hem de öğrencinin öğrenmeyi teşvik eder bir şekilde davranması gerektiği” ilkesine dayanmaktadır (Muller, 2001). Bu da öğretmen öğrenci ilişkisinin eğitim öğretim faaliyetlerine ulaşmada belirleyici bir öneme sahip olduğunun göstergesidir.

Öğrencilerin akademik başarıları üzerinde öğretmen – öğrenci ilişkilerinin etkisi araştırmalarla ortaya konmuştur (O'Connor ve McCartney, 2007; Tatar, 2004). Özellikle okula yeni başlanılan dönemde olan ilkökul öğrencilerinin öğretmenleriyle olumlu ilişki geliştirmesinin ileriki okul yıllarında öğrencilerin daha fazla sosyal beceriye sahip olmalarının yanı sıra akademik alanda daha başarılı olmalarına olanak sağladığı görülmektedir (Hamre ve Pianta, 2001). Yapılan boylamsal bir araştırmada, öğrenci öğretmen etkileşiminde öğretmenin duygusal ve eğitimsel desteğinin öğrencinin akademik başarısındaki gelişmeyi yordadığı belirlenmiştir (Curby, Rimm-Kaufman ve Pointz, 2009).

Sınıf içinde ortaya çıkan öğretmen – öğrenci arasındaki etkileşimler yalnızca öğrenmeye etki etmemektedir. Öğretmenler ve öğrenciler arasında sınıf içinde ortaya çıkan etkileşimler çocuk gelişimindeki ekolojik modelin temel sürecidir (Bronfenbrenner, 2005) ve çocukların okula eğitsel ve sosyal uyumlarının şekillenmesinde kayda değer bir etkisi bulunmaktadır (Griggs, Gagnon, Huelsman, Kidder-Ashley, ve Ballard, 2009). Yapılan bir çalışmada öğretmen – öğrenci ilişkisinin, öğrencinin utangaçlığı ve okula uyumu arasındaki ilişkide düzenleyici rolü olduğu ortaya konmuştur (Wu, Wu, Chen, Han, Han, Wang, ve Gao, 2015). Bununla birlikte öğretmenlerle kurulan ilişkinin kalitesinin, öğrencinin iyi oluşunu, olumlu benlik algısını, sosyal ve akademik konularda uyum sağlamadaki motivasyonunu ve bu konulardaki becerilerinin gelişmesini desteklediğine inanılmaktadır (Wentzel, 2009).

Olumlu bir öğrenci öğretmen ilişkisi aynı zamanda öğrencilerin okula karşı olumlu tutum geliştirmelerine yardımcı olabilmektedir (İpek ve Terzi, 2010). Bununla beraber özellikle, duyuşsal ve sosyal açıdan hazır olmayan çocuklara karşı, öğretmen tutum ve davranışı; çocuğun bütün okul yaşamını olumlu ya da olumsuz yönde etkileyecek kadar yaşamsal bir öneme sahiptir (Yapıcı, 2004). Kolburan-Geçer (2002), yapmış olduğu çalışmasında da öğretmenin öğrenciye gösterdiği yakınlık derecesinin öğrencinin okula karşı tutumunu etkilediğini ortaya koymuştur.

Öğretmen – öğrenci ilişkileri; her ilişkide olduğu gibi iletişimden, çatışmaya, yakınlıktan bağımsızlığa kadar birçok ögeyi içermektedir. Öğretmenin sınıf içinde koyduğu kurallar ve bunları öğrenciye aktarma biçimi, sınıf yönetimi stratejileri gibi birçok davranış öğretmen – öğrenci ilişkisi kapsamında ele alınabilir. Yapılan araştırmalar da öğretmenlerin sınıf içindeki davranışlarının öğrenciler üzerindeki etkisini ortaya koymuştur. Örneğin; Ogelman ve Ersan (2014) tarafından yapılan bir araştırmada okul öncesi öğretmenlerin sınıf yönetimi stratejilerinin öğrencilerin akran ilişkilerinde etkisi olup olmadığı test edilmiştir. Çalışma sonuçları, öğretmenlerin sınıf yönetimi stratejilerinin öğrencilerin başkalarına yardım amaçlı sosyal davranışlarını, akranlarına karşı korkulu kaygılı olmalarını ve akran şiddetine maruz kalmalarını anlamlı olarak yordadığını göstermiştir.

Öğrencilerin sosyal ve akademik gelişimlerini destekleyen öğretmen – öğrenci ilişkisinin riskli davranışlar gösteren öğrenciler açısından da koruyucu bir faktör olabileceği araştırmayla (Rudasill, Reio, Stipanovic ve Taylor, 2010) desteklenmektedir. Başka bir araştırmada öğretmen-öğrenci ilişkisinin kalitesi arttıkça, problem davranışlara sahip öğrencilerin sosyal, davranışsal yönden ve uyum açısından olumlu yönde değiştiği belirlenmiştir (Decker, Dona ve Christenson, 2007). Ayrıca öğretmen öğrenci ilişkisindeki yakınlık, öğrencinin davranışsal problemlerinden olan dışsallaştırma problemlerindeki azalmayla da ilişkili bulunmuştur (Silver, Measelle, Armstrong ve Essex, 2005). Skalická, Belsky, Stenseng ve Wichstrøm (2015), yapmış oldukları çalışmalarında anaokulunda öğretmeninden daha az yakınlık gören öğrencilerin, ilkökul birinci sınıfta öğretmenleriyle daha çok çatışma yaşadıklarını ve bunun sonucunda da gösterilen problem davranışların arttığını ifade etmişlerdir.

Literatürde olumlu öğretmen-öğrenci ilişkisinin öğrenciler üzerindeki etkisinin değerlendirilmesinin yanı sıra öğretmenlere olan etkisi de değerlendirilmiştir. Örneğin Spilt, Koomen ve Thijs (2011), tarafından yürütülen araştırmada olumlu öğretmen-öğrenci ilişkisi öğretmenlerin iyi oluş hali ile ilişkili bulunmuştur. Bir başka deyişle öğretmenin öğrenciyle kurduğu ilişkinin niteliği sadece öğrencisinin duygusal, sosyal, akademik gelişimini değil ayrıca öğretmenin psikolojik durumunu da etkilemektedir.

Öğretmen – öğrenci ilişkisi öğrencinin akademik başarısını, duygusal ve sosyal uyumunu ve problem davranışlarını etkileyen çok boyutlu bir davranışlar ağıdır. Bunun yanı sıra söz konusu ilişki, sadece öğrencinin değil öğretmenin de duygusal-sosyal durumunu etkileyebilmektedir. Karşılıklı etkisi bulunan bu ilişkinin niteliğini ve boyutlarını ortaya koymak için geçerli ve güvenilir ölçme araçlarına ihtiyaç duyulmaktadır.

Alanyazın ayrıntılı olarak incelendiğinde öğretmen – öğrenci ilişkilerini ölçmek için Türkiye’de sınırlı sayıda ölçme aracının bulunduğu saptanmıştır. Bunlardan biri İpek ve Terzi (2010), tarafından geliştirilen “Öğretmen-Öğrenci İlişkisi Değerlendirme Ölçeği”dir. Ölçek 22 maddeden ve sınıf içi ilişki, okul içi ilişki ve okul dışı ilişki olmak üzere üç alt ölçekten oluşmaktadır. Ölçeğin alt ölçekleri genel ilişki niteliğini ölçmekte ve öğretmen – öğrenci ilişkisinin boyutlarına ilişkin bilgi vermemektedir. Bununla birlikte ölçek ilkökul, ortaokul ve lise öğretmenlerinden toplanan verilerle geliştirilmiş ve eğitim kademeleri arasında herhangi bir ayırım yapılmamıştır. İlkokul, ortaokul ve lisedeki öğretmen – öğrenci ilişki boyutunun değişebileceği ve her kademenin kendine özgü özellikleri olduğu göz önünde bulundurulduğunda her kademeye ayrı ölçek kullanılmasının daha sağlıklı olduğu düşünülmektedir.

Bu konudaki alanyazın incelendiğinde özellikle Pianta (2001), tarafından öğretmen-öğrenci ilişkisini ölçen Öğretmen – Öğrenci İlişki Ölçeği adlı bir ölçeğin geliştirildiği ve bu ölçeğin ilkökul düzeyinde öğretmen – öğrenci ilişkisini ölçmek amacıyla sıklıkla kullanılan ölçeklerden biri olduğu görülmektedir. ÖÖİÖ’nün farklı kültürlerde de uygulanabilmesi için uyarlama çalışmaları yapılmıştır. Ölçeğin İtalya’da (Fraire, Longobardi ve Sclavo, 2008), Hollanda’da (Koomen, Verschueren, Schooten, Jak ve Pianta, 2012), Yunanistan (Gregoriadis ve Tsigilis, 2008) ve Norveç (Drugli ve Hjemdal, 2013) ve Çinde (Mi-young ve Neuharth-Pritchett, 2011) adaptasyon çalışmalarının yapıldığı görülmüştür. Özetle, alanyazındaki çalışmalara bakıldığında ÖÖİÖ’nün anaokulu ve ilkökul düzeyinde öğretmenlerin öğrenciyle olan ilişkilerine yönelik algılarını ölçmek için araştırmalarda sıklıkla kullanılan bir ölçek olduğu görülmektedir (Ahnert, Harwardt-Heinecke, Kappler, Eckstein-Madry ve Milatz, 2012; Eisenhower, Baker ve Blacher, 2007; McIntyre, Blacher ve Baker, 2006; Murray ve Murray, 2004; Murray, Waas, ve Murray, 2008; Murray ve Zvoch, 2011; Rudasill, Reio, Stipanovic ve Taylor, 2010; Thijs, Westhof ve Koomen, 2012; Trentacosta ve Izard, 2007; Verschueren ve Koomen, 2012). Farklı kültürlerde de uyarlaması yapılan ve yaygın bir kullanıma sahip olan bu ölçeğin uyarlama çalışmasının yapılmasının ilgili alanyazında bir boşluğu dolduracağı ve gelecekte de kültürler arası çalışmalar yürütmeye olanak tanıyabileceği düşünülmektedir.

Araştırmanın Amacı

Bu çalışmanın amacı Pianta (2001), tarafından geliştirilen Öğretmen – Öğrenci İlişki Ölçeğinin Türkiye uyarlamasını gerçekleştirmektir.

YÖNTEM

Çalışma Grubu

Öğretmen Öğrenci İlişkileri Ölçeği (ÖÖİÖ)’nün uyarlama çalışması için 2014 – 2015 Eğitim Öğretim yılında toplanan veriler 46 öğretmenin 560 öğrenci için doldurduğu ölçeklerden elde

edilmiştir. Çalışmaya katılan öğretmenlerin 31'i kadın (% 67) ve 15 erkektir (% 33). Öğrencilerin ise 263'ü kız (% 47) ve 282'si (%50) erkektir. 15 öğrencinin (% 3) ise cinsiyeti belirtilmemiştir. Bununla birlikte öğrencilerin 133'ü anaokuluna (% 23), 78'i birinci sınıfa (% 13), 134'ü ikinci sınıfa (% 23), 108'i üçüncü sınıfa (% 19) ve 104'ü dördüncü sınıfa (% 18) devam etmektedir. Üç öğrencinin (% 0.4) ise sınıfı belirtilmemiştir.

Veri Toplama Araçları

Çalışmada kullanılan veri toplama araçları ile ilgili bilgiler aşağıda verilmiştir.

Öğretmen – Öğrenci İlişki Ölçeği Orijinal Form: Öğretmenlerin bakış açısından öğretmen öğrenci ilişkisinin kalitesini değerlendirmeyi amaçlayan Öğretmen – Öğrenci İlişki Ölçeği (ÖÖİÖ), Pianta ve Nimetz (1991) tarafından anaokulu öğretmenleri üzerinde geliştirilmiştir. Ölçeğin ilk sürümünde 16 madde ve üç açık uçlu soru bulunmaktadır. Ölçeğin geçerlik ve güvenilirlik çalışmaları başka çalışmalarla da (Pianta ve Steinberg, 1992; Pianta, Steinberg ve Rollins, 1995; Saft ve Pianta, 2001) desteklenmiş ve ölçeğe Pianta (2001), tarafından son hali verilmiştir. Tüm bu çalışma sonuçlarına göre ÖÖİÖ hem anaokulu hem de ilkökul öğretmenlerinin kullanabileceği üçlü likert derecelendirmeye sahiptir. ÖÖİÖ, üç boyutta toplam 28 maddeden oluşmaktadır. Çatışma boyutunda 12 madde; yakınlık boyutunda 11 madde ve bağımlılık boyutunda ise beş madde bulunmaktadır. (Örn; “Öğrencinin bana karşı duyguları aniden değişebilir,” “Öğrenci hatasını düzelttiğimde incinir.”, “Öğrenci benimle olan ilişkisine değer verir.”, “Öğrenci duygu ve yaşantılarını açık bir şekilde benimle paylaşır.”, “Öğrenciyle baş etmeye çalışmak enerjimi tüketir.”)

Ölçeğin güvenilirliğinin hesaplanması için test tekrar test güvenilirliğine ve iç tutarlılık katsayısı hesaplama yöntemlerine başvurulmuştur. Dört hafta arayla yapılan iki uygulamada puan ortalamaları arasındaki Pearson korelasyon katsayısı yakınlık alt boyutu için .88; çatışma alt boyutu için .92 ve bağımlılık alt boyutu için .76 ve tüm ölçek için .89 olarak bulunmuştur. Son hali verilen ölçeğin Cronbach Alpha katsayısı çatışma alt boyutu için .92; yakınlık alt boyutu için .86; alt bağımlılık boyutu için .64 ve tüm ölçek için .89 olarak hesaplanmıştır (Pianta, 2001).

Ölçeğin geçerlik çalışması kapsamında ÖÖİÖ'nün yapı geçerliği incelenmiş ve sonucunda ölçekte yer alan 28 maddenin toplam varyansın % 48.8'ini açıkladığı bulunmuştur. Çatışma alt boyutu toplam varyansın % 29.8'ini, yakınlık alt boyutu % 12.9'unu ve bağımlılık alt boyutu % 6.2'sini açıklamaktadır (Pianta, 2001). Ayrıca boyutlar arası ilişkiler incelenmiş ve buna göre, çatışma ile yakınlık alt boyutları arasında -.45 ($p < .001$), çatışma ile bağımlılık alt boyutları arasında .28 ($p < .001$) ve yakınlık ile bağımlılık arasında .13 ($p < .01$) düzeyinde korelasyonların anlamlı olduğu görülmüştür. ÖÖİÖ'nün geçerliği kapsamında beş yaşından küçük çocuklar ve beş yaşından büyük çocuklar karşılaştırılmış ve öğretmenlerin beş yaşından büyük çocuklarla daha fazla çatışma ve bağımlılık yaşadığı ortaya konmuştur. Ayrıca ÖÖİÖ puanları ile çocukların davranış problemleri ve yetkinliklerle arasındaki ilişkiler incelenmiş ve buna göre davranış problemleri ile ÖÖİÖ puanları arasında -.72 ve öğrencilerin bağımlılık düzeyleri ile ÖÖİÖ puanları arasında .67 düzeyinde ($p < .01$) önemli ilişkiler bulunmuştur (Pianta, 2001). Tüm bu bulgular ÖÖİÖ'nün geçerliğine yönelik bir kanıt olarak yorumlanmıştır.

Kişisel Bilgi Formu: Öğrencilerin cinsiyetlerinin ve devam ettikleri sınıf düzeylerinin öğrenilmesi amacıyla araştırmacılar tarafından hazırlanmıştır.

İşlem

Ölçeğin uyarılma çalışması kapsamında Hambleton ve Patsula (1999), tarafından önerilen basamaklar takip edilmiştir. Orijinal form İngilizce dil bilgisine sahip en az yüksek lisans düzeyinde eğitimi bulunan beş psikolojik danışma ve rehberlik öğretim elemanına gönderilmiş ve her bir maddenin Türkçe çevirisinin yapılması istenmiştir. Yapılan çeviriler incelenerek araştırmacılar tarafından tek bir Türkçe form oluşturulmuştur. Daha sonra bu Türkçe form İngilizce dil bilgisine sahip ve davranış bilimleri alanında en az yüksek lisans derecesine sahip üç uzmana gönderilmiş ve

tekrar İngilizceye çevirisi sağlanmıştır. Gelen çeviri formlar ile orijinal ölçekteki maddeler karşılaştırılmış ve maddelerin anlam içeriği açısından ve Türkçe ifade gücü açısından uygun bir anlam verdiğinden emin olunmuştur. Daha sonra ölçek pilot uygulama için hazır hale getirilmiştir. Ölçek maddeleri araştırmacılar tarafından kültürel uygunluk bakımından da incelenmiştir. Ölçeğin pilot uygulama formu on sınıf öğretmenine uygulanmış ve öğretmenlerden maddelerin anlaşılabilirliği hakkında dönüt alınmıştır. Alınan dönütler çerçevesinde iki maddede anlamı güçleştirmeye yönelik olarak Türkçe dil bilgisi açısından düzeltme yapılmıştır.

Çalışma kapsamında veri toplamak üzere öncelikle Ankara İl Milli Eğitim Müdürlüğü'nden ve Hacettepe Üniversitesi Etik Komisyonu'ndan gerekli izinler alınmıştır. Ankara iline bağlı altı farklı devlet okulluna gidilmiş ve bu okullarda çalışan ve çalışmaya katılmayı gönüllü olarak kabul eden 46 ilkokul öğretmeni ile görüşülmüştür. Öğretmenlere çalışmanın amacı hakkında gerekli açıklamalar yapılmıştır. Bu kapsamda öğretmenlerden sınıflarından rastgele olarak en az beş en fazla 15 öğrenci seçmeleri ve her bir öğrenci için ayrı ayrı ölçeği doldurmaları istenmiştir. Öğretmenlerden elde edilen veriler toplanmış ve araştırmacılar tarafından bilgisayar ortamına girilerek analize hazır hale getirilmiştir.

Verilerin Analizi

ÖÖİÖ'nün geçerlik çalışması kapsamında kapsam geçerliği için uzman görüşüne, yapı geçerliği için ise doğrulayıcı ve açıklayıcı yapısal eşitlik modeli (AYEM) yöntemlerine başvurulmuştur. Dolan (1994) değişkenlerin normal dağılıma sahip olması koşuluyla kategori sayısının en az beş olması durumunda ki kare istatistiğinin kullanılabilirliğini, eğiklik ve basıklık katsayılarının büyüklüğüne bağlı olarak (normalden sapma) yedi kategorili değişkenler için bile pearson korelasyon kullanılmaması gerektiğini, bunun yerine polikorik korelasyonların kullanılmasının daha yansız (unbias) sonuç verebileceğini belirtmiştir. Lubke ve Muthen (2004) de aynı koşullar altında kategori sayısının en az yedi olması gerektiğini ifade etmektedirler. Hatta Jamieson (2004) likert tipi ölçeklerde kategoriler arasındaki aralıkların eşit olmadığını ve bunlara herhangi bir ortalama ya da parametrik bir korelasyon yapılamayacağını söylemektedir. Ayrıca bu tür verilere sadece nonparametrik istatistiklerin kullanılabilirliğini belirtmektedir. Bu açıklamalar ışığında doğrulayıcı faktör analizi polikorik korelasyonlar üzerinde robust ağırlıklandırılmış en küçük kareler yöntemi (WLSMV) kullanılarak gerçekleştirilmesine karar verilmiştir. Güvenirlilik çalışması kapsamında ise Cronbach Alpha iç tutarlılık katsayısı hesaplanmıştır. Çalışmada elde edilen verilerin analizinde Mplus 7 İstatistik Programından yararlanılmıştır.

BULGULAR

ÖÖİÖ orijinal formu üç boyuttan ve toplam 28 maddeden oluşmaktadır. Ölçeğin Türkçe Formu kullanılarak elde edilen verilerin kuramsal yapıyı ne derece doğruladıklarını ya da kuramsal yapı tarafından veriler arasındaki ilişkilerin ne derece açıklandığını incelemek amacıyla doğrulayıcı faktör analizi yapılmıştır. Analiz sonuçları Tablo 1'de verilmiştir.

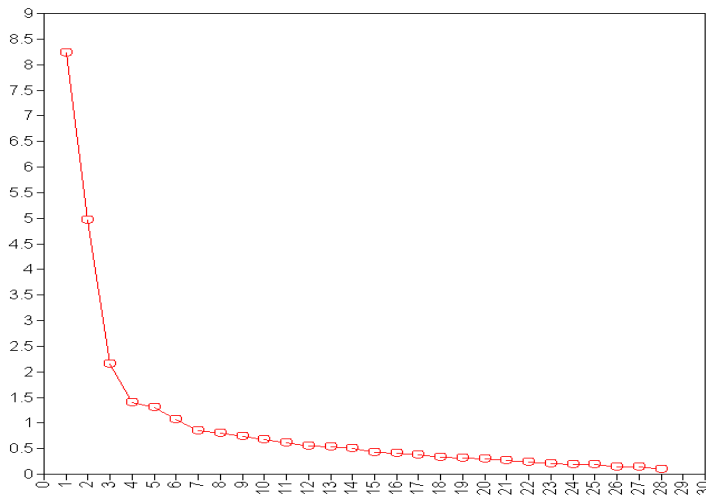
Tablo 1. Model Karşılaştırma Sonuçları

Model	χ^2	sd	p	CFI	RMSEA (%90 GA)	WRMR
1. Önerilen (Kuramsal)	3709.41	321	.00	.68	.14 (.13 -.14)	3.35
2. AYEM	1312.16	297	.00	.81	.08 (.07-.08)	.05

Tablo 1'de görüldüğü gibi, maddeler arasındaki ilişkilerin üç boyutlu kuramsal yaklaşımla açıklanabileceğine ilişkin önerilen modelin uyum katsayıları düşüktür. Hair, Black, Babin ve Anderson (2010), madde sayısı 12-30 arasında ve kişi sayısı 250'den fazla olması durumunda kabul

edilebilir bir model için genel uyum katsayılarına yönelik kesme noktalarını şöyle açıklamaktadır: istatistiksel bakımdan önemli χ^2 , CFI'nın .92'den büyük ve RMSEA'nın .07'den düşük olması. WRMR katsayısının ise .90 ve altında olması istenmektedir (Muthen ve Muthen, 2012). Bu açıklama ışığında önerilen modele ilişkin $\chi^2_{321} = 3.709,41$, $p=.00$, CFI= .68, RMSEA= .14 (.13- .14) ve WRMR 3.35 olması, model uyum katsayılarının kabul edilebilir düzeyde olmadığını göstermektedir.

Kuramsal modelin uyum katsayıları düşük bulunduğundan, bir sonraki aşamada açıklayıcı yapısal eşitlik modeli (exploratory structural equaliton model) kullanılmıştır. Brown (2006) DFA'nın model uyumsuzluğu sonucu değişim indekslerine (MI) dayalı daha iyi uyum elde etmek için sıklıkla keşfedici (explorative) amaçlarla kullanıldığını; bu durumda hatalı yük tanımlamalarının MI'lara dayalı incelenmesi yerine, döndürülmüş faktör matrisinin kullanıldığı açıklayıcı faktör analizi kullanılmasının uygun olduğunu ifade etmektedir. Marsh (2007) geleneksel DFA'nın çok faktörlü ölçeklerin çoğu için genel uyum katsayılarının aşırı katı olduğunu, her biri 5-10 maddeden oluşan 5-10 faktörlü ölçeklerin madde düzeyinde analiz edilmesi durumunda, faktör yapısı "iyi" olduğunda bile neredeyse CFI'nın .90'dan büyük, RMSEA'nın da .05'ten küçük elde etme olasılığının bulunmadığını belirtmektedir. Asparouhov ve Muthén (2009) doğrulayıcı faktör analizindeki uyum problemleri nedeniyle açıklayıcı yapısal eşitlik modelinin kullanılabileceğini, bu analizle tüm SEM parametrelerinin yanı sıra, maddelerin ileri sürülen diğer faktörlerdeki yüklerinin de incelenebileceğini ifade etmektedirler. Bu önerinin, bazı maddelerin yüklerinin düşük, bazılarının birden fazla faktörle ilişkili olduğu ÖÖİÖ için de geçerli olabileceği varsayılarak, araştırma grubundan elde edilen ölçümlerin öngörülen kuramsal yapıyı ne derece doğruladığı AYEM ile de test edilmiştir. Asparouhov ve Muthén (2009) bu yöntemin faktörlerin döndürülmesi ve her bir göstergenin tüm faktörlerle ilişkilendirilmesi bakımından, doğrulayıcı faktör analizinden farklı olduğunu belirtmektedirler. Bu noktada, Asparouhov ve Muthén'in (2009) önerileri doğrultusunda, açıklayıcı yapısal eşitlik modeli yapılarak analize devam edilmiştir. Tablo 1'de görüldüğü gibi AYEM'e ilişkin genel uyum katsayıları da kabul edilebilir düzeyin altında kalmıştır ($\chi^2_{297} = 1312.16$, Ki-kare/sd= 3.35; $p=.00$, CFI= .81, RMSEA= .08(.17- .08) ve SRMR .05). Bu durumda son seçenek olarak alternatif boyut sayılarını tanımlamak için açıklayıcı faktör analizinin (exploratory factor analysis) kullanılmasına karar verilmiştir. Açıklayıcı faktör analizi (AFA) sonucunda elde edilen özdeğer grafiği Şekil 1'de verilmiştir:



Şekil 1. OOİÖ AFA Özdeğer Grafiği

Şekil 1'de görüldüğü gibi, özdeğeri, birden büyük altı boyut bulunmaktadır. Bu altı boyuttan ilk ikisiyle diğerleri arasında bir kopma gözlenmektedir. Bu iki boyutun açıkladığı varyans .47 olarak

görülmüştür (sırasıyla % 29.38 ve % 17.71). Bununla birlikte, üç boyutlu kuramsal model ve üç boyutlu AYEM modelleri maddeler arasındaki ilişkileri açıklamada yeterli olmamıştır. Buradan hareketle ölçeğin üç boyutlu yapısı dışında alternatif modeller olabileceği düşünülmüştür. Mplus İstatistik Programı çıktısından ayrıca özdeğerlere ilişkin standart hatalar da elde edilebilmektedir. Bu hatalar göz önünde bulundurulduğunda, Ölçeğin iki boyutlu yapısının alternatif modellerden biri olabileceği sonucuna varılmıştır. Tablo 2'de iki boyutlu çözüme ilişkin madde faktör ilişkileri verilmiştir.

Tablo 2. ÖÖİÖ İki Boyutlu Model: Faktör Yükleri

Madde	İki Boyut		Madde	İki Boyut	
	F1	F2		F1	F2
M1	.80	-.29	M15	.50	-.49
M2	-.40	.63	M16	-.35	.68
M3	.73	-.02	M17	.54	.44
M4	.31	-.44	M18	.09	.71
M5	.68	-.35	M19	-.34	.38
M6	.08	.41	M20	-.20	.74
M7	.70	-.19	M21	.52	.25
M8	.45	.18	M22	-.01	.62
M9	.76	.00	M23	.02	.76
M10	.39	.15	M24	-.13	.73
M11	-.14	.77	M25	.16	.53
M12	.56	-.19	M26	.14	.57
M13	-.09	.70	M27	.77	-.07
M14	.28	.31	M28	.61	-.44

Tablo 2'de görüldüğü gibi, ölçeğin orijinal formunda yer alan yakınlık alt boyutundaki maddeler (1, 3, 4, 5, 7, 9, 12, 15, 21, 27, 28) ilk; çatışma alt boyutundaki maddeler ise (2, 11, 13, 16, 18, 19, 20, 22, 23, 24, 25, 26) ikinci boyutta yer almaktadır. Yakınlık alt boyutunda yer alan 4. ve 15. maddeler her iki boyuta da yük yüklemektedir. Yakınlık alt boyutuyla ilişkili olan maddelerin faktör yükleri .80 -.39; çatışma alt boyutunda yer alan maddelerin faktör yükleri .38 - .77 arasındadır. Bununla birlikte, bağımlılık alt boyutunda yer alan 6. madde ikinci boyuta; 8. ve 10. maddeler ilk boyuta; 14 ve 17. maddeler ise her iki boyuta yakın yükler yüklemektedir. Hair (2010), faktörlerin uygunluğu konusunda son kararı verenin araştırmacı olması gerektiğini ve bu kararların ampirik temelli değil kuram temelli olmasını önermektedir. Bununla birlikte her maddenin okunması ve hangi maddelerin hedef faktör altında anlamlı ve yararlı bir yapı oluşturduğuna karar vermek de madde seçiminde kriterlerden biridir. Frankel ve Wallen (2006) da madde yazımına başlanırken ve maddelere karar verilirken rasyonel bir temel olması gerektiğini öne sürmektedir. Tüm bu açıklamalardan yola çıkılarak ölçeğin orijinal formunun dayandığı kuramsal yapı temel alınmış, madde seçimleri yapılırken buna sadık kalınmıştır.

Bağımlılık alt boyutundaki maddelerin diğer alt boyutlarda da yüksek madde faktör yükleri almasından ve söz konusu boyutların anlamlandırılmasını sınırlamasından dolayı, bu boyutun ölçekten çıkarılmasına karar verilmiştir. Yakınlık ve çatışma alt boyutları arasındaki korelasyon -.42 ($p < .01$) olarak hesaplanmıştır. Bununla birlikte, ölçek puanlarının güvenilirlik çalışması sonucunda ilk boyutta madde – toplam korelasyonları .31 ile .70 arasında olup, Cronbach Alfa katsayısının .83 olduğu bulunmuştur. İkinci boyutta ise, madde – toplam korelasyonları .33 ile .69 arasında olup, Cronbach Alfa katsayısı .86'dır.

