
Eđitimde ve Psikolojide Ölçme ve Deęerlendirme Dergisi

Journal of Measurement
and Evaluation in
Education and Psychology

ISSN:1309-6575

Yaz 2017
Summer 2017

Cilt: 8- Sayı: 2
Volume: 8- 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
Dr. Sakine GÖÇER ŞAHİN

Assistant Editor

Assist. Prof. Dr. Kübra ATALAY KABASAKAL
Dr. Sakine GÖÇER ŞAHİN

Genel Sekreter

Doç. Dr. Tülin ACAR

Secretary

Doç. Dr. Tülin ACAR

Yayın Kurulu

Prof. Dr. Terry A. ACKERMAN
Prof. Dr. Cindy M. WALKER
Doç. Dr. Cem Oktay Güzeller
Doç. Dr. Neşe GÜLER
Doç. Dr. Hakan Yavuz ATAR
Doç. Dr. Oğuz Tahsin BAŞOKÇU
Yrd. Doç. Dr. Hamide Deniz GÜLLEROĞLU
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Yrd. Doç. Dr. Okan BULUT
Yrd. Doç. Dr. N. Bilge BAŞUSTA
Yrd. Doç. Dr. Derya ÇAKICI ESER
Yrd. Doç. Dr. Mehmet KAPLAN
Dr. Nagihan BOZTUNÇ ÖZTÜRK

Editorial Board

Prof. Dr. Terry A. ACKERMAN
Prof. Dr. Cindy M. WALKER
Assoc. Prof. Dr. Cem Oktay GÜZELLER
Assoc. Prof. Dr. Neşe GÜLER
Assoc. Prof. Dr. Hakan Yavuz ATAR
Assoc. Prof. Dr. Oğuz Tahsin BAŞOKÇU
Assist. Prof. Dr. Hamide Deniz GÜLLEROĞLU
Assist. Prof. Dr. Derya ÇOBANOĞLU AKTAN
Assist. Prof. Dr. Okan BULUT
Assist. Prof. Dr. N. Bilge BAŞUSTA
Assist. Prof. Dr. Derya ÇAKICI ESER
Assist. Prof. Dr. Mehmet KAPLAN
Dr. Nagihan BOZTUNÇ ÖZTÜRK

Dil Editörü

Doç. Dr. Burcu ATAR
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Dr. Gonca YEŞİLTAŞ

Language Reviewer

Assoc. Prof. Dr. Burcu ATAR
Assist. Prof. Dr. Derya ÇOBANOĞLU AKTAN
Dr. Gonca YEŞİLTAŞ

Sekreteryaya

Arş. Gör. İbrahim UYSAL
Arş. Gör. Seçil UĞURLU
Arş. Gör. Nermin KIBRISLIOĞLU UYSAL

Secretarait

Res. Assist. İbrahim UYSAL
Res. Assist. Seçil UĞURLU
Res. Assist. Nermin KIBRISLIOĞLU UYSAL

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

Journal of Measurement and Evaluation in
Education and Psychology (EPOD) is a national
refereed journal that is published four 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.o

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)
Akif AVCU (Marmara Üni.)
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 Üni.)
Burcu ATAR (Hacettepe Üni.)
Burhanettin ÖZDEMİR (Siirt Üni.)
Beyza AKSU DÜNYA (Illinois 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.)
Eren Halil Özberk (Hacettepe Üni.)
Ergül DEMİR (Ankara Üni.)
Esin TEZBAŞARAN (İstanbul Üni.)
Esin YILMAZ KOĞAR (Hacettepe Üni.)
Esra Eminoğlu ÖZMERCAN (MEB)
Evrin ÇETİNKAYA YILDIZ (Erciyes Üni.)
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.)
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.)
Levent YAKAR (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.)
Nükheth DEMİRTAŞLI (Ankara Üni.)
Okan BULUT (University of Alberta)
Onur ÖZMEN (TED Üniversitesi)
Ömer KUTLU (Ankara Üni.)
Ömür Kaya KALKAN (Hacettepe Üni.)
Özge BIKMAZ BİLGİN (Adnan Menderes 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 KİLMEN (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

Sınıf Etkinliklerim Ölçeği'nin (SEÖ) Türk Kültürüne Uyarlanması: Geçerlik ve Güvenirlik Çalışması Adaptation of My Classroom Activities Scale to Turkish Culture: Validity and Reliability Study Kaan Zülfikar DENİZ, Adile Gülşah SARANLI	169
Türkiye'de Eğitim Alanında Yayımlanan Ölçek Geliştirme Çalışmalarının Uygunluğunun Çok Yüzeyle Rasch Modeli ile İncelenmesi Investigation of Scale Development Studies Conducted in Educational Sciences Published in Turkey by Many-Faceted Rasch Model Güliden KAYA UYANIK, Neşe GÜLER, Gülşen TAŞDELEN TEKER, Süleyman DEMİR	183
Kayıp Veriyle Baş Etme Yöntemlerinin Model Veri Uyumu ve Madde Model Uyumuna Etkisi The Effect of Missing Data Techniques on Model Fit and Item Model Fit Duygu KOÇAK, Ömay ÇOKLUK BÖKEOĞLU	200
Lise Öğrencilerinin Meslek Seçimlerini Etkileyen Faktörlerin İkili Karşılaştırmalar Yöntemiyle İncelenmesi Investigation of the Factors Affecting Occupation Choices of High School Students with Paired Comparison Method Funda NALBANTOĞLU YILMAZ	224
Örtük Markov Model: Latent GOLD 5.1'e Genel Bakış Latent Markov Model: An Overview of Latent GOLD 5.1 Duygu GÜNGÖR, Selva ÜLBE	237

Sınıf Etkinliklerim Ölçeği'nin (SEÖ) Türk Kültürüne Uyarlanması: Geçerlik ve Güvenirlik Çalışması*

Adaptation of My Classroom Activities Scale to Turkish Culture: Validity and Reliability Study

Kaan Zülfikar DENİZ **

Adile Gülşah SARANLI ***

Öz

Öğrencilerin sınıf etkinliklerine ilgi duymaları, etkinlik konularından zevk almaları, etkinliğin konusuyla ilgili seçim yapabilmeleri ve etkinlikler sırasında sınırlarını zorlayabilmeleri, üst düzey öğrenmelerin altında yatan temel unsurlardandır. Eğitsel etkinliğin ilgi çekici, zevkli, seçim yapabilmeye ve sınırları zorlamaya olanak tanıyan özelliklerinin olması, aynı zamanda 21. Yüzyılın eğitim sistemindeki neredeyse tüm eğitsel içerik, süreç ve üründe bulunmasını istediğimiz özelliklerdendir. Bu özelliklerin ölçülmesi de dolayısıyla büyük önem taşımaktadır. Bu çalışmanın amacı, Gentry ve Gable (2001) tarafından Amerika Birleşik Devletleri'nde geliştirilen; geliştirildiği kültürdeki farklı öğrenci gruplarına uygulanan ve ayrıca Kore, Çin ve Arap dillerine çevrilen ve Güney Kore kültürüne de uyarlama çalışması yapılmış olan Sınıf Etkinliklerim Ölçeği'nin Türk kültürüne uyarlamasını gerçekleştirmektir. Bu doğrultuda araştırma, 2015-2016 eğitim-öğretim yılında 3., 4., 5., 6., 7. ve 8. sınıflarda öğrenim gören toplam 214 öğrenciden elde edilen veri ile yürütülmüştür. Ölçeğin geçerlik çalışması kapsamında orijinal ölçek geliştirme çalışmasında elde edilen faktör yapısı Doğrulamalı Faktör Analizi (DFA) ile test edilmiştir. Ayrıca geçerlik çalışması olarak madde-toplam korelasyonları ve boyutlar arası korelasyon analizleri gerçekleştirilmiştir. Ölçeğin güvenirlilik çalışmasında Cronbach alfa güvenirlilik katsayısı kestirilmiştir. Elde edilen bulgulara göre ölçeğin faktör yapısı orijinal ölçek geliştirme çalışmasına paralel olarak doğrulanmıştır. Ayrıca Cronbach alfa değerlerinin 0,82 ile 0,90 arasında olduğu görülmektedir. Yapılan bu çalışmalar ile ölçeğin, Türk kültüründe kullanımı için gerekli geçerlik ve güvenirlilik kanıtları sağlanmış ve ölçek farklı çalışmalar için kullanılmak üzere alana kazandırılmıştır.

Anahtar Kelimeler: Sınıf etkinlikleri, ölçek uyarlama, doğrulamalı faktör analizi

Abstract

Student interest in class activities, their enjoyment of activity topics, their ability to make choices about the activity topics, and opportunities for students to challenge themselves during activities are among basic components that support their higher level learning. Properties of educational activities that make them interesting, enjoyable, and challenging while allowing students with choices are also among properties that are known to be necessary in all educational content, processes, and products within educational systems of the 21st century. Consequently, measuring these properties is also of great importance. The goal of this study is to perform the Turkish adaptation of the My Class Activities Scale, developed by Gentry and Gable (2001) in the United States and subsequently adapted to the Korean, Chinese, and Arabic languages. To this end, data was collected from 214 students attending 3rd, 4th, 5th, 6th, 7th, and 8th grades during the 2015-2016 academic year. As part of the validity study for the scale, the factor structure obtained from the original development of the scale was tested using the Confirmatory Factor Analysis (CFA) method. Moreover, item-total correlation and inter-dimensional correlation analyses were also performed as part of the validity study. In studying the reliability of the scale, the Cronbach-Alpha reliability coefficients were estimated (Cronbach alpha values ranged between 0.82-0.90). Based on the results, the factor structure of the scale was verified in parallel with the original development work for the scale. In conclusion, the validity and reliability of using the scale in Turkey was established, contributing a new scale adaptation to the Turkish literature for use in different studies.

Keywords: Classroom activities, scale adaptation, confirmatory factor analysis

*Bu araştırma 115K814 kodlu TÜBİTAK projesi tarafından desteklenmektedir.

** Doç.Dr., Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, zlfkrdnz@yahoo.com

*** Yrd. Doç. Dr., TED Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, gsaranli@gmail.com

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Cilt 8, Sayı 2, Haziran 2017, 169-182.

Journal of Measurement and Evaluation in Education and Psychology, Vol. 8, Issue 2, June 2017, 169-182.

Geliş Tarihi: 12.02.2017

Kabul Tarihi: 07.04.2017

DOI: 10.21031/epod.291825

GİRİŞ

Sınıf ortamında öğrencileri motive etmek ve sınıf disiplinini sağlayabilmek öğretmenler için önemli bir güçlük oluşturmaktadır. Sınıf disiplininin bozulması durumu başta öğretmenin, sonrasında ise öğrencilerin genel motivasyonlarını olumsuz etkiler. Motive olmayan öğrenciler ise genellikle düşük başarı göstermeye eğilimlidirler (Gentry & Gable, 2001) ve sınıftaki ortamı bozarak kendi çalışmalarına odaklanmaya çalışan diğer öğrencileri olumsuz yönde (Gentry, Gable, 2001; Gentry, Rizza & Owen, 2002) etkileyebilirler. Öğrencilerin sınıf ortamında gerçekleştirilen etkinliklere karşı yönlendirilebilmeleri için bu etkinliklerin içeriğinin, uygulanmasındaki yöntemlerin ve ortaya çıkarılan ürünlerin ilgi çekici, zevkli, seçim yapabilmeye olanak sağlayan ve sınırlarını zorlayıcı niteliklerde olması çok önemlidir. Bu nitelikler öğrencilerin eğitim programını yaşamdan ayrı ve yapay bir biçimde algılamayıp, onunla bütünleşmelerini ve eğitim programında onları motive eden konuları bulmalarını sağlayacak şekilde dizayn edildiğinde hem akademik hem de Pratik hayatta en yüksek başarıya ulaşılmaktadır. Öğrencilerin sınıfta gerçekleştirdikleri etkinlikleri nasıl algıladıklarına ve onları öğrenme yolculuğunda nelerin motive ettiğine ilişkin bilginin olması, hem eğitimci hem de araştırmacıların, öğrencilerin öğrenmeye motive olmalarında hangi önemli basamakların yer teşkil ettiğini anlamalarını (Gentry, Gable & Rizza, 2002; Middleton, 1995; Middleton, Littlefield & Lehrer, 1992) ve daha da ötesi başarısızlık problemini önlemeyi (Matthews & McBee, 2007) sağlayacak önemli adımlardan olacaktır. .

Alan yazındaki pek çok köklü eğitim kuramcısı ve araştırmacı, (Dewey, 1916; James, 1890; Renzulli, 1978; Ward, 1980; Whitehead, 1929) öğrencilerin ilgi alanlarının, onları öğrenmeye teşvik etmek için kullanılmasını önermişlerdir. Schiefele (1991) 'ilgi'yi belirli içerik alanlarında performans ve motivasyonu etkileyen yönlendirici bir güç olarak tarif etmiştir. Benzer bir şekilde, Whitehead (1929) ilgi olmadan zihinsel gelişim olamayacağını belirtmiştir. Buna ek olarak, üstün yetenekli çocukların eğitimini destekleyen araştırmacılar da eğitim programlarının belirlenmesinde ilgi kavramının merkezde yer alması gerektiğini savunmuşlardır (Gallagher, 1985; Maker, 1982; Parke, 1989; Passow, 1982; Renzulli, 1994). Ayrıca Good ve Brophy (1987) de tüm öğrencilere okul ortamında ilgilerini keşfederek geliştirmek için fırsat sunulması gerektiğini belirtmiştir. Hootstein (1994) öğrencileri motive etmek için öğrenmenin, öğrencilerin ihtiyaç, ilgi, endişe ve tecrübeleri ile ilişkilendirilmesini ve onların ilgi duydukları konular üzerine çalışmalarının teşvik edilmesini önermiştir. Kısacası, ilgi motivasyona, motivasyon da öğrenmeye bağlıdır ve bunun bir sonucu olarak ilgilerin incelenmesi ve ölçülmesinin öğretim ve öğrenmenin iyileştirilmesi için önemli öngörüler ve fikirler sağlayacağı ortadadır (Deci & Ryan, 1985; Schiefele, 1991; Tobias, 1994).