SONUÇ ve TARTIŞMA

Öğretmen-Öğrenci İlişki Ölçeği'nin uyarlanması amacıyla yapılan bu çalışma çerçevesinde gerçekleştirilen geçerlik ve güvenilirlik çalışmaları kapsamında ölçeğin Türkiye'de kullanımı için gerekli kanıtlar sağlanmıştır. Ölçeğe ilişkin bu çalışmaların yapılması aşamasında maddeler arasındaki ilişkilerin üç boyutlu kuramsal yaklaşımla açıklanabileceğine ilişkin önerilen modelin uyum katsayıları düşük olduğu görülmüştür. Bu nedenle açımlayıcı yapısal eşitlik modeli yapılarak istatistiksel analize devam edilmiştir. AYEM'e ilişkin genel uyum katsayıları kabul edilebilir düzeyin altında kalmıştır. Bu durumda son seçenek olarak alternatif boyut sayılarını tanımlamak için açımlayıcı faktör analizinin (AFA) kullanılmasına karar verilmiştir. AFA sonuçları ölçeğin çatışma ve yakınlık olarak iki boyuttan oluştuğunu göstermiştir. Ölçeğin orijinalinde bulunan bağımlılık boyutu maddelerinin diğer boyutlarda da yüksek madde faktör yükleri almasından dolayı bu boyutun ölçekten çıkarılmasına karar verilmiştir. Whitaker, Dearth-Wesley ve Gooze (2015) ve ABD'de bulunan National Institute of Health (NIH)'a bağlı Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Study of Early Child Care (2010) tarafından yapılan çalışmalarda da ÖÖİÖ'nün kısa formu kullanılmış ve bağımlılık boyutu kısa formdan da çıkarılmış, yakınlık ve çatışma boyutundan bazı maddeler çıkartılarak ölçek 15 maddeye indirilmiştir.

Bununla birlikte aynı kuramsal yaklaşıma dayalı olarak geliştirilen Ebeveyn – Çocuk İlişkisi Ölçeği'nin (Pianta, 1992) Türkiye'ye uyarlanmasında da üç boyutlu yapının çalışmadığı görülmüş ve ölçekten yedi madde çıkartılarak ölçek iki boyut olarak tanımlanmıştır (Akgün ve Yeşilyaprak, 2010). Görülmektedir ki; bu çalışmaya benzer şekilde başka çalışmalarda da çatışma, yakınlık ve bağımlılık boyutlarından oluştuğu ileri sürülen ilişki farklı örneklerde iki boyutlu olarak görülebilmektedir. Dolayısıyla ölçeğin Türkçe formu iki boyuttan oluşmaktadır.

Sonuç olarak; ÖÖİÖ'nün Türkiye uyarlaması sonucunda ölçeğin 23 maddeden ve iki boyuttan oluştuğu görülmüştür. ÖÖİÖ ilkokullarda öğretmen – öğrenci ilişkisini değerlendirmek üzere araştırmalarda kullanılacak bir ölçme aracıdır. Birey hakkında tek başına karar verebilmek için bu aşamada bu ölçeğin tek başına kullanımının uygun olmadığı düşünülmektedir. Ölçeğin çatışma boyutundan alınan yüksek puan öğretmen – öğrenci ilişkisinde çatışmanın; yakınlık boyutundan alınan yüksek puan ise ilişkide yakınlığın varlığına işaret etmektedir.

ÖÖİÖ, okul öncesi öğretmenleri, sınıf öğretmenleri ve okul öncesi ve ilkokul dönemiyle çalışan psikolojik danışmanlar ve diğer ruh sağlığı uzmanları tarafından öğretmen – öğrenci ilişkisini ölçmek amacıyla kullanılabilir. Bireyle psikolojik danışma, grupla psikolojik danışma, sınıf içi rehberlik etkinlikleri gibi öğretmen – öğrenci ilişki kalitesini arttırmaya yönelik yapılan müdahalelerde ÖÖİÖ ön ve son test olarak kullanılabilir ve yapılan müdahalenin etkililiği hakkında bilgi verebilir. Bunun yanı sıra önleyici çalışmalar kapsamında okulda ÖÖİÖ uygulanarak, öğretmeniyle çatışma içerisinde olan çocuklar belirlenebilir ve erken müdahale edilebilir. Alan uygulayıcılarının yanı sıra araştırmacılar da okul öncesi ve ilkokul döneminde yaptıkları araştırmalarında ÖÖİÖ'yü öğretmen – öğrenci ilişkisini belirlemek amacıyla kullanabilirler. Bununla beraber geçerlik çalışmalarına ek olarak ÖÖİÖ'nün iki faktörlü yapısı yeni bir örneklem grubundan toplanan verilerle doğrulayıcı faktör analizi yöntemi kullanılarak test edilebilir.

KAYNAKÇA

- Ahnert, L., Harwardt-Heinecke, E., Kappler, G., Eckstein-Madry, T., & Milatz, A. (2012). Student-teacher relationships and classroom climate in first grade: how do they relate to students' stress regulation?. *Attachment & human development, 14*(3), 249-263.
- Akgün, E., Yeşilyaprak, B. (2010). Çocuk Anababa İlişki Ölçeği Türkçe formunun geçerlik ve güvenilirlik çalışması. *Balıkesir University Journal of Social Sciences Institute, 13*(24), 44-53.
- Asparouhov, T., ve Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438.

- Bowlby, J. (2012). *Bağlanma*. (Çev. Tuğrul Veli Soylu). İstanbul: Pinhan Yayıncılık.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives, *Developmental Psychology*, 22(6), 723-742.
- Bronfenbrenner, U. (2005). *Making human beings human: Bioecological perspectives on human development*. Sage Publication.
- Brown, T. A., (2006). *Confirmatory Factor Analysis for Applied Research* (1st. Edition). New York: The Guilford Press.
- Curby, T. W., Rimm-Kaufman, S. E., Ponitz, C. C. (2009). Teacher–child interactions and children’s achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology*, 101(4), 912.
- Decker, D. M., Dona, D. P., Christenson, S. L. (2007). Behaviorally at-risk African American students: The importance of student–teacher relationships for student outcomes. *Journal of School Psychology*, 45(1), 83-109.
- Demirkaya, P. N. (2013). *Anasınıfına özel gereksinimi olan ve olmayan öğrencilerin öğrenci-öğretmen ilişkisinin incelenmesi*. Yüksek Lisans Tezi, Ankara: Ankara Üniversitesi.
- Dolan, C. V. (1994), Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British journal of Mathematical and Statistical Psychology*, 47, 309-326
- Drugli, M. B., & Hjemdal, O. (2013). Factor structure of the student–teacher relationship scale for Norwegian school-age children explored with confirmatory factor analysis. *Scandinavian Journal of Educational Research*, 57(5), 457-466.
- Eisenhower, A. S., Baker, B. L., & Blacher, J. (2007). Early student–teacher relationships of children with and without intellectual disability: Contributions of behavioral, social, and self-regulatory competence. *Journal of School Psychology*, 45(4), 363-383.
- Fraire, M., Longobardi, C., Scavo, E. (2008). Contribution to validation of the student-teacher relationship scale (STRS Italian Version) in the italian educational setting. *European Journal of Education and Psychology*, 1(3), 49-59.
- Frankel, J., & Wallen, T. (2006). Cross-cultural on organizational commitment: a further review and application of hofstede’s value survey module. *Journal of International Business and Entrepreneurship*, 10(1), 1-26.
- Gregoriadis, A., & Tsigilis, N. (2008). Applicability of the Student—Teacher Relationship Scale (STRS) in the Greek Educational Setting. *Journal of Psychoeducational Assessment*, 26(2), 108-120.
- Griggs, M. S., Gagnon, S. G., Huelsman, T. J., Kidder-Ashley, P., & Ballard, M. (2009). Student-teacher relationships matter: Moderating influences between temperament and preschool social competence. *Psychology in the Schools*, 46(6), 553-567.
- Hair, J. F. (2010). *Multivariate data analysis*. Pearson College Division.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010) *Multivariate data analysis* (7th ed.). Pearson Company.
- Hambleton, R.K. & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1(1), 1-30.
- Hamre, B. K., Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children’s school outcomes through eight grade. *Child Development*, 72(2), 625-638.
- İpek, C., Terzi, A. R. (2010). İlk ve ortaöğretim kurumlarında öğretmen-öğrenci ilişkilerinin öğretmen görüşlerine göre belirlenmesi: Van ili örneği. *Kuram ve Uygulamada Eğitim Yönetimi*, 16(3), 433-456.
- Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*, 38(12), 1217-1218.
- Kolburan-Geçer, A. (2002). *Öğretmen yakınlığının öğrencilerin başarıları, tutumları ve güdülenme düzeyleri üzerindeki etkisi*. Doktora tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Koomen, H. M., Verschueren, K., van Schooten, E., Jak, S., Pianta, R. C. (2012). Validating the Student-Teacher Relationship Scale: Testing factor structure and measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology*, 50(2), 215-234.
- Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11(4), 514-534.
- Marsh, H. W. 2007. “Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology”, *Handbook of Sport Psychology* (3rd Edition). 774-798.
- McIntyre, L. L., Blacher, J., & Baker, B. L. (2006). The transition to school: Adaptation in young children with and without intellectual disability. *Journal of Intellectual Disability Research*, 50(5), 349-361.
- Mi-young, L. W., Neuharth-Pritchett, S. (2011). Examining factorial validity and measurement invariance of the Student–Teacher Relationship Scale. *Early Childhood Research Quarterly*, 26(2), 205-215.

- Muller, C. (2001). The role of caring in the teacher-student relationship for at-risk students. *Sociological Inquiry*, 71(2), 241-255.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide*. Los Angeles, CA: Muthen & Muthen; 1998-2010.
- Murray, C., & Zvoch, K. (2011). Teacher—student relationships among behaviorally at-risk African American youth from low-income backgrounds: student perceptions, teacher perceptions, and socioemotional adjustment correlates. *Journal of Emotional and Behavioral Disorders*, 19(1), 41-54.
- Murray, C., & Murray, K. M. (2004). Child level correlates of teacher—student relationships: An examination of demographic characteristics, academic orientations, and behavioral orientations. *Psychology in the Schools*, 41(7), 751-762.
- Murray, C., Waas, G. A., & Murray, K. M. (2008). Child race and gender as moderators of the association between teacher—child relationships and school adjustment. *Psychology in the Schools*, 45(6), 562-578.
- NICHD Study of Early Child Care. (2000). Teacher's relationship with child; first grade student-teacher relationship scale. *Child Care Data Report - 309*. Research Triangle Park, NC: Research Triangle Institute.
- O'Connor, E., & McCartney, K. (2007). Examining teacher—child relationships and achievement as part of an ecological model of development. *American Educational Research Journal*, 44(2), 340-369.
- Ogelman, H. G., Ersan, C. (2014). Okul Öncesi Öğretmenlerinin Sınıf Yönetimi Stratejilerinin 5-6 Yaş Çocuklarının Akran İlişkilerini Yordayıcıları. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 3(2), 63-84. Doi: 10.14686/BUEFAD.201428172
- Pianta R. C. (2001). *Student-Teacher Relationship Scale: Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc
- Pianta, R. C., Nimetz, S. L. (1991). Relationships between children and teachers: Associations with classroom and home behavior. *Journal of Applied Developmental Psychology*, 12(3), 379-393.
- Pianta, R. C., Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child and Adolescent Development*, 57, 61-80.
- Pianta, R. C., Steinberg, M. S., Rollins, K. B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, 7(02), 295-312.
- Pomeroy, E. (1999). The teacher-student relationship in secondary school: Insights from excluded students. *British Journal of Sociology of Education*, 20(4), 143-148.
- Rudasill, K. M., Reio, T. G., Stipanovic, N., Taylor, J. E. (2010). A longitudinal study of student—teacher relationship quality, difficult temperament, and risky behavior from childhood to early adolescence. *Journal of School Psychology*, 48(5), 389-412.
- Saft, E. W., Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, 16(2), 125.
- Silver, R. B., Measelle, J. R., Armstrong, J. M., Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher—child relationship during the school transition. *Journal of School Psychology*, 43(1), 39-60.
- Skalická, V., Belsky, J., Stenseng, F., Wichstrøm, L. (2015). Preschool- Age Problem Behavior and Teacher—Child Conflict in School: Direct and Moderation Effects by Preschool Organization. *Child Development*, 86(3), 955-964.
- Spilt, J. L., Koomen, H. M., Thijs, J. T. (2011). Teacher wellbeing: The importance of teacher—student relationships. *Educational Psychology Review*, 23(4), 457-477.
- Tatar, K. (2004). *Sınıf içi öğretmen öğrenci iletişiminin öğrenci başarısı üzerine etkileri ve bir uygulama*. Yayınlanmamış Yüksek lisans Tezi. Dumlupınar Üniversitesi, Sosyal Bilimler Enstitüsü, Kütahya.
- Thijs, J., Westhof, S., & Koomen, H. (2012). Ethnic incongruence and the student—teacher relationship: The perspective of ethnic majority teachers. *Journal of School Psychology*, 50(2), 257-273.
- Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion*, 7(1), 77.
- Verschuere, K., & Koomen, H. M. (2012). Teacher—child relationships from an attachment perspective. *Attachment & human development*, 14(3), 205-211.
- Wentzel, K. (2009). Students' relationships with teachers as motivation contexts. In K. Wentzel A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 301–322). Mahwah, NJ: Erlbaum.
- Whitaker, R. C., Dearth-Wesley, T., Gooze, R. A. (2015). Workplace stress and the quality of teacher—children relationships in Head Start. *Early Childhood Research Quarterly*, 30, 57-69. doi: <http://dx.doi.org/10.1016/j.ecresq.2014.08.008>

- Wu, Y., Wu, J., Chen, Y., Han, L., Han, P., Wang, P., Gao, F. (2015). Shyness and School Adjustment Among Chinese Preschool Children: Examining the Moderating Effect of Gender and Teacher-Child Relationship. *Early Education and Development*, 26(2), 149-166.
- Yapıcı, M. (2004). İlköğretim birinci sınıfa başlayan öğrencilerin hazır bulunuşluk düzeyleri. *Uluslararası İnsan Bilimleri Dergisi*, 1(1), 1-8.

EXTENDED ABSTRACT

Introduction

Primary school students not only have more social skills but also become academically more successful in further years provided that if they develop a positive relationship with their teachers upon starting the school (Hamre and Pianta, 2001). Teacher-student relationship is a multi-dimensional factor affecting the academic success, emotional and social adjustment as well as problem behaviors of the student. That relationship can affect the emotional-social situation of both the student and the teacher. A number of valid and reliable assessment instruments are needed to ascertain the characteristics and aspects of this reciprocation.

The related literature shows us that there is a limited number of instruments to assess the teacher-student relationship in Turkey. One of these is the “Teacher-Student Relationships Assessment Scale” developed by Ipek and Terzi (2010). The scale consists of 22 items and three factors described as classroom, school, and out-of-school relationships.

However, when the existing studies are examined, it is seen that the Student-Teacher Relationship Scale (STRS) developed by Pianta (2001) is one of the scales frequently used to assess the teacher-student relationships at primary school level. Several adaptation studies have been done to apply the STRS in different cultures like Italian (Fraire, Longobardi and Sclavo, 2008), Dutch (Koomen, Verschueren, Schooten, Jak and Pianta, 2012), Chinese (Mi-young and Neuharth-Pritchett, 2011).

Therefore, this study aims to adapt Pianta’s Student-Teacher Relationship Scale (2001), to Turkey and to contribute to the works of pre- and elementary school psychological counselors, classroom teachers and researchers.

Method

The data gathered from six different public schools in Ankara city center and interviewed with 46 primary school teachers who were working in those schools and voluntarily accepted to participate in this study. The teachers were duly informed about the aim of the research on an individual basis. They were asked to choose five students randomly from their own classrooms and to fill the scales separately for each of five students. The data needed to adapt the STRS were accordingly obtained through these scales filled by 46 teachers for 560 students in the 2014-2015 education year.

As for the validity of STRS, the experts were asked to rate its content validity. Its construct validity was assessed with confirmatory and exploratory factor analyses. Lastly, Cronbach’s Alpha coefficient of internal consistency was calculated to determine its reliability.

Results and Discussion

The confirmatory factor analysis was conducted with categorical data, in other words, by using the method of robust weighted least squares (WLSMV) over polychoric correlations. The coefficients of concordance for the suggested model were not at acceptable levels considering $\chi^2_{321} = 3.709,41$, $p=.00$, CFI= .68, RMSEA= .14 (.13- .14) and WRMR 3.35. We therefore proceeded to the analysis through the exploratory structural equation model (ESEM) in accordance with Asparouhov and Muthén’s (2009) and Kelloway’s suggestions (2015). The coefficients of concordance for ESEM

also remained below the acceptable level ($\chi^2_{297} = 1312.16$, $p=.00$, CFI= .81, RMSEA= .08(.17- .08) and SRMR .05). As a last resort, we decided to use an exploratory factor analysis (EFA) in order to identify the alternative number of dimensions.

According to the results of EFA, there are six dimensions with an eigenvalue greater than one. A disconnection is observed between the first two and the remaining four of these six dimensions. The variance the two dimensions explained was found as .47 (29.38 % and 17.71 %, respectively). However, the three-dimensional theoretical model and the three-dimensional ESEM models failed to explain the relationships between items. Standard errors for eigenvalues can be also obtained from Mplus output. Considering these errors, we concluded that the two-dimensional form could be one of the alternative models. The dimension of dependency was excluded from the scale as the items related to it had high factor loadings in the other dimensions and limited the explanation of those dimensions.

Separately from this study, the reliability analysis of the scale points indicated that item-total correlations for the first dimension ranged from .31 to .70 and Cronbach's Alpha coefficient was .83. Item-total correlations for the second dimension were between .33 and .69, and Cronbach's Alpha coefficient was .86.

In conclusion, the Turkish adaption of STRS consists of 23 items and two dimensions. STRS is an assessment instrument usable to assess the teacher-student relationships in primary schools. Scoring high at the conflict dimension of the scale indicates the existence of conflict and scoring high at the intimacy dimension indicates the existence of intimacy in a given teacher-student relationship.

STRS can be used to assess teacher-student relationships by preschool and classroom teachers as well as psychological counsellors and other mental health professionals dealing with pre- and primary school children. It can be used as a pre- or post-test during interventions to enhance teacher-student relationships like individual or group counselling and classroom guidance activities. STRS can inform about the effectiveness of those interventions. Furthermore, it can be applied in schools within the scope of preventive works. Students in conflict with their teachers can be determined and treated with an early intervention in this way. In addition to field practitioners, researchers can also use STRS to assess teacher-student relationships in their studies related to the pre- and primary school period.

Vee Diyagramından Elde Edilen Puanların Güvenirliğinin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre İncelenmesi*

The Examination of Reliability of Vee Diagrams According to Classical Test Theory and Generalizability Theory

Betül POLAT DEMİR**

Öz

Araştırmada, vee diyagramından elde edilen puanların güvenirlüğünün incelenmesi amaçlanmıştır. Bu amaçla araştırmada klasik test kuramı ve genellenebilirlik kuramından yararlanılmıştır. Araştırmada betimsel araştırma modeli kullanılmıştır. Çalışma Niğde ilinde öğrenim gören 32, 7. ve 8. sınıf öğrencisi ile yürütülmüştür. Vee diyagramları puanlama anahtarları kullanılarak puanlanmıştır. KTK’da, Cronbach alfa iç tutarlılık güvenilirlik katsayısı; puanlayıcılar arası güvenilirlik için Kendall’in uyum katsayısı (W), G kuramında ise tek değişkenli modellerde E_p^2 (G) ve Φ (Phi) katsayıları kestirilmiştir. Cronbach alfa güvenilirlik katsayıları birinci puanlayıcı için 0,77, ikinci puanlayıcı için 0,76 ve üçüncü puanlayıcı için 0,76 bulunmuştur. Puanlayıcılar arası uyum ise Kendall’in uyum katsayısı ile belirlenmiş; birinci, ikinci, üçüncü ve dördüncü vee diyagramları için Kendall W katsayısı sırasıyla; 0,959, 0,924, 0,963 ve 0,954 olarak bulunmuştur. Genellenebilirlik kuramı analizinde varyansları ve varyans yüzdelerini hesaplamak amacıyla, tümüyle çaprazlanmış desen uygulanmıştır. G Kuramına göre matematik başarısının ölçülmesiyle vee diyagramından elde edilen puanların Genellenebilirlik (G) katsayısı 0,918 ve Güvenirlik (Phi) katsayısı 0,899 bulunmuştur. G çalışması sonucunda, öğrenci (ö) ana etkisi için kestirilen varyans bileşenin toplam varyansın % 42,4’ünü açıkladığı ve toplam varyans içinde en yüksek değere sahip varyans bileşeni olduğu belirlenmiştir. Puanlayıcı (p) ana etkisi için kestirilen varyans bileşenin ise toplam varyansın %0,1’ini açıkladığı ve bu varyans bileşenin ana etkiler içinde en düşük değere sahip varyans bileşeni olduğu görülmektedir. Elde edilen bulgular, kullanılan ölçme aracının, öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar verdiğini göstermektedir.

Anahtar Kelimeler: Vee diyagramı, Güvenirlik, Klasik Test Kuramı, Genellenebilirlik Kuramı

Abstract

The aim of this study was to investigate reliability of score obtained from vee diagrams. Therefore in this study, the Classical test theory and generalizability theory were used. The study was carried out with descriptive research model. The participants of this research were composed of 32 students at 7th and 8th grades in Niğde. The responses given to the vee diagrams were scored through the holistic rubric. In CTT, Cronbach’s alpha reliability coefficient and Kendall’s concordance coefficient for interrater reliability; in G theory, E_p^2 (G) and Φ (Phi) coefficients with univariate model were computed. The Cronbach alpha reliability coefficients estimated were 0,77, 0,76, and 0,76 for each rater, respectively. For determination of interrater consistency, Kendall’s concordance coefficient was calculated as $W=0,959$ for the first vee diagram, $W=0,924$ for the second one, $W=0,963$ for the third one, $W=0,954$ for the fourth one. For the purpose of estimating variances and percent of variances in G study, fully crossed design was applied. Generalizability coefficient for vee diagram scores was 0,918 and phi coefficient was 0,899. As a result of the G study, it was determined that the variance component estimated for the student (s) main effect accounted for 42,4% of the total variance, and it was the variance component that had the highest value. The variance component estimated for raters (p) main effect, which is the lowest variance component in all main effect, accounted for 0,1% of the total variance.

*Bu çalışma IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde (Hacettepe Üniversitesi, 9-13 Haziran 2014) sözlü bildiri olarak sunulmuştur.

** Arş. Gör. Dr., Ömer Halisdemir Üniversitesi, Eğitim Fakültesi, Niğde-Türkiye, betul.polat6006@gmail.com

According to all results, it was seen that vee diagram was reliable for determination of students' mathematics achievement.

Keywords: Vee diagram, Reliability, Classical Test Theory, Generalizability Theory

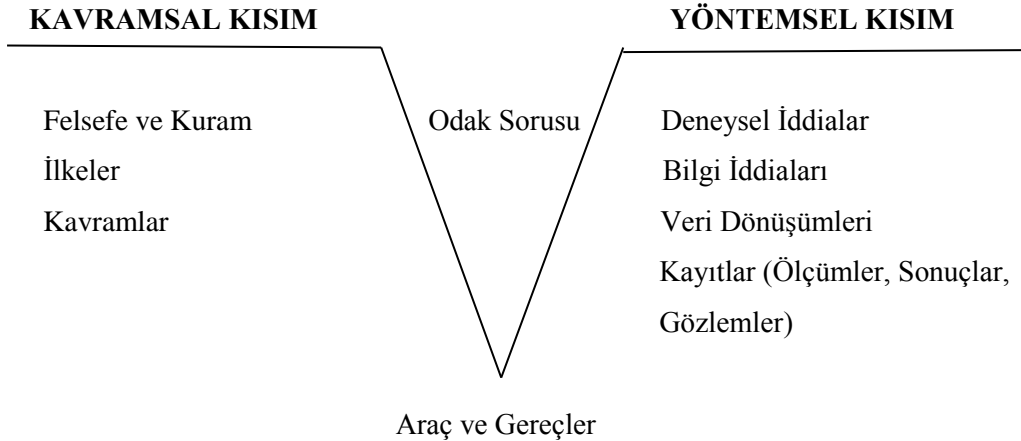
GİRİŞ

21. yüzyıl teknoloji çağında bilginin öneminin hızla artması ile gelişmiş toplumlarda bireylerin sahip olması gereken özellikler değişmiştir. Günümüzde bilgileri olduğu gibi depolayan bireylere değil; yeni bilgilere doğru ve hızlı bir biçimde ulaşarak, öğrendikleri bilgi ve becerileri yaşamlarının her alanında etkin bir biçimde kullanabilen bireylere ihtiyaç duyulmaya başlanmıştır (Kutlu, Doğan ve Karakaya, 2010). Gereksinim duyulan bireylerin özelliklerinin değişmesi ile eğitim sistemleri yapılandırmacı yaklaşıma göre yeniden düzenlenmiştir. Bu kapsamda, 2005–2006 öğretim yılında ilköğretim 1–5 sınıflarının ve aşamalı olarak 2006–2007 öğretim yılından itibaren ise diğer sınıflar için hazırlanan yeni öğretim programlarında değişiklikler meydana gelmiştir (Gelbal ve Kelecioğlu, 2007).

Yapılandırmacı yaklaşım temele alınarak hazırlanan matematik programında, kavramsal yaklaşım benimsenerek kavramlara ve kavramlar arası ilişkilere ağırlık verilmiştir (Baykul, 2007). Bu bağlamda, öğrencilerin matematiği öğrenebilmeleri için sadece işlemsel bilginin değil kavramsal bilginin de önemli olduğuna vurgu yapılmıştır. Matematiksel anlama, öğrencinin sadece formülleri bilmesine, hesaplamaları doğru yapmasına bağlı değildir; aynı zamanda kavramları ve kavramlar arasındaki ilişkileri, işlemleri anlamasına ve matematiksel düşünmesinin gelişmesine dayanmaktadır. Bu doğrultuda, okul matematiğinde işlemsel çözüm yollarının öğretilmesinin yanında kavram ve bu kavramlar arasındaki ilişkilere önem verilmesinin gerekliliği ortaya konulmuştur (Baki ve Kartal, 2004).

Matematik problemlerinin çözümünde işlemsel bilginin yanında kavramsal bilginin de önemine vurgu yapılmasına rağmen TIMSS 2011 matematik sonuçları, Türkiye'deki 8. sınıf öğrencilerinin bilme düzeyi ortalamasının, uygulama ve muhakeme düzeylerinin ortalamasından daha düşük olduğunu göstermektedir. Bu durum, ülkemizdeki öğrencilerin bilme düzeyinde yeterliliğe sahip olmadan bile uygulama ve muhakeme yapabildiklerinin bir göstergesi olarak ele alınmaktadır. Bilme düzeyinin diğer düzeyler arasında en düşük ortalamaya sahip olması ise ülkemizde matematik öğretiminde temel kavramların öğretimine ve ölçülmesine yeterince ağırlık verilmemesi ile açıklanabilir (Yücel, Karadağ ve Turan, 2013). Bu nedenle ülkemizde matematik öğretiminde işlemsel bilginin yanında kavramsal bilgiyi sınavan ölçme ve değerlendirme araçlarına gereksinim duyulmaktadır.

Kavramsal bilgiyi ölçebilen ölçme ve değerlendirme yaklaşımları, yapılandırmacı yaklaşımın temele alındığı programlarda yerini almıştır. Öğretim ve değerlendirme süreçlerini birbirine yakınlaştıran ve kavramsal bilgiyi de sınavan bu yeni yaklaşımlardan biri de vee diyagramlarıdır. Vee diyagramı, öğrencilerin bilgi yapılarını ve bilgi oluşturma süreçlerini anlamak amacıyla Gowin tarafından geliştirilmiştir (Alvarez ve Risko, 2007). Vee diyagramı "V" şeklinde bir diyagramdır ve iki temel kısımdan oluşur. Vee diyagramının sol tarafı kavramsal kısımdır. Bu kısım düşünme boyutunu içerir ve bu kısımda hipotezler geliştirmek için kullanılan kavramsal ve yapısal bilgi sunulur. Bu kısma kuram ve ilkeler ile kavramlar yazılır. Vee diyagramının sağ tarafı yönetsel kısımdır. Vee diyagramının sağ kısmı ise uygulama boyutunu içerir ve burada yönetsel ve işlemsel süreçler yer alır (Novak ve Gowin, 1984). "V" şeklinin ortasında odak sorusu ve şeklin sivri ucunda ise bilgi oluşturma sürecinde ihtiyaç duyulan araç ve gereçler yer alır. (Thiessen, 1993). Vee diyagramının genel gösterimi Şekil 1'de sunulmuştur (Novak ve Gowin, 1984).



Şekil 1. Vee Diyagramı

Vee diyagramı, problemin nasıl çözüldüğünü ya da bilginin nasıl ispat edildiğini, bu süreçte mevcut bilgilerle ihtiyaç duyulan yeni bilgilerin nasıl ilişkilendirildiğini ve bilgilerin nasıl yeniden yapılandırıldığını ortaya koymaktadır (Afamasaga-Fuata’i, 2004a). Bu bağlamda, vee diyagramı öğrencilerin problemleri çözerken var olan kavramsal bilgilerini sistemleştirmelerine, örgütlemelerine ve kullanmalarına yardımcı olmaktadır (Thiessen, 1993). Aynı zamanda vee diyagramı öğrencilerin kavram yanlışlarını belirlemeye ve gidermeye; kuram ve uygulama arasındaki boşluğu doldurmaya ve var olan kavramları organize etmeye de katkı sağlamaktadır. Vee diyagramı öğrenme sürecine sağladığı bu katkılardan dolayı hem öğretim hem de değerlendirme aracı olarak kullanılabilir (Afamasaga-Fuata’i, 2004b).

Matematik probleminin çözümü için gerekli olan işlemsel bilginin yanında kavramsal bilgiyi de sınavan vee diyagramlarından elde edilen sonuçlara dayalı olarak öğrenciler hakkında doğru kararlar verebilmek için bu puanların güvenilir olması gerekmektedir. Güvenirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesidir (Baykul, 2000). Güvenirlik, ölçme aracının ölçtüğü özelliği ya da özellikleri, ne derecede bir kararlılıkta ölçmekte olduğunun bir göstergesidir (Tekin, 2000). Bir ölçme aracıyla yapılan ölçmelerden elde edilen puanların güvenilirliği, ölçmenin üç temel kuramı olan klasik test kuramı (KTK), madde tepki kuramı (MTK) ve genellenebilirlik (G) kuramına dayalı yöntemlerle belirlenmeye çalışılmaktadır. Bu çalışmada KTK ve G kuramına dayalı yöntemler kullanılmıştır.

Güvenirlik, tesadüfi ölçme hatalarından etkilenen psikometrik bir özelliktir. Güvenirliği etkileyen hataların kaynakları dört başlık altında toplanabilir; a) ölçmenin yapıldığı kişiler, b) ölçmeyi yapan kişi ya da kişiler, c) ölçme ortamı ve d) ölçme aracıdır (Baykul, 2000; Cohen ve Swerdlik, 2009; Crocker ve Algina, 1986; Erkuş, 2003; Turgut ve Baykul, 1992). Eğitim ve psikoloji alanlarındaki gerçek dünyada var olan bazı yapıları ölçme yoluyla kuramsal yapılara bağlanmasına ve bunların açıklanmasına yardımcı olan KTK’da güvenilirlik belirlenirken ölçme hatalarına karışan hatalar, sadece tek bir kaynaktan gelen hatalar olarak ele alınmaktadır. Ele alınan bu varyans (değişkenlik) kaynağına göre de güvenilirlik farklı adlar almakta ve farklı anlamlar taşımaktadır. Bu doğrultuda hata kaynaklarına göre güvenilirlik belirleme yöntemleri de farklılıklar göstermektedir (Baykul, 2000).

Vee diyagramlarıyla yapılan ölçmelerde, öğrencilerin farklı puanlayıcılar tarafından değerlendirmesi sonucunda puanlayıcılar ve vee diyagramları ölçme sonuçlarına karışan farklı hata kaynakları olarak ele alınmıştır. Bu nedenle çalışmada, vee diyagramlarından elde edilen puanların güvenilirliğini belirlemek için KTK’ya dayalı yöntemlerden Cronbach alfa iç tutarlık katsayısı ile puanlayıcılar arası güvenilirlik için Kendall’in uyum katsayısı hesaplanmıştır.

Cronbach alfa katsayısı, ölçme aracında bulunan maddelerin ya da görevlerin birbirleri ve testin tamamı ile ilişkisinin bir ölçüsüdür ve “iç tutarlılık” anlamında güvenilirliği vermektedir. Bu

anlamdaki güvenilirlikte hata kaynağı maddeler ya da görevlerdir ve zaman bağlı olarak ortaya çıkabilecek hata kaynağı göz ardı edilmektedir (Brennan, 2001; Crocker ve Algina, 1986; Güler, 2008). İki ya da daha fazla puanlayıcı ya da yargıcının ölçülen aynı özelliklere ya da farklı maddelere ilişkin yaptıkları puanlamalar arasındaki tutarlılığın derecesi ise “puanlayıcı güvenilirliği” olarak adlandırılmaktadır ve bu durumda hata kaynağı puanlayıcıdır (Cohen ve Swerdlik, 2009; Güler, 2008). Kendall’ın uyuşum katsayısı, ikiden fazla puanlayıcının olduğu durumda güvenilirliğin hesaplanmasında kullanılan bir korelasyon katsayısıdır ve parametrik olmayan istatistiksel bir tekniktir (Martin ve Bateson, 1992; akt: Tavşancıl, 2002). Kendall’ın uyuşum katsayısında, puanlayıcılar arasındaki tutarlılık derecesinin en az 0,80 düzeyinde olması istenmektedir (Szymanski ve Linkowski, 1993).

G kuramı davranış ölçümlerinde; genellenebilirlik (G) ve karar (K) çalışmalarıyla güvenilir gözlemlerin tasarımını, araştırılmasını sağlayan ve temelinde varyans analizine (ANOVA) dayanan istatistiksel bir kuramdır. G kuramı, birçok varyans kaynağını tek bir analizde ele alır ve dolayısıyla her bir varyans kaynağının büyüklüğünün hesaplanmasını sağlar. Genellenebilirlik (G) kuramında KTK’nın aksine aynı anda bütün hata kaynakları dikkate alınarak tek bir hesaplama ile kapsamlı bir güvenilirlik elde edilmektedir (Brennan, 2001; Shavelson ve Webb, 1991).

G kuramında puanlayıcı, zaman, test formu, madde, görev gibi bir ölçme içerisinde yer alabilen bütün hata kaynakları göz önüne tutularak güvenilirlik için genellenebilirlik (G veya $E\rho^2$) ve güvenilirlik (Phi veya Φ) katsayıları elde edilmektedir (Atılğan, 2004). G katsayısı göreceli modellerde ve bağlı değerlendirmeler için kullanılır. G katsayısı, bir bireyin gözlenen puanından, o bireyin evren puanının ne derece doğru genellendiğinin bir göstergesidir. G katsayısı gerçek varyansın, gerçek varyans ile göreceli varyansın toplamına bölünmesiyle hesaplanır. KTK’ndaki güvenilirlik katsayısı gibi G katsayısı da bireylerin puanlarındaki çeşitliliğin oranını yansıtır. Phi katsayısı ise mutlak modellerde ve mutlak değerlendirmeler için kullanılır. Phi katsayısı gerçek varyansın gerçek varyans ile mutlak varyansın toplamına bölünmesiyle hesaplanır (Atılğan, 2005; Güler, 2008). Dolayısıyla, bu iki katsayı hatanın nasıl ele alındığına bağlı olarak değişiklik gösterir (Güler, 2008). Brennan (2001), G ve Φ katsayılarının kabul edilebilir yeterlik sınırlarının isteğe bağlı olarak değişebildiğini ancak bazı araştırmacıların G ve Φ katsayılarının 0,80’den büyük olduğu durumda güvenilirliğin “yüksek” olarak değerlendirilebileceğini belirtmektedir. Shavelson ve Webb (1991) ise G ve Φ katsayılarının en az 0,80 olması gerektiğini vurgulamaktadırlar.

Alanyazın incelendiğinde, vee diyagramlarının bir ölçme aracı olarak güvenilirliğinin ortaya konduğu sadece bir çalışmanın olduğu göze çarpmaktadır. Çalışkan (2014) çalışmasında puanlayıcılar arası güvenilirliği kestirmek amacıyla iki puanlayıcıdan elde edilen puanlar arasındaki korelasyon katsayısını hesaplamıştır. Çalışmada vee diyagramlarının iki puanlayıcı tarafından puanlanması sonucunda 0,87 güvenilirlik katsayısı elde edilmiştir. Bu nedenle işlemsel bilginin yanında kavramsal bilginin öğretiminde ve ölçülmesinde önemli bir yere sahip olan vee diyagramından elde edilen puanların güvenilirliğinin belirlenmesi gerekli görülmüştür.

Araştırmanın Amacı

Araştırmanın amacı, vee diyagramlarından elde edilen puanların güvenilirliğinin KTK ve G kuramı ile incelemektir. Bu amaç doğrultusunda KTK’ya dayalı yöntemlerden Croanbach alfa iç tutarlık katsayısı ile puanlayıcılar arası güvenilirlik için Kendall’ın uyuşum katsayısı hesaplanmıştır. G kuramında ise tek değişkenli modelle $E\rho^2$ (G) ve Φ (Phi) katsayıları kestirilmiştir. Bu doğrultuda araştırmada aşağıda verilen araştırma sorularına yanıt aranmıştır.

1. Vee diyagramlarından elde edilen puanlara ait iç tutarlığın derecesi nedir?
2. Vee diyagramlarına ait yapılan puanlamada puanlayıcılar arası tutarlılığın derecesi nedir?
3. Öğrenci, puanlayıcı, vee diyagramı ve bunların etkileşimlerine ait kestirilen varyans bileşenleri nasıldır?
4. Üç puanlayıcının 32 öğrencinin her bir vee diyagramlarına verdikleri yanıtları puanlanmasıyla elde edilen güvenilirlik katsayıları nasıldır?

5. Puanlayıcı sayısının bir artırılıp bir azaltıldığı ve vee diyagramı sayısının 2, 4, 6 ve 8 olduğu durumlarda elde edilen güvenilirlik katsayıları nasıldır?

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama aracı ve verilerin analizi ile ilgili bilgiler sunulmuştur.

Araştırmanın Modeli

Vee diyagramlarından elde edilen puanların güvenirliğinin klasik test kuramına ve genellenebilirlik kuramına dayalı yöntemlere göre belirlemeyi amaçlayan bu araştırma betimsel araştırma niteliğindedir. Betimsel araştırmalarda amaç, bir olgunun ya da durumun kesin tanımlamalarını ortaya koymaktır (Johnson ve Christensen, 2004).

Çalışma Grubu

Araştırmanın çalışma grubunu Niğde iline bağlı bir devlet ortaokulunun 7. ve 8. sınıfına giden toplam 32 öğrenci oluşturmuştur. Puanlayıcı sayısının üç olmasından dolayı aracın puanlama açısından kullanılabilirliği göz önüne alınarak 32 öğrenci araştırma kapsamına alınmıştır. Araştırmanın amacına daha iyi hizmet edebilecek bir çalışma grubu oluşturmak için matematik başarısı açısından farklı seviyelerde olan farklı sınıf düzeylerinden öğrenciler araştırmaya dâhil edilmiştir. Öğrencilerin tümü vee diyagramları hakkında bilgilendirilmiştir. Öğrencilerin sınıf ve cinsiyetlerine ilişkin dağılımları Tablo 1’de sunulmuştur.

Tablo 1. Öğrencilerin Sınıf ve Cinsiyete Göre Dağılımları

Sınıf	Cinsiyet	N	%	Toplam
7. Sınıf	Kız	9	56,3	16
	Erkek	7	43,7	
8. Sınıf	Kız	10	62,5	16
	Erkek	6	37,5	

Veri Toplama Aracı

Vee diyagramları “Geometri ve Ölçme” alt öğrenme alanlarında yer alan “Doğruda Açılar” ve “Çokgenlerde Açılar” konularında yer alan kazanımlara dayalı olarak hazırlanmıştır. Bu doğrultuda vee diyagramlarının yanıtlama süresi de göz önüne alınarak dört farklı vee diyagramı hazırlanmıştır. Hazırlanan vee diyagramları kazanıma ve öğrenci seviyelerine uygunluğu, ifadelerin anlaşılabilirliği açısından matematik öğretmenin görüşüne sunulmuş ve dönütler doğrultusunda gerekli düzeltmeler yapılmıştır. Öğrencilerin vee diyagramlarına verdikleri yanıtlar vee diyagramı puanlama anahtarıyla puanlanmıştır. Vee diyagramı puanlama anahtarı hazırlanırken Novak ve Gowin (1984) tarafından önerilen puanlama anahtarı dikkate alınmıştır. Vee diyagramı puanlama anahtarı, vee diyagramlarının kavramsal kısmı ile yöntemsel kısmı için iki farklı bütünsel dereceli puanlama anahtarının hazırlanmasıyla oluşturulmuştur. Vee diyagramının her bir kısmı için hazırlanan bütünsel puanlama anahtarları 0 ile 4 arasında derecelendirilmiştir. Öğrencilerin vee diyagramlarından aldıkları puanlar, vee diyagramının her iki kısımdaki öğrenci yanıtlarının ayrı ayrı puanlanması ve puanların toplanmasıyla elde edilmiştir.

Verilerin Analizi

Verilerin analizi iki aşamada gerçekleştirilmiştir. İlk aşamada, üç puanlayıcıdan elde edilen puanlar için açımlayıcı faktör analizi yapılmıştır. Açımlayıcı faktör analizi (AFA) sonucunda, her bir

puanlayıcıdan elde edilen puanlarda öz değeri birden büyük tek bir faktörün olduğu gözlenmiştir. Diğer bir deyişle, her bir puanlayıcıdan elde edilen puanlarda vee diyagramlarının tek bir faktör altında toplandığı ve bu puanlarda benzer sonuçların elde edildiği gözlenmiştir. Bu bulgular vee diyagramlarının ölçmek istenilen tek bir ortak yapıyı ölçtüğünü ortaya koymaktadır. Tablo 2’de AFA sonuçları sunulmuştur.

Tablo 2. Açımlayıcı Faktör Analizi Sonuçları

Vee diyagramı	Faktör yükleri			Ortalama puan
	Puanlayıcı 1	Puanlayıcı 2	Puanlayıcı 3	
VD1	,927	,907	,931	,932
VD2	,830	,842	,844	,852
VD3	,820	,803	,821	,827
VD4	,457	,458	,376	,443
Öz değer	2,429	2,386	2,395	2,474
Açıklanan varyans	60,722	59,652	59,872	61,838

Vee diyagramlarından elde edilen puanların güvenilirliğinin kestirildiği ikinci aşamada, KTK’dan ve G kuramından faydalanılmıştır. KTK kapsamında, Cronbach alfa katsayısı ve Kendall’in uyum katsayısı (W) kestirilmiş ve verilerin analizinde SPSS 13 programından yararlanılmıştır. G kuramına göre, ana ve ortak etkiler için varyans bileşenlerinin kestirilmesinde, puanların güvenilirliği için G ve Phi katsayılarının hesaplanmasında ve K çalışmasında EduG programı kullanılmıştır. Çalışmada farklı ilköğretim okullarından üç matematik öğretmeni puanlayıcı olarak görev almıştır. Vee diyagramı ve puanlanması hakkında öğretmenlere bilgi verilmiştir. Her bir öğrenci tüm vee diyagramlarını yanıtlamış ve tüm öğrencilerin yanıtları üç puanlayıcı tarafından aynı koşullar altında bağımsız olarak puanlanmış olduğu için değişkenlik kaynaklarına ilişkin tümüyle çaprazlanmış (öğrenci x vee diyagramı x puanlayıcı) desen kullanılmıştır.

BULGULAR VE YORUMLAR

Bu bölümde öncelikle vee diyagramlarından alınan puanlara ait betimsel istatistikler hesaplanmıştır. Puanlara ait betimsel istatistikler Tablo 3’te sunulmuştur.

Tablo 3. Vee Diyagramından Alınan Puanlara İlişkin Betimsel İstatistikler

Vee Diyagramı	Puanlayıcı	\bar{X}	Sx	En küçük değer	En büyük değer	Çarpıklık	Basıklık
VD1	P1	2,91	2,44	1,00	8,00	1,024	-,612
	P2	2,78	2,35	1,00	8,00	1,137	-,278
	P3	2,88	2,42	1,00	8,00	1,132	-,351
VD2	P1	1,72	2,23	0,00	7,00	1,194	,303
	P2	2,07	2,15	0,00	7,00	1,153	,192
	P3	2,16	2,34	0,00	7,00	,939	-,557
VD3	P1	2,72	1,58	1,00	7,00	1,299	1,802
	P2	2,91	1,72	1,00	7,00	1,062	,637
	P3	2,88	1,76	1,00	7,00	1,081	,416
VD 4	P1	1,66	1,86	0,00	6,00	,796	-,724
	P2	1,50	1,75	0,00	5,00	,684	-1,163
	P3	1,56	1,78	0,00	6,00	,939	-,225
Toplam	P1	9,00	6,29	2,00	25,00	,976	-,052
	P2	9,25	6,15	2,00	24,00	1,083	,005
	P3	9,47	6,39	2,00	25,00	,937	-,208

Her bir vee diyagramından alınabilecek en düşük puan “0”, en yüksek puan ise “8” dir. Öğrencilerin hem her bir vee diyagramında hem de tamamından aldıkları puanların düşük olduğu ve en düşük ortalamının dördüncü vee diyagramına ait olduğu görülmektedir. Puanlayıcıların verdikleri puanlara

ilişkin ortalamalar incelendiğinde ise puanlayıcılar arasında Vee Diyagramı 2’de daha çok farklılıkların olduğu gözlenmektedir. Puanlara ilişkin çarpıklık değerleri incelendiğinde ise, puanların sağa çarpık bir dağılım gösterdiği görülmektedir.

Araştırmanın amacı doğrultusunda öncelikle KTK’ya ilişkin güvenilirlik analizi bulguları, daha sonra G Kuramına ilişkin analiz sonuçları aşağıda sunulmuştur.

Klasik Test Kuramı

Vee diyagramından elde edilen puanların güvenilirliğini kestirmek için KTK’ya dayalı yöntemlerden Croanbach alfa iç tutarlık katsayısı ile Kendall’ın uyuşum katsayısı hesaplanmıştır. Cronbach alfa katsayısı her bir puanlayıcının öğrencilere verdikleri puanlar için ayrı ayrı hesaplanmıştır. Ayrıca üç puanlayıcının öğrencilere verdikleri puanların ortalaması alınarak Cronbach alfa değeri hesaplanmıştır. Birinci, ikinci ve üçüncü puanlayıcıların vee diyagramlarını değerlendirmeleri ile elde edilen Cronbach alfa katsayıları Tablo 4’te sunulmuştur.

Tablo 4. Cronbach Alfa Güvenirlik Katsayıları

Puanlayıcı	Cronbach alfa
P1	0,77
P2	0,76
P3	0,76
Ortalama puan	0,78

Her bir puanlayıcının vee diyagramlarına verdiği puanlar için hesaplanan Croanbach alfa güvenilirlik katsayıları birinci puanlayıcı için 0,77, ikinci puanlayıcı için 0,76 ve üçüncü puanlayıcı için 0,76 bulunmuştur. Puanlayıcıların öğrencilerin yanıtlarına verdikleri puanların ortalaması alınarak hesaplanan Cronbach alfa güvenilirlik katsayısı ise 0,78 bulunmuştur. Cronbach alfa katsayıları incelendiğinde bu değerlerin birbirine yakın olduğu gözlenmektedir. Nunnally (1978)’ a göre 0,70 ve üzeri güvenilirlik katsayısı kabul edilebilir olarak görülmektedir. Bu doğrultuda elde edilen bu katsayılar göre, vee diyagramlarının öğrenci matematik başarısını birbirleriyle tutarlı bir şekilde ölçtüğü sonucuna ulaşılabilir.

Üç puanlayıcının aynı koşullar altında her bir öğrenciyi bağımsız olarak puanlamasına ilişkin elde edilen puanlar arasındaki tutarlılık derecesi parametrik olmayan bir teknik olan Kendall’ın uyuşum katsayısı ile hesaplanmıştır. Bulgular Tablo 5’te sunulmuştur.

Tablo 5. Kendall’ın Uyuşum Katsayıları

Vee diyagramı	W	χ^2	p
VD1	0,959	89,222	,000*
VD2	0,924	88,754	,000*
VD3	0,963	89,511	,000*
VD4	0,954	88,754	,000*
Toplam	0.987	91.795	.000*

*p<0,05

Tablo 5 incelendiğinde dört vee diyagramından elde edilen puanlar için uyuşum katsayısı W=0,987 ($\chi^2=91,795$; p<0,05) bulunmuştur. Birinci vee diyagramı için W=0,959 ($\chi^2=89,222$; p<0,05), ikinci vee diyagramı için W=0,924 ($\chi^2=85,972$; p<0,05), üçüncü vee diyagramı için W=0,963 ($\chi^2=89,511$; p<0,05) ve dördüncü vee diyagramı için W=0,954 ($\chi^2=88,754$; p<0,05) olduğu ve bu katsayıların oldukça yüksek olduğu görülmektedir. Bu bulgulara dayalı olarak her bir puanlayıcının aynı koşullar altında vee diyagramlarını değerlendirmeleri arasında uyumun yüksek olduğu, dolayısıyla

puanlayıcıların öğrencilere verdikleri puanlarda katılık/cömertlik düzeylerinin birbirine benzer olduğu sonucuna ulaşılabilir.

Genellenebilirlik Kuramı

Üç puanlayıcıdan her birinin öğrencilerin her bir vee diyagramına vermiş oldukları yanıtları değerlendirmeleri doğrultusunda çaprazlanmış (öxvxp) desen ile yapılan G çalışması sonucunda kestirilen varyans bileşenleri ve açıklama yüzdeleri Tablo 6'da verilmiştir. Çaprazlanmış (öxvxp) desen Tablo 6'da da görüldüğü üzere sekiz varyans kaynağına ayrılmaktadır. Her bir varyans kaynağına ait kestirilen varyans bileşenleri aşağıda açıklanmıştır.

Tablo 6. Vee Diyagramlarına İlişkin Puanlamada öxvxp Modeli İçin Kestirilen Varyans Bileşenleri

Varyans kaynağı	Kareler toplamı	sd	Kareler ortalaması	Varyans	%
Öğrenci(ö)	902,372396	31	29,108787	2,567484	42,4
Vee diyagramı (v)	6,320312	3	2,106771	0,000262	0,3
Puanlayıcı (p)	2,255208	2	1,127604	0,000152	0,1
öxv	93,429688	93	1,004620	2,173787	38,3
öxp	17,244792	62	0,278142	0,000178	0,2
vxp	113,515625	6	18,919271	0,000437	0,5
öxvxp,e	554,984375	186	2,983787	1,165230	18,2
Toplam	1690,122396	383			100

G çalışması sonucunda kestirilen varyans bileşenleriyle bu bileşenlerin toplam varyansı açıklama yüzdeleri incelendiğinde, öğrenci (ö) ana etkisi için kestirilen varyans bileşeninin toplam varyansın %42,4'ünü açıkladığı görülmektedir. Öğrenciler için kestirilen varyans bileşeni, toplam varyans içinde en yüksek değere sahip varyans bileşenidir. Genellenebilirlik çalışmalarında, öğrenci (yanıtlayıcı) ana etkisi evren puanı varyansı olarak değerlendirilmekte ve ölçülen özellik bakımından öğrenciler arasındaki farklılığı ifade etmektedir (Brennan, 2001). Öğrenciler için kestirilen varyansın toplam varyans içindeki oranının en büyük olması istenilen bir durum olduğundan, araştırma sonucunda elde edilen bu bulgu, vee diyagramlarının ölçülen özellik bakımından öğrenciler arasındaki farklılıkları ortaya çıkarılabildiğini ortaya koymaktadır.

G çalışması sonucunda kestirilen vee diyagramı (v) ana etkisi varyans bileşeni toplam varyansın %0,3'ünü açıklamaktadır. Vee diyagramı ana etkisi varyans bileşeninin büyüklüğünün, ana etkiler içinde ikinci sırada yer aldığı görülmektedir. Bu bulgu, vee diyagramlarının güçlük düzeylerinin birbirinden çok fazla farklılaşmadığını ortaya koymaktadır. Puanlayıcı (p) ana etkisi için kestirilen varyans bileşeninin ise toplam varyansın %0,1'ini açıkladığı ve bu varyans bileşeninin ana etkiler içinde en düşük değere sahip varyans bileşeni olduğu görülmektedir. G çalışması sonucunda kestirilen puanlayıcı ana etkisinin toplam varyans içindeki oranının çok düşük çıkması, puanlayıcıların tüm öğrenciler için yaptıkları puanlamalar arasında büyük farklılıkların bulunmadığını, yapılan puanlamalar arasında bir uyumun olduğunu ortaya koymaktadır.

Öğrenci-vee diyagramı ortak etkisi (öxv) için kestirilen varyans bileşeni, toplam varyans içinde en büyük ikinci değere sahiptir. Bu varyans bileşeni toplam varyansın %38,3'ünü açıklamaktadır. Bu bulgu, öğrencilerin bir vee diyagramından diğerine bağlı konularının değiştiğini ortaya koymaktadır. Öğrenci-puanlayıcı ortak etkisi (öxp) için kestirilen varyans bileşeni ise toplam varyansın %0,2'sini açıklamaktadır. Bu oranın sıfıra çok yakın olması bir puanlayıcının belli bir öğrenciyi diğer puanlayıcılara göre daha katı/cömert puanlamadıklarını göstermektedir. Vee diyagramı-puanlayıcı ortak etkisi (vxp) varyans bileşeni ise, toplam varyansın %0,5'ini açıklamaktadır. Bu doğrultuda, bu oranın da sıfıra çok yakın olması puanlayıcıların bir vee diyagramını diğerleriyle aynı kararlılıkta puanladıklarını göstermektedir. Kalan etki varyans bileşeninin (öxvxp) toplam varyansı açıklama oranı %18,2'dir ve toplam varyans içinde en büyük

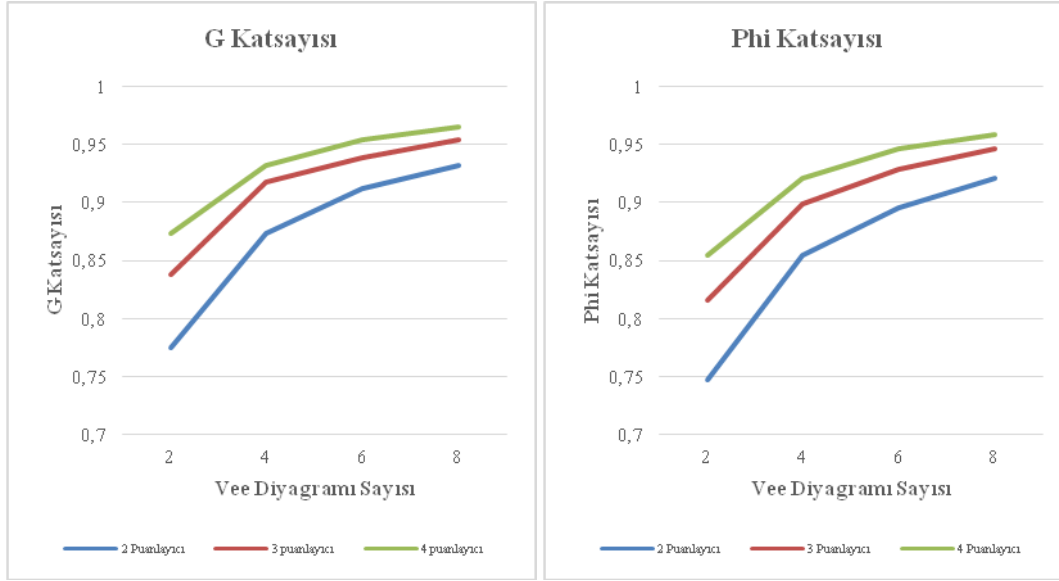
üçüncü değere sahiptir. Bu bulguya göre öğrenci-vee diyagramı-puanlayıcı ortak etkisi ve/veya tesadüfi hataların yüksek olduğu sonucuna ulaşılabilir.

Vee diyagramlarından elde edilen verilere göre orijinal puanlayıcı ve vee diyagramı sayılarına göre G ve Phi(Φ) katsayıları kestirilmiştir. Ayrıca puanlayıcı sayısının bir artırılıp bir azaltılması ve vee diyagramı sayılarının iki arttırılıp iki azaltılması senaryolarına göre yapılan karar çalışmaları sonucunda G ve Phi (Φ) katsayıları kestirilmiştir. Bulgular Tablo 7’de sunulmuştur.

Tablo 7. Vee Diyagramlarından Elde Edilen Verilere Göre Kestirilen G ve Phi Katsayıları

Puanlayıcı sayısı	Vee diyagramı sayıları							
	2		4		6		8	
	$E\rho^2$	Φ	$E\rho^2$	Φ	$E\rho^2$	Φ	$E\rho^2$	Φ
2	0,775	0,747	0,873	0,855	0,912	0,895	0,932	0,921
3	0,838	0,816	0,918	0,899	0,939	0,929	0,954	0,947
4	0,873	0,855	0,932	0,921	0,954	0,947	0,965	0,959

Dört vee diyagramının üç puanlayıcı tarafından değerlendirilmesi sonucunda elde edilen puanlardan kestirilen G ($E\rho^2$) katsayısı 0,918; Phi (Φ) katsayısı 0,899 bulunmuştur. Bu değerler incelendiğinde değerlerin birbirine çok yakın ve oldukça yüksek olduğu görülmektedir. Vee diyagramı sayısının sabit tutularak (4) puanlayıcı sayısının bir azaltılması (2) durumunda kestirilen G katsayısı 0,873 ve Φ katsayısı 0,855, puanlayıcı sayısının bir arttırılması (4) durumunda ise G katsayısı 0,932 ve Φ katsayısı 0,921 olarak kestirilmiştir. Puanlayıcı sayısının sabit tutularak (3) vee diyagramı sayısının iki azaltılması (2) durumunda kestirilen G katsayısı 0,838 ve Φ katsayısı 0,816, iki arttırılması (6) durumunda ise G katsayısı 0,939 ve Φ katsayısı 0,929 olarak hesaplanmıştır. Tablo 6’dan görüleceği üzere, puanlayıcı sayısının sabit tutularak vee diyagramı sayısının iki artması durumunda, G katsayısındaki artış 0,021 iken, vee diyagramı sayısının sabit tutularak puanlayıcı sayısının bir arttırılması durumunda G katsayısındaki artış 0,014 olmaktadır. Benzer durum, Φ katsayısı için de incelendiğinde, puanlayıcı sayısının sabit tutularak vee diyagramı sayısı iki arttırıldığında meydana gelen artış 0,030 iken, vee diyagramı sayısının sabit tutularak puanlayıcı sayısı bir arttırıldığında gözlenen artış 0,021 olduğu gözlenmektedir. Tablo 7 ve Şekil 2 incelendiğinde, vee diyagramı sayısı sabit tutulduğunda puanlayıcı sayısındaki artışın ya da azalmanın G ve Φ katsayılarını fazla etkilemediği görülmektedir. Puanlayıcı sayısı sabit tutulduğunda vee diyagramı sayısının arttırılması durumunda da G ve Φ katsayılarında daha fazla bir artışın olduğu fakat bu artışın giderek azaldığı gözlenmektedir. Diğer bir deyişle, hem G katsayısı, hem de Φ katsayısı için vee diyagramı sayılarını iki arttırmanın, puanlayıcı sayısını bir arttırmaya göre daha büyük artışa sebep olduğu söylenebilir. Dolayısıyla bir ölçme aracının kullanışlılığı dikkate alındığında, puanlayıcı sayısını arttırmanın emek ve zaman açısından ekonomik olmadığı söylenebilir. Vee diyagramları işlemsel bilginin yanında kavramsal bilgiyi de değerlendirildiği için yanıtlama süresinin göz önüne alınarak vee diyagramı sayısının arttırılması ölçme aracının güvenilirliği, kapsam geçerliği ve kullanışlılığı açısından daha uygun olacağı söylenebilir.



Şekil 2. K Çalışması ile Vee Diyagramı ve Puanlayıcı Sayıları Senaryolarına Göre G ve Φ Katsayıları

SONUÇ, TARTIŞMA VE ÖNERİLER

Öğrencilerin matematik başarısının ölçülmesinde kullanılan dört vee diyagramının KTK'ya göre iç tutarlılık düzeyinin belirlenmesi amacıyla Cronbach alfa güvenilirlik katsayısı hesaplanmıştır. Üç puanlayıcı için ayrı ayrı ve tüm puanlayıcıların vee diyagramlarına verdikleri puanların ortalamalarından elde edilen güvenilirlik katsayıları 0,76 ile 0,78 arasına bulunmuştur. Bu katsayılara göre, matematik başarısının ölçülmesinde kullanılan vee diyagramlarının matematik başarısını birbirleriyle tutarlı bir şekilde ölçtüğünü ortaya koymaktadır. Üç puanlayıcı arasında tutarlılığın olup olmadığını belirlemek için KTK'ya göre hesaplanan Kendall'in uyum katsayısı sonuçları puanlayıcıların vee diyagramlarına verilen yanıtları puanlamadaki katılık/cömertlik düzeylerinin birbirine benzer olduğunu göstermektedir.