Öğrenme ve motivasyona etki eden bir başka faktör olan sınırları zorlama (challenge) kavramının değerlendirilmesinde, Vygotsky'nin (1962) çocukların kendi becerilerinin az da olsa ötesinde olan etkinlikleri tercih etmeleri ve böylece de zihinsel gelişimin zor görevler gerektirdiği gözlemi önemli bir yer taşımaktadır. Öğrencilerin sınırlarını zorlayıcı, daha iyiye ulaşmaları için onları teşvik edici görevlere ve eğitsel etkinliklere duyulan ihtiyaç ortadadır. Öğrenciler için onların potansiyellerini ortaya çıkartıcı ve onların sınırlarını zorlayan bir eğitim programı oluşturulmasının öğrenme fırsatlarını daha üst seviye getireceği söylenebilir. Buna rağmen, birçok sınıfta öğrenciler için sınırlarını zorlamalarını, daha iyiyi yapmalarını teşvik eden etkinliklere rastlanmamaktadır ve bu da sıkılmış ve bıkmış, bunun sonucunda da potansiyellerine ulaşamayan öğrenciler yaratmaktadır (Archambault vd. 1993; Feldhusen & Kroll, 1991; Goodlad, 1984; Reis vd., 1993; Westberg, Archambault, Dobyns & Salvin, 1993). Eccles ve Midgley (1989) öğretmenlere gençlerin akademik performansları ile ilgili yüksek beklentilere sahip olmalarını ve onları buna uygun şekilde zorlamalarını önermiştir. Ayrıca Clifford (1990) da başarının motivasyon ile ilgili olduğunu ve öğrencilerin sınırlarını zorlayıcı görevlerin varlığı ile motivasyonun ve başarının elde edilebileceğini belirtmiştir. Üstün yetenekliler alanındaki araştırmacılar ve program geliştirme uzmanları eğitimcilerin pek çok farklı yöntem kullanarak öğrencilerin yeteneklerinin sınırlarını zorlayacak ve bu yolla onları geliştirecek çeşitli içerikler yaratabileceklerini belirtmişlerdir. Bu yollardan bazıları; etkinlikler sırasında karmaşık içeriklere odaklanmak; etkinliklere ileri düşünme becerilerini entegre etmek; etkinlikler sırasında orjinal yöntemler kullanıp gerçek durumlar ve problemler için ürünler ve

yöntemler geliştirmek ve son olarak da müfredatı sıkılaştırmak, hızlandırmak ve farklılaştırmaktır. (Bloom, 1985; Reis vd., 1993; Renzulli, 1994; Schlichter, 1986; Tomlinson, 1992, 1999; Treffinger, 1986). Doğru zorluğa sahip bir müfredat ve etkili öğretim yöntemleri geliştirerek, eğitimciler tüm başarı seviyelerindeki öğrencilere yüksek kalitede bir eğitim verebilirler (Bloom, 1985; Shore, Cornell, Robinson & Ward, 1991; Vygotsky, 1962).

Renzulli (1994), öğrencilerin derinden ilgi duydukları ve kendileri yapmayı seçtikleri projeler ile ilgili çalıştıklarında daha anlamlı öğrenmeler gerçekleştiğini belirtmiştir. Çeşitli araştırmacılar (Bloom, 1985; Dewey, 1913, 1916; Gardner, 1991; Goodlad, 1984; Renzulli & Reis, 1997; Shore vd., 1991; Wang & Lindvall, 1984) kendi eğitimleri konusunda öğrencilere seçenek vermenin öğrenmeyi teşvik eden motivasyonel bir araç olduğunu belirlemişlerdir. Pintrich ve DeGrot (1990) ortaokul öğrencileri için kendi kontrollerinde olan ve seçim olanakları içeren öğrenmenin daha yüksek etkililik ve akademik performansa yol açtığını bulmuşlardır. Buna ek olarak, Faske ve Grubb (1997) öğrencilerin, öğretmenlerinin öğrenci merkezli yöntemler kullandıklarını algıladıkları durumlarda başarıda, kendine yeterli ve etkililikte pozitif etkiler görüldüğünü bulmuşlardır. Deci (1995) öğrencilerin kendi kendilerine yetebilmeleri yolunda onlara seçenekler sunulmasının önemine değinmiştir. Seçeneklerin varlığı, öğrencilerin kendi başlarına öğrenmelerini destekleyici bir unsurdur. Buna uyumlu olarak hem Deci (1995) hem de Glasser (1996) sınıf ortamında seçenekler sunmanın öğretmenlerin güçlerini öğrenciler ile paylaşmasını gerektirdiğini ve böylece de öğrencilerin kendi öğrenmelerini sahiplendiklerini ve bu doğrultuda kararlar almalarını sağladığını gözlemlemişlerdir. Benzer bir şekilde, Kerka (1994) hedefler, amaçlar, katılım şekli, içerik, yöntem ve değerlendirme üzerinde söz sahibi olmanın öz düzenleyici öğrenme için önemli olduğunu iddia etmiştir. Bandura (1997) da kendi amaçlarını belirlemesine izin verilen öğrencinin onlara ulaşmada da daha aktif olacağını net bir şekilde belirtmiştir. Birey tarafından belirlenen, serbestçe seçilen ve birey tarafından kontrol edilen öz düzenleyici davranışlar, sonuçta üstlenilen göreve karşı yüksek ilgi, yaratıcılık, bilişsel esneklik, olumlu duygular ve sebat sağlayacaktır (Betts & Kercher, 1999; Deci & Ryan, 1985). Kısaca, öğrencilere eğitim etkinlikleri sırasında seçim yapabilmeye şansı vermek ilgi, başarı ve sahiplik hissini artırmaya fayda sağlamaktadır.

Eğitsel etkinliklerden zevk alınmasının öğrenme sürecinde motive edici bir faktör olduğu gözardı edilemez. Lepper ve Chabay (1985) öğrencilerin kontrol duygularını artırmak, sınırlarını zorlayıcı etkinlikler sunmak, merak uyandırmak ve etkinliklerin sahip oldukları fonksiyonları ön plana çıkarmak gibi yollarla sınıfta motivasyonu artırmanın mümkün olduğunu söylemiştir. Bu yolla aynı zamanda öğrenmeyi hem eğlenceli hem de verimli yapmanın mümkün olduğundan bahsetmiştir. Ayrıca, diğer araştırmacılar da (Csikszentmihalyi, 1990; Dewey, 1916; Renzulli, 1994; Schiefele, 1991) etkili eğitim uygulamaları için çekici ve eğlenceli öğrenme tecrübeleri sunmanın gerekli olduğunu söylemişlerdir. Renzulli (1994) en iyi öğrenmenin çocukların yaptıklarından keyif aldıkları zaman gerçekleştiğini gözlemlemiştir. Buna ek olarak, yüksek yaratıcılığa sahip üretken kişilerin yaptıklarından zevk aldıkları zaman en yüksek performans seviyelerine ulaştıklarını da söylemiştir. Öğrenmenin keyifli ve ilgi çekici olması kavramına bağlı olarak, öğrencilerin öğretmekten zevk alan (Csikszentmihalyi & McCormack, 1986) ve öğrettikleri konulara tutkuyla bağlı olan (Renzulli, 1988) öğretmenlerden daha verimli bir şekilde öğrenebildikleri düşüncesi de bulunmaktadır. Günlük sınıf etkinliklerinden zevk alınmasını sağlayacak öğeler dahil edildiğinde okulun hem öğrenciler hem de öğretmenler için daha iyi bir öğrenme ortamı haline gelmesi mümkün olacaktır.

Buraya kadar aktarılan bilgilerden de görüldüğü üzere, ilgi, konudan zevk alma, sınırları zorlama ve seçeneklerin olması gibi faktörler, hem öğrenme ortamının kalitesini hem de öğrenenin motivasyonunu artıran unsurlardır. Gentry ve Gable (2001) Sınıf Etkinliklerim Ölçeği'nde tüm bu faktörleri bir arada ölçebilecek bir araç ortaya koymuştur. Türkiyede'ki alan yazın incelendiğinde bu dört önemli boyutu birden ölçen bir ölçme aracına rastlanmamıştır. Yurtdışı alan yazında ölçeğin geliştirildiği zamandan beri öğrenme ortamlarındaki bu dört farklı boyutu ölçmek amacıyla pek çok farklı araştırmada da kullanılmış olduğu görülmektedir (Aryan & Shahrokhi, 2015; Gentry & Gable, 2001; Gentry, Gable & Rizza, 2002; Gentry, Gable & Springer, 2000; Gentry, Rizza & Gable, 2001; Gentry, Rizza & Owen, 2002; Pereira, Peters & Gentry, 2010; Yang, Gentry & Choi, 2012). Gentry ve Gable (2001) tarafından Amerika'da geliştirilen bu ölçek, geliştirildiği kültürdeki farklı öğrenci

gruplarında da uygulanmış (Pereira, Peters & Gentry, 2010), aynı zamanda Güney Kore (Yang, Gentry & Choi, 2012) kültürüne uyarlama çalışması yapılmış, Arapça (Aryan & Shahrokhi, 2015) ve Çince'ye (Yang & Gentry, 2011) ise çevirisi gerçekleştirilmiştir. Oldukça yaygın bir kullanım alanına sahip olan bu ölçeğin alanyazındaki önemli bir boşluğu dolduracağı ve çeşitli disiplinlerde araştırmalar yürütmekte olan pek çok araştırmacının kullanacağı düşünülmektedir.

Araştırmanın Amacı

Bu çalışmada öğrencilerin sınıf etkinliklerine yönelik ilgi duyma, etkinlik konularından zevk alma, etkinliğin konusuyla ilgili seçim yapabilme ve etkinlik sırasında sınırlarını zorlamalarına ilişkin algılarını ölçen “Sınıf Etkinliklerim Ölçeği”nin geçerlik ve güvenirlik çalışmasının yapılarak Türk kültürüne uyarlanması amaçlanmıştır.

YÖNTEM

Araştırmanın Modeli

Araştırma tarama modelinde betimsel bir araştırmadır. Karasar (2016)'a göre, betimsel çalışmalar; olayların, objelerin, varlıkların, kurumların, grupların ve çeşitli alanların “ne” olduğunu betimlemeye, açıklamaya çalışan çalışmalardır.

Çalışma Grubu

Araştırmanın çalışma grubu, Ankara ili Mamak İlçe Milli Eğitim Müdürlüğü Büyük Elçi Nazım Belger İlkokulu ve Şehitlik Ortaokulu, Çankaya İlçe Milli Eğitim Müdürlüğü Rauf Orbay İlkokulu ve Rauf Orbay Ortaokulu, Ankara Üniversitesi Geliştirme Vakfı Özel İlkokulu ve Ankara Üniversitesi Geliştirme Vakfı Özel Ortaokulu'nda 2015-2016 eğitim-öğretim döneminde 3.,4., 5., 6.,7. ve 8. sınıflarında öğrenim gören ve çalışmaya katılmaya gönüllü 214 öğrenciden oluşmaktadır. Çalışma grubunun betimsel özelliklerini içeren bilgiler Tablo 1'de yer almaktadır.

Tablo 1. Çalışmaya Katılan Öğrencilerin Betimsel Özelliklerinin Dağılımı

Özellik	n	Yüzde
Cinsiyet		
Kız	108	50,5
Erkek	106	49,5
Sınıf Düzeyi		
3	40	18,7
4	38	17,8
5	37	17,3
6	32	15
7	35	16,4
8	32	15
Toplam	214	100

Tablo 1'e göre çalışmaya katılan öğrencilerin %50,5'i kız, %49,5'i erkektir. Öğrencilerden yaşı 10 olanlar (%22,4) ve 3.sınıfa devam edenler (%18,7) diğer öğrenciler arasında en fazla orana sahip olan öğrenci grubudur. Çalışmaya katılan en yüksek sayıdaki öğrenci grubu Ankara Üniversitesi Geliştirme Vakfı İlkokulu ve Ortaokulu'ndandır. Öğrencilerin yaşları ve örneklemedeki yüzdeleri; 8 yaş (% 0,9), 9 yaş (12,6), 10 yaş (% 22,4), 11 yaş (% 15,4), 12 yaş (%15,4), 13 yaş (%18,2), 14 yaş (% 13,6) ve 15 yaş (% 1,4) şeklindedir. Çalışmaya katılan öğrencilerin % 30,8'i Büyük Elçi Nazım Belger İlkokulu ve Şehitlik Ortaokulu'ndan, %32,2'si Rauf Orbay İlkokulu ve Rauf Orbay

Ortaokulu'ndan ve % 36,9 u Ankara Üniversitesi Geliştirme Vakfı Özel İlkokulu ve Ortaokulundan gelmektedir.

Ölçeğin Deneme Uygulamasına İlişkin Çalışma Grubu ve Süreç

Hazırlanan ölçeğin ilk deneme uygulaması, Çankaya İlçe Milli Eğitim Müdürlüğü Hamdullah Suphi İlkokulu'nda öğrenim gören 30 öğrenci ile yapılmış ve öğrenciler tarafından net anlaşılmayan 3 madde tespit edilerek düzeltilmiştir. Ölçeğin deneme uygulamasında ayrıca sınıf öğretmenlerinden de ölçeğin çocukların gelişim düzeyine uygun olup olmadığı hakkında uzman görüşü alınmış, gelen uzman görüşleri doğrultusunda gerekli düzeltmeler yapılarak ölçeğin deneme formu revize edilerek uygulama formu hazırlanmış ve çalışma grubuna uygulanmıştır. Araştırma hakkında verilen ön bilgiden sonra gönüllü olan aile ve öğrenciler araştırmaya katılmıştır.