Araştırmada yer alan dört vee diyagramına tümüyle çaprazlanmış övxp modeli uygulanarak yapılan G çalışması sonucunda öğrenci (ö) ana etkisi için kestirilen varyans bileşeninin toplam varyansın büyük bir kısmını açıkladığı gözlenmektedir. Bu bulgu, vee diyagramlarının öğrenciler arasındaki farklılıkları ortaya çıkarabildiğini ortaya koymaktadır. Puanlayıcı (p) ana etkisi için kestirilen varyans değişkeninin açıklama oranının ise en düşük değere sahip varyans bileşeni olduğu ve sıfıra oldukça yakın olduğu gözlenmektedir. Bu bulgu ise puanlayıcıların yaptıkları puanlamalar arasında bir tutarlılığın olduğunu ortaya koymaktadır. G çalışması sonucunda dört vee diyagramı sayısına ve üç puanlayıcıya göre elde edilen G katsayısı ile Φ katsayısı birbirlerine çok yakın ve çok yüksek bulunmuştur. Yapılan K çalışması sonucunda puanlayıcı sayısındaki artış ya da azalma vee diyagramlarındaki artış ya da azalmaya göre G ve Φ katsayılarını daha az etkilemektedir. Dolayısıyla zaman ve emek açısından ölçme aracının kullanışlılığı ile aracın kapsam geçerliği dikkate alınarak vee diyagramlarının sayısı artırılarak daha güvenilir ölçmeler yapılabilir.

Yapılan güvenilirlik çalışmaları sonucunda elde edilen tüm bu bulgular, vee diyagramının öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar ortaya koyduğunu göstermektedir. Bu açıdan vee diyagramlarından elde edilen puanlarla öğrenciler hakkında alınan kararların güvenilir olduğu sonucuna varılabilir. Bu bağlamda, öğretmenlere öğrencilerinin hem kavramsal hem de işlemsel bilgilerinin ölçmelerinde vee diyagramlarını kullanmaları önerilmektedir. Araştırmacılar ise matematik dersinden farklı bir disiplinde vee diyagramları hazırlayarak benzer çalışmalar yürüterek sonuçları karşılaştırabilirler.

KAYNAKÇA

- Afamasaga-Fuata'i, K. (2004a, September). *An undergraduate's understanding of differential equations through concept maps and vee diagrams*. In A. J. Canäs, J. D. Novak & Gonázales (eds). *Concept Maps: Theory, Methodology, Technology*. Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain.
- Afamasaga-Fuata'i, K. (2004b, September). *Concept maps and vee diagrams as tools for learning new mathematics topics*, In A. J. Canäs, J. D. Novak and Gonázales (eds). *Concept Maps: Theory, Methodology, Technology*. Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain.
- Alvarez, M.C. & Risko, V.J. (2007). The use of vee diagrams with third graders as a metacognitive tool for learning science concepts. Retrieved from: <http://e-research.tnstate.edu/pres/5>
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2005). Genellenebilirlik kuramı ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama*, 4(7), 95-108.
- Baki, A., ve Kartal, T. (2004). Kavramsal ve İşlemsel bilgi bağlamında lise öğrencilerinin cebir bilgilerinin değerlendirilmesi. *Türk Eğitim Bilimleri Dergisi*, 2(1), 13-26.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme Klasik Test Teorisi ve uygulaması*. Ankara: ÖSYM Yayınları
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Cohen, J. R., & Swerdlik, E. M. (2009). *Psychological testing and assessment: An Introduction to Tests and Measurement*. (7th Ed.). New York: McGraw-Hill Book Co.
- Crocker, L. M., & Algina, L. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Çalışkan, İ. (2014). The perceptions of pre-service science teachers' about using vee diagrams and electronic portfolios in physics laboratory course. *Educational Research and Reviews*, 9(6), 173-182.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği
- 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.
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. NY: Pearson/Allyn & Bacon.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A primer*. Newbury Park, CA: Sage Publications.
- Szymanski, E. M., & Linkowski, D. C. (1993). Human resource development: An examination of perceived training needs of certified rehabilitation counselors. *Rehabilitation Counseling Bulletin*, 37 (2), 163-176.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*, Ankara: Nobel Yayıncılık.
- Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları
- Turgut, M. F., ve Baykul, Y. (1992). *Ölçme teknikleri*. Ankara: ÖSYM Yayınları.
- Thiessen, R. (1993). The Vee Diagram: A Guide for Problem Solving, *Aims Newsletter*, May/June.
- Yücel, C., Karadağ, E., ve Turan, S. (2013, Şubat). *TIMSS 2011 ulusal ön değerlendirme raporu*. Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitimde Politika Analizi Raporlar Serisi I, Eskişehir.

EXTENDED ABSTRACT

Introduction

With the increase in the importance of knowledge in the technology period of 21st century, the characteristics of individuals needed by societies have changed. Today, not the individuals storing information but the ones using the information and skills learnt at school in real life are needed. In this context, education systems have been reconstructed in accordance with the constructivist approach. Accordingly, a variety of changes were made in the curriculums depending on this approach. In the math curriculum, by adapting a conceptual approach, the concepts and the relationships between concepts were emphasized. Moreover, not only the procedural knowledge but also the conceptual knowledge was emphasized in the teaching of math.

Despite the emphasis on the conceptual knowledge along with the procedural knowledge, TIMSS 2011 mathematic results showed that eighth graders' mean of the level of knowing is lower than the mean of the level of applying and reasoning in Turkey. This situation can be explained as not enough importance was given to teaching and measuring of basic concepts in mathematics education in Turkey. Therefore, measurement and evaluation instruments testing the conceptual knowledge along with the procedural knowledge are required in our country.

Measurement and evaluation methods testing the conceptual knowledge took place in these curriculums. One of these methods is Vee diagrams. Vee diagrams reveal how information is proved or how a problem is solved, how the current information and the required new information is associated. Depending on the results obtained from vee diagrams, which test the conceptual knowledge along with the procedural knowledge needed for the solution of a mathematics problem, it can be stated that these scores should be reliable in order to make accurate decisions about students.

When the literature is reviewed, no study was observed about the reliability of scores obtained from vee diagrams. Therefore, it was considered that investigating the reliability of the scores obtained from vee diagram is necessary. In this direction, the reliability of the scores obtained from vee diagram is aimed to predict by computing Cronbach alpha internal consistency coefficients and Kendall's concordance coefficient for interrater reliability based on Classical test theory (CTT) and the $E\rho^2$ (G) and Φ (Phi) coefficients with univariate model based on generalizability (G) theory.

Method

This research aiming at determining the reliability of the scores obtained from vee diagrams based on the techniques of CTT and G theory is a descriptive study. The participants of this research were composed of 32 students at seventh or eighth grades of a state school in Niğde. Four different vee diagrams based on the objectives in the topics of "Angles of a line" and "Angles of a polygon" in "Geometry and Measurement" learning domain were prepared. A mathematic teacher was asked to assess these vee diagrams in terms of students' levels and objectives and the comprehensibility of the statements. The responses given to the vee diagrams were scored through a holistic rubric,

The analysis of the data was conducted in two steps. In the first step, an exploratory factor analysis was conducted on the scores obtained from three raters. As a result of the exploratory factor analysis, it was observed that the scores given by each rater on vee diagrams gathered under one factor. In the second step, reliability of scores obtained from vee diagrams were predicted. SPSS 13 Software was used for estimating the Cronbach Alpha coefficient and Kendall's Concordance Coefficient (W). EduG Software was used for estimating the variance components, G and Phi coefficients, and Decision (D) study. The fully crossed design was used for source of variation since each student responded to all items and three raters rated these responses.

Results and Discussion

The estimated Cronbach alpha coefficients were 0,77, 0,76, and 0,76 for each rater, respectively. Cronbach alpha reliability coefficient was calculated as 0,78 by using the mean scores of grades given by three raters. These values were determined to be close. According to coefficients obtained, vee diagrams were determined to measure the mathematics achievement consistently. The degree of consistency among the scores given by three raters under the same conditions was estimated using Kendall's Concordance coefficient. Kendall's concordance coefficient for four vee diagrams was estimated to be $W=0,987$. Kendall's concordance coefficient were estimated as $W=0,959$ for the first vee diagram, $W=0,924$ for the second one, $W=0,963$ for the third one, $W=0,954$ for the fourth one. These coefficients were observed to be high. Depending on these findings, it can be stated that the consistency among the assessments made by three raters under the same conditions was high.

As a result of the G study, it was determined that the variance component estimated for the student (s) main effect accounted for 42,4% of the total variance, and it was the variance component that had the highest value. This finding indicates that vee diagrams have a capacity to reveal the differences among the students in terms of the characteristics which are measured. The variance component estimated for the vee diagram (v) main effect accounted for 0,3% of the total variance. This is the second among the main effect. This finding shows that the difficulty levels of vee diagrams were similar. The variance component estimated for raters (p) main effect accounted for 0,1% of the total variance. This is the lowest value among the main effects. This finding shows that there was consistency between raters' scores.

The variance component estimated for the student-vee diagram common effect accounted for 38,3% of the total variance. This value is the second highest among the total variance. This finding shows that relative position of the students varied from one vee diagram to the other. The variance component estimated for the student-rater common effect accounted for 0,2% of the total variance. This finding shows that all raters are same level of lenient/severe across all students. The variance component estimated for the vee diagram-rater common effect accounted for 0,5% of the total variance. This rate is close to zero, which shows that the raters graded the vee diagrams with the same stability. The residual effect variance component accounted for 18,2% of the total variance. This is the third highest value in the total variance. This finding shows that student-vee diyagram-rater common effect and/or random error had a high value.

As a result of the scenarios where the number of raters were added and deducted one and the number of vee diagrams were added and deducted two, G and Phi (Φ) coefficients were estimated. Estimated from the assessments of four diagrams by three raters, G ($E\rho^2$) coefficient was determined to be 0,918, and Phi (Φ) coefficient was determined to be 0,899. As a result of the D study, it was observed that an increase or decrease in the number of raters barely had an effect on G and Φ coefficients when the number of vee diagrams was fixed. On the other hand, when the number of raters was fixed, an increase in the number of vee diagrams caused a higher increase in G and Φ coefficients; however, this increase diminished gradually. In other words, it can be stated that increasing the number of vee diagrams by two causes a bigger increase in both G and Φ coefficients compared to increasing the number of raters by one. Therefore, when the practicality of a measurement tool is considered, it can be expressed that increasing the number of raters is not economical in terms of time and effort. Increasing the number of vee diagrams by considering the response time since the vee diagrams assess the conceptual knowledge along with the procedural knowledge is better in terms of reliability, content validity, and practicality.

According to the findings, it can be stated that vee diagrams yield reliable results in assessment of mathematics achievement. In this direction, the teachers are recommended to use vee diagrams to measure the students' both conceptual and procedural knowledge. The researchers can conduct similar studies by preparing vee diagrams for courses other than mathematics and compare the results.

The Examination of Model Fit Indexes with Different Estimation Methods under Different Sample Sizes in Confirmatory Factor Analysis*

Doğrulayıcı Faktör Analizinde Farklı Örneklem Büyüklüklerinde Farklı Kestirim Yöntemleriyle Hesaplanan Uyum İndekslerinin İncelenmesi

Ayfer SAYIN**

Abstract

In adjustment studies of scales and in terms of cross validity at scale development, confirmatory factor analysis is conducted. Confirmatory factor analysis, multivariate statistics, is estimated via various parameter estimation methods and utilizes several fit indexes for evaluating the model fit. In this study, model fit indexes utilized in confirmatory factor analysis are examined with different parameter estimation methods under different sample sizes. For the purpose of this study, answers of 60, 100, 250, 500 and 1000 students who attended PISA 2012 program were pulled from the answers to two dimensional “thoughts on the importance of mathematics” dimension. Estimations were based on methods of maximum likelihood (ML), unweighted least squares (ULS) and generalized least squares (GLS). As a result of the study, it was found that model fit indexes were affected by the conditions, however some fit indexes were affected less than others and vice versa. In order to analyze these, some suggestions were made.

Keywords: confirmatory factor analysis, sample size, fit index, scale development and adjustment

Öz

Ölçme araçlarının uyarlama çalışmalarında ve ölçme aracı geliştirirken çapraz geçerlik çalışmaları kapsamında doğrulayıcı faktör analizi gerçekleştirilmektedir. Çok değişkenli bir istatistik olan doğrulayıcı faktör analizinin hesaplanmasında kullanılan birçok parametre kestirim yöntemi ve model uyumunun değerlendirilmesinde kullanılan farklı uyum indeksleri mevcuttur. Bu çalışmada, doğrulayıcı faktör analizinde kullanılan model-veri uyum indekslerinin farklı örneklem büyüklüklerinde farklı parametre kestirim yöntemleri ile hesaplama sonuçları incelenmiştir. Bu doğrultuda PISA 2012 çalışmasına katılan ve iki boyuttan oluşan “matematiğin önemine yönelik görüşler” maddelerine cevap veren öğrenciler içerisinde 60, 100, 250, 500 ve 1000 kişilik veri setleri belirlenmiştir. Hesaplamalarda en çok olabilirlik (EÇO), ağırlıklandırılmamış en küçük kareler (AEKK), genelleştirilmiş en küçük kareler (GEEK) hesaplama yöntemleri kullanılmıştır. Araştırmanın sonucunda oluşturulan koşullardan bazı uyum indekslerinin daha fazla, bazılarının ise daha az etkilendiği belirlenmiştir. Bu doğrultuda analizler için önerilerde bulunulmuştur.

Anahtar Kelimeler: doğrulayıcı faktör analizi, örneklem büyüklüğü, uyum indeksi, ölçek uyarlama ve geliştirme

INTRODUCTION

Observing humans, as they are one of the most basic subjects of measurement, can lead to information about humans, objects, actions, or processes (DeVellis, 2003). However, objects, actions, or behaviors cannot always be observed directly in the studies concerning social sciences. Psychological constructs that are subject to research via indirect methods bring about problems with

* . This study was presented at the V. Measurement and Evaluation in Education and Psychology Conference organized by Akdeniz University in Antalya, Turkey, on September 1–3, 2016.

** Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme, Ankara-Türkiye, ayfersayin@yahoo.com

the improvement of the measurement devices. For instance, psychological measurements are always based on limited behavior patterns and there is always a chance for faulty measurements. Units that are not defined precisely on the measurement tools and methods create a different problem for measurements. Psychological constructs should be demonstrated with their relative underlying constructs and observed phenomena because they cannot be defined as standalone, thus require observed responses (Crocker & Algina, 1986).

Confirmatory factor analysis (CFA), which in terms of structural equation modeling explores the relationships between latent and observable variables, analyses the measurement model where factor count and relations indicated beforehand. CFA focuses on one variable explaining another variable (or variables) and its error variances from all its sources. Error in measurement varies from each other and unbound to the factors. Relations between factors don't have to be completely analyzed (factors are assumed to differ as a set) (Kline, 2011).

Since CFA is a multivariate statistics model; normality of the data set, single and multivariate outliers, linearity, multicollinearity and homogeneity of variances should be analyzed before the estimation (Tabachnick & Fidell, 2007). When multivariate normality assumption is not supported, unweighted least squares method can be utilized.

In the estimation process of CFA, firstly the description of the model is presented, then the degrees of freedom in the model is explored by identifying the model. After that, in accordance with the features of the data set, a parameter estimation method is settled upon and used to determine the path coefficients and error variances in the observed-latent variables. Lastly, to evaluate the model-data fit, fit indexes are calculated and necessary adjustments are made according to the modification suggestions.

CFA, which is a special instance of structural equation modeling that all the equations involving multiple regression equations being calculated at the same time, is a multivariate statistics and it depends on certain assumptions. These assumptions are level of scale, outlier, missing value, normality, multiple connections, variable count, linearity, uncorrelated error term and sample size (Schumacker & Lomax, 2004; Tabachnick & Fidell, 2007; Byrne, 2010). Kaplan (2001) suggests that basic assumptions of structural equation modeling are sufficient sample size, multivariate normality, missing data and proper building of the model. Fan, Thompson & Wang (1999) focused in their studies that effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Researchers state that although there are so many fit indexes which were developed under different theoretical rationales, it is not known which fit indexes are the ideal. Therefore, it is necessary to investigate the variables in which fit indexes are sensitive (Fan and others, 1999). Fan and Sivo (2007) examined the sensitivity of model fit indices to different types of models while controlling for the severity of model misspecification. At each level of misspecification, the severity of misspecification is comparable across the three different models. Ximénez (2009) first examined recovery for models correctly specified with the known number of factors, and then investigated recovery for models incorrectly specified by underfactoring. Savalei (2012) was studied with RMSEA under different types of misspecification.

Within the scope of this research, two models consisting of two dimensions were created. The difference between the first model and the second model is the items of the dimensions. Within the scope of research, three questions are answered:

- What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 1?
- What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 2?
- How do fit indexes differ within the Model 1 and Model 2, under different sample sizes and different parameter estimation methods?

Purpose of the study

CFA, which is used commonly to test a theoretical construct, used mainly in scale development, scale adjustment, cultural comparison, and comparing groups. CFA should be properly managed in areas like development and adaptation of scales used in the education field. When the literature in the field is examined, it can be seen that confirmatory factor analysis along with structural equation modeling is used frequently. However, in those studies sample size assumption is not properly utilized or not determined in accordance with assumptions of parameter estimation methods. This study aims to investigate the model fit indexes with different parameter estimation methods under different sample size conditions.

METHODOLOGY

Research model

In this research, the effect of varying parameter estimation methods and different sample sizes on fit indexes is meant to be explored. At the same time the effect of models set on the model-data fit indexes are meant to be determined. From this point of view, the research can be classified as a “basis research”.

Research Group

The research group consists of students who attended the PISA 2012 in Turkey. In PISA, 4848 students answered the questions in the “views on the importance of mathematics” dimension. The missing or invalid and n/a data (n=1693) are removed from the data set and also analyses of the assumptions were done. From the remaining student answers (n=3155), different clusters of 60, 100, 250, 500 and 1000 answers were pulled randomly and processed within the research.

Data Collection Tool

In PISA 2012, six items in the mathematics dimension of the research were scored in a four-point likert scale. First three of these items ask about their friends’ view on mathematics, and the remaining three asks about their families’ views on mathematics. Exploratory factor analysis was used with the answers of 3155 students who responded to the survey items in Turkey. At the end of the calculation, it was concluded that the dimension regarding the views of the friends of the students accounted for %37.193 of the total variance, and the family dimension accounted for %32,740 of the total variance. The total variance of the six items was calculated as %69,933. Cronbach’s alpha of the first dimension was calculated as 0,824; reliability coefficient of the second dimension was calculated as 0,705. The items can be seen in Table 1 below.

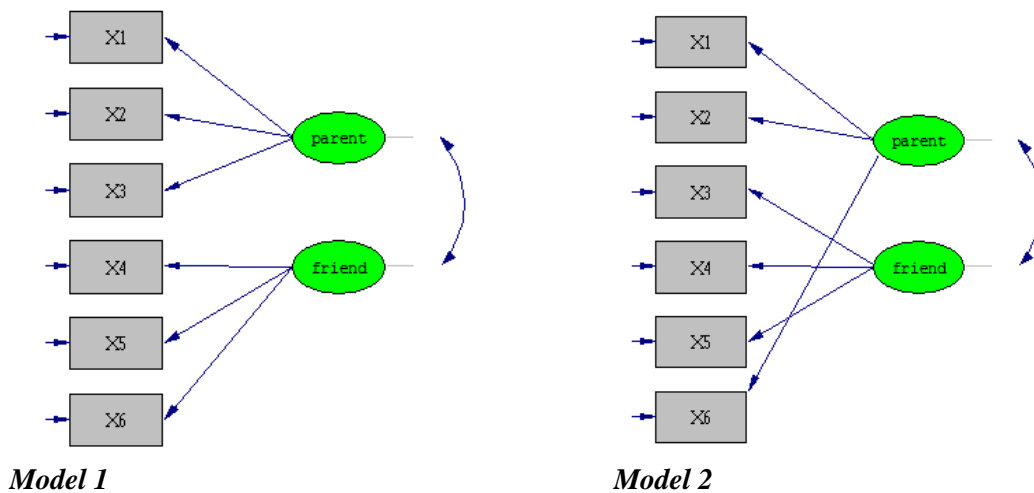
Table 1. Items and Dimension

Dimension	Items
Friends	X1. Most of my friends do well in mathematics.
	X2. Most of my friends work hard on mathematics.
	X3. My friends enjoy taking mathematics tests.
Parents	X4. My parents believe it’s important for me to study mathematics.
	X5. My parents believe that mathematics is important for my career.
	X6. My parents like mathematics.

Analysis of The Data

In the research, analyses of assumptions were carried out before analyses of the data sets. Firstly, the students with missing answers to the items were removed from the data set. Following that, univariate (determined with z statistics) and multivariate (calculated with Mahalanobis coefficient) observations with extreme values were removed. In order to examine the normality assumption, coefficient of skewness and kurtosis were utilized. It was calculated that the answers to the items were between -1 and +1 in the given values. In addition, histogram graphics were examined and no extreme deviation was detected in the values. In order to check if there is a multicollinearity, correlation coefficients of the answers to the items were calculated and found to be between 0,074 and 0,709. Box's M test was carried out and calculated to check the homogeneity of variances ($p>0,05$). The aforementioned assumptions were analyzed in the groups of 60, 100, 250, 500, 1000 sample sizes that were pulled from the whole data set. In all data sets, it was found that there is no significant differences between items correlation coefficients ($p>0,05$).

In the analysis of the data, Model 1 which was reached via calculations with different sample sizes using exploratory factor analysis was calculated first. Then the factors with the least factor loads were determined and replaced to reach to Model 2, which the calculations were based upon. Estimations were carried out via LISREL packaged program.



Model 1: The model is specified in the first stage of DFA. Model 1 is specified in terms of AFA results and survey assumption.

Model 2: Factors with the least factor loads (X3-X6) were determined and replaced to reach to Model 2.

FINDINGS

1. What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 1?

Analyses for Model 1 which were based on maximum likelihood, unweighted least squares and generalized least squares estimation methods were carried out with sample sizes of 60, 100, 250, 500, and 1000. The fit indexes based on those estimations and their results are shown in Table 2.

When Table 2 is examined, it can be seen that X^2/df fit index tend to increase with sample size for all three estimation methods. RMSEA fit index tend to decrease as the sample size increases to 250 and to increase with a sample size of 500. CFI, which is a goodness of fit index, tend to increase as the sample size increases to 250 and it shows a stable value (similar in 500 and 1000 sample size) with increasing sample size. NFI and GFI show a similar pattern as CFI.

It is observed that overall model-data fit is achieved when the sample size is at least 250 in ML estimation method, 60 in ULS estimation method, and 250 in GLS estimation method.

Table 2. Estimation Results for the Model 1

Estimation methods	Sample Size	X ² /df	RMSEA	CFI	NFI	GFI
ML	60	2,78	0,174	0,88	0,83	0,89
	100	2,87	0,138	0,93	0,91	0,93
	250	3,27	0,095	0,97	0,96	0,97
	500	6,76	0,107	0,97	0,96	0,97
	1000	12,20	0,106	0,97	0,96	0,97
ULS	60	1,34	0,088	0,97	0,92	0,96
	100	1,46	0,058	0,99	0,96	0,99
	250	1,49	0,045	0,99	0,98	0,99
	500	3,01	0,063	0,99	0,98	0,99
	1000	7,06	0,075	0,98	0,98	0,98
GLS	60	1,53	0,111	0,87	0,77	0,81
	100	1,72	0,073	0,94	0,87	0,88
	250	1,60	0,049	0,97	0,94	0,95
	500	3,09	0,065	0,96	0,94	0,95
	1000	6,16	0,072	0,95	0,94	0,95

2. What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 2?

Analyses for Model 2 were carried out with the same three parameter estimation methods under five sample sizes. Results are shown in Table 3.

Table 3. Estimation Results for the Model 2

Estimation methods	Sample size	X ² /df	RMSEA	CFI	NFI	GFI	Statistically insignificant article count
ML	60	Model does not converge					
	100	Model does not converge					
	250	269,29	0,345	0,51	0,51	0,74	
	500	564,80	0,362	0,52	0,52	0,73	
	1000	959,99	0,373	0,54	0,54	0,76	
ULS	60	14,13	0,114	0,97	0,94	0,95	
	100	63,93	0,266	0,77	0,75	0,97	2
	250	183,71	0,297	0,72	0,71	0,88	
	500	435,92	0,327	0,70	0,70	0,87	
	1000	316,27	0,278	0,77	0,76	0,89	
GLS	60	15,39	0,125	0,85	0,77	0,66	
	100	164,78	0,441	0,00	-0,81	0,27	3
	250	301,43	0,445	0,00	-0,49	0,40	
	500	784,54	0,384	0,00	-0,82	0,33	
	1000	565,00	0,374	0,00	-0,42	0,41	

When Table 3 is examined, it can be seen that model did not converge with 60 and 100 sample sizes in ML method. In ULS method, with 100 sample size (2 items) and in GLS method, with 100 sample size (3 items), some items did not produce statistically significant results.

When the information on Table 3 is analyzed, it can be seen that generalized least squares and maximum likelihood method are affected the most from Model 2, and unweighted least squares method was the least affected among the three. At the same time, it is observed that ULS method overall manages to achieve data model fit in 60 sample size (relative to bigger fit indexes).

According to the information on Table 3, fit indexes of X^2/df showed increase with the sample size and a model data fit was not achieved in any sample size. RMSEA fit index similarly increased with the sample size in ML overall. In ULS and GLS estimation methods,, RMSEA showed a decline in 250 and 500 sample sizes.

In ML parameter estimation method, CFI and NFI fit indexes produced similar results in every sample size, whereas GFI fit index showed a relatively bigger estimation among others. In ULS method, apart from 60 sample size, a similar change is observed. In GLS parameter estimation method, GFI fit index produced higher estimations relative to other fit indexes.

3. How does fit indexes differ within Model 1 and Model 2, different sample sizes and different parameter estimation methods?

Fit indexes that reduced in value after the estimations of Model 1 and Model 2 and differences between them are analyzed. The differences between the reduced fit indexes are shown in Figure 1.

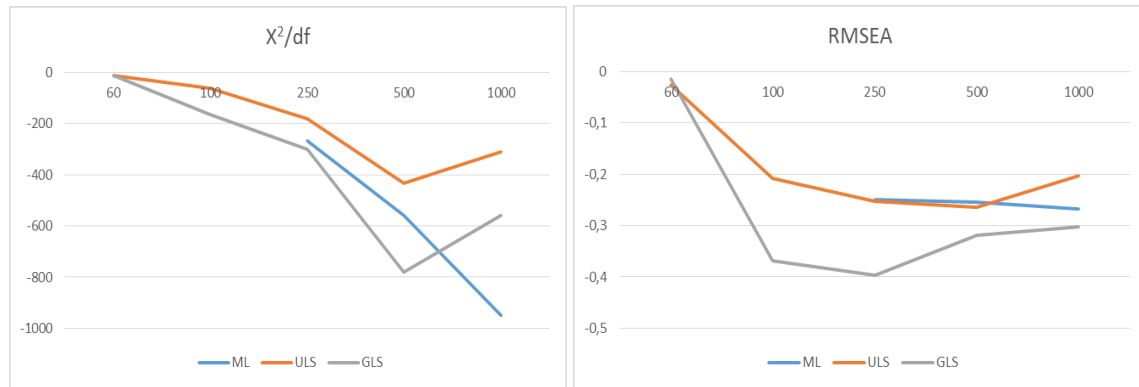


Figure 1. The Difference Among the Decremental Fit Indexes

As seen in Figure 1, X^2/df fit index showed the most increase in GLS parameter estimation method, and the least in ULS parameter estimation method. The biggest difference between the Model 1 and Model 2 model was estimated with ML parameter estimation method in 1000 sample size. RMSEA fit index is observed to not change in ULS and GLS in 60 sample size. ULS parameter estimation method is found to be the least affected from the Model 2 setting of the model.

As seen in Figure 2, the least affected parameter estimation method from the Model 2 setting of the model is ULS; whereas the most affected is GLS. In 60 sample size, the incremental fit indexes of CFI, NFI and GFI show little or no sign of being affected by Model 1 and Model 2 setting of the model. GFI fit index is found to be the least affected by the setting of the model among other incremental fit indexes.

The fit indexes showing increase with the Model 1 and Model 2 model are shown in Figure 2 below.