Veri Toplama Araçları

Sınıf Etkinliklerim Ölçeği (My Class Activities Survey)

Bu ölçek Gentry ve Gable (2001) tarafından geliştirilmiştir. Ölçekte, üst düzey öğrenmelerin temelinde yatan alt boyutlar olan eğitsel etkinliğe ilgi duyma, etkinlik yapılan konudan zevk alma, konuyla ilgili seçim yapabilme ve etkinlik sırasında sınırlarını zorlama boyutlarının sınıftaki etkinlikler sırasında ne kadar gerçekleştiğine ilişkin öğrenci algıları ölçülmeye çalışılmaktadır. Ölçek 3.-8. sınıf arasındaki öğrenciler için geliştirilmiş bir ölçektir ve orijinali 31 maddeden oluşmaktadır. Yukarıda da belirtildiği gibi eğitsel etkinliğe ilişkin İlgi Duyma, Seçim Yapabilme, Sınırlarını Zorlama ve Konudan Zevk Alma olmak üzere ölçeğin toplam 4 alt boyutu bulunmaktadır ve bunlar ölçeği geliştirenler tarafından aşağıdaki gibi tanımlanmışlardır (Gentry & Gable, 2001):

1. İlgi Duyma (Interest): Bu alt boyut öğrencinin belli konulara, içeriğe veya etkinliklere karşı olumlu duygular beslemesi, onlara ilgi duyması ve onları tercih etmesi olarak açıklanmaktadır.
2. Seçim Yapabilme (Choice): Bu alt boyut, öğrenciye öğrenmesi sırasında eğitim ortamında ona sunulanlar arasından seçim yapabilme gücünü ve şansını vererek kendi öğrenmelerini kendisinin yönlendirebilmesini sağlaması olarak açıklanmıştır.
3. Sınırlarını Zorlama (Challenge) Öğrencinin zihnini meşgul etmesi ve bu sırada her zaman harcadığından farklı ve daha üst düzeyde bir zihinsel çabayı aynı zamanda bundan keyif alarak harcaması kastedilmektedir.
4. Konudan Zevk Alma (Enjoyment): Öğrencinin öğrenme ortamında ilgilendiği her ne ise bundan zevk alması ve tatmin olması kastedilmektedir (Gentry & Gable, 2001).

Ölçekte Likert tipi cevap formatı kullanılmıştır. Bu format "Hiçbir Zaman" (1) ile "Her Zaman" (5) arasındadır. Öğrenciler ölçekte yer alan 31 maddenin her birinin sınıf etkinlikleri sırasında ne kadar yer aldığını düşünüyorlarsa ona göre derece vererek değerlendirirler. Ölçeğin puanlarının tüm alt boyutlarda yüksek çıkması öğrencinin sınıfındaki etkinlikleri yüksek oranda ilgi çekici, seçim yapmaya müsait, zihinsel olarak zorlayıcı ve zevkli olarak algıladığı anlamına gelmektedir. Düşük skorlar da tahmin edileceği üzere bu alt boyutların öğrenciler tarafından sınıf etkinlikleriyle ilgili düşük düzeyde algılandığını göstermektedir. Örneğin "Zihinsel Zorlanma" alt boyutunun ortalamasından alınan 4,5 ortalama, öğrencilerin sınıftaki etkinlikleri sıklıkla ve olumlu anlamda zihinsel olarak zorlayıcı, fazladan çaba sarf ettirici fakat cazip/çekici şekilde algıladıkları anlamına gelecektir. Öte taraftan "Zevk Alma" boyutundan ortalama olarak 1 alınmış olması o sınıftaki etkinliklerin öğrenciler tarafından keyifli veya tatmin edici olarak algılanmadığını gösterecektir (Gentry & Gable, 2001). Ölçeğin geliştirme çalışmasında Doğrulayıcı Faktör analizi yapılmış Uyum iyiliği değerlerinin Tucker-Lewis için 0,92 ve RMSEA için 0,04 olduğu görülmüştür. Ayrıca boyutlar arası korelasyon değerleri 0,36 ile 0,77 arasında elde edilmiştir. Ölçeğin güvenilirlik çalışması olarak elde edilen Cronbach alfa katsayısı İlgi Duyma, 0,89; Sınırları Zorlama, 0,78; Seçim Yapabilme, 0,75; Keyif Alma, 0,92'dir (Gentry & Gable, 2001).

Süreç

Ölçeğin çeviri çalışmalarının gerçekleştirilebilmesi için öncelikle ölçeği geliştiren Amerikalı araştırmacılardan yazılı izin alınmıştır. Sonrasında uyarılama ile ilgili uygulamaların yapılabilmesi için TED Üniversitesi Etik Kurulu'ndan, Milli Eğitim Bakanlığı'ndan ve Ankara Üniversitesi Geliştirme Vakfı Okulları'ndan yazılı izin alınmıştır. Daha sonra çeviri aşamasıyla ilgili aşağıda aktarılan adımlara geçilmiştir.

Çeviri Aşamasına İlişkin Yapılan Çalışmalar

Ölçeğin ülkemize uyarlamasını yapmak üzere ilk olarak Sınıf Etkinliklerim Ölçeği'nin Türkçe'ye çevirisi için bir ön çalışma yapılmıştır. Bu ön çalışma çeviri ve geri-çeviri olmak üzere iki aşamadan oluşmuştur. İlk aşamada ölçeğin maddeleri İngilizce'den Türkçe'ye dört uzman (Üstün yetenekliler alanında yurtdışında eğitim almış iki uzman ve yabancı diller eğitimi alanında uzman olup üstün yetenekliler konusunda doktora bulunan iki uzman) tarafından çevrilmiştir. Çeviride daha iyi sonuç alınması için çeviri yapacak akademisyenler panel şeklinde toplanarak beyin fırtınası yöntemiyle maddeler üzerinde çalışmışlardır. Daha sonra bu taslak çeviri örnekleri araştırmacılar tarafından değerlendirmeye alınmıştır. Değerlendirmede her bir ölçek maddesi için önerilen alternatifler arasından en uygun olanının seçilmesi veya eğer uygunsa yeni bir alternatif önerilmesi istenmiştir. Bu işlem sonucunda uzmanlardan ortak onay alan maddelerden oluşan Türkçe formu oluşturmuştur. Sonraki aşama olan geri-çeviri aşamasında, elde edilen ölçek maddeleri farklı üç uzman tarafından yeniden İngilizce'ye çevrilmiştir. Bu sonuçla elde edilen maddeler üzerinde gerekli düzenlemeler yapılarak ölçeğin deneme formu oluşturulmuştur (Deniz, 2007).

Ölçeğin Orijinal Formunun Metninde Yapılan Değişikliklere İlişkin Çalışmalar

Ölçeğin asıl formunda 3. madde olan " What I do in my class gives me interesting and new ideas." ifadesi deneme uygulamasının Türkçe formunda iki madde olarak bölünmüş 3. (*Sınıfta yaptıklarım ilgimi çeker.*) ve 4. (*Sınıfta yaptıklarım bana yeni fikirler verir.*) madde olarak iki madde olmuştur. Bu değişimin nedeni ölçeğin asıl formunda iki ayrı tutumun aynı ifade içinde verilmiş olmasından kaynaklanmaktadır. Ölçeğin asıl formunda 13. madde "*My work can make a difference*" ifadesi Türkçe karşılığı "*Yaptığım çalışmalar günlük yaşama katkı sağlar.*" şeklini almıştır. Araştırmaya katılacak çocukların gelişim özellikleri ve ülkemizde fark yaratmanın çeşitli algılara neden olacağı düşünülerek bu değişiklik yapılmıştır. Ölçeğin asıl formunda 17. madde "*This class is difficult*" ifadesi Türkçe karşılığı "*Sınıf etkinliklerimiz yeteneklerimizi zorlamamızı sağlar.*" şeklini almıştır. Araştırmaya katılan çocukların gelişim özellikleri dikkate alındığında "*Bu sınıf zordur*" ifadesinin genel ve belirsiz olması dolayısıyla bu değişiklik yapılmıştır. Ölçeğin deneme uygulamasında 6 sınıf öğretmeninden ölçeğin çocukların gelişim düzeyine uygun olup olmadığı hakkında uzman görüşü alınmıştır.

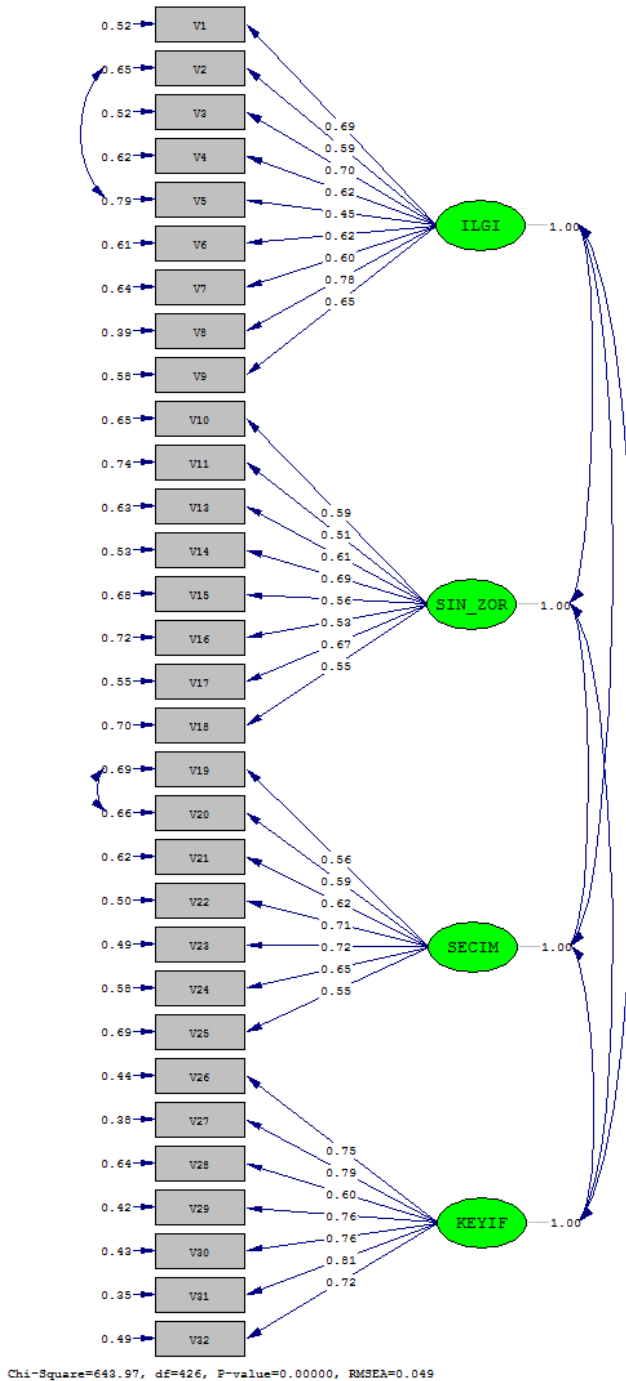
Verilerin Analizi

Verilerin analizinde, SPSS 23 ve LISREL 8.7 programlarından faydalanılmıştır. Ölçeğin geçerlik çalışmasında z puanları alınmış 5 kişinin z puanı 2,5'ten büyük olduğu için (Uç değer belirlemede farklı yöntemler olmakla birlikte 2,5 z puanı tüm grubun %99'unu kapsadığı için uç değerler çıkartılmıştır.) uç değer kabul edildiğinden 4 kişi de tamamına olumlu yanıt verdiği için datalardan silinmiş, analizler 3. sınıf ile 8. sınıf arası toplam 214 öğrenci üzerinden yapılmıştır. Öncelikle orijinal kültürdeki yapının doğrulanıp doğrulanmadığını belirlemek için Doğrulayıcı Faktör Analizi (DFA) yapılmış ve madde toplam korelasyonları hesaplanmıştır. Daha sonra alt boyutlar için

Cronbach alfa güvenirlilik katsayısı hesaplanmıştır. Bir başka geçerlik kanıtı olarak ölçeğin alt boyutları arasındaki korelasyon katsayıları Pearson korelasyon katsayısı ile hesaplanmıştır.

BULGULAR

Sınıf Etkinliklerim Ölçeği'nin Türk kültürüne uyarlama çalışmasında geçerlik kanıtı olarak elde edilen DFA, madde-toplam korelasyonları, boyutlar arası korelasyon ve güvenirlilik kanıtı olarak elde edilen Cronbach alfa katsayısına ilişkin bulgular bu bölümde yer almaktadır. Öncelikle orijinal ölçekteki madde ve boyutlarla uyumlu olarak DFA yapılmıştır. DFA'ya yönelik elde edilen path diyagramı Şekil 1'de yer almaktadır.



Şekil 1. Sınıf Etkinliklerim Ölçeği DFA Bulguları

Şekil 1 incelendiğinde maddelerle boyutları arasındaki yol katsayılarının İlgi için 0,45-0,78; Sınırları Zorlama için 0,51-0,67; Seçim Yapabilme için 0,55-0,72 ve Keyif Alma için 0,60-0,81 arasında değiştiği gözlenmiştir. Bu maddeler en az 0,20 varyans açıklamaktadırlar. Açıkladıkları varyans ve ilişki değerlerinin orta ve üzeri olması nedeniyle bu değerlerin yeterli olduğu kabul edilmiştir (Büyüköztürk, 2016). Şekil 1’de görüldüğü gibi İlgi boyutunda yer alan ikinci ve beşinci madde ile Seçim Yapabilme boyutunda yer alan 19. ve 20. maddelerin hata varyansları ki-kare değerini önemli ölçüde düşürdüğü için eşitlenmiştir. Bu maddelerin içerikleri incelendiğinde ikinci madde “Sınıfta ilgimi çeken konular üzerinde çalışma fırsatım olur.”, beşinci madde ise “Sınıfta ilgi çekici konular üzerinde çalışırım.” şeklindedir. Her iki madde de ilgi çekici konu üzerinde çalışmayla ilgili birbirini tamamlayan ve paralel görünen ifadelerdir. Bunun yanı sıra 19. madde “Sınıf etkinliklerimizde grupta çalışıp çalışmayacağımı kendim seçebilirim.” ve 20. madde “Sınıf etkinliklerimizde tek başıma çalışıp çalışmayacağımı kendim seçebilirim.” şeklinde olup bu iki madde de birbirini tamamlayıcı niteliktedir. Dolayısıyla DFA’da yapılan hata varyanslarının eşitlenmesi işleminin yerinde olduğu sonucuna varılabilir.