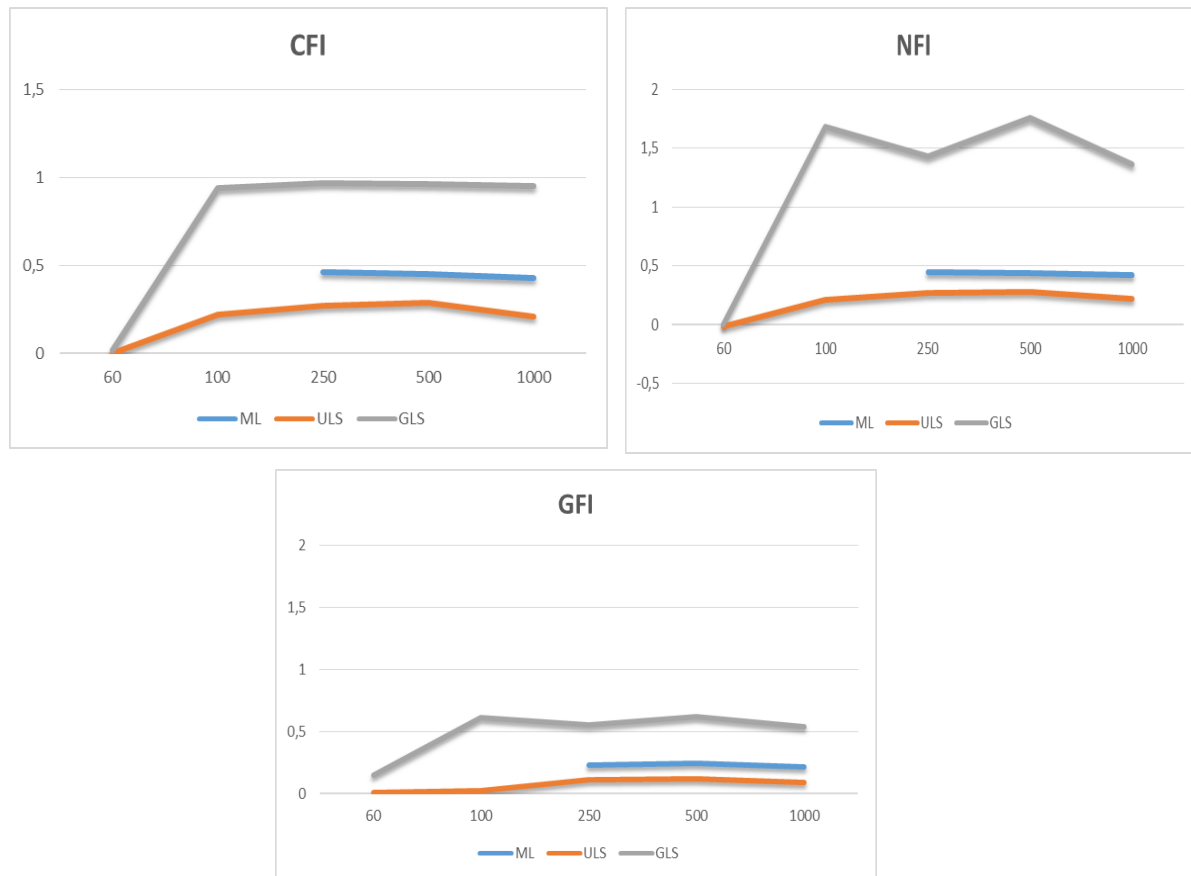


Figure 2. The Difference Among the Incremental Fit Indexes

RESULTS and DISCUSSION

In this study, the aim was to study the effect of different parameter estimation methods and sample sizes on the fit indexes for two models in confirmatory factor analysis. Towards this, sample data sets of 60, 100, 250, 500 and 1000 were pulled from answers to the “thoughts on the importance of mathematics” section in PISA 2012 data. Six items in the mathematics section of the research were scored in a four-point likert scale . Data sets were firstly analyzed in exploratory factor analysis and it was found that scale factors were two dimensionally gathered in friends (%37,193) and family (%32,740). Cronbach’s alpha reliability coefficients were estimated to be 0,824 and 0,705 in dimension 1 and dimension 2 of the scale, respectively.

Firstly, Model 1 was formed with six items and two dimensions, and ML, ULS and GLS parameter estimation methods were carried out with five different sample sizes. Results of the estimations showed that X^2/df fit index shows increase with the sample size. RMSEA fit index, on the other hand, shows increase with sample size up to a point, then decreases with some parameter estimation methods. Those which were referred to as decremental fit indexes showed a tendency to rise with the sample size, meaning that they may be misleading with bigger sample sizes. Research shows that X^2/df fit index can be statistically underpowered in small sample sizes (Kenny & McCoach, 2003). On the other hand, Kline (2011) suggests X^2 fit index shows increase along with the sample size. Sayın & Gelbal (2016) also states in their study that fit index shows increase with increasing sample size, that model data fit cannot be achieved in especially greater sample sizes. On the other hand X^2 fit index may be quite affected from the Model 2 model. Kenny, Kaniskan & McCoach (2014) states that RMSEA, which is the most used to evaluate model data fits, fit index is not fit to be used in small df models. They state that in especially small sample sizes, model parameters are estimated lower than they should be with RMSEA.

The fit indexes that showed an increase, CFI, GFI and NFI fit indexes produced and increase until 250 sample size, and remained stable after that point, overall. Accordingly, it can be said that 250 sample size is enough to measure the model data fit. Boomsma (1982) states that at least a sample size of 200 is needed for structural equation modeling. Molwus, Erdogan, & Ogunlana (2013) suggests, when the effect of sample size on fit indexes are explored, small sample sizes produce low fit indexes, and as the sample sizes get bigger, fit indexes also do. The also suggest that is vital to consider this aspect when studying with small and big sample sizes at the same time.

When parameter estimation methods are observed, ML estimation method achieved model data fit at 250 sample size, ULS at 60 sample size, and GLS at 250 sample size. Fan and others (1999), state that ML and GLS should not produce different statistics from each other in theory, but there can be variations in practice against prediction. Sayın & Gelbal (2016) states that ML estimation method produces higher model fit index estimations than GLS. In this case if the prerequisite assumptions are met, It is suggested that ML is used.

In the research, in order to determine the effect of modeling on fit indexes, the items with the least factor loads were replaced and a Model 2 model was created. Same analysis procedures were carried out for Model 2. In the Model 2, one of the decremental fit indexes, X^2/df fit index showed an increase with the sample size and it showed no model data fit in any sample size. RMSEA fit index showed an overall increase with ML estimation method, along with the sample size. In ULS and GLS estimation methods, at 250 and 500 sample sizes, RMSEA showed a decline. In light of those informations, it can be said that X^2/df and RMSEA decremental fit indexes can be utilized if the model is set as Model 2 but they should not be evaluated for their results in different sample sizes. Fan and others (1999) states in their studies that RMSEA is more sensitive than GFI to model misspecification.

In ML parameter estimation method, CFI and NFI fit indexes produced similar results in all sample sizes, whereas GFI fit index was estimated to be higher than other incremental fit indexes. In ULS method, a similar pattern was observed over 60 sample size. In this case, it can be said that GFI fit index is the least affected from the Model 2 setting of the model, and whenever the model is inconsistent, it should not be used. Fan (1996) states in his study that a completely wrong model setting affects GFI fit index more than CFI and NFI. Schermelleh-Engel, Moosbrugger, & Müller (2003) also suggest in their study that GFI fit index estimates are higher than other incremental fit indexes, and it is not affected by a Model 1 or Model 2 setting of the model. Sharma, Mukherjee, Kumar & Dillon (2005) states in their studies that despite the percent of times misspecified model is accepted is quite low for higher degrees of misspecification, RMSEA index performs better than GFI. Also they noticed that the use of GFI should be discouraged.

GLS and ML methods are found to be the most affected from the Model 2 setting of the model, whereas ULS method was affected the least. At the same time ULS method achieves model data fit at 60 sample size (relative to the incremental fit indexes). In light of this information, it is suggested that when the model data fit is achieved preliminarily, ULS should not be used. Fan and others (1999) state that parameter estimation methods affect all fit indexes strongly. And they have found similar results to these research results that fit index values based on GLS invariably appeared to indicate a better fit than those based on ML. And also the research can be repeated with the simulation study.

REFERENCES

- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. (Eds. K. G. Jöreskog & H. Wold). Systems under indirect observations: Causality, structure, prediction, Amsterdam: North-Holland.
- Byrne, B. M. (2010). Structural equation modeling with AMOS basic concepts, applications, and programming. New York: Taylor & Francis Group.
- Crocker, L. M. & Algina, L. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

- DeVellis, R. F. (2003). Scale development theory and applications, applied social research methods series. California: SAGE Publications.
- Fan, X. (1996). Structural equation modeling and canonical correlation analysis: What do they have in common?. *Structural Equation Modeling: A Multidisciplinary Journal*, 4, 64-78.
- Fan, X., Thompson B. & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6(1), 56-83.
- Fan, X. and Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509-529.
- Kaplan (2001), Structural equation modeling. *International Encyclopedia of the Social & Behavioral Sciences*, 15215-15222.
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333-351.
- Kenny, D. A., Kaniskan, B. & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, 1-22.
- Kline, R. B. (2011). *Principals and practice of structural equation modeling*. New York: The Guilford Press.
- Molwus, J. J., Erdogan, B. & Ogunlana, S. O. (2013). Sample Size and Model Fit Indices for Structural Equation Modelling (SEM): The Case of Construction Management Research. *ICCREM*, 338-347.
- Savalié, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, 72(6), 910-932.
- Sayın, A. & Gelbal, S. (2016). Yapısal eşitlik modellemesinde parametrelerin klasik test kuramı ve madde tepki kuramına göre sınırlandırılmasının uyum indekslerine etkisi. *International Journal of Education Science and Technology*, 2(2), 57-71.
- Schermelleh-Engel, K., Moosbrugger, H. & 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.
- Schumacker, R.E. & Lomax, R.G. (2004). *A beginner's guide to structural equation modeling*. Lawrence Erlbaum Associates, New Jersey.
- Sharma, S., Mukherjee, S., Kumar, A. & Dillon, W.R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, 58, 935-943.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.

GENİŞ ÖZET

Giriş

Ölçmenin temel konularından birini oluşturan insanlar gözlemlenerek; nesnelere, olaylar ve süreçler hakkında bilgi edinilebilir (DeVellis, 2003). Ölçülmesi gereken psikolojik yapılar gözlenen değişkenler aracılığı ile belirlenebilir. Ancak daha çok sosyal bilimlerde doğrudan gözlemler yoluyla gerekli bilgiler sağlanamayabilir ve bir hata ortaya çıkabilir. (Crocker, & Algina, 2006).

Gözlenen cevaplar doğrultusunda gizil yapıların belirlendiği ve daha çok teorik bir yapının test edildiği Doğrulamalı Faktör Analizi (DFA); ölçek uyarlama, geliştirme, kültürel karşılaştırma, grupları karşılaştırma gibi alanlarda sıklıkla kullanılmaktadır. Eğitimde kullanılan ölçeklerin geliştirme ve uyarlama sürecinde DFA'nın doğru bir şekilde kullanılması gerekmektedir. Alan yazında gerçekleştirilen araştırmalar incelendiğinde, yapısal eşitlik modellemesi ile doğrulamalı faktör analizi çalışmalarının sıklıkla kullanıldığı ancak bu çalışmalarda örneklem büyüklüğü varsayımının incelenmediği, parametre kestirim yöntemlerinin varsayımları doğrultusunda belirlenmediği görülmektedir. Bu çalışma ile örneklem büyüklüğü ile parametre kestirim yöntemlerinin uyum iyiliği indeksleri üzerindeki ne düzeyde bir etkisi olduğu belirlenmesi amaçlanmaktadır. Bunun yanı sıra DFA'nın kullanıldığı çalışmalarda, özellikle geliştirme çalışmalarında, modelden tam olarak emin olunamadığı durumlarla karşılaşmaktadır. Bu doğrultuda araştırmada aynı zamanda oluşturulan iki farklı modelin hesaplama sonuçları karşılaştırılmıştır. Sosyal bilimlerde ve davranış bilimlerinde doğrulamalı faktör analizi sıklıkla kullanılmasına rağmen böyle bir çalışmanın yaygınlaşmaması araştırmayı ayrıca önemli kılmaktadır.

Yöntem

Bu araştırmada, farklı parametre kestirim yöntemleri ve farklı örneklem büyüklüklerinin doğrulayıcı faktör analizindeki uyum indeksleri üzerindeki etkisinin araştırılması amaçlanmaktadır. Aynı zamanda model-veri uyum indekslerinin oluşturulan farklı modellerde nasıl hesaplandığının belirlenmesi amaçlanmaktadır. Bu doğrultuda araştırmamızın “temel araştırma” niteliğinde olduğu söylenebilir.

Araştırmanın çalışma grubunu Türkiye’de PISA 2012 çalışmasına katılan öğrenciler oluşturmaktadır. PISA çalışmasında “matematiğin önemine yönelik görüşler” maddelerine 4848 öğrenci cevap vermiştir. Eksik ve geçersiz veriler, veri setinden çıkarılmış ve varsayımların incelemesi gerçekleştirilmiştir. Kalan 3155 öğrencinin cevapları içerisinde rastgele seçilen 60, 100, 250, 500 ve 1000 öğrencinin görüşleri araştırma kapsamında incelenmiştir.

PISA 2012 çalışmasında öğrencilerin cevaplandığı matematik anketinde kişilerin matematiğin önemine yönelik 4’lü likert tipinde derecelendirilmiş altı madde yer almaktadır. Bu maddelerden ilk üçü arkadaşlarının, diğer üç madde de ailesinin matematiğe yönelik algılarını ifade etmektedir. Türkiye’de çalışmaya katılan öğrencilerin cevapları (n=3155) doğrultusunda öncelikle açılımlı faktör analizi hesaplanmıştır. Hesaplama sonucunda arkadaş boyutundaki maddelerin toplam varyansın %37,193’üne; aile boyutundaki maddelerin de varyansın %32,740’ına açıklık getirdiği tespit edilmiştir. Toplam altı maddenin varyansın %69,933’ünü açıkladığı belirlenmiştir. Birinci boyuttaki maddelerin Cronbach alfa katsayısı 0,824; ikinci boyuttaki maddelerin güvenilirlik katsayısı da 0,705 olarak hesaplanmıştır.

Bu doğrultuda PISA 2012 çalışmasına katılan ve iki boyuttan oluşan “matematiğin önemine yönelik görüşler” maddelerine cevap veren öğrenciler içerisinde 60, 100, 250, 500 ve 1000 kişilik veri setleri belirlenmiştir. Belirtilen şekilde iki boyutlu oluşturulan Model 1 ile en çok olabilirlik (EÇO), ağırlıklandırılmamış en küçük kareler (AEKK), genelleştirilmiş en küçük kareler (GEKK) hesaplama yöntemleri kullanılarak hesaplamalar gerçekleştirilmiştir. Ardından boyutlar içerisinde en düşük faktör yük değerine sahip maddelerin yerleri değiştirilerek Model 2 oluşturulmuştur. Model 2 ile de hesaplamalar gerçekleştirilerek sonuçlar karşılaştırılmıştır.

Sonuç ve Tartışma

Araştırma kapsamında öncelikle altı madde ve iki boyuttan oluşan Model 1 kurulmuş ve EÇO, AEKK ve GEKK parametre kestirim yöntemleri ile beş farklı örneklem büyüklüğünde hesaplamalar gerçekleştirilmiştir. Hesaplama sonucunda X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği belirlenmiştir. RMSEA uyum indeksinin ise parametre kestirim yöntemlerine göre belirli örneklem büyüklüğüne göre azalma, daha sonra ise artma eğiliminde olduğu tespit edilmiştir. Söz konusu uyum indekslerinin örneklem büyüklüğüne bağlı olarak artış göstermesi, büyük örneklem büyüklüklerinde yanıtıcı olabileceklerini göstermektedir.

Artan uyum indeksleri olan CFI, GFI ve NFI uyum indekslerinin tüm parametre kestirim yöntemlerinde genel olarak 250 örneklem büyüklüğüne kadar artış gösterdiği, daha sonra ise genel olarak sabitlendiği belirlenmiştir. Bu doğrultuda 250 örneklem büyüklüğünün model-veri uyum tespitinde yeterli olduğu söylenebilir.

Parametre kestirim yöntemleri bazında incelemeler gerçekleştirildiğinde ise EÇO hesaplama yöntemi ile en az 250 örneklem büyüklüğünde, AEKK ile 60 örneklem büyüklüğünde ve GEKK hesaplama yöntemi ile de 250 örneklem büyüklüğünde model-veri uyumunun genel olarak sağlandığı tespit edilmiştir.

Model içerisinde maddelerin boyutlarındaki yerleri değiştirilerek Model 2 oluşturulmuştur. Model 2 ile gerçekleştirilen hesaplamalar sonucunda X^2/sd ve RMSEA azalan uyum indekslerinin değişim gösterdiği belirlenmiştir. Model 2 hesaplanmasında en çok GEKK yöntemi ile EÇO yönteminin, en az ise AEKK’nin Model 1’den farklı sonuçlar verdiği belirlenmiştir.

Araştırma kapsamında gerçekleştirilen hesaplamalar sonucunda X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği belirlenmiştir. RMSEA uyum indeksinin ise parametre kestirim yöntemlerine göre belirli örneklem büyüklüğüne göre azalma, daha sonra ise artma eğiliminde olduğu saptanmıştır. Araştırmalar X^2/sd uyum indeksinin küçük örneklem büyüklüklerinde istatistiksel gücünün düşük olacağını ifade etmektedir (Kenny & McCoach, 2003). Buna karşın Kline (2011) X^2 uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiğini ifade etmektedir. Sayın (2014) da çalışmasında X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği, özellikle de geniş örneklem büyüklüklerinde iyi uyum gösteren modelde model-veri uyumunun sağlanamadığına yönelik bilgiler verdiğini belirlemiştir. Kenny, Kaniskan & McCoach (2014) model-veri uyum değerlendirmesinde en çok kullanılan RMSEA uyum indeksinin serbestlik derecesinin küçük olduğu modellerinde kullanılmasının uygun olmadığını ifade etmektedir.

Artan uyum indeksleri olan CFI, GFI ve NFI uyum indekslerinin tüm parametre kestirim yöntemlerinde genel olarak 250 örneklem büyüklüğüne kadar artış gösterdiği, daha sonra ise genel olarak sabitlendiği belirlenmiştir. Bu doğrultuda 250 örneklem büyüklüğünün model-veri uyum tespitinde yeterli olduğu söylenebilir. Bu çalışma kapsamında 250 örneklem büyüklüğü aynı zamanda serbestlik derecesinin katımı da ifade etmektedir. Bu doğrultuda hesaplamalarda serbestlik derecesine bağlı bir örneklem büyüklüğünün de belirlenebileceği söylenebilir. Boomsma (1982) yapısal eşitlik modellemesinde en az 200 örneklem büyüklüğüne ihtiyaç duyulduğunu belirtmektedir. Molwus, Erdogan, & Ogunlana (2013) örneklem büyüklüğünün uyum indekslerine etkisini incelediği çalışmasında küçük örneklem büyüklüklerinde uyum indekslerinin düşük, geniş örneklem büyüklüklerinde ise yüksek çıktığını belirlemiştir.

EÇO parametre kestirim yönteminde CFI ve NFI uyum indeksinin tüm örneklem büyüklüklerinde benzer sonuçlar ürettiği, GFI uyum indeksinin ise diğer artan uyum indekslerinden daha yüksek kestirildiği belirlenmiştir. Bu durumda modelin farklı tanımlanmasından GFI uyum indeksinin diğer indekslere göre daha az etkilendiği; modelden emin olunmadığı durumlarda söz konusu uyum indeksinin raporlanmaması gerektiği önerilmektedir. Fan, Thompson, & Wang (1996) ise yaptığı çalışmada ise modelin tam olarak hatalı kurulmasında GFI uyum indeksinin CFI ve NFI uyum indeksinden daha duyarlı olduğunu belirtmiştir. Schermelleh-Engel, Moosbrugger, & Müller (2003) çalışmalarında bu çalışma sonuçlarına benzer şekilde GFI uyum indeksinin diğer artan uyum indekslerine göre daha yüksek kestirimler gerçekleştirdiği; farklı tanımlanan modellerdeki hesaplama sonucunun çok fazla değişiklik göstermediği belirlenmiştir. Benzer şekilde Sharma, & diğerleri (2005) de GFI uyum indeksinin kullanılmasının uygun olmadığını belirtmektedir.

A Scaling of Preservice Teachers' Motivational Factors in Choosing the Teaching Profession*

Öğretmen Adaylarının Öğretmenlik Mesleğini Seçme Nedenlerinin Ölçeklenmesi

Sakine GÖÇER ŞAHİN**

Selahattin GELBAL***

Abstract

In this research, the reasons of preservice teachers' choice of the teaching profession has been studied using the method of pair-wise comparison. Data was acquired by 1137 preservice teachers studying in various departments of seven different universities in Turkey. Overall, it was found that preservice teachers have chosen the teaching profession due to easy work conditions. It was found that female preservice teachers chose the teaching profession because of their sympathy towards it, whereas male preservice teachers chose it since their university admission exam scores were just enough for teaching programs. According to the departments of faculties of education, it was determined that psychological counseling and guidance department preservice teachers chose the teaching profession since it ensured employment, Turkish and English teaching department preservice teachers chose it since they liked it, primary school teachers and primary Math preservice teachers chose it because the work conditions were convenient. According to the places they lived in and the departments, there was no hierarchy regarding reasons why they choose the teaching profession. When the reasons of choosing the teaching profession according to level of income were examined the results that the first reason was the same in all three departments. While this was the seventh reason for the preservice teachers who had teachers in their families, it was the eighth reason for the preservice teachers who didn't have any relatives who are teachers.

Keywords: scaling, pair-wise method, teaching profession, preservice teachers, reasons.

Öz

Bu çalışmanın amacı, öğretmen adaylarının öğretmenlik mesleğini seçme nedenlerini ikili karşılaştırmalar yöntemiyle ölçeklemektir. Araştırmanın verileri, Türkiye'nin 7 farklı üniversitesinde eğitim gören 1137 öğretmen adayından toplanmıştır. Herhangi bir değişken olmaksızın genel olarak bakıldığında öğretmen adaylarının öğretmenlik mesleğini öncelikli olarak çalışma koşulları rahat olduğu için seçtiği bulgusuna ulaşılmıştır. Kadın öğrencilerin, öncelikli olarak mesleği sevdiği için, erkek öğrencilerin ise puanı yettiği için öğretmenliği seçtiği görülmektedir. Bölümlere göre bakıldığında, RPD öğrencilerinin öncelikli olarak iş bulma olanağı yüksek olduğu için, Türkçe ve İngilizce Öğretmenliği öğrencilerinin öğretmenliği sevdiği için, Sınıf ve İlköğretim Matematik Öğretmenliği öğrencilerinin ise çalışma koşulları rahat olduğu için öğretmenliği seçtikleri görülmektedir. Öğrencilerin yaşamlarının en çok geçtiği yerleşim birimi ile öğretim grubuna göre öğretmenlik mesleğini seçme nedenlerinin başında yine çalışma koşullarının rahatlığı gelmekte olup; sıralamaları arasında bir fark görülmemiştir. Gelir düzeylerine göre öğretmen adaylarının öğretmenlik mesleğini seçme nedenleri incelendiğinde her üç grup için ilk nedenin aynı olduğu görülmektedir. Ailelerinde geçmişte veya hala öğretmen olan ve olmayan öğretmen adaylarının öğretmenlik mesleğini seçme nedenleri neredeyse aynı olup bu nedenlerin başında “çalışma koşulları rahat olduğu için” nedeni gelmektedir. Öğretmen adaylarından Anadolu Öğretmen Lisesi mezunu olan ve olmayan öğrencilerin her ikisinde de öğretmenlik mesleğini seçme nedenlerinin başında yine çalışma koşullarının rahat olması gelmektedir.

Anahtar Kelimeler: ölçekleme, ikili karşılaştırmalar, öğretmenlik mesleği, öğretmen adayları, nedenler.

* This study was presented at the 21st National Educational Sciences Conference organized by Marmara University in Istanbul, Turkey, on September 12–14, 2012.

** Corresponding Author: Research Assistant, Dr., Hacettepe University, Educational Science Department, Ankara, Turkey, sgocersahin@gmail.com

*** Prof. Dr., Hacettepe University, Educational Science Department, Ankara, Turkey, gelbal@hacettepe.edu.tr

INTRODUCTION

The drawbacks associated with Turkish education system are among the most debated issues. There is an abundance of studies carried out at all stages of this process in order to highlight those drawbacks. Of all educational questions, teacher preparation has been heavily discussed recently. Though various institutions were appealed for teacher preparation since the beginning, the Ministry of National Education (MONE) introduced a selection criterion with the Law 1739. But this criterion was not put into practice during selecting prospective preservice teachers prior to 1985. The selection of teachers by means of an examination was started in 1985, and then graduates of both educational institutions and other institutions could apply. The responsibility of bringing up teachers was transferred from the MONE to universities as of 1982.

Preservice teachers were admitted to such institutions against differing criteria throughout the years (Celikten, Sanal, & Yeni, 2005). Besides how those institutions select preservice teachers, the reason why preservice teachers choose the teaching profession can be addressed as another aspect of the teacher preparation issue. For example, some preservice teachers opt for Anatolian Teacher High Schools, as a teacher preparation institution in Turkey, due to the high quality of education in those schools rather than their positive attitude towards teaching (Ozoglu, 2010). This fact alone can be considered as a real matter in teacher preparation. Ozguven (1999) lists influential factors on the choice of profession as individual's physical characteristics including height, weight, gender, age and level of education, both mental and physical health, talents and interests, future plans, communication skills, emotional and social maturity, academic achievement, study habits, work experience, family, community's expectations and needs, the profession's employment potential and future, prestige, and contemporary and popular status. However, the reason of preservice teachers' choice of the profession turns into a debated issue taking into consideration the economic conditions, living standards and frequently changing education system in Turkey.

The reason of the preservice teachers' choice of the teaching profession has been subject to research in many countries around the world. Richardson, and Watt (2005) examined why teacher candidates chose this profession using open-ended questions to 74 preservice teachers who studied in the first grades of universities in Australia. In this study, factors related to social status, career fit, prior considerations, financial reward, and time for family were identified based on factor analysis. Again in research carried out by Klassen, Al-Dhafri, Hannok, and Betts (2011), the reasons why the teacher candidates choose the profession of teaching were determined with the data collected from 200 teacher candidates in Canada and Oman and these reasons were compared to one another. The results of this study show that Canadian participants made more self-references, and expressed higher levels of individual-focused motivation and social utility value as career motivators than Omani participants did. Participants from Oman expressed greater endorsement of teaching as a fall back career and higher levels of socio-cultural influences than Canadian participants. The motivation to choose the teaching profession in the USA was also studied in a research carried out by Thomson, Turner, and Nietfield (2012). Data was collected using surveys and semi-structured interviews. The findings of this study indicate that all participants expressed altruistic, intrinsic, and extrinsic reasons for teaching at any level.

Besides these studies, the reasons of teacher candidates' choice of this profession were investigated by various researches in many countries including Singapore, Germany, Norway (Krecic, & Grmek 2005; Kyriacou, Hultgen & Stephens, 1999; Johnston, McKeown & McEwen, 1999; Richardson, & Watt, 2010; Sinclair, 2008; Stiegelbauer, 1992). The studies concerning similar questions were also conducted in Turkey. The reasons why Chemistry Teacher candidates chose the profession of teaching were studied in a research carried out by Boz and Boz (2008). In this study it was found that both intrinsic and extrinsic reasons were the main factors affecting prospective chemistry and mathematics teachers' preferences to become teachers. Inspiration from primary and secondary chemistry and mathematics teachers, love for teaching, and chemistry or mathematics content knowledge were among the prior preferences.

The reasons why English Teacher candidates chose the profession of teaching were studied using correlation, regression and structural modeling in another research carried out by Eren, and Tezel

(2010). Results show that the prospective teachers' futures played a significant mediating role in relationships regarding planned effort, planned persistence, ability, intrinsic career value, social contributions, and career choice satisfaction. In another research by Yuce, Sahin, Kocer, and Kana (2013), the reasons why lead teacher candidates to choose the teaching profession were asked; the responses were analyzed using qualitative data analysis approach and descriptive statistics. It was found that extrinsic, altruistic, and intrinsic motivations play a role when individuals choose teaching as a career. In addition, although altruistic motives were very dominant for females, mercenary-based extrinsic motives were dominant for males. In another research carried out by Ozsoy, Ozsoy, Ozkara, and Memis (2010), the factors that have influence in teacher candidates' choosing of this profession were attempted to be determined with the help of a Likert-type scale, whether the findings changed according to various variables was studied. According to this study's findings, the reasons for choosing teaching were ranked in this order: 1) teaching is sacred; 2) possess required qualifications; 3) best fit profession; 4) enjoy children; and 5) teaching is an ideal profession. The roles of the family and the gender in teacher candidates' choosing of this profession were also studied in a research carried out by Akbayir (2003). It was found that the parental education of preservice teachers who selected the teaching profession was low and their income levels were moderate. Especially their mothers claimed that the teaching profession was a suitable occupation for their children. Another important result of the study is that mothers' education level increased the probability of selecting the teaching profession was increased. It was also found that while income of the preservice teachers increased, the choice of teaching decreased. Another important result was related to field selection. While male students tended to choose subjects such as mathematics, physics, classroom, chemistry and biology education, female students usually tended to choose kindergarten education. The effectuality of personal, social and economic characteristics was studied in another research carried out by Ozbek (2007). Apart from these, it is seen that similar studies with various variables were conducted on different paradigms (Cermik, Dogan, & Sahin, 2010; Eskicumali, 2002). In the study by Cermik, Dogan, & Sahin, it was found that mercenary, extrinsic, intrinsic and altruistic factors underlie prospective classroom teachers' career choices; however, the rate of intrinsic and altruistic motives increases at graduation. Gender also played an important role in career choice.

There are many studies investigating the reasons of preservice teachers' choice of the teaching profession. Forty-four articles published to determine the introduction features of teacher candidates were studied in research carried out by Brookhart, and Freeman (1992), and characteristic features of teacher candidates were attempted to be put forth with this literature review. According to this study four major categories of variables have been studied in these articles: (a) demographics and high-school background; (b) motivation to teach and career expectations; (c) confidence and optimism or anxiety and concerns about teaching; and (d) perceptions of the roles and responsibilities of teachers. Most of the mentioned research employed a survey methodology.

In review of some of the previous studies, it was found out that they did not go beyond calculating frequency; and other quantitative studies did not use scaling by means of the pair-wise comparison method (Cermik, Dogan, & Sahin, 2010; Eskicumali, 2002; Johnston, McKeown, & McEwen, 1999; Stiegelbauer, 1992; Watt, & Richardson, 2007).

Purpose of the Study

Scaling is an academic subject with the aim of putting basic principles and methods of "increasing from observations to measurements". In this study, it was aimed to investigate the reasons for which preservice teachers choose the teaching profession by pair-wise comparison method. In other words, the aim is to convert qualitative data to quantitative by identifying a zero point and unit. Overall, the purpose was to scale the qualitatively reported reasons.

METHOD

This study employed a pair-wise comparison method to investigate the reasons why preservice teachers in the faculties of education choose the teaching profession. This scaling study is carried out with variables such as gender, graduated high school, whether having a teacher relative in the family, income, education type (day or evening classes), settlement unit independent variables and aims at exposing an existing situation. Therefore, it can be categorized as a descriptive study.

Participants

Data was collected from 1137 preservice teachers in various departments from seven public universities in Turkey. Statistics regarding the study sample are presented in Table 1.

Table 1. Frequency and Percentages of Participants by University and Gender

		Gender			
		Female	Male	Total	
Universities	Erciyes	N	238	104	342
	University	%	21.6	9.4	31
	Hacettepe	N	114	24	138
	University	%	10.3	2.2	12.5
	Kastamonu	N	75	19	94
	University	%	6.8	1.7	8.5
	Mersin	N	206	167	373
	University	%	18.7	15.1	33.8
	Adiyaman	N	44	35	79
	University	%	4.0	3.2	7.2
	Kirikkale	N	21	13	34
	University	%	1.9	1.2	3.1
	Aydin Adnan Menderes	N	34	9	43
	University	%	3.1	0.8	3.9
Total	N	732	371	1103	
	%	66.4	33.6	100	

As seen in Table 1, the total number of participants was 1103. Gender information for 34 participants wasn't available. Departments and respective frequency of the study participants are as follows: English Language Teaching (124), Turkish Language Teaching (312), Science Teaching (67), Elementary Math Teaching (211), Religious Culture Teaching (30), Social Sciences Teaching (57), Preschool Teaching (72), Classroom Teaching (138), Guidance and Psychological Counselling (110), and German Language Teaching (13). Of the 1137 participants, 641 were attending day classes, while 495 attended evening classes. Only one (1) preservice teacher did not specify her/his type of education. Six hundred and twenty-two (59%) preservice teachers stayed in cities for most of their lives, 301 (28.6%) of them stayed in a sub province, and 52 (4.9%) stayed in a village. of the participants, 7.5% did not provide information about this item.

Data Collection Instruments

Firstly, existing studies on the reasons for choosing the teaching profession in the literature have been examined (Akbayir, 2003; Cermik, Dogan, & Sahin, 2010; Eskicumalı, 2002; Johnston, McKeown, & McEwen, 1999; Krecic, & Grmek 2005; Kyriacou, Hultgen, & Stephens, 1999; Ozbek, 2007; Sinclair, 2008; Stiegelbauer, 1992; Watt, & Richardson, 2007).

Aside from those 24 preservice teachers in various departments of the faculties of education were asked to specify the reason why they chose the teaching profession by responding open-ended questions. Then, the reasons in the literature and the answers of the preservice teachers were evaluated together and 8 new reasons were written. These reasons are compared using the didactic comparison method. The preservice teachers chose the one most suitable for themselves. Finally, the instrument had two parts. Part one was comprised of demographic information about participants, and part two was comprised of 56 items resulting from pair-wise comparison.

Data Analysis

Data analysis was conducted by using State III and State V equations for the absolute data matrix of Thurstone's pair-wise comparison method. First, the frequency matrix (F), which expresses the preference of either pair over the other, was prepared. Then, matrix P was constructed from matrix F. The element p'_{jk} is observed proportion of times stimulus k was judged greater than stimulus j. From matrix P was constructed in turn matrix X, the basic transformation matrix. The element x'_{jk} is the unit normal deviate corresponding to the element p'_{jk} . The equation given in formula (1) was used for State V.

$$S_K = D_{(K-1)(K-2)} + \dots + D_{21} \quad (1)$$

where $D = \bar{d} = \bar{s}_j - \bar{s}_k$,

\bar{s}_j : Scale value of stimulus U_j

\bar{s}_k : Scale value of stimulus U_k

\bar{d} : Difference of scale values of U_j and U_k

Thurstone's pair-wise comparison law is based on certain assumptions. The assumptions are as follows; 1) In the end of the discrimination process, an observer pairs the stimulus with a point on the scaling dimension; 2) Distribution of such points on the scaling dimension is normal; 3) Mean of the distribution is U_j , and scale value and standard deviation of the corresponding stimulus is the standard deviation of the discrimination distribution (Turgut, & Baykul, 1992). Internal consistency of the scaling done with State V was checked in order to find out if the scaling built on the obtained data meets those assumptions and if the observers were careful enough in making judgments. The internal consistency of the data tested based on the formula 2.

$$\chi^2 = \frac{\sum_{j>k} |\theta''_{jk} - \theta'_{jk}|}{821/N} \quad (2)$$

$$df = \frac{K(K-1)}{2}$$

where

$$\theta''_{jk} = \sin^{-1} \sqrt{p''_{jk}} \quad (3)$$

$$\theta'_{jk} = \sin^{-1} \sqrt{p'_{jk}} \quad (4)$$

χ^2 : Chi-square value

df= Degree of freedom

p'_{jk} : The observed proportion of times stimulus k was judged greater than stimulus j.

p''_{jk} : The theoretical proportion of times stimulus k was judged greater than stimulus j.

K: Number of stimulus

N: Total number of judgements for each stimulus

The internal consistency for most of the variables was low. Low internal consistency means that the data do not hold assumptions of State V or assumptions of the pair-wise comparison method or both (Turgut, & Baykul, 1992). To this end, scaling by using State III was preferred for scaling the eight reasons since the internal consistency was found low. For State III, standard deviation of the values in the column of matrix X, was calculated by formula 4, 5, and 6.

$$C = \frac{2K}{\sum_{i=1}^K \frac{1}{V_i}} \quad (4)$$

$$\sigma_j = \frac{c}{V_j} - 1 \quad (5)$$

Where,

$$V_j = \frac{1}{K} \sqrt{K \cdot \sum_{k=1}^K x_{kj}^2 - (\sum_{k=1}^K x_{kj})^2} \quad (6)$$

V_j : Standard deviation of x values of j column in the matrix X.

V_i : Standard deviation of x values of i column in the matrix X.

σ_j : The estimated variance of stimulus U_j ,

K: Number of stimulus

After calculating standard deviation of stimulus' scale values a rational zero point must be chosen. It is chosen arbitrarily. The method itself cannot determine an absolute zero point, since it operates only on judgements of differences between stimuli (Torgerson, 1965). The scale values were thus determined in the following equation.

$$r_j = ax_i \sigma_i \quad (7)$$

where

r_j is the scale value of stimulus j referred to the rational zero point.

x_i is the unit normal deviate corresponding to p_i , the theoretical proportion of times j is judged greater than zero.

σ_i is discriminal dispersion of stimulus j.

a is a constant of proportionality to allow for the possibility that the dispersion is greater for this type of judgement than for the straight paired- comparison judgement.

The transaction was repeated for all of the independent variables studied. Internal consistency was found high for data obtained from classroom teaching preservice teachers and pre service teachers with an income level between 2000 and 2749 Turkish Lira (TL) only. For those data obtained the scaling assumptions with State V, scaling was not performed with State III.

RESULTS

In the first instance, the reasons why pre service teachers choose the teaching profession were scaled by processing the data collected from 1137 participants.

Table 2. Scale Values and Stimulus Order of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	Scale Value	Stimulus Order
The profession has social prestige	0.3488	5
My university entrance score was just enough for it	0.1450	3
It has easy work conditions	0.0000	1
My family/relatives/friend favored it	0.7787	8
I am interested in working with children/adolescents/adults	0.2441	4
It is an easy subject	0.7463	7
It offers job security	0.4380	6
I like teaching	0.0571	2

As provided in Table 2, preservice teachers chose the profession primarily because it has relatively "easy work conditions". It is followed by other reasons as "I like teaching"; "My university entrance score was just enough for it"; "I am interested in working with children/adolescents/adults"; "It has social prestige"; "It offers job security" and "My family/relatives/friend favored it".

Moreover, reasons of choosing the teaching profession were examined by gender of the preservice teachers.

Table 3. Scale Values and Stimulus Order by Gender of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	Female		Male	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.4027	5	0.3868	5
My university entrance score was just enough for it	0.2664	4	0.0000	1
It has easy work conditions	0.0232	2	0.2369	2
My family/relatives/friend favored it	0.8124	7	0.8997	8
I am interested in working with children/adolescents/adults	0.2425	3	0.3707	4
It is an easy subject	0.8535	8	0.7038	7
It offers job security	0.4561	6	0.5465	6
I like teaching	0.0000	1	0.2847	3

According to Table 3, female preservice teachers chose the teaching profession due to the following reasons in a decreasing order: "I like teaching", "It has easy work conditions", "I am interested in working with children/adolescents/adults", "My university entrance score was just enough for it", "It has social prestige", "It offers job security", "My family/relatives/friend favored it" and lastly "It is an easy subject". The primary reason for which male preservice teachers chose the teaching profession was found as "My university entrance score was just enough for it". It is followed by other reasons such as "I has easy work conditions", "I like teaching", "I am interested in working with children/adolescents/adults", "I has social prestige", "I has job security", and "My family/relatives/friend favored it".

Similarly, the reasons for which preservice teachers chose the teaching profession was investigated by considering if they were graduates of Anatolian Teacher High Schools.

Table 4. Scale Values and Stimulus Order by Graduation of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	Graduation from Anatolian Teacher High Schools		Other	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.3513	6	0.359	5
My university entrance score was just enough for it	0.1148	2	0.174	3
It has easy work conditions	0.0000	1	0.000	1
My family/relatives/friend favored it	0.8050	8	0.796	7
I am interested in working with children/adolescents/adults	0.2814	4	0.242	4
It is an easy subject	0.5744	7	0.817	8
It offers job security	0.3128	5	0.482	6
I like teaching	0.1647	3	0.056	2

For both groups of participants, Anatolian Teacher High Schools graduates and others, the main reason for choosing the teaching profession was found as easy work conditions. This reason is followed by “My university entrance score was just enough for it” for graduates of Anatolian Teacher High Schools, while followed by “I like teaching” for the others. In addition, the last reason for those coming from Anatolian Teacher High Schools on the scale was “My family/relatives/friend favored it”, whereas for the others, it was found as “It is an easy subject”.

Table 5. Scale Values and Stimulus Order by Department of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	PCG		Classroom Teaching		Turkish Language Teaching		English Language Teaching		Elementary Maths Teaching	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order	Scale Value	Stimulus Order	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	1.08	5	0.19	4	0.68	5	0.37	6	0.64	4
My university entrance score was just enough for it	0.72	3	0.06	2	0.61	4	0.35	4	0.30	2
It has easy work conditions	0.06	2	0.00	1	0.40	3	0.29	3	0.00	1
My family/relatives/friend favored it	1.57	8	0.29	6	1.37	8	1.00	8	0.88	8
I am interested in working with children/adolescents/adults	0.85	4	0.35	7	0.37	2	0.36	5	0.60	5
It is an easy subject	1.28	6	0.41	8	1.34	7	0.97	7	0.85	7
It offers job security	0.000	1	0.10	3	1.23	6	0.24	2	0.84	6
I like teaching	1.38	7	0.20	5	0.00	1	0.00	1	0.45	3

Though the data was collected from ten departments in education faculties, scaling could not be conducted for the departments which had less than 100 participants. The number of preservice teachers by department is provided under “the study participants”.

As seen in Table 5, the primary reason of choosing the profession was found as “It has job security” among preservice teachers in the Psychological Counseling and Guidance (PCG) Department; “It has easy work conditions” among those in classroom teaching and elementary math teaching departments; and “I like teaching” for preservice teachers from departments on teaching language

skills such as Turkish and English Language Teaching. Ordering of the reasons for preservice teachers' choice varied different from department to department.

Table 6. Scale Values and Stimulus Order by Existence of a Teacher in the Family of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	Have a teacher family member		Don't have a teacher family member	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.243	5	0.481	5
My university entrance score was just enough for it	0.135	3	0.158	3
It has easy work conditions	0.000	1	0.000	1
My family/relatives/friend favored it	0.660	7	0.930	8
I am interested in working with children/adolescents/adults	0.230	4	0.268	4
It is an easy subject	0.666	8	0.854	7
It offers job security	0.374	6	0.517	6
I like teaching	0.018	2	0.114	2

As seen in Table 6, the reasons for choosing the teaching profession hardly differ between those with and without a teacher family member. The primary reason was found as "It has easy work conditions". Different results were obtained only for reasons 7 and 8 between these two groups of participants. The reason "My family/relatives/friend favored it" was ranked 7th for those who has a teacher family member, while was ranked as 8th for the others.

Table 7. Scale Values and Stimulus Order by Income Level of Reasons for Preservice Teachers' Choosing the Teaching Profession

Reasons	500 – 1249 TL		1250 – 1999 TL		2000 – 2749 TL	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.349	5	0.461	6	0.312	5
My university entrance score was just enough for it	0.163	3	0.183	3	0.119	2
It has easy work conditions	0.000	1	0.000	1	0.000	1
My family/relatives/friend favored it	0.847	8	0.872	8	0.505	7
I am interested in working with children/adolescents/adults	0.267	4	0.253	4	0.287	3
It is an easy subject	0.862	7	0.777	7	0.544	8
It offers job security	0.503	6	0.443	5	0.306	4
I like teaching	0.005	2	0.135	2	0.369	6

The reasons of preservice teachers' choices of their profession based on the income level of their families are displayed in Table 7. Although monthly income level of families was divided into 5 subgroups in the instrument used for data collection, there were not enough under "2750 – 3499 TL" and "3500 TL and above"; therefore, the reasons were scaled by categories of income given in the table only. For all of the preservice teachers with a monthly income of "500 – 1249 TL" and "1250 – 1999 TL" and "2000 – 2749 TL", the primary reason for choosing the teaching profession was "It has easy work conditions". For those with a monthly income level of "500 – 1249 TL" and "1250 – 1999 TL" and with a "relatively lower income level" than others, the 2nd most common reason for choosing the profession was "I like teaching", while it was replaced by "My university entrance score was just enough for others with income level of "2000 – 2749 TL". Ordering of the reasons of choosing the teaching profession among those with a monthly income level of "500 – 1249 TL" and

“1250 – 1999 TL” was found similar, but it was different for pre service teachers with a monthly income of 2000 – 2749 TL.

Table 8. Scale Values and Stimulus Order by Education Type of Reasons for Preservice Teachers’ Choosing the Teaching Profession

Reasons	Day classes		Evening classes	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.349	5	0.341	5
My university entrance score was just enough for it	0.185	3	0.078	3
It has easy work conditions	0.000	1	0.000	1
My family/relatives/friend favored it	0.773	8	0.783	8
I am interested in working with children/adolescents/adults	0.219	4	0.268	4
It is an easy subject	0.703	7	0.796	7
It offers job security	0.431	6	0.442	6
I like teaching	0.059	2	0.051	2

According to Table 8, there were no differences between the ordering of reasons for which preservice teachers chose the teaching profession by education type. In this case, primary reason for choosing the teaching profession for both groups was found as “It has easy work conditions”, while the last reason was “My family/relatives/friend favored it”.

Table 9. Scale Values and Stimulus Order by Settlement Unit Reasons for Preservice Teachers’ Choosing the Teaching Profession

Reasons	Urban		County	
	Scale Value	Stimulus Order	Scale Value	Stimulus Order
The profession has social prestige	0.355	5	0.356	5
My university entrance score was just enough for it	0.188	3	0.162	3
It has easy work conditions	0.000	1	0.000	1
My family/relatives/friend favored it	0.820	8	0.735	8
I am interested in working with children/adolescents/adults	0.260	4	0.226	4
It is an easy subject	0.797	7	0.706	7
It offers job security	0.452	6	0.410	6
I like teaching	0.111	2	0.008	2

Since there were only a limited number of participants who marked “village” and “town” as their settlement unit type, only provincial and sub provincial centers were taken into consideration. As participants were compared by settlement unit, there was no difference between orderings of the reasons of teachers’ choices. For both groups, the primary reason was found as “It has easy work conditions”, and the last one was “My family/relatives/friend favored it”.

DISCUSSION and CONCLUSION

This study was carried out to investigate the reasons for which preservice teachers choose the teaching profession using pair-wise comparison method including multiple variables. From the “Teacher preparation issue in Turkey” point of view, the finding that preservice teachers chose this profession mostly because it has easy work conditions intensifies the issue. In a study carried out by Cermik, Dogan, and Sahin (2010), it was investigated why preservice classroom teachers choose the teaching profession during entrance to and departure from the university. The easy work conditions

offered by the profession were not found as the most important reason in this study. Still, it was opted as primary reason by approximately 4% of the respondent group. In another study by Johnson, McKeown, and McEwen (1999), the leading reason for both female and male trainee teachers choosing the teaching profession was found as interest in working with children. In general, it can be said that preservice teachers choose the teaching profession due to the extrinsic motivation factors such as job security and re-compensation in underdeveloped countries, while they are replaced by intrinsic motivation factors such as interest in the profession and serving the largest community (Bastick, 2000). From this point of view, we can infer that preservice teachers in Turkey choose the teaching profession primarily due to “extrinsic” motivation factors. Specifically, they choose the profession because “due to easy work conditions, which could be explained with Maslow’s Hierarchy of Needs and it deserves priority on the hierarchy. However, it doesn’t seem probable to fill personal needs before filling basic ones on the hierarchy. Thus, it can be advocated that living standards of preservice teachers need to be improved at first.

The teaching profession was found to be preferential among female preservice teachers due to the reasons such as their positive attitude to teach, easy work conditions and interest in working with children/adolescents/adults in a descending order. The ordering for male participants was as “My university entrance score was just enough for it”, “It has easy work conditions” and “I like teaching”. There are some studies categorizing the reasons for choosing the teaching profession in subgroups as “intrinsic” and “extrinsic” (Bastick, 2000; Cermik, Dogan, & Sahin, 2010; Johnston, McKeown, & McEwen, 1999). From this perspective, it can be argued that intrinsic motivation factors take precedence for female preservice teachers compared to male, and extrinsic factors take precedence for males than for females. Likewise, in a study by Johnston, McKeown, and McEwen (1999), it was released that the teaching profession is chosen by female preservice teachers under the influence of intrinsic reasons, whereas it is chosen by males due to extrinsic factors. In a research carried out by Richardson, and Watt (2005), whether the reasons of choosing the profession like "career fit", "prior considerations", "financial reward" and "time for family" change according to factors like gender, having children or not and graduation grade was studied. There was a difference in only terms of the gender factor in the prior considerations variable; this difference resulted in favorability by males. It was reported that this difference stemmed from the different number of males and females and that this finding was not in parallel with the findings of similar research.

Furthermore, it was mentioned in many studies that the teaching profession is perceived as a “feminine profession” and preferred by women more (Foster, & Newman, 2005; Johnston, McKeown, & McEwen, 1999; Saban, 2003; Smedley, 2007). In light of these findings, it can be said that our study has supportive findings for other studies in literature.

Regardless of the type of high school, preservice teachers choose the teaching profession firstly due to the “easy work conditions” it offers. Among graduates of Anatolian Teacher High Schools, the item “I like teaching” was the third most influential factor, while it was in the second reason among respondents not graduated from Anatolian Teacher High Schools. As it may be known, preservice teachers graduated from Anatolian Teacher High Schools could get additional points at university entrance once they chose relevant departments. For this reason, it was an expected result to see the item “My university entrance score was just enough for it” in the second rank for participants who did not graduate of Anatolian Teacher High Schools. In a study carried out by Ovet (2006), no difference was found between respondents graduated from Anatolian Teacher High Schools and the others in terms of reasons for choosing the teaching profession. Based on the study findings, it could be said that Anatolian Teacher High Schools failed to accomplish their foundation objectives. Those schools were founded from 1989 – 1990 with objectives such as preparing preservice teachers for higher education institutions accepting preservice teachers in Turkey; making the teaching profession preferential among preservice teachers, and teaching knowledge, skills, attitudes and behaviors required by the teaching profession (MONE, 2012). However, since an amendment made in 2012, graduates of secondary education could not get additional points any longer when they were placed in a higher education program relevant to their background. Following the law, graduates of

Anatolian Teacher High Schools cannot be granted extra points for the university entrance examination, either.

It seems the ordering of reasons for choosing the teaching profession by department of study is quite revealing. Only those attending PCG chose the profession because it offered good employment opportunities. On the other hand, preservice teachers in departments like Turkish Language and English Language Teaching, which are mainly about teaching linguistic skills, chose the profession because they liked teaching. Those in Classroom Teaching and Elementary Mathematics Teaching indicated the fact that the profession “has easy work conditions”. The latter found that in the middle of a discussion over teacher preparation not all of the preservice teachers chose the teaching profession consciously and those with low academic success levels opted for such departments as motivated by employment consideration (Ozoglu, 2010). According to the examination of base points recorded during teacher placement during August 2011, graduates of PCG were employed despite their lowest university entrance scores among all of the five departments. This fact supports the finding that those admitted to this department chose it since it provides a high chance of employment. All in all, the selection of a profession is not merely a way of making one’s life, rather, and most importantly, it is the route whereby an individual expresses and actualizes her/himself (Kuzgun, 1982). On the other hand, it would be sensible to remember the fact that one needs to fulfil his basic needs for self-actualization. In the context of Turkey, we suggest contemplating how conflicting motives affect preservice teachers in choosing their profession.

The ordering of the reasons for choosing the teaching profession was found to be almost the same for those who had a teacher family member. The only difference was found with item “My family/relatives/friend favored it”. For participants with a teacher in their families in the past or today, this reason was ranked as 7th, while it was ranked 8th for the others. Moran et al. (2001) indicated that preservice teachers preferred the teaching by being inspired by since they were teachers or parents. In Akbayir’s study (2003), it was demonstrated that 62% of the participant preservice teachers chose the teaching profession despite the lack of a teacher in their families, while only 9.4% of the participants did so due to a teacher family member. Our result is supportive of this study’s finding. In a study by Erjem (2000), it was found that the existence of a teacher in the family is not the sole influential factor on choosing the teaching profession by pre service teachers, yet was still important. For those with a teacher family member, it was not specified if s/he chose the profession due to considering that person as a model or just as a consequence of the pressure applied by that person. Thus, it is hardly possible to discuss this finding from a positive or negative point of view.

Examination of the reasons for which preservice teachers choose that profession by income levels reveals that the first reason remains the same for three of the groups. The teaching profession faces considerable challenges related with income level and work conditions (Akar, 1998; Erjem, 2000; Sonmez, 1989; Tezcan, 1981) and is usually preferred by those belonging to the lower and middle class (Akbayir, 2003). Taking this finding into consideration, among participants with a relatively low income level of 500 – 1249 TL and 1250 – 1999 TL, the reason “I like teaching” was ranked 2nd. The same reason was on the 6th rank for those from other income groups. In this study, the reason “It offers job security” was found less influential for the two groups with a low level of income than for the others. This is contrary to the finding that intrinsic and extrinsic motivation factors affect preservice teachers in well-off and underdeveloped countries, respectively (Bastick, 2000).

Evening education is a kind of formal education carried out after daytime education in higher education institutions (Council of Higher Education, 2012). Evening education fees are higher in comparison to daytime education. The leading reason for choosing the teaching profession was the same for both types of education, which is “It has easy work conditions”. In addition, there was no difference between the orders of their reasons. It was not surprising that preservice teachers placed in daytime and evening education with university entrance scores close to each other had the same order of reasons for choosing the profession.

Bearing in mind the fact that the teaching profession is mostly preferred by those born in urban areas or counties (Akbayir, 2003; Erjem 2000; Tatlidil, 1993), again it was not surprising that preservice

teachers who spent most of their lives in provincial or sub-provincial settlements provided the same order of reasons.

This study was carried out in order to investigate reasons for choosing the teaching profession by pairwise comparison method among pre service teachers. A similar study can be done by using other scaling methods. The reasons preservice teachers choose the teaching profession were studied with respect to several variables. However, scaling could not be done on some variables as sufficient number of participants could not be reached specifically for those items. We are of the opinion that considerable findings might be obtained for those variables. A future study to be carried out with the necessary number of participants would likely contribute to the present study as well.

REFERENCES

- Akar, I. (1998). Cumhuriyetin 75. yılına girerken Türkiye'de öğretmen yetiştirme ve istihdama genel bakış. *Yeni Türkiye*, 23(27), 2127–2132.
- Akbayır, K. (2003). Öğretmenlik mesleğine yönelmede ailenin ve branş seçiminde cinsiyetin rolü. Retrieved from: www.fedu.metu.edu.tr/ufbmek-5/b_kitabi/PDF/ÖğretmenYetiştirme/Bildiri/t271d.pdf. (Retrieve date: 03.12.2012).
- Bastick, T. (2000). Why teacher trainees choose the teaching profession? Comparing trainees in metropolitan and developing countries. *International Review of Education*, 46(3/4), 343-349.
- Boz, Y., & Boz, N. (2008). Prospective chemistry and mathematics teachers' reasons for choosing teaching as a profession. *Kastamonu Education Journal*, 16(1), 137–144.
- Brookhart, S. M., & Freeman, D. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, 62(1), 37-60.
- Celikten, M., Sanal, M., & Yeni, Y. (2005). Öğretmenlik mesleği ve özellikleri. *Erciyes University Journal of Social Sciences Institute*, 19(2), 207-237.
- Cermik, H., Dogan, B., & Sahin, A. (2010). Prospective elementary classroom teachers' motives for selecting teaching profession. *Pamukkale University Journal of Education*, 28, 201-212.
- Council of Higher Education, (2012). <http://www.yok.gov.tr/content/view/433/183/lang,tr/> retrieve date: 16.08.2012
- Eren A., & Tezel K.V. (2010). Factors influencing teaching choice, professional plans about teaching, and future time perspective: A mediational analysis. *Teaching and Teacher Education*, 26(7), 1416-1428.
- Erjem, Y. (2000). Öğretmenlik mesleğine yönelmede ailenin işlevi. *Ç.Ü. Eğitim Fakültesi Dergisi*, 2(19), 70-79.
- Eskicumali, A. (2002). *Eğitim, öğretim ve öğretmenlik mesleği öğretmenlik mesleğine giriş*. Ankara: Pegem Yayıncılık.
- Foster, T., & Newman, E. (2005). Just a knock back? Identity bruising on their route to becoming a male primary school teacher. *Teachers and Teaching: Theory and Practice*, 11(4), 341-358.
- Johnston, J., McKeown, E., & McEwen, A. (1999). Choosing primary teaching as a career: The perspectives of males and females in training. *Journal of Education for Teaching*, 25(1), 55-64.
- Klassen, R.M., Al-Dhafri S., Hannok W., & Betts, S. M. (2011). Investigating pre-service teacher motivation across cultures using the Teachers' Ten Statements Test. *Teaching and Teacher Education*, 27(3), 579-588.
- Krecic, M., & Grmek, M. (2005). The reasons preservice teachers choose teaching professions. *Educational Studies*, 31, 265–274.
- Kuzgun, Y. (1982). Mesleki rehberliğin bireylerin yetenek ve ilgilerine uygun meslekleri tanınmalarına etkisi. Ankara: *Ankara Üniversitesi Eğitim Fakültesi Yayınları*, No:118.
- Kyriacou, C., Hultgren, A., & Stephens, P. (1999). Preservice teachers' motivation to become a secondary school teacher in England and Norway. *Teacher Development*, 3, 373–381.
- MONE, (2012). Anadolu öğretmen liselerini tanımak ister misiniz? Retrieved from: http://oyegm.meb.gov.tr/egitim/aol_tanitim.htm (retrieve date: 05.12.2012).
- Moran, A., Kilpatrick, R., Abbott, L., Dallat, J., & McClune, B. (2001). Training to teach: Motivating factors and implications for recruitment. *Evaluation and Research in Education*, 15, 17–32.
- Ovet, O. (2006). *Eğitim fakültesi öğrencilerinin öğretmenlik mesleğini tercih etmelerinde etkili olan faktörlerin belirlenmesi*. Unpublished master dissertation. İstanbul: Yeditepe University, Social Science Institute.
- Ozbek, R. (2007). Öğretmen adaylarının öğretmenlik mesleğini tercih etmelerinde kişisel, ekonomik ve sosyal faktörlerin etkililik derecesine ilişkin algıları. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 17(1), 145–160.

- Ozguven, İ.E. (1999). *Çağdaş eğitimde psikolojik danışma ve rehberlik*. Ankara: PDREM Yayınları.
- Ozoglu, M. (2010). Türkiye’de öğretmen yetiştirme sisteminin sorunları. SETA: Ankara. retrieved from: http://www.setav.org/document/SETA_Analiz_17_Turkiyede_Ogretmen_Yetistirme_Sisteminin_Sorunlari_Murat_Ozoglu_son.pdf
- Ozsoy, G., Ozsoy, S., Ozkara, Y., & Memis, A. D. (2010). Factors affecting pre-service teachers’ choice of teaching as a profession. *Elementary Education Online*, 9(2), 910–927.
- Richardson, P.W., & Watt, H.M.G. (2005). ‘I’ve decided to become a teacher’: Influences on career change. *Teaching and Teacher Education*, 21(5), 475-489.
- Richardson, P. W., & Watt, H. M. G. (2010). *Current and future directions in teacher motivation research*. In T. C. Urdan, & S. A. Karabenick (Eds.), *The decade ahead: Applications and contexts of motivation and achievement* (pp. 139–173). Emerald: Bingley.
- Saban, A. (2003). A Turkish profile of prospective elementary school teachers and their views of teaching. *Teaching and Teacher Education*, 19, 829-846.
- Sinclair, C. (2008). Initial and changing preservice teachers’ motivation and commitment to teaching, *Asia-Pacific Journal of Teacher Education*, 36(2), 79–104.
- Smedley, S. (2007). Learning to be a primary school teacher: reading one man’s story. *Gender and Education*, 19(3), 369-385.
- Sonmez, V. (1989). Türkiye’de öğretmenin ekonomik durumu. *H.U. Journal of Education*, 77–84.
- Stiegelbauer, S. (1992). Why we want to be teachers: new teachers talk about their reasons for entering the profession. Paper presented at the Annual Meeting of the American Research Association, San Francisco, CA (ERIC Document Reproduction Service ED 348 367).
- Tatlidil, E. (1993). Toplum, eğitim ve öğretmen. *E.Ü. Edebiyat Fakültesi Dergisi Yayını*.
- Tezcan, M. (1981). Eğitim sosyolojisine giriş. *A.Ü. Edebiyat Fakültesi Yayını*.
- Thomson, M. M., Turner, J. E., & Nietfeld, J. L. (2012). A typological approach to investigate the teaching career decision: Motivations and beliefs about teaching of prospective teacher candidates. *Teaching and Teacher Education*, 28(3), 324–335.
- Torgerson, W.S. (1965). *Theory and methods of scaling*. Newyork: John Wiley & Sons, Inc.
- Turgut, M.F., & Baykul, Y. (1992). *Ölçekleme teknikleri*: Ankara: ÖSYM Yayınları.
- Watt, H. M. G., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice scale. *Journal of Experimental Education*, 75, 167–202.
- Yuce, K., Sahin, E., Kocer, O., & Kana, F. (2013). Motivations for choosing teaching as a career: a perspective of pre-service teachers from a Turkish context. *Asia Pacific Educ. Rev.* 14, 295–306.