Sınıf Etkinliklerim Ölçeği’nin DFA sonuçlarının uyumuna ilişkin istatistikler ise Tablo 2’de verilmiştir.

Tablo 2. Sınıf Etkinliklerim Ölçeği İçin Uyum İyiiliği Değerleri

Ki-Kare	sd	Ki-Kare /sd	CFI	NFI	GFI	IFI	NFI	RMSEA
643,97	426	1,51	0,98	0,94	0,84	0,98	0,94	0,049

Tablo 2 incelendiğinde ki-kare değerinin serbestlik derecesine oranı iki katın altında olduğu görülmektedir. RMSEA değeri incelendiğinde 0,049 olduğu görülmektedir. CFI, AGFI ve NFI değerleri incelendiğinde tüm alt boyutlarda bütün bu değerlerin 0,90’ın üzerinde olduğu görülmektedir. GFI değerinin (0,84) 0,80’in üzerinde olduğu görülmektedir. Byrne (1998)’ye göre bu uyum indeksleri kabul edilebilir uyumu göstermektedir. Sınıf Etkinliklerim Ölçeği’nin Düzeltilmiş Madde Toplam Korelasyonu ve Cronbach Alfa Değerleri Tablo 3’te yer almaktadır.

Tablo 3. Sınıf Etkinliklerim Ölçeğinin Düzeltilmiş Madde Toplam Korelasyonu ve Cronbach Alfa Değerleri

n=214	Ölçek Madde No	Nihai Ölçek Madde No	Düzeltilmiş Madde Toplam Korelasyonu
İlgi Duyuma	1	1	0,64
	2	2	0,59
	3	3	0,61
	4	4	0,58
	5	5	0,47
	6	6	0,56
	7	7	0,50
	8	8	0,72
	9	9	0,57
	10	10	0,59
Sınırları Zorlama	11	11	0,44
	13	12	0,59
	14	13	0,56
	15	14	0,51
	16	15	0,52
	17	16	0,52
	18	17	0,55

	19	18	0,58
	20	19	0,60
	21	20	0,57
Seçim Yapabilme	22	21	0,61
	23	22	0,58
	24	23	0,57
	25	24	0,48
	26	25	0,69
	27	26	0,74
	28	27	0,56
Keyif Alma	29	28	0,70
	30	29	0,71
	31	30	0,76
	32	31	0,67
Cronbach Alfa	İlgi Duyma: 0,86 Sınırları Zorlama:0,82 Seçim Yapabilme: 0,83 Keyif Alma: 0,90		

Tablo 3 incelendiğinde, ölçeğin iç tutarlılığı için hesaplanan düzeltilmiş madde toplam korelasyonları İlgi Duyma için 0,47-0,72, Sınırları Zorlama için 0,44-0,59, Seçim Yapabilme için 0,48-0,61 ve Keyif Alma için 0,56-0,76 aralığındadır. Gentry ve Gable (2001)'nin ölçeği geliştirdikleri çalışmalarında düzeltilmiş madde toplam korelasyonları İlgi Duyma için 0,61-0,74, Sınırları Zorlama için 0,32-0,64, Seçim Yapabilme için 0,28 - 0,54 ve Keyif Alma için 0,68-0,81 aralığındadır.

Sınıf Etkinliklerim Ölçeği'nin alt boyutlarına göre Cronbach alfa katsayıları, İlgi Duyma için 0,86; Sınırları Zorlama için 0,82; Seçim Yapabilme için 0,83 ve Keyif Alma için 0,90 olarak kestirilmiştir. Buna göre ölçek alt boyutları kabul edilebilir derecede güvenilirdir. Gentry ve Gable (2001)'nin ölçeği geliştirdikleri çalışmalarında, Sınıf Etkinliklerim Ölçeği'nin alt boyutlarının Cronbach alfa katsayısı (İlgi Duyma, 0,89; Sınırları Zorlama, 0,78; Seçim Yapabilme, 0,75; Keyif Alma, 0,92) en düşük 0,75 ve en yüksek 0,92 aralığında olduğunu belirtmişlerdir. Çalışma bulguları Gentry ve Gable (2001)'nin çalışma bulguları ile benzerlik göstermektedir. Sınıf Etkinliklerim Ölçeği'nde yer alan boyutlar arası korelasyonlar Tablo 4'te verilmiştir.

Tablo 4. Sınıf Etkinliklerim Ölçeğinin Alt Boyutları arasındaki Pearson Korelasyon Katsayıları

	n=214	İlgi Duyma	Sınırları Zorlama	Seçim Yapabilme
Sınırları Zorlama	Pearson K.K.	0,77**		
	p	0,000		
Seçim Yapabilme	Pearson K.K.	0,45**	0,50**	
	p	0,000	0,000	
Keyif Alma	Pearson K.K.	0,72**	0,70**	0,36**
	p	0,000	0,000	0,000

** : p<0,01

Tablo 3 incelendiğinde, Sınıf Etkinliklerim Ölçeği'nin İlgi Duyma alt boyutu ile Sınırları Zorlama, Keyif Alma, Seçim Yapabilme alt boyut puanları arasında pozitif yönde anlamlı düzeyde korelasyonlar olduğu görülmektedir (p<0,01). İlgi Duyma alt boyutu ile pozitif en yüksek ilişki gösteren alt boyutun Sınırları Zorlama alt boyutu olduğu (0,77, p<0,01; ölçeğin orjinal formunun geliştirme çalışmasında 0,50, p<0,01 elde edilmiştir (Gentry ve Gable, 2001); en düşük ilişkinin Seçim Yapabilme alt boyutuyla olduğu (0,45, p<0,01 Ölçeğin orjinal formunun geliştirme çalışmasında 0,39, p<0,01 elde edilmiştir (Gentry ve Gable, 2001)) göze çarpmaktadır. Sınırları Zorlama alt boyutu ile Seçim Yapabilme (0,50, p<0,01; Ölçeğin orjinal formunun geliştirme çalışmasında 0,39, p<0,01 elde edilmiştir (Gentry ve Gable, 2001)) ve Keyif Alma (0,70, p<0,01;

Ölçeğin orjinal formunun geliştirme çalışmasında 0,40, $p < 0,01$ elde edilmiştir (Gentry ve Gable, 2001) alt boyut puanları arasında anlamlı düzeyde korelasyonlar olduğu görülmektedir. Seçim Yapabilme alt boyutu ile Keyif Alma (0,36, $p < 0,01$; Ölçeğin orjinal formunun geliştirme çalışmasında 0,49, $p < 0,01$ elde edilmiştir (Gentry ve Gable, 2001) alt boyut puanları arasında anlamlı düzeyde korelasyonlar olduğu görülmektedir. Büyüköztürk (2016)'e göre, korelasyon değerleri düşük (0,00-0,29), orta (0,30-0,69) ve yüksek (0,70-1,00) olarak sınıflandırabilir. Buna göre, Sınıf Etkinliklerim Ölçeği'nde yer alan boyutlar arası korelasyonların orta ve yüksek düzeyde pozitif yönde olduğu görülmektedir. Çalışma bulguları Gentry ve Gable (2001)'nin çalışma bulguları ile benzerlik göstermektedir.

SONUÇLAR ve TARTIŞMA

Bu araştırma kapsamında Gentry ve Gable (2001) tarafından Amerika Birleşik Devletleri'nde geliştirilmiş olan Sınıf Etkinliklerim Ölçeği'nin uyarlama çalışması gerçekleştirilmiş ve ölçeğin Türk kültüründe kullanımı için gerekli olan geçerlik ve güvenirlik kanıtları ortaya konmuştur. Geçerlik çalışması kapsamında öncelikle DFA yapılmıştır. DFA sonucunda ölçeğin Türk kültüründe orijinal ölçekteki faktör yapısıyla aynı şekilde kullanılabilmesi sonucuna ulaşılmıştır. DFA için elde edilen uyum iyiliği değerleri dört faktörlü yapının kabul edilebilir bir uyum gösterdiği sonucunu ortaya koymuştur.

İlgili alan yazında ölçeğin farklı kültürlere ve aynı kültür içindeki farklı öğrenci gruplarına uygulanmış olduğu daha önce de belirtilmişti. Örneğin Pereira vd. (2010) tarafından yapılan çalışmada sınıf düzeyleri üç ila sekizinci sınıf arasında değişen 826 üstün yetenekli çocuk üzerinde ölçek uygulanmıştır. Amerikan kültüründeki üstün yetenekli çocuk grubunun örneklem olarak alındığı bu çalışmada, 16 ve 17. maddeler önemli bir modifikasyon verdiği için bu maddeler ölçekten çıkartılmıştır. Ölçeğin dört faktörlü yapısı bu çalışmada da doğrulanmıştır. Yang vd. (2012)'nin ölçeği Güney Kore kültürüne uyarlama çalışmasında 3. Sınıftan 6. Sınıfa kadar Güney Korenin dört farklı şehrinde seçilmiş toplam 564 öğrenci örnekleme dahil edilmiştir. Bu çalışmada da 17. madde ölçekten çıkartılarak dört faktörlü yapı doğrulanmıştır. Benzer şekilde bu çalışmada da toplamda dört maddenin hata varyansları eşitlenerek model doğrulanmıştır. Aryan ve Shahrokhi (2015) tarafından yapılan çalışmada ölçek İran'da uygulanmış ancak uyarlama çalışması gerçekleştirilmemiştir. Uyarlamak yerine İngilizce olan orijinal ölçek İngilizce öğrenen grup üzerinde kullanılmıştır. Belirlenen faktör yapıları dahilinde her maddenin ilgili faktör yapısıyla olan korelasyonu madde-toplam korelasyonları ile hesaplanmış ve bu değerlerin de DFA sonucunda elde edilen yapının uygunluğunu gösterir şekilde olduğu görülmüştür. Bununla birlikte boyutlar arası korelasyon değerleri incelendiğinde boyutları birbirinden yeterince ayrılmış bir ölçek için beklenenden yüksek değerler verdiği sonucu da ortaya çıkmıştır. Ancak Gentry ve Gable (2001)'in orijinal ölçek geliştirme çalışmasında da boyutlar arası korelasyon 0,39 ile 0,76 arasında değer almıştır. Yang vd. (2012)'nin Güney Kore kültürü'ne uyarlama çalışmalarında ise 0,67 ile 0,88 arasında korelasyon değerleri elde edilmiştir. Boyutlar arası korelasyon değerlerinin yüksek olmasının nedeni, tüm maddelerin öğrencinin sınıf etkinlikleri ile ilgili olmasından kaynaklı olabileceği şeklinde açıklanabilir. Ölçeğin güvenirlik kanıtı olarak elde edilen Cronbach alfa değerleri de yeterli düzeydedir. Yang vd. (2012)'nin çalışmasında, bu çalışmaya paralel olarak, Cronbach alfa değerleri 0,74 ile 0,93 arasında elde edilmiştir. Sonuç olarak; Sınıf Aktivitelerim Ölçeği'nin Türk kültürüne uyarlaması sonucunda ölçeğin 32 maddeden ve dört boyuttan oluştuğu görülmüştür. Sınıf Etkinliklerim Ölçeği ile ölçülen boyutlar eğitim alanındaki pek çok teori ve pratikten güç almaktadır. Ayrıca ilgili boyutlar öğrenci motivasyonu ve öğrenme ile doğrudan ilgilidir. Sınıf Etkinliklerim Ölçeği kullanılarak, ilkökul ve ortaokul öğrencilerinin sınıflarındaki aktiviteleri nasıl algıladıkları değerlendirilebilir. Bu bilgiler sayesinde pratik anlamda eğitimciler gerçekleştirdikleri eğitsel etkinliklerin öğrenciler tarafından nasıl algılandığını, onları motive edecek düzeyde ilgi çekici, zevkli, seçenekli ve sınır zorlayıcı olup olmadığını görebilirler. Ölçeğin sonuçlarına göre eğitim ortamları ve etkinlikleri için amaçlar oluşturabilir ve önemli görülen alanlarda geliştirme planları oluşturabilirler. Bu konularda çalışan araştırmacılar Sınıf Etkinliklerim Ölçeği'ni, yapılan bir müdahalenin öğrencilerin okula karşı olan tavırlarında bir gelişmeye yol açıp açmadığını veya

öğrenci tavırları ile başarı seviyeleri arasında bir bağlantı olup olmadığını ölçmek ve anlamak amacıyla kullanılabilirler. Ayrıca değişik öğrenci grupları arasında okulda gerçekleştirilen etkinliklere yönelik düşünceler ile ilgili bir fark olup olmadığını veya diğer birçok araştırma sorusunu değerlendirmek için kullanılabilirler. Örneğin ölçek normal gelişim gösteren öğrenciler dışında üstün yetenekli çocuklar veya özel gereksinimli diğer öğrenci gruplarına uygulanarak farklı öğrenci gruplarının okuldaki sınıf etkinliklerine ilişkin görüşleri incelenebilecektir. Öğrencilerin sınıf etkinliklerini değerlendirmelerinin incelenmesi hem araştırmacıların hem de eğitimcilerin etkili öğretim için kullanılabilecek yöntemleri incelemeleri ve oluşturmaları için yardımcı olacaktır. Bundan başka ölçeğin yaygın olarak kullanılması ülke genelinde eğitim programında gerçekleştirilebilecek değişiklik ve yeniliklerin hangi alanlarda olması gerektiği konusunda Milli Eğitim Bakanlığı yetkililerine ve bu konularda politika yapan kişilere ışık tutabilecektir. Diğer çalışmalarda test tekrar test gibi farklı bir güvenilirlik kanıtının elde edilmesi de uyarlanmış olan bu ölçeğe katkı sağlayacaktır. Bu çalışmada bulunan bazı sınırlılıklara değinmek gereklidir. Öncelikle uyarlama çalışmasında kullanılan örneklem rastgele örnekleme ile seçilmemiştir. Örneklemedeki vakıf okulundan gelen öğrenci sayısının devlet okulu öğrenci sayısından fazla olması bir başka sınırlılığı oluşturmaktadır.