GENİŞ ÖZET

Giriş

Öğretmen yetiştiren kurumlara adayların seçilmesi yıllar içerisinde çeşitli ölçütlere göre yapılmış, bu ölçütler zaman içerisinde değişiklik göstermiştir (Çelikten, Şanal ve Yeni, 2005). Kurumların öğretmen adaylarını nasıl seçtiklerinin yanı sıra öğretmen adaylarının öğretmenlik mesleğini neden seçtikleri, öğretmen yetiştirme sorununun bir başka parçası olarak düşünülebilir. Öğretmen adaylarının öğretmenlik mesleğini neden tercih ettikleri dünyanın birçok ülkesinde araştırma konusu olmuştur.

Öğretmen adaylarının öğretmenlik mesleğini seçme nedenlerine ilişkin birçok çalışma bulunmaktadır. Daha önce yapılan bazı çalışmalarda frekans saymanın ötesine gidilmediği; nicel verilerin elde edildiği diğer çalışmalarda ise ikili karşılaştırmalarla ölçekleme gibi bir yönteminin kullanılmadığı görülmüştür. Bu çalışmada, öğrencilerin öğretmenlik mesleğini seçme nedenlerinin ikili karşılaştırmalar yoluyla incelenmesi dolayısıyla bu nedenleri başlangıç noktası ve birimi olan bir ölçek düzeyine getirerek, verilere nicel olma özelliği kazandırılması kısacası nedenlerin ölçeklenmesi amaçlanmaktadır.

Bu çalışmada eğitim fakültesi öğrencilerinin öğretmenlik mesleğini seçme nedenleri ikili karşılaştırmalar yapılarak ölçeklenmiştir. Bu işlem çeşitli bağımsız değişkenlere göre yapılmış olup var olan durumu ortaya çıkarma, bir çeşit betimleme yapma amacı taşımaktadır. Bu nedenden dolayı çalışmanın betimsel bir çalışma olduğu söylenebilir.

Yöntem

Verilerin analizinde Thurstone'nun ikili karşılaştırmalar yönteminin tam veri matrisi için III. Hal ve V. Hal denklemlerinden yararlanılmıştır. Bu yöntemde öncelikle her bir ikiliden birinin diğerine tercih edilme sıklığını belirten frekans matrisi (F) oluşturulmuştur. Bu frekanslar matrisindeki her bir hücre toplam kişi sayısına (N) bölünerek oranlar matrisi (P) elde edilmiştir. Oranlar matrisindeki her bir orana karşılık gelen birim normal dağılım değeri bulunarak birim normal sapma matrisi (Z) oluşturulmuştur. Ardından önce birim normal sapma matrisindeki hücreler sütun bazında toplanmış; daha sonra da sütun toplamları her bir sütundaki gözlem sayısına bölünerek 8 nedenin diğer bir ifadeyle 8 uyarıcının ölçek değerleri bulunmuştur. Tüm uyarıcıların ölçek değerine, negatif olan en küçük ölçek değeri kadar belli bir sayı eklenmiştir. Böylece en küçük ölçek değeri başlangıç noktası olmuş ve buna dayanarak diğer uyarıcıların ölçek değerleri hesaplanmıştır.

Sonuç ve Tartışma

Günümüzde “Türkiye’de öğretmen yetiştirme sorunu” çerçevesinden bakıldığında öğretmen adaylarının bu mesleği öncelikli olarak çalışma koşulları rahat olduğu için seçmiş olmaları, bu sorunun varlığını bir daha pekiştirmektedir. Genel olarak gelişmekte olan ülkelerde öğretmenlerin, iş güvenliği, maaş gibi dışsal motivasyon kaynaklardan; gelişmiş, refah düzeyi yüksek ülkelerde ise mesleği sevmeye, ülkeye topluma faydalı olma gibi içsel veya özgeci motivasyon kaynaklarından etkilenerek meslek seçiminde bulunduğu söylenebilir (Bastick, 2000). Bu açıdan bakıldığında ülkemizde öğretmen adaylarının öncelikli olarak “dışsal” motivasyon kaynaklarından etkilenerek öğretmenliği seçtikleri ve “çalışma koşulları rahat olduğu için” nedeninin Maslow’un temel ihtiyaçlar teorisine göre daha temel olduğu ve dolayısıyla öncelikli olarak karşılanması gereken bir ihtiyaç olduğu söylenebilir.

Kadın öğretmen adaylarının öğretmenlik mesleğini seçme nedenlerinin başında “öğretmenliği sevmeleri”, sonra “çalışma koşullarının rahat olması” ve üçüncü olarak da “çocuklarla/gençlerle/çocuklarla çalışmak istemesi” gelmektedir. Bu sıralama erkeklerde ise “puanım yettiği için”, “çalışma koşulları rahat olduğu için” ve “öğretmenliği sevdiğim için” şeklindedir. Öğretmenlik mesleğini seçme nedenlerinin “içsel” ve “dışsal” gibi çeşitli kategoriler altında toplandığı çalışmalar bulunmaktadır (Bastick, 2000; Çermik ve ark., 2010; Johnston ve ark., 1999). Bu açıdan bakıldığında kadın öğretmen adaylarının öncelikli olarak öğretmenliği seçerken içsel motivasyon kaynaklarından; erkeklerin ise daha çok dışsal motivasyon kaynaklarından etkilendiği söylenebilir.

Mezun olunan lise türü arasında bir fark olmaksızın öğretmen adaylarının öğretmenliği seçme nedenlerinin başında “çalışma koşulları rahat olduğu için” nedeni gelmektedir. Anadolu Öğretmen Lisesi mezunu öğrencilerinin öğretmenlik mesleğini seçme nedenlerinden “öğretmenliği sevdiğim için” nedeni 3. sırada iken; bu neden Anadolu Öğretmen Lisesi mezunu olmayan öğrenciler için 2.sıradadır. Bilindiği gibi Anadolu Öğretmen Lisesi mezunu öğrencilerin öğretmenlik mesleğini tercih etmeleri durumunda üniversiteye giriş puanlarına ek puan dahil edilmekteydi. Bu çerçeveden bakıldığında “puanım yettiği için” nedeninin Anadolu Öğretmen Lisesi mezunu öğrencilerinin öğretmenlik mesleğini seçme nedenleri arasında 2. sırada yer alması beklenen bir sonuçtur. Türkiye’de öğretmen yetiştiren yükseköğretim kurumlarına öğrenci hazırlamak, öğrencilerine öğretmenlik mesleğini sevdirmek, öğretmenlik mesleğinin gerektirdiği bilgi, beceri, tutum ve davranışları kazandırmak (MEB, 2012) gibi amaçlar doğrultusunda 1989 – 1990 yıllarında kurulmuş olan Anadolu Öğretmen Liselerinin araştırma bu sonuçlara göre belirtilen amaçları yerine getirmediği söylenebilir. Nitekim 2012 yılında yapılan bir yasa değişikliği ile ortaöğretim kurumlarından mezun olan öğrencilerin aynı meslek dalında yer alan yükseköğretim programlarına yerleşmeleri durumunda üniversiteye giriş puanlarına bir ek puan eklenmemektedir. Bu yasa gereği Anadolu Öğretmen Lisesi mezunlarının öğretmenliği tercih etmeleri durumunda üniversite giriş puanlarına ek puan dahil edilmemektedir. Son yapılan değişiklikler ile Anadolu Öğretmen Liselerinin öğretmen yetiştirme işlevi tamamen kaldırılmıştır.

Bölgümlere göre öđretmenlik mesleđini tercih etme nedenleri sıralamalarının çarpıcı olduđu söylenebilir. Yalnızca RPD bölümü öğrencilerinin öncelikli olarak iş bulma olanađı yüksek olduđu için; dile ilişkin becerileri kazandırmaya dayalı olan Türkçe Öđretmenliđi ve İngilizce Öđretmenliđi bölümü öğrencilerinin öđretmenliđi sevdiđi için, Sınıf Öđretmenliđi ve İlköđretim Matematik Öđretmenliđi Bölümü öğrencilerinin çalışma koşulları rahat olduđu için öđretmenliđi seçtiđi görülmüştür.

Ailesinde geçmişte veya hala öđretmen olan ve olmayan öđretmen adaylarının öđretmenlik mesleđini seçme nedenleri hemen hemen aynı olup tek fark “ailem/yakınım/arkadaşım istediđi için” nedeninde ortaya çıkmıştır. Ailesinde geçmişte veya hala öđretmen olan öğrenciler için bu neden 7. sırada iken öđretmen olmayan bireyler için 8. sıradadır. Ailesinde öđretmen olan bir bireyin öđretmenlik mesleđini, ailedeki öđretmen olan bireyin model olmasından mı kaynaklı yoksa bireye öđretmenlik mesleđini seçmesi yönünde baskı yapmasından dolayı mı seçtiđi bilinmediđi için bu bulguyu iyi ya da kötü şeklinde deđerlendirme olanađı bulunmamaktadır.

Gelir düzeylerine göre öđretmen adaylarının öđretmenlik mesleđini seçme nedenleri incelendiđinde her üç grup için ilk nedenin aynı olduđu görülmektedir. Gelir ve çalışma koşulları açısından önemli sorunları olan öđretmenlik mesleđi (Akar, 1998; Erjem, 2000; Sönmez, 1989; Tezcan, 1981) daha çok alt ve orta sosyo-ekonomik düzeydeki ailelere üye olan kişiler tarafından seçilmektedir (Akbaýır, 2003).

İkinci öđretim, Yükseköđretim kurumlarında normal örgün öđretimin bitimini takiben yapılan örgün öđretimi olarak tanımlanmaktadır (YÖK, 2012). İkinci öđretim harç ücretleri normal öđretime göre biraz daha fazla olup öđretim akşam saatlerinde yapılmaktadır. İkinci öđretim ve normal öđretim öğrencilerinin öđretmenlik mesleđini seçme nedenlerinin başında “çalışma koşulları rahat olduđu için” nedeni gelmekte olup nedenlerin sıraları arasında hiçbir farklılık görülmemektedir. Üniversiteye giriş puanları arasında çok bir farklılık bulunmayan I. öđretim ve II. öđretimde okuyan öđretmen adaylarının öđretmenlik mesleđini seçme nedenlerinin sıralamalarında bir farklılık çıkmaması beklendik bir sonuçtur.

Öđretmenlik mesleđini daha çok şehir ve ilçe doğumlu bireylerin seçtiđi bulgusu göz önünde bulundurulduğunda (Akbaýır, 2003; Erjem 2000; Tatlıdil, 1993) yaşamlarının büyük çođunluđunun il veya ilçe merkezlerinde geçiren öđretmen adaylarının öđretmenlik mesleđini seçme nedenleri sıralamasının aynı olması yine beklenen bir sonuçtur.

Çocuklardaki Kısa Süreli Anlık Bellek İşlevinin Değerlendirilmesi

Evaluation of Short Term Memory Span Function In Children

Bariş ERGÜL*

Arzu ALTIN YAVUZ**

Ebru GÜNDOĞAN AŞIK***

Öz

Kısa süreli bellek, kodlanan bilgilerin geçici olarak depolandığı yer olmakla birlikte bu bilgiler sonraki aşamada çalışma belleğine kayıt edilir. Zihinsel olarak tekrar edilen bilgiler bellekte uzun süre kalmaktadır. Okuma becerisi ile kısa süreli bellek ilişkisini ele alan çalışmalar, daha çok çalışma ve kısa süreli bellek süreçleri ile okuduğunu anlama arasındaki ilişkiyi inceleyen çalışmalardır. Bu çalışmada, kısa süreli belleğe gelen bilgiler ile bu belleğin çalışmasını etkileyen değişkenler arasında ilişki regresyon modeli ile araştırılmıştır. Araştırmanın amacı, çocuklarda kısa süreli belleği etkileyeceği düşünülen yaş, IQ (zeka derecesi) ve okuma becerisi arasındaki ilişkiyi regresyon analizi yardımıyla kurmaktır. Regresyon Analizi varsayımlarından bir tanesi hata terimlerinin normal dağılımının normal dağılım göstermediği durumlarda etkin sonuçlar veren sağlam regresyon teknikleri tanıtılmış ve uygulanmıştır. Ayrıca her bir teknik için bağımsız değişkenlerin bağımlı değişkeni açıklama oranları verilmiştir. Elde edilen bulgulara göre, yaş, IQ ve okuma becerisindeki artış, kısa süreli bellek üzerinde artışa neden olmaktadır. Sağlam regresyon teknikleri uygulandıktan sonra yaş, IQ ve okuma becerisi değişkenlerinin kısa süreli anlık bellek değişkenini açıklama yüzdesi en yüksek değerini Winsorized En Küçük Kareler Tekniği (WLS) vermiştir.

Anahtar Kelimeler: kısa süreli bellek, yaş, IQ, regresyon, en küçük kareler, sağlam regresyon

Abstract

Although details of the information encoded in the short-term memory where it is stored temporarily be recorded in the working memory in the next stage. Repeating the information mentally makes it remain in memory for a long time. Studies investigating the relationship between short-term memory and reading skills that are carried out to examine the relationship between short-term memory processes and reading comprehension. In this study information coming to short-term memory and the factors affecting operation of short term memory are investigated with regression model. The aim of the research is to examine the factors (age, IQ and reading skills) that are expected the have an effect on short-term memory in children through regression analysis. One of the assumptions of regression analysis is to examine which has constant variance and normal distribution of the error term. In this study, because the error term is not normally distributed, robust regression techniques were applied. Also, for each technique; coefficient of determination is determined. According to the findings, the increase in age, IQ and reading skills caused the increase in short term memory in children. After applying robust regression techniques, the Winsorized Least Squares (WLS) technique gives the highest coefficient of determination.

Keywords: short term memory, age, IQ, regression, ordinary least squares, robust regression

GİRİŞ

Zeka, karmaşık bir süreçtir ve kesin bir tanımı yoktur. Zihnin karşılaştırma, planlama, problem çözme ve öğrenme ile ilgili özelliklerini kapsamaktadır. Başka bir deyişle; zeka, kişinin bilgiyi alma hızı olarak da tanımlanabilir (Doruk ve Öngören, 2005). IQ (Intelligence Quotient) ise, zeka derecesidir ve farklı insanlar arasında zeka farklılıkları olduğunu varsayar, bir skor ile kişinin aynı

* Arş. Gör., Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, Eskişehir-Türkiye, e-posta: bergul@ogu.edu.tr

** Doç. Dr., Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, Eskişehir-Türkiye, e-posta: aaltin@ogu.edu.tr

***Arş.Gör., Karadeniz Teknik Üniversitesi, Fen-Edebiyat Fakültesi, Trabzon-Türkiye, e-posta: ebrugundogan@ktu.edu.tr

yaş grubundan kişiler ile zihinsel performansını karşılaştırılması işlemi olduğu belirtilmektedir (Köksal ve Boran, 2015). Başka bir deyişle IQ, sözel ve analitik zihinsel kabiliyetlerin ölçüldüğü psikometrik bir testten alınan skor olarak da ifade edilmektedir (Fer, 2004).

Zekayı belirlediği öne sürülen iki kaynak bulunmaktadır. Bu kaynaklardan ilki olan bilgi işleme hızı, zihinsel becerilerin analitik olarak incelenmesi sürecidir. Çalışma belleği olarak adlandırılan diğer kaynak, bilginin depolanması sürecini kapsamaktadır (Alp ve Özdemir, 2007). Kısa süreli bellek ise, bilginin geçici olarak depolandığı birimdir. Aynı zamanda zihinsel işlemlerin ve bilgi işleme sürecinin yapılmasında görev aldığı için, işleyen bellek olarak da adlandırılmaktadır (Miller, 1956; Banikowski, 1999). Seamon ve Kenrick (1994), kısa süreli anlık belleği, çalışma belleğinin bir bilşeni olarak tanımlamıştır. Bu özellikleri ile kısa süreli bellek, hızlı düşünme işlemlerinin de yapıldığı yerdir. Bu nedenle hep uyanıktır (Senemoğlu, 2002). Kısa süreli bellekte muhafaza edilen bilgiler kodlama yoluyla çalışma belleğine gönderilir (Korkmaz ve Mahiroğlu, 2007). Cowan (1995), çalışan belleğin yapısının karmaşık olduğunu belirterek, bilgilerin merkezde çalışan bir süreç ile değerlendirilmesi gerektiğini söylemiştir. Saeed (2011), çalışma belleğini, kısa süreli anlık belleğin alt kümesi olarak tanımlamıştır. Son yıllarda yapılan çalışmalar ile çalışma belleği ve kısa süreli bellek arasında fark olduğu ve bu farkın kontrol edilebilen dikkat ile ilgili olduğu belirlenmiştir (Alp ve Özdemir, 2007).

Peterson ve Peterson'un (1959) yaptığı bir araştırmayla kısa süreli bellekte bilginin kalış süresinin 20 saniye civarında olduğu bulunmuştur. Araştırmaya katılan kişilerden 3, 6, 9, 12, 15 ve 18 saniye içerisinde verilen bilgileri hatırlamaları istenmiştir. Sonuç olarak; 3 saniye sonra %80, 6 saniye sonra %50, 18 saniye sonra %10'dan daha az bilgi hatırlanmıştır. Bu bulgular ışığında, bilginin kısa süreli bellekte daha uzun süre kalması, zihinsel tekrar ile mümkündür. Kısa süreli belleğe iletilen bilgi, zihinsel tekrar yapılmadığı takdirde kısa sürede unutulmaktadır (Senemoğlu, 2002).

Baddeley ve Hitch (1974) kurdukları çalışma belleği modellerine göre, merkezi yürütücü dikkatin kontrolü, strateji seçimi ve bunların yürütülmesinden sorumludur. Bunun yanında, bilginin sistemde tutulmasını sağlayarak merkezi yürütücünün işini kolaylaştıran iki köle sistem bulunmaktadır. Bunlar, sözel bilgiyi sistemde tutan ses bilgisel döngü ile görsel bilgiyi sistemde tutan görsel-uzamsal yazboz tahtasıdır. Ses bilgisel döngü, bilginin ses bilgisel şekilde tutulduğu bir tekrarlama süreci ile işlemektedir (Sayar ve Turan, 2012).

Okuma becerisi, bir süreçtir ve bu süreç birkaç farklı sürecin eşanlı olarak gerçekleşmesi sonucu ortaya çıkar. Bu beceri, genel olarak söylenen sözcükleri eşleştirme ile ilişkilidir. Bu yönü ile zihinsel bir süreç içerdiğinden, okuma becerisinin çeşitli işaretleri tanıyarak bunlardan anlam çıkarma etkinliği olduğu belirtilmektedir (Özkan ve Doğan, 2013). Sözü edilen okuma becerisi, dilin ses bilgisel yapısının çalışma belleği ile ilgili bölümü üzerine kurulmuş olan süreci ifade eder (Cain et al., 2004 ; Scarborough ve Brady, 2002; Akoğlu ve Turan, 2012). Ses bilgisel farkındalık, çeşitli becerileri içeren farkındalık olarak ortaya çıkar. Bunlar; kelimelerin ilk ünsüzlerini eşleştirme, kelime içindeki sesleri sayma, uyaklı (kelimelerin son heceleri aynı) sesleri eşleştirmedir (Stahl ve Murray, 1994; Sayar ve Turan, 2012). Adams (1990), ses bilgisel farkındalığın doğuştan gelmediğini ve zaman, çaba harcayarak geliştirilebileceğini söylemektedir.

IQ ve okuma becerisi puanları arasındaki korelasyonun yaş ile arttığına ilişkin bulgular, IQ testleri ile ölçülen okuma ve genel bilişsel yeteneğin arasındaki karşılıklı etkinin bulunduğu hipotezini destekler niteliktedir (Stanovich, 1986).

Okuma becerisi ve kısa süreli bellek kavramları arasındaki ilişkinin ele alındığı çalışmalar genellikle, çalışma belleği ve kısa süreli bellek süreçleri ile okuduğunu anlama arasındaki ilişkiyi inceleyen çalışmalardır. Daneman ve Carpenter (1980), çalışma belleğinin okuma performansı ile ilgili ilişkisini incelemişler ve çalışma belleği ile okuma becerisi arasında yüksek korelasyon olduğunu tespit etmişlerdir. Kyllonen ve Christal (1990), yaptıkları çalışmada benzer sonuçlar bulmuşlardır.

Türkiye'de kısa süreli bellek ile okuma becerisi ve zeka arasındaki çalışmalar son on yılda artarak yapılmaya başlamıştır. Korkmaz ve Mahiroğlu (2007), etkili öğrenmeye katkı sağlayacağını düşündükleri bellek destekleyicileri açıklamışlardır. Alp ve Özdemir (2007), yetişkinlerde gözlenen bilgi işleme hızının akıcı zekayı etkilemediği bulgusunun, çocuklar için de geçerli olduğunu

belirlemişlerdir. Babacan (2012), analitik okuma stratejileri ile zeka alanları arasındaki ilişkinin anlamlı olduğunu saptamışlardır.

Çalışmanın 2. Bölümünde Regresyon Analizi ile giriş yapılmış ve Sağlam Regresyon teknikleri tanıtılmış, ardından gelen 3. Bölümde ise kısa süreli belleği etkileyeceği düşünülen değişkenler ile regresyon analizi yapılmıştır. Hata terimlerinin dağılımının normal dağılım olmamasından dolayı sağlam regresyon teknikleri uygulanmıştır. 4. Bölüm olan Sonuçlar bölümünde ise sonuçlar değerlendirilmiştir.

Araştırmanın Amacı

Literatürde incelenen benzer çalışmaların çoğunda, kısa süreli belleğin IQ'yu etkilediği düşünülmüştür. Ancak Yapısal Eşitlik Modeli (YEM) yardımıyla yapılan çalışmalarda bir değişken hem içsel hem de dışsal değişken olarak modele dahil edilmektedir. Bir başka ifade ile, aynı değişken hem bağımsız hem de bağımlı değişken olarak model katılmaktadır. Bu çalışmada ise, kısa süreli anlık belleği etkileyen etmenlerin belirlenmesi amaçlanmıştır. Literatürdeki çalışmalardan hareketle, kısa süreli anlık bellek ile IQ ve okuma becerisi değişkenleri göz önünde bulundurulmuştur. Kurulan modelde, hata terimlerinin dağılımı incelendiği zaman ilgili dağılımın normal dağılım göstermediği belirlendiğinden robust regresyon teknikleri uygulanmıştır.

YÖNTEM

Örneklem/Çalışma Grubu

Eskişehir’de özel bir kuruluş ile birlikte 2015 yılında yürütülen bu çalışmada 6-8 yaş aralığında okuma-yazma bilen 10 kız ve 10 erkek toplam 20 çocuk, çalışmanın örneklemini oluşturmaktadır. Özel kuruluşun zaman ve maliyet gibi durumlarından dolayı örneklem 20 çocuk ile sınırlandırılmıştır. Örnekleme oluşturan çocuklar, uygulamaya geçilmeden önce, ailelerine uygulanacak test hakkında gerekli bilgiler aktarılmış ve sonrasında, gönüllük esasına uygun olarak katılmak isteyenlerden bilgilendirilmiş onay formu ile izin alınmıştır.

Y : Kısa süreli bellek,

X_1 : Yaş,

X_2 : IQ

ve

X_3 : Okuma becerisi.

Çocuklardan elde edilen veriler için betimleyici istatistik değerleri Tablo 1 ‘de gösterilmiştir.

Tablo 1. Çocuklarda yaş, IQ, kısa süreli bellek ve okuma becerisi için betimsel istatistikler

Değişken	n	Min.	Maks.	Ortalama	Standart Sapma
Yaş	20	6,00	8,00	7,40	0,75
Kısa Süreli Anlık Bellek	20	3,60	5,50	4,53	0,58
IQ	20	88,00	102,00	95,25	4,39
Okuma Becerisi	20	5,00	7,30	6,30	0,69

Veri Toplama Araçları/Veri Toplama Yöntemleri / Veri Toplama Teknikleri

Araştırmaya katılmaya gönüllü olan çocuklara, Temel Okuma/Yazma Becerileri Ölçeği uygulanarak, okuma becerisi puanları tespit edilmiştir. Ölçek, Yangın tarafından 2002 yılında geliştirilmiştir. Ölçeğin güvenirliği 0,87’dir. Temel okuma-yazma becerilerini belirleyen ölçek, okuma için 54, yazma için de 54 olmak üzere toplam 108 sorudan oluşmaktadır.

Çocuklara, Cohen (1997) tarafından geliştirilen ve Özyürek (2009) tarafından Türkçe uyarlama çalışması yapılan Çocuklarda Bellek Süreçlerini Değerlendirme Ölçeği 5-8 yaş formu uygulanmıştır. İlgili ölçek, altı yaşındaki çocuklara Özyürek (2009) tarafından uygulanmış ve toplam güvenilirlik katsayısı olan Cronbach alfa değeri, 0.98 olarak bulunmuştur. Ölçekte yer alan Görsel Anlık İndeks değerleri, kısa süreli bellek puanını; Görsel Ertelenmiş İndeks değerleri de uzun süreli bellek puanını belirtmektedir.

Wechsler (1974), Çocuklar İçin Zeka Ölçeği/Geliştirilmiş Formu kullanılarak IQ değerlerine ulaşılmıştır. Wechsler Çocuklar için Zeka Ölçeği/Geliştirilmiş Formu (1974), toplam 12 testten ve 2 bölümden oluşan standart hale getirilmiş ve 6-16 yaş aralığındaki çocuklara uygulanan bir zeka ölçeğidir. Bu testlerden, Sözel Zeka Bölümü, Performans Zeka Bölümü ve Genel Zeka Bölümü olmak üzere üç farklı puan türü elde edilmektedir.

Veri Analizi

İlgili çalışma için regresyon analizi tekniği kullanılarak, veriler analiz edilmiştir. Regresyon analizinin temelinde, gözlenen bir olayın fonksiyonel yapısının belirlenmesi yatmaktadır. Regresyon analizi yapılırken gözlem değerlerinden yola çıkarak ve bu değerler kullanılarak, olayların bir matematiksel gösterimle yani bir fonksiyon yardımıyla ifade edilmesi gerekmektedir. Kurulan bu modele “Regresyon Modeli” denilmektedir. Regresyon Analizi, değişkenler arasındaki ilişkiyi belirleyerek parametre kestirimini sağlayan istatistiksel bir yöntemdir (David, 1981).

Regresyon analizi, ilk olarak astronomi alanında Legendre (1805) ve Gauss (1809) tarafından gezegenlerin yörüngelerini belirlemek amacıyla kullanılmıştır.

p adet bağımsız değişken için çoklu doğrusal regresyon modeli aşağıdaki gibidir:

$$Y = Xb + \varepsilon \quad (1)$$

Burada Y ; Bağımlı değişken vektörü, X_1, X_2, \dots, X_p : bağımsız değişkenler (X veri matrisi) ve b_i ($0 \leq i \leq p$) regresyon katsayıları, ε ise rassal hata vektörüdür. b_i ($0 \leq i \leq p$) regresyon katsayılarının tahmini EKK yöntemi yardımıyla (2) formülü kullanılarak,

$$\hat{b} = (X'X)^{-1}X'Y \quad (2)$$

bulunur.

Regresyon katsayılarının tahmin amaçlı kullanılabilmesi için temel bazı varsayımlar vardır. Bu varsayımlar genel olarak hata terimleri ile ilgilidir. Özetle, hata terimleri ε_i bağımsız ve özdeş, ortalaması sıfır, varyansı σ^2 olan normal dağılıma sahiptir. Bu varsayımların bozulması durumunda tutarlı olmayan ve yanlı tahminlere yol açılabilir (Mosteller ve Tukey, 1977). Ancak hata terimleri dağılımı normal dağılıma sahip değilse, uygulanacak teknikler, sağlam regresyon teknikleridir. Sağlam regresyon teknikleri, aykırı değer adı verilen ve verinin geri kalan kısmından oldukça büyük farklılıklar gösteren değerlerden etkilenmeyen tekniklerdir. Varsayımlara bağlı olmayan, özellikle normallik varsayımına duyarsız yaklaşımlar “robust (sağlam)” olarak adlandırılmıştır (Öztürk, 2003; Ergül, 2006).

Bu çalışmada literatürde yer alan bazı sağlam regresyon teknikleri tanıtılacaktır. Bunlar; EKK, M, Winsorized, LMS, LTS, Theil, R-Regresyon, KR teknikleridir.

En Küçük Kareler Tekniği (EKK)

EKK tekniği, regresyon analizinde parametre tahmini için en yaygın kullanılan tekniktir. Bu teknikte amaç, hata terimlerinin karelerinin minimizasyonudur. Bu durum kısaca Eşitlik (3) ile verilmiştir. Hata kareler toplamını minimum yapan \hat{b} tahminleri regresyon modelindeki b katsayılarının EKK tahmin edicileridir. değerlerine eşit ağırlık vererek kareler toplamını minimize etmektir:

$$\min \sum_{i=1}^n (\varepsilon_i^2), i = 1, 2, \dots, n \quad (3)$$

Burada, ε_i hata terimlerini göstermektedir.

M-Regresyon Tekniği

Hata terimlerinin ana kitle için normallik varsayımı geçersiz olduğunda EKK tekniğinin uygulanması ve elde edilen modelin tahmin için kullanılması doğru olmaz. M-Regresyon bu varsayımına uygun olarak geliştirilmiştir. M-Regresyon, EKK tekniğine alternatif olarak geliştirilmiştir. M-Regresyon tekniğinde, hata terimlerinin bir fonksiyonu minimize edilir (Tukey, 1970; Huber, 1973; Andrews, 1974; Birkes ve Dodge, 1993; Ergül, 2006). Parametre tahmini için

ise, ağırlıklandırılmış EKK yöntemini kullanılmaktadır. $\min \sum_{i=1}^n \rho(\varepsilon_i)$ ($i = 1, 2, \dots, n$) minimize edilir. Burada ρ ; amaç fonksiyonunu göstermektedir.

Winsorized En Küçük Kareler Tekniği (WLS)

WLS tekniği, hata terimlerinin büyüklüklerinde önemli farklılıklar olduğu zaman kullanılır. Bu metodun amacı, parametre tahmin değerlerini etkileyen aykırı değerlerin etkilerini azaltmaktır. WLS tekniğinde parametre tahmini yapmak için iterasyon metodu diye adlandırılan algoritma kullanılır. Bunun dışında seviyelendirme metodu ve artan seviyelerde iterasyon diye bilinen metodlar da kullanılabilir. Hata terimleri sıralandıktan sonra baştan ve sondan g adet hata terimi atılarak, baştaki g adet hata teriminden sonraki hata terimi ve sondaki g adet hata teriminden önceki hata terimi atılan hata terimlerinin yerine alınmaktadır. Daha sonra hesaplanan bu yeni y değerlerini kullanarak yeni parametre değerleri bulunur. İkinci iterasyon için yeniden hata terimi değerleri hesaplanır. Ardından uyarlanmış y değerleri üçüncü iterasyon için hesaplanır. Bu iterasyonlar böylece devam eder. Nevitt ve Tam (1998), 5 iterasyondan sonra Winsorized regresyon parametre tahminlerinin birbirine çok benzediğini göstermiştir. Bu nedenle 5 iterasyon yapmak regresyon parametreleri hakkında karar vermek için yeterlidir (Yale ve Forsythe, 1976).

En Küçük Medyan Kare Tekniği (LMS)

LMS, EKK tekniğinde olduğu gibi hata terimlerinin kareler toplamını minimize etmek yerine, kareleri alınmış hata terimlerinin medyan değerini minimize etmeyi amaçlayan bir tekniktir (Rousseeuw, 1984). Başka bir ifadeyle LMS tekniğinde minimize edilecek fonksiyon,

$$\min \text{med}(\varepsilon_i^2) \quad (4)$$

şeklinde tanımlanır. LMS tekniği için geliştirilen algoritma, k adet farklı gözlem değerinin alt kümelerini elde etmekle başlar. k adet gözlem değeri için regresyon parametreleri hesaplanır. Daha sonra tüm gözlemler kullanılarak hata terimlerinin karelerinin medyan değeri hesaplanır. Sonuçta minimum medyan değeri veren parametre değeri, en iyi parametre değeri olarak kabul edilir (Rousseeuw ve Leroy, 1987).

Budanmış En Küçük Kareler Tekniği (LTS)

LTS tekniği, Rousseeuw (1984) tarafından geliştirilmiştir. Bu metodun kullanılmasındaki amaç aşağıdaki fonksiyonu minimize etmektir:

$$\min \sum_{i=1}^h (\varepsilon_i^2) \quad (5)$$

Bu teknikte hata terimlerinin kareleri, ε_i^2 , EKK metodu kullanılarak hesaplanır ve küçükten büyüğe doğru sıralanır. Hata terimlerinin küçükten büyüğe sıralanmasından sonra $0 < \alpha < 1/2$ olmak koşulu ile en büyük αn tane hata terimine karşılık gelen gözlem değeri veri setinden silinerek geriye kalan

gözlem değerleri için EKK metodu uygulanır (Nevitt ve Tam, 1998). Uygulamalarda α değeri sıklıkla 0,05 veya 0,1 olarak seçilir. Bu algoritmadan farklı algoritmalar da mevcuttur.

Theil Tekniği

Theil (1950) metodu parametrik olmayan bir metottur ve literatürde yaygın bir kullanıma sahiptir. Hata terimleri bir dağılıma uymadığında ya da veri setinde aykırı değerlerin varlığından şüphelenildiğinde bu metodun kullanılması önerilmiştir. Çoklu doğrusal regresyon modelinin eğim katsayılarını elde etmek için, veri setindeki tüm ikili çiftler kullanılarak Theil'in basit doğrusal regresyon eğim katsayısı tahmin edicisi bilgisinden hareket edilir. Çeşitli yollardan elde edilen tahmin edici değerleri çoklu regresyon için kullanılır (Zhou ve Serfling, 2006).

R-Regresyon Tekniği

R-kestiricileri, Hodges ve Lehman (1963)' in çalışmaları sayesinde ortaya çıkmıştır. R-Regresyon tekniğinde, parametre tahmini yapılırken hata terimlerinin sıralandırılması gerekmektedir.

Bu teknikte amaç,

$$\min \sum_{i=1}^n a_n(R_i) \varepsilon_i \quad (6)$$

fonksiyonunu minimize etmektir.

Burada R_i , ε_i hata teriminin sırasını, $a_n(i)$ ise $\sum_{i=1}^n a_n(i) = 0$ eşitliğini sağlayan monoton bir skor fonksiyonunu göstermektedir (Heikkilä, 2006). Skor fonksiyonları Tablo 2 'de gösterildiği biçimlerde tanımlanırlar (Zhou, 1992):

Tablo 2. Skor Fonksiyonları

Skor Fonksiyonu	$a_n(i)$
Wilcoxon	$i - \frac{n+1}{2}$
Van Der Waerden	$\Phi^{-1}\left(\frac{i}{n+1}\right)$
	(Φ^{-1} birikimli normal dağılım fonksiyonunun tersidir.)
Median	$\text{sgn}\left(i - \frac{n+1}{2}\right)$

Kantil Regresyon (KR)

Alternatif regresyon tekniklerinden biri de Kantil Regresyon (KR)'dur ve diğer regresyon modellerinde olduğu gibi bu yöntemde amacı değişkenler arasındaki ilişkiyi açıklamaktır (Kurtoğlu, 2011). Koşullu kantil fonksiyonlarının tahmini için uygun bir yöntem olarak Koenker ve Basset (1978) tarafından geliştirilmiştir. Kantil regresyon, özellikle koşullu kantillerin değişkenlik gösterdiği durumlarda kullanışlıdır. Kantillere bağlı olarak regresyon katsayıları belirlenir (Chen ve Wei, 2005). Klasik regresyon modelinin aksine hataların varyansının homejenliği ve hataların dağılımı hakkında herhangi bir varsayımı yoktur. Klasik regresyon bağımlı değişkenin koşullu

beklenen değeri için bir model ararken, kantil regresyon bağımlı değişkenin koşullu dağılımında seçilen kantiller için model belirler.

Kantil regresyonunun amaç fonksiyonu, mutlak sapmaların ağırlıklandırılmış toplamlarıdır. θ 'inci kantil regresyon aşağıdaki fonksiyonun minimizasyonu ile tahmin edilir (Kuo ve Yu, 2013):

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{\theta y \geq x_i \beta} \theta |y_i - x_i \beta| + \sum_{\theta y < x_i \beta} (1-\theta) |y_i - x_i \beta| \right\} \text{ dir ve} \quad (7)$$

$$\hat{\beta}_{\theta} = \frac{1}{n} \sum_{i=1}^n \rho_{\theta}(y_i - x_i \beta) = \sum_i \rho_{\theta}(u_{\theta i}) \quad (8)$$

Kantil regresyonun bu formda gösterim, doğrusal programlama gösterimidir. Burada, ρ_{θ} fonksiyonu;
 $u_{\theta i} < 0$ yani $(y_i < x_i \beta)$ ise $(1-\theta)u$
 $u_{\theta i} \geq 0$ yani $(y_i \geq x_i \beta)$ ise θu olarak belirlenir.

BULGULAR

Çalışmanın ilk aşamasında çoklu doğrusal regresyon analizi uygulanmış ve sonuçlar Tablo 3 ve 4 'de gösterilmiştir. Sonraki aşamada hata terimlerinin normal dağılıp-dağılmadığı araştırılmış ve hata terimleri için aykırı değer grafiği çizilmiştir. Bu grafik Şekil 1 'de gösterildiği gibidir. Son aşamada ise, sağlam regresyon teknikleri uygulanmış ve tüm sonuçlar Tablo 4 'de gösterilmiştir.

Tablo 3. EKK Tekniğine Ait Çoklu Regresyon Analizi Sonuçları

Değişkenler	Kestirim	Standart Hata	t	
			İstatistik	p-değeri
Sabit	-2,987	1,390	-2,149	0,047 *
X ₁	0,267	0,121	2,207	0,042 *
X ₂	0,037	0,017	2,190	0,044 *
X ₃	0,359	0,140	2,572	0,020 *

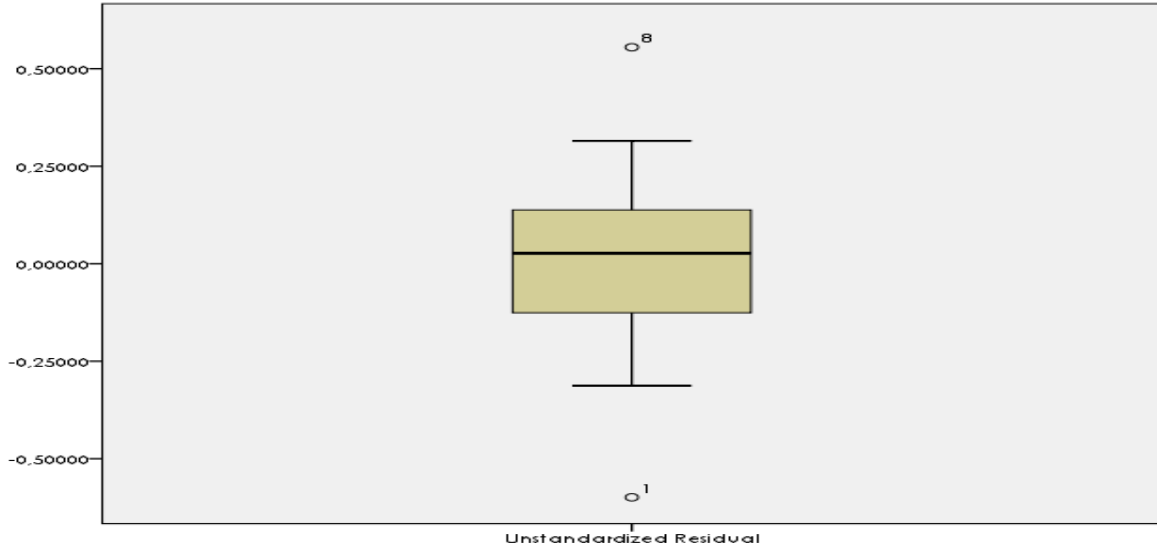
*0,05 anlam düzeyinde anlamlı

Tablo 4. EKK Tekniğine Ait Çoklu Regresyon Analizi için Varyans Analizi Tablosu

Değişim Kaynağı	Kareler Toplamı	s.d.	Kareler Ort.	F	p-değeri	R ²
Model	5,101	3	1,700	22,010	0,000 *	0,805
Hata	1,236	16	0,077			
Toplam	6,338	19				

*0,05 anlam düzeyinde model genel olarak anlamlı

Tablo 3 ve 4 'de çoklu regresyon modeli için sonuçlar görülmektedir. Bağımsız değişkenlerin bağımlı değişkeni açıklama oranının %80,5 olduğu görülmektedir. Parametrelerin %5 anlamlılık düzeyinde anlamlı oldukları Tablo 3'den hareketle söylenebilir (p -değeri $< 0,05$). IQ değerindeki bir birimlik artışın kısa süreli bellek üzerinde yaratacağı ortalama artış miktarının 0,037 birim olduğu aynı zamanda okuma becerisindeki bir birimlik artışın kısa süreli bellek üzerinde yaratacağı ortalama artış miktarının da 0,357 birim olacağı görülmektedir.



Şekil 1. Hata terimleri için aykırı değer grafiği

EKK tekniğinin uygulanabilmesi için, hata terimleri, ε_i ' lerin bağımsız ve özdeş dağılıma sahip ortalaması sıfır, varyansı σ^2 olan normal dağılıma sahip olması şartını sağlaması gerekmektedir. Normal dağılım varsayımının sağlanıp sağlanmadığını görsel olarak kontrol edebilmek için hata terimlerinin aykırı değer grafiğine bakılır. Hata terimleri grafiği Şekil 1'de gösterilmiştir. Şekil 1 incelendiğinde hata terimleri içerisinde aykırı gözlemlerin 1 ve 8. gözlem değerleri (çocuklar) olabileceği ve bu nedenle sağlam regresyon tekniklerinin uygulanması gerektiği söylenebilir.

Sağlam Regresyon teknikleri uygulanan kısa süreli bellek tahminlerine ait denklemler ve denklemler için elde edilen R^2 değerleri Tablo 5 'te gösterilmiştir. Tablo 5' ten de görüleceği gibi bağımsız değişkenlerin bağımlı değişkeni açıklama yüzdesi bakımından en yüksek değeri WLS tekniği vermiştir. Daha sonra sırasıyla EKK, Huber-M, Logistics, Andrews, Tukey, Kantil Q50, R Regresyon, LMS, Kantil Q25, Kantil Q75, Theil ve LTS teknikleri gelmektedir.

Tablo 5. Kısa Süreli Bellek tahmini için Regresyon Analizi Sonuç Tablosu

Teknik	Denklem	R^2
EKK	$\hat{y}_i = -2,987 + 0,267x_1 + 0,037x_2 + 0,359x_3$	0,805
Huber-M	$\hat{y}_i = -2,959 + 0,275x_1 + 0,034x_2 + 0,398x_3$	0,803
Andrews	$\hat{y}_i = -2,977 + 0,265x_1 + 0,034x_2 + 0,412x_3$	0,802
Tukey	$\hat{y}_i = -2,976 + 0,265x_1 + 0,034x_2 + 0,412x_3$	0,802
Logistic	$\hat{y}_i = -3,040 + 0,255x_1 + 0,035x_2 + 0,408x_3$	0,803
LMS	$\hat{y}_i = -5,825 + 0,400x_1 + 0,100x_2 - 0,250x_3$	0,772
LTS	$\hat{y}_i = -4,940 + 0,198x_1 + 0,056x_2 + 0,457x_3$	0,403
Theil	$\hat{y}_i = -3,870 + 0,250x_1 + 0,045x_2 + 0,400x_3$	0,607
WLS	$\hat{y}_i = -2,939 + 0,282x_1 + 0,034x_2 + 0,385x_3$	0,816
R Regression	$\hat{y}_i = -1,894 + 0,180x_1 + 0,030x_2 + 0,375x_3$	0,781
Kantil Q25	$\hat{y}_i = -1,966 + 0,299x_1 + 0,027x_2 + 0,289x_3$	0,680
Kantil Q50	$\hat{y}_i = -2,693 + 0,216x_1 + 0,029x_2 + 0,491x_3$	0,790
Kantil Q75	$\hat{y}_i = -2,71 + 0,32x_1 + 0,02x_2 + 0,55x_3$	0,640

SONUÇLAR ve TARTIŞMA

Araştırmada, çocuklarda kısa süreli belleği etkileyeceği düşünülen yaş, IQ ve okuma becerisi arasındaki model regresyon analizi yardımıyla kurulmuştur. Ancak, regresyon analizinde gerçek veri seti ile çalışılırken en büyük sıkıntı hata terimleri dağılımının normal dağılıma uymamasıdır. Çünkü EKK tekniğinin çıkarsama amaçlı kullanılabilmesini sağlayan en önemli varsayımlarından birisi, hata terimlerinin dağılımının normal dağılım olmasıdır. Bu çalışmada psikoloji alanında regresyon analizi uygulamalarına farklı bir bakış açısı sağlanması için EKK regresyonuna alternative olan sağlam regresyon yöntemleri tanıtılmış ve kısa süreli belleği etkileyen etmenlerin incelenmesinde uygulanmıştır.

EKK tekniği kullanılmış olduğunda, IQ değerindeki bir birimlik artışın kısa süreli bellek üzerinde yaratacağı ortalama artış miktarının 0,037 birim olacağı belirtilmiştir. Hata terimleri dağılımı normal dağılıma sahip olmadığından, EKK tekniği ile ilgili parametre değerleri, güven aralıkları ve hipotez testleri de gerçeği yansıtmayacaktır. Benzer şekilde okuma becerisi ve yaş ile ilgili parametre değerleri, güven aralıkları ve hipotez yanlı olacaktır. Ancak sağlam regresyon teknikleri uygulandıktan sonra bağımsız değişkenlerin bağımlı değişkeni açıklama yüzdesi ile en yüksek değeri WLS tekniği vermiştir.

Hata terimleri dağılımının normal dağılıma uymadığı durumlarda, sağlam regresyon teknikleri uygulanarak bağımsız değişkenlerin bağımlı değişkeni açıklama oranı üzerinde yükselme sağlanacağı görülmüştür. Ayrıca bu çalışma ile eğitim alanında bir çok araştırmada kullanılan regresyon analizine alternatif regresyon teknikleri tanıtılarak, yapılacak çalışmalar için zemin oluşturulmuştur.

Literatür çalışmaları incelendiğinde, yaş ile birlikte artan anlık bellekten söz edilmektedir (Logie, 1986). Çalışmada, kısa süreli anlık belleği etkilediği düşünülen yaş değişkeninin katsayısının anlamlı olarak bulunması ilgili sonucu desteklemektedir.

Literatür çalışmalarına göre; kısa süreli anlık bellek ve IQ birbiri içine geçmiş kavramlar olarak yer almaktadır ve daha çok kısa süreli bellek, IQ testlerinin alt boyutu olarak yer almaktadır (Cohen, 1997; Conway, 2002). Çoğu çalışmada yapısal eşitlik modeli kullanılarak, kurulan modellerde kısa süreli belleğin IQ'yu etkilediği varsayılmıştır (Colom et al., 2005; Ackerman et al., 2005). Yapısal eşitlik modellerinin avantajları olmasına rağmen, bir takım dezavantajları da vardır. Yapısal Eşitlik Modellerinde, modeli etkileyen bir çok değişken içsel ve dışsal değişkenler olarak birbirini etkilemektedir. Ackerman et al. (2005), çalışmasında bu durumu belirtmiş ve IQ ile kısa süreli anlık bellek ve çalışma belleği arasındaki ilişkiyi belirlerken, IQ ve kısa süreli anlık belleğin benzerlik ve farklılıklarını ortaya koymuştur. Daha açık bir ifade ile, IQ ve kısa süreli anlık bellek arasında ilişki aranırken, bu değişkenlerin bir biri yerine kullanılabileceğini ifade etmişlerdir.

Bu çalışmada ise, farklı bir durum olarak bu durum alınmış, IQ'nun kısa süreli belleği etkilediği düşünülmüştür. Elde edilen sonuçlar da bunu desteklemektedir. WLS tekniğine göre, IQ değerindeki bir birimlik artışın kısa süreli bellek üzerinde yaratacağı ortalama artış miktarının 0,034 birim olduğu söylenebilir. Kısaca, IQ arttığı zaman, kısa süreli bellek değerinde artış meydana gelmektedir. Hem EKK, hem de diğer sağlam regresyon teknikleri ile bu ortaya konulmuştur.

Okuma becerisi ve kısa süreli bellek kavramları arasındaki ilişkinin ele alındığı çalışmalar genellikle, çalışma belleği ve kısa süreli bellek süreçleri ile okuduğunu anlama arasındaki ilişkiyi inceleyen çalışmalar olduğundan bahsedilir (Cain et al., 2004) Çalışmada elde edilen sonuçlara göre, WLS tekniğine göre, okuma becerisi değerindeki bir birimlik artışın kısa süreli bellek üzerinde yaratacağı ortalama artış miktarının 0,385 birim olduğu söylenebilir. Daha açık bir ifade ile, okuma becerisindeki artış, kısa süreli bellek değerinde artışı meydana getirmektedir. Okuduğunu anlama, kısa süreli anlık bellek üzerinde etkili bir değişkendir.

Bu çalışmada, kısa süreli anlık belleğin önemi vurgulanmaya çalışılmıştır. Kısa süreli anlık bellek değişkenini etkileyen değişkenler olarak, yaş, IQ ve okuma becerisi ele alınmış ve modellenmeye çalışılmıştır. Elde edilen bulgulara göre, yaş, IQ ve okuma becerisindeki artış, kısa süreli bellek üzerinde artışa neden olmaktadır. Bu durumda, eğitim ile ilgilenen öğretilere, öğrencilerinin okuma

becerilerini arttırmaya yönelik çalışmalar yapmalarını tavsiye edilir. Çünkü okuma becerisindeki artış, kısa süreli bellek üzerinde etkili olduğu söylenebilir.

Kısa süreli anlık bellek değişkeni için tahminlerin yapıldığı bu çalışmada, genel IQ değerleri kullanılarak çalışma yürütülmüştür. IQ değerleri için geliştirilen Wechsler (1974) 'in ölçeğinde sözel ve performans IQ değerleri ayrı ayrı kullanılarak bu çalışma ile veya literatürde bulunan diğer çalışmalar ile ilgili karşılaştırmalar yapılabilir. Aynı şekilde okuma ve yazma becerileri ölçeğinde kullanılan yazma becerileri göz önüne alınarak ilgili çalışma devam ettirilebilir. Ayrıca psikoloji yada eğitim alanında yürütülen çalışmalarda klasik EKK regresyonuna alternatif olarak tanıtılan sağlam regresyon analizi yöntemlerinin kullanılabilmesi için örnek bir çalışmadır. Bu çalışma, araştırmacılara yol gösterici bir çalışma olabilir.

KAYNAKÇA

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs?. *Psychological Bulletin*, Vol: 131, No: 1, 30-60.
- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Akoğlu, G., & Turan, F. (2012). Eğitsel Müdahale Yaklaşımı Olarak SesBilgisel Farkındalık: Zihinsel Engelli Çocuklarda Okuma Becerilerine Etkileri. *H.Ü. Eğitim Fakültesi Dergisi*, 42, 11-22.
- Alp, İ., & Özdemir, B. (2007). Çocuklarda Akıcı Zekanın Bilgi İşleme Hızı, Kısa Süreli Bellek ve Çalışma Belleği Kapasitesi İle İlişkisi. *Türk Psikoloji Dergisi*, 22(60), 1-15.
- Andrews, D. (1974). A robust method for linear regression. *Technometrics*, 16, 523-531.
- Babacan, T. (2012). *Sınıf Öğretmeni Adaylarının ÜstBilişsel Okuma Stratejileri ile Çoklu Zeka Alanları Arasındaki İlişkinin İncelenmesi*. Cumhuriyet Üniversitesi, Yüksek Lisans Tezi, Sivas.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and Motivation*, 8, 47-89.
- Banikowski, A.K., & Mehring, T.A. (1999). Strategies to Enhance Memory Based on Brain Research. *Focus on Exceptional Children*, 32, 1-16.
- Birkes, D., & Dodge, Y. (1993). *Alternative Methods of Regression*. NY: Wiley.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, Vol: 96, No:1, 31-42.
- Chen, C., & Wei, Y. (2005). Computational Issues for Quantile Regression. *Special Issue on Quantile Regression and Related Methods*, Volume 67, Part 2, 399-417.
- Cohen, J. M., (1997). *Children's Memory Scale*. Manual, San Antonio: The Psychological Corporation, Harcourt Brace & Company.
- Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, 33, 623-642.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19: 450-466.
- David, A. (1981). *Introduction to Statistics*. St. Paul: West Publishing Company.
- Doruk, E., & Öngören, H. (2005). İnsan İlişkilerinde Ve İş Yaşamında Duygusal Zeka. *İstanbul Üniversitesi İletişim Fakültesi Dergisi*, 22(1), 157-163.
- Ergül, B. (2006). *Robust Regression ve Uygulamaları*. Eskişehir Osmangazi Üniversitesi, Yüksek Lisans Tezi, Eskişehir.
- Fer, S. (2004). Qualitative Evaluation of Emotional Intelligence In-Service Program for Secondary School Teachers. *The Qualitative Report*, 9(4), 562-588.
- Gauss, C. (1809). *Theoria Motus Corporum Coelestium* (Reprinted 1963 by Dover). Hamburg: Perthes et Esser.
- Heikkila, J. (2006). *Graudate course on Advanced statistical signal processing*. Retrieved from <http://www.ee.oulu.fi/~jth/robust.pdf>.
- Hodges, J., & Lehmann E. (1963). Estimates of location based on rank tests. *Ann. Math. Stat.*, 34, 598-611.
- Huber, P. (1973). Robust regression: Asymptotics, conjections and monte carlo. *Ann. Stat.*, 1, 799-821.
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *The Econometric Society*, Vol. 46, No.1, 33-50.

- Korkmaz, Ö., & Mahiroğlu, A. (2007). Beyin, Bellek ve Öğrenme. *Kastamonu Eğitim Dergisi*, 15(1), 93-104.
- Köksal, M.S., & Boran, A.İ. (2015). Üstün Yetenekli Öğrencilerin IQ Puanlarının Anne-Baba Eğitimi Ve Aile Gelirleri Değişkenleri Açısından Karşılaştırılması. *ESOGÜ SBF Dergisi*, 16(1), 109-121.
- Kuo, C., & Yu, S. (2013). The non-uniform pricing effect of employee stock options using quantile regression. *North American Journal of Economics and Finance*, 26, 400-415.
- Kurtoğlu, F. (2011). *Quantile Regresyon: Teorisi Ve Uygulamaları*. Çukurova Üniversitesi, Yüksek Lisans Tezi, Adana.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, 14, 389-433.
- Legendre, A. (1805). *Nouvelles Methodes pour la Determination des Orbites des Cometes* (Reprinted 1959 by Dover). Paris: Courcier.
- Logie, R. H. (1986). Visuo-Spatial Processing In Working Memory. *The Quarterly Journal of Experimental Psychology*, 38(2), 229-247.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, 81- 97.
- Mosteller, F., & Tukey, J. W. (1977). *Data Analysis and Regression*. Philippines: Addison-Wesley Publishing Company.
- Nevitt, J., & Tam, H. (1998). A comparison of robust and nonparametric estimators under the simple linear regression model. *Multiple Linear Regression Viewpoints*, 25, 54-69.
- Özkan, Y., & Doğan, B. (2013). İlköğretim 8. Sınıf Öğrencilerinin Okuma Becerilerinin Kestirilmesinde Etkili Olan Değişkenlerin Belirlenmesi. *JASSS*, 6(4), 667-680.
- Öztürk, L. (2003). *Doğrusal Regresyonda Sağlam Kestirim Yöntemleri ve Karşılaştırılmaları*. Mimar Sinan Üniversitesi, Doktora Tezi, İstanbul.
- Özyürek, A. (2009). *Okul Öncesi Eğitim Kurumuna Devam Eden Altı Yaş Grubu Çocukların Bellek Gelişimine Bellek Eğitiminin Etkisinin İncelenmesi İncelenmesi (Basılmamış Doktora Tezi)*. Gazi Üni. Eğitim Bilimleri Enstitüsü.
- Peterson & Peterson (1959). *Duration of Short Term Memory*. Retrieved from <http://www2.qeliz.ac.uk/psychology/P%20and%20P1959.htm>
- Rousseuw, P. J. (1984). Least median of squares regression. *J. Am. Stat. Assoc.*, 79, 871-880.
- Rousseuw, P. J., & Leroy, A. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons Inc.
- Saeed, T. (2011). *A Comparative Study of Working Memory in Children with Neuro developmental Disorders*. PhD Thesis, National University of Ireland Maynooth.
- Sayar, F., & Turan, F. (2012). Okuma Gelişiminde Üst Dil Farkındalığı, Sesbilgisel Süreçler ve Bellek Süreçlerinin Etkisi: Kısa Süreli Bellek ve Çalışma Belleği. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 13(2) 49-64.
- Scarborough, H.S., & Brady, S.A. (2002). Toward a common terminology for talking about speech and reading: A glossary of the “phon” words and some related terms. *Journal of Literacy Research*, 34, 299-334.
- Seamon, J. G., & Kenrick, D. T. (1994). *Psychology*. Englewood Cliffs, NJ: Prentice Hall.
- Senemoğlu, N. (2002). *Gelişim, Öğrenme ve Öğretim*, Gazi Kitabevi, Ankara.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86, 221-234.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.
- Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Ned. Akad. Wetensch Proc.* 53, 386-392.
- Tukey, J. (1970). *Exploratory Data Analysis*. MA: Addison-Wesley.
- Wechsler, D. (1974). *WISC-R manual for the Wechsler Intelligence Scale for Children-Revised*. New York. Psychological Corporation.
- Yale, C., & Forsythe, A. (1976). Winsorized regression. *Technometrics*, 18(3), 291-312.
- Yangın, B. (2002). *Kuramdan uygulamaya türkçe öğretimi*. Ankara: Dersal Yayıncılık.
- Zhou, J. (1992). *Bounded Influence R-Estimators In The Linear Model*, Thesis of Master of Science, University of Alberta.
- Zhou, W., & Serfling, R. (2006). *Multivariate spatial U-quantiles: a BahadurKiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator*. Manuscript

EXTENDED ABSTRACT

The term, short-term memory is often used to refer to a concept quite similar to working memory. Short-term memory is the type of memory that we use when we wish to retain information for a short time to think about it. The short-term store has a working memory component, a sort of mental workspace in the mind that is used to manipulate information in consciousness. Working memory is a more complex construct than short-term memory. Repeating the information mentally makes it remain in memory for a long time. Studies investigating the relationship between short-term memory and reading skills that are carried out to examine the relationship between short-term memory processes and reading comprehension. In this study information coming to short-term memory and the factors affecting operation of short term memory are investigated with regression model. The aim of the research is to examine the factors (age, IQ and reading skills) that are expected the have an effect on short-term memory in children through regression analysis. One of the assumptions of regression analysis is the normal distribution of the error term. In this study, the error term was not normally distributed, therefore robust regression techniques were applied. Robust Regression techniques are M-Regression, Winsorized Regression, Trimmed Regression, Median Squares Regression, R Regression and Quantiles Regression. Also, for each technique; coefficient of determination is determined.

Introduction

Intelligence; how quickly people can receive information and expertise on how the tasks are not doing well. Intelligence can be captured by a number obtained by an IQ test. The short-term memory is a sort of mental workspace in the mind that is used to manipulate information in consciousness. Reading skills are generally closely associated with matching the spoken words.

These studies between short-term memory and reading skills and intelligence in Turkey that began to be made after 10 years.

The purpose of the research is to examine the factors (age, IQ and reading skills) that are expected the have an effect on short-term memory in children through regression analysis.

Method

Regression analysis is a statistical technique for parameter estimation to determine the relationships between variables. There are some basic assumptions made in order to estimate the parameters of the OLS technique in Regression Analysis. These assumptions are generally related to the error term.

However, when the error term distribution is not normally distributed, robust regression techniques are applied. Robust regression techniques are not affected by outliers. It disrupts the normality of observation is the most important factor, which is considered as the error terms have a great amount of other error term value, are outliers.

Robust Regression techniques are M-Regression, Winsorized Regression, Trimmed Regression, Median Squares Regression, R Regression and Quantiles Regression.

Results and Discussion

In this study, which is expected to affect short-term memory in children, the relationship between age, IQ and reading skills is to establish through regression analysis.

However, when working with real data sets of regression analyzes, the biggest problem is failure to comply with the normal distribution of the error terms of distribution.

OLS technique is used, a unit increase in the value of the average IQ increase will have on the amount of short-term memory is specified to be 0,037 units. After applying robust regression techniques, the highest R^2 value is given by the WLS technique. According to WLS technique, a unit

increase in the value of the average IQ increase will have on the amount of short-term memory it can be said to be 0,034 units.