KAYNAKÇA

- Archambault, E. X., Westberg, K., Brown, S. B., Hallmark, B. W., Emmons, C. L., & Zhang, W. (1993). *Regular classroom practices with gifted students: Results of a national survey of classroom teachers*. Storrs, CT: National Research Center on the Gifted and Talented.
- Aryan, E., & Shahrokhi, M. (2015). Students' perceptions of class activities: An investigation into the role of gender and grade level. *Mediterranean Journal of Social Sciences*, 6(4), 19-26.
- Bandura, A. (1997). *Self efficacy*. New York: W. H. Freeman.
- Betts, G. T., & Kercher, J. K. (1999). *Autonomous learner model optimizing ability*. Greeley, CO: Autonomous Learning Publications & Specialists.
- Bloom, B. S. (Ed.) (1985). *Developing talent in young people*. New York: Ballantine.
- Büyüköztürk, Ş. (2016). *Veri analizi el kitabı*. Pegem Akademi.
- Byrne, B. M. (1998). Structural equation modeling with LISREL, PIRELIS and SIMPLIS: Basic concepts, applications and programming. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Csikszentmihalyi, M. (1990). Literacy and intrinsic motivation. *Daedalus*, 119, 115-140.
- Csikszentmihalyi, M., & McCormack, J. (1986). The influence of teachers. *Phi Delta Kappan*, 67(6), 415-419.
- Clifford, M. (1990). Students need challenge, not easy success. *Educational Leadership*, 48, 22-26.
- Deci, E. L. (1995). *Why we do what we do: The dynamics of personal autonomy*. New York: G. P. Putnam's Sons.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self determination in human behavior*. New York: Plenum.
- Deniz, K. Z. (2007). Psikolojik ölçme aracı uyarlama, *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 40(1), 1-16.
- Dewey, J. (1913). *Interest and effort in education*. New York: Houghton Mifflin.
- Dewey, J. (1916). *Democracy and education*. New York: The Free Press.
- Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 3 Goals and cognitions* (pp. 13-44). New York: Academic.
- Faske, D., & Grubb, D. J. (1997). *Implications of the learner-centered battery for new teacher standards and teacher education reform in Kentucky*. Paper presented at the Annual Meeting of the American Psychological Association, Chicago, IL.
- Feldhusen, J. E., & Kroll, M. D. (1991). Boredom or challenge for the academically talented in school. *Gifted Education International*, 7, 80-81.
- Gallagher, J. J. (1985). *Teaching the gifted child* (3rd ed.). Boston: Allyn & Bacon
- Gardner, H. (1991). *The unschooled mind: How children think and how schools should teach*. New York: Basic Books.
- Gentry, M., & Gable, R. K. (2001). From the students' perspective - my class activities: An instrument for use in research and evaluation. *Journal for the Education of the Gifted*, 24, 322-343.
- Gentry, M., Gable, R. K., & Rizza M. K. (2002). Students' perceptions of classrooms activities: Are there grade level and gender differences? *Journal of Educational Psychology*, 94, 539-544.

- Gentry, M., Gable, R. K., & Springer, P. (2000). Gifted and non-gifted middle school students: Are their attitudes toward school different as measured by the new affective instrument, my class activities? *Journal for the Education of the Gifted*, 24, 74-96.
- Gentry, M., Rizza, M. G., & Gable, R. K. (2001). Gifted students' perceptions of their class activities: Differences among rural, urban, and suburban student attitudes. *Gifted Child Quarterly*, 45, 115-129.
- Gentry, M., Rizza, M. G., & Owen, S. V. (2002). Examining perceptions of challenge and choice in classrooms: The relationship between teachers and their students and comparison between gifted students and other students. *Gifted Child Quarterly*, 46, 145-155.
- Glasser, W. (1996). Then and now. The theory of choice. *Learning*, 25, 20-22.
- Good, T. L., & Brophy, J. E. (1987). *Looking in classrooms* (4th ed.). New York: Harper & Row.
- Goodlad, J. (1984). *A place called school*. New York: McGrawHill.
- Hootstein, E. W. (1994). Motivating middle school students. *Middle School Journal*, 25(5), 31-35.
- James, W. (1890). *The principles of psychology*. London: MacMillan.
- Karasar, N. (2016). *Bilimsel araştırma yöntemi*. Nobel Akademi.
- Kerka, S. (1994). *Self directed learning myths and realities*. (ERIC Document Reproduction Service No. ED 365818).
- Lepper, M. R., & Chabay, R. W. (1985). Intrinsic motivation and instruction: Conflicting views on the motivational processes in computer-based education. *Educational Psychologist*, 20, 217-230.
- Middleton, J. A. (1995). A study on intrinsic motivation in the mathematics classroom: A personal constructs approach. *Journal for Research in Mathematics Education*, 26, 254-279.
- Middleton, J. A., Littlefield, J. & Lehrer, R. (1992). Gifted students' conceptions of academic fun: An examination of a critical construct for gifted education. *Gifted Child Quarterly*, 36, 38-44.
- Matthews, M. S., & McBee, M. T. (2007). School factors and the underachievement of gifted students in a talent search summer program. *Gifted Child Quarterly*, 51, 167-181.
- Maker, C. J. (1982). *Curriculum development for the gifted*. Austin, TX: Pro-Ed.
- Pereira, N., Peters, S. J., & Gentry, M. (2010). The my class activities instrument as used in Saturday enrichment program evaluation. *Journal for Advanced Academics*, 21(4), 568-593.
- Parke, B. (1989). *Gifted students in regular classrooms*. Boston: Allyn and Bacon.
- Passow, A. H. (1982). The relationship between the regular curriculum and differentiated curricula for the gifted/talented. *Selected proceedings of the First National Conference on Curricula for the Gifted/Talented*. Ventura, CA: Ventura Superintendents of Schools Office.
- Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40.
- Reis, S. M., Westberg, K. L., Kulikowich, J., Caiillard, R., Hebert, T., Plucker, J., Purcell, J. H., Rogers, J. B., & Smist, J. M. (1993). *Why not let high ability students start school in January? The curriculum compacting study*. Storrs, CT: National Research Center on the Gifted and Talented.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60(3), 180-184, 261.
- Renzulli, J. S. (1988). The multiple menu model for developing differentiated curriculum for the gifted and talented. *Gifted Child Quarterly*, 32(3), 298-309.
- Renzulli, J. S. (1994). *Schools for talent development: A comprehensive plan for total school improvement*. Mansfield Center, CT: Creative Learning.
- Renzulli, J. S., & Reis, S. M. (1997). *The schoolwide enrichment model: A how-to guide for educational excellence* (2nd ed.). Mansfield Center, CT: Creative Learning.
- Schiefele, U. (1991). Interests, learning and motivation. *Educational Psychologist*, 26(3-4), 299-323.
- Schlichter, C. (1986). Talents unlimited: Applying the multiple talent approach to mainstream and gifted programs. In J. S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented* (pp. 352-390). Mansfield Center, CT: Creative Learning.
- Shore, B. M., Cornell, D. G., Robinson, A., & Ward, V. S. (1991). *Recommended practices in gifted education*. New York: Teachers College.
- Tobias, S. (1994). Interest, prior knowledge and learning. *Review of Educational Research*, 64(1), 37-54.
- Tomlinson, C. A. (1992). Gifted education and the middle school movement: Two voices on teaching the academically talented. *Journal for the Education of the Gifted*, 15(3), 206-238.
- Tomlinson, C. A. (1999). *Differentiation of curriculum: A guide*. Arlington, VA: ASCD.
- Treffinger, D. (1986). Fostering effective independent learning through individualized programming. In J. S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented* (pp. 429-460). Mansfield Center, CT: Creative Learning.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: M.I.T.

- Wang, M. C., & Lindvall, C. M. (1984). Individual differences in school learning environments: Theory, research and design. In E. W. Gordon (Ed.), *Review of research in educational* (pp. 161-225). Washington, DC: American Educational Research Association.
- Ward, V. (1980). *Differential education for the gifted*. Ventura, CA: Ventura Superintendent of Schools Office.
- Westberg, K. L., Archambault, E. X., Dobyms, S. M., & Salvin T. J. (1993). *An observational study of instructional and curricular practices used with gifted and talented students in regular classrooms*. Storrs, CT: The National Research Center on the Gifted and Talented.
- Whitehead, A. N. (1929). The rhythm of education. In A. N. Whitehead (Ed.), *The aims of education* (pp. 46-59). New York: MacMillian.
- Yang, Y., & Gentry, M. (2011). *Gifted and general elementary students' perceptions in China and the United States: A cross-national study*. Manuscript submitted for publication.
- Yang, Y., Gentry, M., & Choi, Y. O. (2012). Gifted students' perceptions of the regular classes and pull-out programs in South Korea. *Journal of Advanced Academics*, 23(3), 270-287.

EXTENDED ABSTRACT

Introduction

Motivating students is among the most important first steps in ensuring their learning. Classroom activities need to possess certain qualities in order to motivate students and focus their learning. It is very important for the contents of activities, methods for their application and associated end products to be interesting, enjoyable and challenging, while also providing students with opportunities to make choices. Factors such as student interest, enjoyment of topics, and the presence of challenges, and different choices increase both the quality of the learning environment, as well as the motivation of the student. Gentry and Gable (2001) proposed a tool that can measure all of these factors in their Class Activities Scale. A review of the Turkish literature reveals that there are currently no tools that can jointly measure all of these four dimensions. Moreover, the Class Activities Scale has been successfully used in the international literature in a number of different studies (Aryan and Shahrokhi, 2015; Gentry and Gable, 2001; Gentry, Gable and Rizza, 2002; Gentry, Gable and Springer, 2000; Gentry, Rizza and Gable, 2001; Gentry, Rizza and Owen, 2002; Pereira, Peters ve Gentry, 2010; Yang, Gentry ve Choi, 2012). It is evident that an adaptation of this widely used scale would fill an important gap in the Turkish literature, and provide a useful tool for many researchers in different disciplines. This study aims to perform an adaptation of the "Class Activities Scale" to the Turkish culture, including its validity and reliability, allowing measurement of the perceptions of students in their interest and enjoyment of class activities as well as their opportunities for making choices and how much they were challenged during these activities.

Method

This research is a descriptive study based on a survey method. Descriptive studies attempt to describe and explain "what" events, objects, entities, institutions, groups, and areas are. The study group for the research consists of 214 volunteer students attending grades 3 through 8 in the Mamak Directorate of National Education Büyük Elçi Nazım Belger Primary School and Şehitlik Middle School in Ankara, Çankaya Directorate of National Education Rauf Orbay Primary and Middle schools in Ankara, and Ankara University Development Foundation Primary and Middle schools in Ankara during the 2015-2016 school year. The "My Class Activities Survey" was used as the data collection method in this research. This scale was developed by Gentry and Gable (2001), and attempts to measure perceptions of students regarding their interest in the educational activity, their level of enjoyment, opportunities they had for making choices and how much they were challenged, which are among sub-dimensions that underlie their high level learning. An initial study was conducted as in the Turkish adaptation of the "My Class Activities Survey", including first a translation of the original scale to Turkish, followed by a translation back to English. The first translation was performed by a total of four experts, two of which were trained abroad on gifted children, and two of which were trained in foreign languages with doctoral degrees related to gifted children. These experts worked in a panel structure to improve the quality of the translation.

Subsequently, this translation was evaluated by the researchers. In this evaluation, the most appropriate candidate from among suggested translations for each item in the scale was chosen, or additional alternatives were requested if necessary. As a result of this process, the final form in Turkish consisted of items that were jointly approved by all the experts. The next step of back-translation to English was performed by three new experts. Together with additional revisions that were found necessary after this step, the first trial version of the form was finalized.

The first pilot application of the scale was performed on 30 students attending the Çankaya Directorate of National Education, Hamdullah Suphi Primary School, identifying and revising items that were not clearly understood. Moreover, opinions from classroom teachers on whether the scale was appropriate for the developmental levels of students were collected during this pilot study, with corresponding revisions performed on the form. As a result, the scale was finalized for the validity and reliability study and was applied to the study group. Following an initial provision of information on the research, volunteer parents and students participated in the study. As explained above, following revisions to the items to improve items that were difficult to understand, the scale was applied to 223 students in Gülen Muharrem Pakoğlu Middle School who had not participated in the pilot study that was performed in the Hamdullah Suphi Primary School.

The analysis of data was performed using SPSS 23 and LISREL 8.7. In studying the validity of the scale, 5 subjects with z scores greater than 2.5 were excluded, being considered as outliers, and 4 subjects were excluded since they had answered all items affirmatively. The analysis was hence performed on a total of 214 students in grades 3 through 8. In order to determine whether the structure in the original culture could be verified, a Confirmatory Factor Analysis (CFA) was performed and item-total correlations were calculated. Subsequently, Cronbach-alpha reliability coefficient was calculated for sub-dimensions of the scale. As additional evidence for the validity of the study, correlation coefficients between sub-dimensions of the scale were calculated using the Pearson correlation coefficient.

Results and Discussion

In this research, a Turkish adaptation of the “My Class Activities Survey” developed by Gentry and Gable (2001) in the United States was performed and evidence towards the validity and reliability of the adaptation was provided. In studying the validity of the adaptation, a CFA was performed first, establishing that the adapted scale could be used in Turkey with the same factor structure as the original scale. The goodness-of-fit values obtained from the CFA study showed that the four-factor structure exhibited acceptable fits. Within the factor structure obtained in this fashion, the correlation of each item with the corresponding factor structure was computed using item-total correlations and were found to support the suitability of the structure obtained through the CFA. On the other hand, an investigation of inter-dimensional correlations revealed values that were higher than what would be expected from a scale with sufficiently differentiated dimensions. These high values for inter-dimensional correlations could be explained with the fact that all items were related to the class activities of students and were hence related to each other. Finally, Cronbach-alpha values for the results were also found to be sufficiently high, supporting the reliability of the scale.

Türkiye’de Eğitim Alanında Yayımlanan Ölçek Geliştirme Çalışmalarının Uygunluğunun Çok Yüzeyle Rasch Modeli ile İncelenmesi*

Investigation of Scale Development Studies Conducted in Educational Sciences Published in Turkey by Many-Faceted Rasch Model

Gülden KAYA UYANIK** Neşe GÜLER*** Gülşen TAŞDELEN TEKER****
Süleyman DEMİR*****

Öz

Bu çalışmanın amacı Türkiye’de 2010-2015 yılları arasında sosyal bilimler atıf indeksinde (Social Sciences Citation Index- SSCI) taranan bilimsel dergilerde yayımlanmış ölçek geliştirme çalışmalarının, ölçek geliştirme sürecindeki adımları ne derece sağlayabildiklerini incelemektir. Araştırma kapsamında Türkiye’de eğitim alanındaki üç dergide 2010-2015 yılları arasında yayınlanan 57 ölçek geliştirme makalesi, Ölçme ve Değerlendirme alanında dört uzman tarafından araştırmacıların geliştirdiği “Ölçek Geliştirme Makaleleri İnceleme Formu” ile puanlanmıştır. Ölçek Geliştirme Makaleleri İnceleme Formu alan yazında olan genel görüşler dikkate alınarak 19 maddeden oluşan bir form olarak düzenlenmiştir. Analizler çok yüzeyle Rasch modeli (ÇYRM) için FACETS ve frekans analizi için SPSS programı kullanılarak gerçekleştirilmiştir. Çalışmada yer alan üç yüzeyle (makalelerin belirlenen ölçütleri karşılama düzeyi, değerlendirme formunda kullanılan ölçütler ve puanlayıcılar) ve değerlendirmedeki ölçütlerin puanlama düzeylerine ait bulgular ayrı ayrı elde edilmiştir. Analizler sonucunda, Logit cetvele göre çalışmada esas alınan ölçütleri karşılayabilme düzeyi en yüksek 14 numaralı makalenin ve en düşük 27 numaralı makalenin olduğu gözlenmektedir. Değerlendirme formunda yer alan ölçütler incelendiğinde güçlük düzeyi en yüksek (karşılanması en zor) 16. ölçüt (*Ölçüt geçerliği analizi sonucu elde edilen değerler uygundur*) iken güçlük düzeyi en düşük (karşılanması en kolay) 3. ölçüt (*Ölçeğin amacına uygun bir hedef kitle belirlenmiştir.*) olduğu belirlenmiştir. Puanlayıcılara ait ölçümler incelendiğinde ise 4. puanlayıcının en katı (en yüksek pozitif değer), 1. puanlayıcının ise en cömert (en düşük negatif değer) puanlayıcı olduğu gözlenmektedir. Ayrıca incelenen 57 makalenin 27 tanesinde (%47,4) geliştirilen ölçekler makale içeriğinde verilirken 30 makalede (%52,6) bu durum gözlenmemiştir. Geliştirilen ölçeklerin diğer araştırmacılar tarafından kullanıldığında puanlamanın nasıl yapılacağına dair bilginin verilme durumu ise makalelerin %56,1’inde (32 makale) gözlenirken % 43,9’unda (25 makale) puanlama ile ilgili bir bilgiye rastlanmamıştır.

Anahtar Kelimeler: Ölçek geliştirme, çok yüzeyle rasch modeli, eğitim bilimleri

Abstract

The aim of this study is to investigate the scale development studies that were conducted in Turkey from 2010 to 2015 and published in the journals indexed in Social Sciences Citation Index (SSCI) by means of their verifications about scale development steps. In the study, 57 scale development articles published in between years 2010-2015 in three different journals scope of whom are educational sciences were rated according to “Scale Development Article Investigation Form” by four measurement and evaluation expert. The 19-item Scale Development Article Investigation Form was composed according to the view of the literature. The many-facet Rasch model analysis were done by using FACETS and frequencies were gathered by using SPSS programs. The analysis results of the three facets (the level of articles by means of verification of determined

* Çalışmanın bir kısmı Eğitimde ve Psikolojide Ölçme Ve Değerlendirme (Antalya / 2016) kongresinde sunulmuştur.

**Yrd. Doç. Dr., Sakarya Üniversitesi, Sakarya, Türkiye, guldenk@sakarya.edu.tr

***Doç. Dr., Sakarya Üniversitesi, Sakarya, Türkiye, gnguler@gmail.com.tr

****Yrd. Doç. Dr. Hacettepe Üniversitesi, Ankara, Türkiye, gulsentasdelen@gmail.com

*****Araş. Gör., Sakarya Üniversitesi, Sakarya, Türkiye, suleymand@sakarya.edu.tr

criteria, criteria used in investigation form, and the raters) used in the study and the rating levels of investigated criteria were gathered separately. As a result of the study it was seen that the 14th article was cover the investigated criteria mostly and the 27th article was the least cover one according to the logit scale. When the criteria used for investigation of the articles were examined, the 16th criteria was the most difficult one. On the other hand, the third criteria (There were a suitable sample for the aim of scale.) was the easiest one to meet. When the ratings of raters were investigated, it was observed that the fourth rater was the strictest one and the first rater was more lenient one. Moreover, the developed scales were given in just 27 of 57 investigated articles (47,4%) and there were no such a situation on 30 articles (52,6%). While the scoring information of developed scales for other researchers who want to use the developed scales in their future studies was given only 56,1% (f=32) of investigated articles, there were no such information for 43,9 % of investigated articles.

Keywords: Scale development, many facet Rasch model, educational sciences

GİRİŞ

Sosyal bilimlerdeki alan çalışmalarında, bireylerin farklı yetenek, algı ve tutumlarını ölçmek, bunlara dayanarak bazı kararlar almak ve farklı yapılar arasındaki ilişkileri açıklamak amacıyla veri toplama aracı olarak sıklıkla ölçek adı verilen araçlardan yararlanılmaktadır (Stone, 1978). Ölçme aracı olarak ölçeklerin kullanılmasının temeldeki sebebi, eğitimde ve psikolojide var olduğu kabul edilen ancak doğrudan gözlenemeyen değişkenlerin ölçülmesi istenmesidir. Bireyde var olduğu kabul edilen bu tür değişkenler yapı olarak adlandırılmakta ve bu yapılar bireylerde gözle görülür davranışlara neden olmaktadır. Ölçekler ile gözle görülür davranışların dereceli puanlama anahtarları kullanılarak bireylerin ilgili yapıya ne kadar sahip olduklarını belirlemek amaçlanmaktadır (DeVellis, 2003; Erkuş, 2012; Tezbaşaran, 2008).

Ölçeklerin bireylere uygulanması sonucu elde edilen bulguların genellenebilirliği, işlevselliği ve sağlamlığı ise kullanılan bu araçların güvenilirliği ve geçerliğiyle paralellik göstermektedir (Erkuş, 2007). Örneğin aynı ölçeklerin veya aynı özelliği ölçmek üzere geliştirilen farklı ölçeklerin kullanıldığı çalışmalarda elde edilen sonuçlar birbirinden tutarsız olabilmektedir. Bu durumun temelinde en büyük sebebinin güvenilir ve geçerli ölçümler sağlayan ölçme araçlarının kullanılmaması olmasından kaynaklandığı düşünülmektedir (Barrett, 1972; Cook, Hepworth, Wall ve Warr, 1981; Hinkin, 1995; Schriesheim, Powers, Scandura, Gardiner ve Lankau, 1993). Bu sebeple doğrudan gözlenemeyen değişkenleri ölçmek için kullanılan bu araçların niteliği oldukça önemlidir.

Araştırmalarda kullanılan ölçekler belirli bir özelliği ölçmeye yönelik olarak geliştirilebileceği gibi daha önceden araştırmacının ilgilendiği yapıyı ölçmek amacıyla başka bir dilde ve kültürde geliştirilmiş ölçekler de araştırmacının ilgilendiği örnekleme uygun olarak uyarlanarak kullanılabilir. Bu süreçlerden ilki ölçek geliştirme olarak tanımlanırken ikincisi ölçek uyarlama olarak ifade edilmektedir. Araştırmacılar bu iki süreçten araştırmalarında ölçmeyi düşündükleri değişkene göre uygun olanı seçerek çalışmalarını yürütmelidirler. Eğer ölçmeyi planladıkları değişkene ilişkin daha önce geliştirilmiş herhangi bir ölçek bulunmuyorsa ya da daha önce geliştirilmiş olan ölçekler, araştırmacının amacına tam anlamıyla hizmet etmiyorsa yeni bir ölçek geliştirmeleri gerekmektedir. Ancak eğer daha önce başka bir dilde, kültürde veya araştırmacının uygulamak istediği gruptan farklı bir gruba yönelik olarak geliştirilmiş bir ölçek bulunuyorsa, bu ölçeğin araştırmacının uygulamak istediği dile, kültüre ve örnekleme uygun hale getirilmesi gerekmektedir. Bir diğer ifadeyle uyarlanması gerekmektedir. Bu iki sürecin de birbirlerine göre avantaj ve dezavantajları bulunmaktadır. Bu çalışma kapsamında ölçek geliştirme makaleleri incelendiğinden ölçek uyarlama üzerinde fazla durulmamış ölçek geliştirme ise daha detaylı olarak irdelenmiştir.

Araştırmacının amacına uygun bir ölçek bulunmadığı durumda alanyazında belirtilen ölçek geliştirme adımlarını takip ederek yeni bir ölçek oluşturulmaya çalışılır. Bir diğer ifadeyle güvenilir ve geçerli ölçme araçlarının geliştirilmesi için dikkat edilmesi ve eksiksiz bir şekilde izlenmesi gereken bazı işlem basamakları bulunmaktadır. Bu aşamalar alanyazında aşağıdaki gibi özetlenmiştir (Clark ve Watson, 1995; DeVellis, 2003; Hinkin, 1995; Hinkin, Tracey ve Enz, 1997; Tezbaşaran, 2008):

- Ölçme aracına ilişkin açık ve net bir yönergenin hazırlanması,

- Ölçülmek istenen özelliğin açık ve net olarak ifade edilmesi,
- Madde havuzunun oluşturulması,
- Madde havuzunun uzman görüşüne sunulması,
- Ön deneme uygulamasının yapılması,
- Deneme uygulamasının gerçekleştirilmesi,
- Faktör yapısının belirlenmesi,
- Güvenirliğinin belirlenmesi
- Ortaya atılan modelin doğrulanması

Özellikle son yıllarda ölçek geliştirme ve uyarlama çalışmalarında bir artış gözlenmektedir. Erkuş (2007) bu artışın sebebi olarak akademisyenlerin üzerindeki yayın baskısını göstermektedir. Bu durum ise niteliği düşük çalışmalara yol açmaktadır. Çünkü yürütülen çalışmalar incelendiğinde süreçte yapılan ve yapılmaya devam edilen bir takım sorunların ve eksiklerin olduğu gözlenmektedir. Bu sorunları ve eksiklikleri belirleyip ölçek geliştirmek veya uyarlama yapmak isteyen araştırmacılara ışık tutmak adına alanyazında yürütülmüş bazı çalışmalar bulunmaktadır. Bu çalışmaların bir kısmı hem ölçek geliştirme hem de uyarlama çalışmalarını incelerken (Acar Güvendir ve Özer Özkan, 2015; Çüm ve Koç, 2013; Erkuş, 2007) sadece uyarlama (Boztunç Öztürk, Eroğlu ve Kelecioğlu, 2014) ya da sadece ölçek geliştirme çalışmalarına odaklanan (Tavşancıl, Güler ve Ayan, 2014) çalışmalarda bulunmaktadır. Bu çalışma kapsamında özellikle ölçek geliştirmeye odaklanan çalışmalar incelenmiştir.

Çüm ve Koç (2013) 2005-2013 yılları arasında yayımlanan 29 ölçek geliştirme makalesini, ölçek geliştirme adımları açısından incelemişlerdir. Araştırma sonucunda, çalışmaların %67’sinde ölçek geliştirme adımlarına uygun bilgilerin rapor edildiği belirtilmiştir. Erkuş (2007) çalışmasında, ölçek geliştirme ve uyarlama çalışmalarındaki genel sorunlar üzerinde durmuştur. Çalışmanın amacını ise sadece bu konuda bir farkındalık yaratıp bu yönde bir takım girişimlerin başlatılabilmesi olarak ifade etmiştir. Çalışma kapsamında çalışmanın başlığı, ölçülmek istenen örtük değişkenin kavramsal tanımı, madde üretilmesi, deneme uygulaması, madde analizi, güvenilirlik, geçerlik, puanların yorumlanması ve çalışmanın raporlaştırılması başlıkları ile ilgili sorunlar belirtilip olası nedenler ve çözüm önerilerinden de bahsedilmiştir.

Tavşancıl, Güler ve Ayan (2014) eğitim ve psikoloji alanlarında yürütülen çalışmalardaki tutum ölçeklerinin geliştirilme adımlarının karşılanma durumlarını ortaya koymaya çalışmışlardır. Bu amaç doğrultusunda 54 tutum ölçeği belirlenen ölçütlere göre incelenmiş ve alanyazında yer alan çok sayıda yanlış ve eksik tutum ölçeğinin olduğu sonucuna ulaşılmıştır. Bunun yanı sıra, yürütülen bu çalışma ile aynı konuda geliştirilmiş birden fazla ölçeğin tespiti de araştırmacılar için emek ve zaman kaybı olarak belirtilmiştir.

Acar Güvendir ve Özer Özkan (2015) ise 2006-2014 yılları arasında “Social Science Citation Index” (SSCI) dizininde taranan Türkiye’de eğitim alanında yayımlanan üç dergideki 26 ölçek geliştirme makalesini süreçte izlenen adımlara göre incelemişlerdir. Elde edilen bulgulara göre çalışmaların tümünde iç tutarlılık güvenilirlik belirleme yöntemlerinden Cronbach alfa güvenilirlik katsayısının kullanıldığı tespit edilmiştir. Ayrıca, incelenen çalışmaların çok azında geliştirilen ölçeğin uygulama yönergesinin bulunduğu belirlenmiştir.

Bu çalışmada da Türkiye’de eğitim alanında SSCI dizininde yer alan üç dergide 2010-2015 yılları arasında yayınlanan ölçek geliştirme makaleleri, “ölçek geliştirme süreci”ndeki adımları ne derece sağladıkları açısından, dört ölçme ve değerlendirme alan uzmanı tarafından incelenmiştir. Uzmanlardan elde edilen ölçme sonuçlarının güvenilirliğinin belirlenmesinde çok yüzeyle Rasch modelinden yararlanılmıştır.

Çok Yüzeyle Rasch Modeli

Madde Tepki Kuramı (MTK)'na dayalı modellerden biri olan Rasch Modeli bir parametrelili (iki yüzeyle) bir modeldir. Ölçme sonuçlarının, birey ve madde yüzeyleri dışında farklı değişkenlik kaynakları (puanlayıcılar, farklı ölçme oturumları vb.) tarafından etkilendiği ölçme durumlarında ise "Çok Yüzeyle Rasch Modeli"nden yararlanır. Çok yüzeyle Rasch Modeli (ÇYRM)'de yer alan her bir değişkenlik kaynağına "yüzeyle" adı verilir. Bu sebeptendir ki birden fazla değişkenlik kaynağının yer aldığı ölçme durumlarında ele alınan bu model "Çok Yüzeyle Rasch Modeli" olarak ifade edilir. ÇYRM, Klasik Test Kuramı (KTK)'nın sınırlı kaldığı noktaların üstesinden gelebilen bir ölçme modelidir (Anshel ve diğerleri., 2009; Baştürk ve Işıkoğlu, 2008; Güler, 2014; Güler ve Gelbal, 2010; Kim, Park ve Kang, 2012). MTK'ya dayalı tüm modellerde olduğu gibi ÇYRM'de de her bir yüzeyle (birey, madde, puanlayıcı, durum vb.) ilişkin kestirimler, diğer değişkenlik kaynaklarından bağımsız yapılabilmektedir. Örneğin; bireylere ilişkin elde edilen ölçümler, ölçme aracındaki maddelerin güçlük düzeylerinden, puanlayıcıların katılık/cömertliklerinden vb. değişkenlik kaynaklarından; madde parametreleri uygulandığı bireylerin ölçülen özelliğinden, puanlayıcıların katılık/cömertlik düzeylerinden ve ölçme sonuçlarını etkileyebilecek diğer değişkenlik kaynaklarından bağımsız olarak kestirilebilir (Engelhard, 1994). KTK'da bireylerin bir testteki yetenek düzeyleri, test maddelerinden aldıkları puanların toplamı ile kestirilir. Testteki her bir maddenin güçlük düzeylerinin (ya da bir maddeye katılma ya da katılmama olasılıklarının) birbirine eşit ve/veya toplam puana olan katkısının aynı olduğu varsayılır. Bununla birlikte her madde, ölçülen özellikte farklı bir katkıya sahipse her bir maddenin katkısını, toplam puanda eşit kabul etmek yanlış sonuçlara sebep olur ve bu kabul üzerinden yapılan istatistikler şüphe içerir (Anshel ve diğerleri, 2009, Brinthaup ve Kang, 2012). KTK'da ham puanlardan elde edilen sonuçlar üzerinden bireyler yetenek vb. düzeylerine göre ancak sıralanabilir; sıralama ölçüğünde elde edilen bu puanlar ise toplanabilir değildir. Hâlbuki ÇYRM'nin dayandığı matematiksel model bu sınırlılığın üstesinden gelmeye yardımcı olur. Bu matematiksel model ile ham verilerin doğal logaritması alınır (log-odds) ve böylece ölçme sonuçları, eşit aralıklı ölçek (logit) düzeyine dönüştürülür. KTK ile kayıp verilerin bulunduğu ölçme durumlarında, ham veriler üzerinden yapılan analizler hatalı sonuçlar üretebilirken; ÇYRM'de kayıp veriler göz önünde bulundurulmaz, sadece gözlenen verilere dayalı analizler gerçekleştirilir. Bunlara ek olarak, ÇYRM'de tek bir analiz ile her bir değişkenlik kaynağı için uyum istatistikleri (INFIT ve OUTFIT) belirlenebilir. Uyum istatistikleriyle, model-veri arasındaki uyumun derecesini inceleyebilmek mümkün olmaktadır (Revesz, 2012; Smith ve Kulikowich, 2004). Ayrıca, her bir yüzeyle ilişkin parametre kestirimleri ortak bir cetvel (logit cetvel) üzerinde birlikte yorumlanabilir (Linacre, 1989). Ortak bir metriğe sahip bu cetvelde yüzeylerin görece yerleri incelenebilir. Böylece, örneğin; maddelerin dağılımı gözlenerek, yetenek düzeyi boyunca hangi düzeyde madde yok/eksik kalmış, hangi düzeyde çok sayıda madde yer almış belirlenebilir (Brinthaup ve Kang, 2012). Bunun yanı sıra, çalışmada yer alan diğer yüzeylerle (örneğin; puanlayıcılarla) ilgili de açıklayıcı bilgiler sunar. Örneğin, birden fazla puanlayıcının yer aldığı bir ölçme durumunda, bir puanlayıcı diğerlerinden daha cömert puanlama yaparken; diğer tüm puanlayıcıların yüksek puan verip, bu puanlayıcının daha düşük puan verdiği bu "beklenmedik puanlama durumu" tespit edilebilir. Ayrıca, ÇYRM ile puanlayıcıların katılık/cömertlik düzeylerine ilişkin bilgi almak mümkün olduğu gibi halo etkisi, merkeze yönelme ve/veya yanlılık gibi diğer puanlayıcılardan kaynaklanan hatalar da belirlenebilir (İlhan, 2016). Böylece bu tür incelemelerle ÇYRM; puanlamada alınabilecek önlemlere, puanlayıcılara verilecek eğitime vb. ilişkin aydınlatıcı bilgiler sunar.

ÇYRM'de her bir yüzeyle için elde edilen diğer istatistikler ise ayırma indeksi ve güvenilirlik katsayısıdır. Ayırma indeksi, 1 ile ∞ arasında değer alabilirken güvenilirlik katsayısı KTK'da da olduğu gibi 0 ile 1 arasında değişmektedir (Sudweeks, Reeve ve Bradshaw, 2005). Birey ve madde yüzeyi için ayırma indeksi ve güvenilirlik katsayısının yüksek değerlere sahip olması istenilen bir durum olup puanlayıcı yüzeyi için bu değerlerin düşük olması istenir. Birey yüzeyi için güvenilirlik değeri, Cronbach α değerine benzer yorumlanmaktadır ve her iki istatistiğin yüksek değerlere sahip olması bireylerin birbirinden ayırt edilebildiğine işaret eder. Madde yüzeyi için elde edilen ayırma indeksi ve güvenilirlik katsayısının yüksek çıkması farklı özellikleri ölçen maddelerin birbirinden

ayrıtılabildiğini gösterirken; puanlayıcı yüzeyi için hesaplanan bu iki değerin yüksek olması ise puanlayıcıların bireyleri oldukça farklı puanladıklarını ifade eder (İlhan, 2016).

Araştırmanın Amacı

Bu çalışmada, Türkiye’de 2010-2015 yılları arasında SSCI’da yer alan bilimsel dergilerde yayımlanmış ölçek geliştirme çalışmalarının, “ölçek geliştirme süreci”ndeki adımları ne derece sağlayabildiklerine ilişkin bilgi sunmak amaçlanmıştır. Bu amaç için öncelikle, elde edilen verilerin ÇYRM ile her bir yüzey için güvenilirliği incelenmiş, sorun teşkil eden durumlar gözden geçirilmiş ve sonrasında elde edilen yeni veriler analiz edilmiştir. Böylece, makaleleri değerlendiren puanlayıcı yüzeyinin, puanlamada kullanılan ölçütlerin ve incelenen makalelerin düzeyleri ve güvenilirlikleri belirlenmiştir. Ardından, çalışmada yer alan “ölçek geliştirme” makalelerinin belirlenen ölçütler açısından eksik ve güçlü yönleri tanımlanmaya çalışılmıştır. Alanyazında yer alan, ölçek geliştirme makalelerinin incelendiği çalışmalara destekleyici bilgiler sunmanın yanı sıra yapılan bu çalışmada ÇYRM ile tüm yüzeylerin analiz edilerek; özellikle çalışmada yer alan makalelerin ve belirlenen ölçütlerin eşit aralıklı ölçek düzeyindeki yerleri belirlenmiştir. Elde edilen bu bulgular ile incelenen tüm çalışmalar aldıkları puanlar açısından, ayrıca incelenen ölçütler karşılama oranları bakımından sıralanmıştır. Çalışmaların ve ölçütlerin cetvel üzerinde yerlerinin belirlenmesinin alan yazına bu ölçek geliştirme çalışmalarının değerlendirilmesi açısından katkı sağlayacağı düşünülmektedir. Buna ek olarak, ÇYRM’nin benzer ölçme durumlarında kullanılabileceğine ilişkin örnek bir çalışma olabilmesi amaçlanmıştır.

YÖNTEM

Bu çalışmada Türkiye’de eğitim konusunu temel alan ve SSCI kapsamında taranan üç derginin (Eğitim ve Bilim, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Kuram ve Uygulamada Eğitim Bilimleri Dergisi) 2010-2015 yılları arasında yayınladığı ölçek geliştirme makaleleri incelenmiştir. Araştırmada yayımlanmış çalışmalar var olduğu şekilde tanımlandığı için çalışma tarama modelinde betimsel bir araştırmadır (Karasar, 1998).

Örnekleme/Çalışma Grubu

Araştırma kapsamında Türkiye’de eğitim alanında SSCI dizininde yer alan üç dergide 2010-2015 yılları arasında yayınlanan ölçek geliştirme makaleleri çalışmanın evrenini oluşturmaktadır. Bu evrenden amaçlı örnekleme yöntemlerinden biri olan ölçüt örnekleme yolu ile örneklem belirlenmiştir. Ölçüt örnekleme yönteminde bir takım ölçütler önceden belirlenir ve bu ölçütleri karşılayan tüm durumlar çalışmaya katılır (Yıldırım ve Simsek, 2008). İncelenen makalelerin çevrim içi tam metninin ulaşılabilir olması belirlenen ölçütlerden biridir. Çalışmaya, esas amacı ölçek geliştirmek olmayan ancak içinde veri toplama aracı olarak geliştirilmiş ölçek olan çalışmalar, bilgilerin sınırlı olabileceği gerekçesi ile dâhil edilmemiştir. Çalışmaya sadece tipik davranışları ölçmek amacıyla geliştirilen ölçekler ve bu ölçeklerin geliştirilme süreçlerinin rapor edildiği çalışmalar dâhil edilmiştir. Bu bağlamda SSCI kapsamında olan ve eğitim içerikli üç dergide (Eğitim ve Bilim Dergisi, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi ve Kuram ve Uygulamada Eğitim Bilimleri Dergisi) 2010-2015 yılları arasında tam metnlerine ulaşılabilen 57 ölçek geliştirme makalesi, çalışmanın örneklemini oluşturmaktadır. Çalışmaların yıllara göre dağılımı Tablo 1’de verilmiştir.

Tablo 1. Çalışmada kullanılan makalelerin yıllara göre dağılımı

	2010	2011	2012	2013	2014	2015	Toplam
Eğitim ve Bilim Dergisi	4	2	2	3	10	2	23
Hacettepe Üniversitesi Eğitim Fakültesi Dergisi	3	4	2	2	1	2	14
Kuram ve Uygulamada Eğitim Bilimleri Dergisi	-	6	5	7	2	-	20
Toplam	7	12	9	12	13	4	57

Tablo 1’de ölçek geliştirme çalışmaları en çok 2014 yılında yapılmışken, en az çalışmanın 2015 yılında yayımlandığı görülmektedir. Ayrıca çalışmada incelenen 57 makalenin 23’ü Eğitim ve Bilim 14’ü Hacettepe Üniversitesi Eğitim Fakültesi ve 20’si Kuram ve Uygulamada Eğitim Bilimleri dergisinde yayınlanmıştır.

Veri Toplama Araçları

Araştırmada, ölçek geliştirmenin adımları ve gereklilikleri konusunda yayınlanmış kaynakların (Cohen ve Swerdlik, 2010; Edenborough, 1999; Erkuş, 2012; Murphy ve Davidshofer, 2005; Tezbaşaran, 2008) ve ölçek geliştirme makalelerinin incelendiği çalışmaların (Acar Güvendir ve Özer Özkan, 2015; Tavşancıl, Güler ve Ayan, 2014; Çüm ve Koç, 2013; Erkuş, 2007) taraması yapılmıştır. Ölçek geliştirme adımları hakkında alanyazında olan genel görüşler dikkate alınarak araştırmacılar tarafından “Ölçek Geliştirme Makaleleri İnceleme Formu” oluşturulmuştur. Bu formun hazırlanmasında alanyazında bulunan ölçek geliştirme makalelerinin incelendiği çalışmalarda kullanılan ölçütlerden de yararlanılmıştır. Ancak bu ölçütlerden bir kısmının her ölçek geliştirme çalışmasında rapor edilmesinin zorunlu olmadığı düşünüldüğünden daha önceki çalışmalarda esas alınan ölçütlerin bir kısmı bu çalışma kapsamında geliştirilen forma dâhil edilmemiştir. Örneğin, Acar Güvendir ve Özer Özkan’ın (2015) yürüttükleri çalışmada “iki eş değer yarı güvenilirliğinin hesaplanmış olması” ölçek geliştirme çalışmalarının incelenmesinde kullanılan ölçütlerden biri olarak belirlenmiştir. Bu çalışmada ise iki yarı güvenilirliğinin her ölçek geliştirme çalışmasında rapor edilmesinin zorunlu olmadığı düşünülmüş ve bu sebeple ölçek geliştirme çalışmalarının incelemek üzere geliştirilen formda böyle bir ölçüte yer verilmemiştir. Benzer şekilde, Delice ve Ergene’nin (2015) matematik eğitimi alanındaki ölçek geliştirme çalışmalarını inceledikleri araştırmada, ölçeğin geneline ilişkin Cronbach alfa güvenilirlik katsayısının rapor edilmesi gibi bir ölçüt esas alınmaktadır. Ancak tüm ölçeklerde her zaman toplam puan almanın mümkün olmayabileceği ve dolayısıyla ölçeğin geneline ilişkin Cronbach alfa katsayısının rapor edilmesinin her ölçek için bir zorunluluk olmadığı düşünüldüğünden bu çalışma kapsamında ölçek geliştirme çalışmalarının incelenmesinde kullanılan forma, bu tür bir ölçüt dâhil edilmemiştir.

Oluşturulan formun kapsamını değerlendirmesi için dört ölçme ve değerlendirme, dilsel uygunluğu açısından değerlendirme yapması için ise iki dilbilgisi uzmanından görüş alınmıştır. Alınan görüşler doğrultusunda form tekrar düzenlenerek kullanıma hazır hale getirilmiştir.

Elde edilen ölçek geliştirme makaleleri inceleme formu 19 maddeden oluşmaktadır ve ilk 17 maddesi 0 - 3 arası puanlanabilir ölçütlerdir. Bu ölçütlerdeki puanlama, “(0) bilgi yok; (1) bilgi var ancak hiç açıklanmamış; (2) bilgi var ve kısmen açıklanmış; (3) bilgi var ve gerekli şekilde açıklanmış” olarak tanımlanmıştır. Diğer iki madde de ise geliştirilen ölçeğin çalışmada yer alıp almadığı ve ölçeğin puanlanması ile ilgili bilginin verilip verilmediği sorularak; bu maddeler evet-hayır şeklinde cevaplanabilir biçimde oluşturulmuştur.

Verilerin Analizi

Çalışmada Çok Yüzeysel Rasch Analizi kullanılmıştır. ÇYRM kavramsal olarak regresyon analizine benzerdir. Bağımlı değişken, bireyin (ya da ölçmeye konu olan objenin) bir maddeden alabileceği

puanların olasılıkları oranının lojistik dönüşümü, bağımsız değişkenler ise ölçmede yer alan madde güçlük düzeyi, puanlayıcı katılık/cömertlik düzeyi gibi diğer yüzeylerdir (Randall ve Engelhard, 2009). Böylece, bu çalışmadaki gibi makale, ölçütler ve puanlayıcı gibi üç yüzeyin yer aldığı bir ÇYRM Eşitlik 1’deki gibi ifade edilebilir:

$$\text{Log} (P_{mkpc} / P_{mkpc-1}) = B_m - D_k - C_p - F_c \quad (1)$$

P_{mkpc} : m. makalenin k. ölçütte p. puanlayıcı tarafından c kategorisinde puan alma olasılığı

P_{mkpc-1} : m. makalenin k. ölçütte p. puanlayıcı tarafından c-1 kategorisinde puan alma olasılığı

B_m : m. makalenin esas alınan ölçütü karşılama düzeyi

D_k : k. ölçütün güçlük düzeyi

C_p : p. puanlayıcının katılık/cömertlik düzeyi

F_c : c. kategoriden c-1. kategoriye geçişteki güçlük düzeyi

ÇYRM’de bireylerin maddeleri cevaplama olasılıkları (bu çalışmada, makalelerin esas alınan ölçütleri karşılama olasılıkları), yukarıda da değinildiği gibi, “log-odds” olarak ifade edilir ve lojistik cetvelde “log-odds” ya da “logit” birimlerle gösterilir. Lojistik cetvelde artan pozitif değerler bireyler için yüksek yetenek düzeyine (bu çalışma için makalelerin, belirlenen ölçütleri karşılama düzeylerinin yüksek oluşu), maddeler için yüksek güçlük düzeyine ve puanlayıcılar için yüksek katı puanlama düzeyine karşılık gelir. Lojistik cetvelin, her bir yüzeyi görsel olarak inceleme olanağı sağlıyor olması; her yüzeydeki elemanların sırasının, yüzeylerin birbirleriyle olan göreceli durumlarını ve bir yüzeydeki elemanlar arası farklılıkların ne büyüklükte olduğunu gözleme imkânı sunar (Güler, 2014; Hetherman, 2004).

Çalışmada yer alan 57 makalenin tümü dört ölçme ve değerlendirme uzmanı tarafından ölçek geliştirme makaleleri inceleme formu kullanılarak puanlanmıştır. Formda yer alan ilk 17 değerlendirme düzeyinin analizi, Çok Yüzeyle Rasch Modeli (ÇYRM) ile Linacre (2007) tarafından geliştirilen FACETS programı kullanılarak gerçekleştirilmiştir. Analiz sonucunda, çalışmada yer alan üç yüzeyin (“makalelerin belirlenen ölçütleri karşılama düzeyi”, “değerlendirme formunda kullanılan ölçütler” ve “puanlayıcılar”) ve değerlendirmedeki ölçütlerin puanlama düzeylerinin birlikte yer aldığı lojistik cetvel sunulmuştur. Ardından tüm yüzeylere ilişkin ölçüm raporları verilmiştir. Ölçüm raporunda yer alan tüm yüzeylere dair uyum-içi ve uyum-dışı istatistiklerinin istenilen değer aralığı olan 0.5 ile 1.5 arasında olması (Turner, 2003; Wright ve Linacre, 1992), ölçüm sonuçlarının tek boyutlu bir yapı gösterdiğine işaret etmektedir (Anshel ve diğerleri; 2009; Lee ve diğerleri, 2010). Uyum istatistiklerine ek olarak, her bir yüzey için elde edilen güvenilirlik değerleri ve ayırma oranları da yorumlanmıştır.

ÇYRM’nin FACETS programı ile analizinde “beklenmedik tepkilerin (unexpected responses)” incelenmesini sağlayan bir tablo da yer alır. Bu tabloda var olan bilgiler, güvenilirliği olumsuz etkileyen beklenmeyen puanların incelenebilmesine imkân tanır (Güler, 2014). Örneğin, dört puanlayıcının yer aldığı bir ölçme çalışmasında, üç puanlayıcı bir makalenin incelendiği formdaki bir maddesine çok yüksek puan verirken diğer puanlayıcı aynı makalenin aynı maddesine en düşük puanı vermişse, bu durumu gözlemek bu tablo ile mümkün olmaktadır. Bu tabloda ayrıca standartlaştırılmış artık değerler de rapor edilir. Bu değerler, model ile verinin uyumlu olup olmadığı hakkında karar vermemize yardımcı olur (İlhan, 2016). Alanyazında, model-veri uyumunun olabilmesi için ± 2 sınırları dışında kalan standartlaştırılmış artık değerler sayısının toplam veri sayısının yaklaşık %5’ini ve ± 3 sınırları dışında kalan standartlaştırılmış artık değerleri sayısının ise toplam veri sayısının yaklaşık %1’ini geçmemesi gerektiği belirtilmektedir (Linacre, 2014). Bu çalışmada, yer alan toplam veri sayısı 3876 olup (57 ölçek x 17 madde x 4 puanlayıcı = 3876); ± 3 aralığı dışında kalan standartlaştırılmış artık veri sayısı (31) oranı %0.8’dir. Bu sonuca göre, çalışmada yer alan verilerin model ile uyumlu olduğu söylenebilir. Ancak, ölçekler ile ilgili yapılacak çıkarımların en güvenilir sonuçlar üzerinden yapılabilmesi için analiz sonucu elde edilen 31 beklenmedik veri tekrar incelenmiş, problemleri puanlamaların görüldüğü makaleler tekrar puanlayıcılara bildirilmiş ve yeniden puanlamaları istenmiştir. Yeni elde edilen veriler üzerinden

yapılan analizle de (alanyazında da yer aldığı üzere) beklenmedik verilerin çok az olmasından dolayı, tüm bulguların aynı değerlere sahip olduğu, sadece makalelere ilişkin elde edilen ölçüm raporundaki güvenilirlik değerinin 0.85'ten 0.88'e yükseldiği gözlenmiştir.

Bu analizlerin ardından, çalışmanın kapsamında yer alan 57 ölçeğe ilişkin elde edilen logit puanlar, yöntem kısmında detaylı olarak açıklandığı şekilde değerlendirme düzeyleri olan 0-3 değerlerine dönüştürülerek yorumlanmıştır. Benzer süreç, ayrıca değerlendirme formunda yer alan ölçütler için de gerçekleştirilmiş ve her ölçüt [0, 1] puan aralığında karşılık gelen değeri üzerinden incelenmiştir. Bu işlemler için Excel bilgisayar programından yararlanılmıştır.

Son olarak çalışmada kullanılan makale inceleme formunda yer alan "Geliştirilen ölçek çalışmada yer almış." ve "Geliştirilen ölçeğin nasıl puanlanacağı hakkında bilgi verilmiş." ifadelerine puanlayıcıların verdikleri *Evet/Hayır* cevaplarına ilişkin verilere SPSS 20 programı kullanılarak frekans ve yüzde hesaplanmıştır.

BULGULAR

Verilerin ÇYRM ile Analizi

Bu bölümde araştırma sonucu elde edilen bulgulara yer verilmiştir. Öncelikle, FACETS programı ile gerçekleştirilen ÇYRM analizi sonuçları verilmiştir. Şekil 1'de, düzeltilmiş verilere ilişkin elde edilen logit-cetvel sunulmuştur.

Measr	+MAKALE	-MADDE	-PUANLAYICI	Scale
2				(3)
1	14			
	3			
	16 36	16		---
	38 44			
	53	14		
	24 25 28 40 43	5	4	2
	5 12 39 50 57	7		
	1 8 10 13 29 37 41 46 47 51		3	
*	6 15 21 22 42 45 49 55 56	*	* 2	* --- *
	7 19 30 32 52	9		
	9 18 20 48			
	17 31 34 35	15		1
	54		1	
	2 11 23 26			
	4	8		---
-1	33	11		
	27	6		
		4		
		2		
		17		
		13		
		1	10 12	
		3		
-2				(0)
Measr	+MAKALE	-MADDE	-PUANLAYICI	Scale

Şekil 1. Üç yüzey için Logit Cetvel

Şekil 1'in ilk sütununda logit-cetvelin ölçüm birimi (logit) gözlenmektedir. Çalışmada yer alan tüm düzeyler, bu ölçüm birimi üzerinden yorumlanır. Cetvelin ikinci sütununda ölçekler, belirlenmiş olan

ölçütleri karşılayabilme düzeyleri; yukarıdan aşağıya doğru en yüksekten en düşüğe doğru sıralanmıştır. Böylece, çalışmada esas alınan ölçütleri karşılayabilme düzeyi en yüksek 14 numaralı makalenin (1.13) ve en düşük 27 (-1.01) numaralı makalenin olduğu gözlenmektedir. Makalelerin düzeylerinin logit cetvelinde geniş bir aralıkta dağılım göstermiş olması birbirlerinden güvenilir bir şekilde ayrılabildiklerinin de bir göstergesi olmaktadır. Cetvelin üçüncü sütununda ise değerlendirme formunda yer alan ölçütlerin güçlük düzeylerine göre yukarıdan aşağıya doğru, güçlük düzeyi en yüksek olandan en düşük olana doğru sıralandığı görülmektedir. Buna göre, 16. ölçütün (*Ölçüt geçerliği analizi sonucu elde edilen değerler uygundur.*) güçlük düzeyi en yüksek (karşılanması en zor) (0.72) iken 3. ölçütün (*Ölçeğin amacına uygun bir hedef kitle belirlenmiştir.*) güçlük düzeyi en düşüktür (karşılanması en kolay) (-1.58). Şekil 1’in dördüncü sütununda puanlayıcılara ait ölçümler yer almaktadır. Bu sütunda yer alan veriler, en yüksek pozitif değere karşılık en katı puanlayıcı, en düşük negatif değere karşılık en cömert puanlayıcı gelecek şekilde yorumlanır. Böylece, çalışmada yer alan 4. puanlayıcının (0.35 logit) en katı, 1. puanlayıcının (-0.51 logit) ise en cömert puanlayıcı olduğu gözlenmektedir.

Logit cetvel, tüm yüzeylerin göreceli durumlarının doğrusal bir metrik üzerinden incelenmesine imkan tanımakla birlikte, yüzeylerin her birine ilişkin daha detaylı bilgiler ölçüm raporları ile aşağıda yer alan tablolar üzerinden açıklanmıştır. Tablo 1’de makaleler için elde edilen ölçüm raporları sunulmuştur.

Tablo 2. Makalelere İlişkin Ölçüm Raporu

	Logit Ölçüsü	Std. Hata	Uyum-İçi	Uyum-dışı
Ortalama	0.00	0.14	1.02	0.91
Standart Sapma	0.40	0.02	0.33	0.39
Güvenirlik = 0.88	Ayrırma Oranı = 2.65	Ki-kare = 444.9	sd = 56	p = 0.00

Tablo 2’deki değerler incelendiğinde, makalelerin belirlenen ölçütlere sahip olma düzeyleri ortalamasının 0 ve standart sapmasının 0.40 logit puan olduğu gözlenmektedir. Uyum-İçi ve uyum-dışı istatistikleri ise sırasıyla, 1.02 ve 0.91 olarak elde edilmiştir. Uyum istatistikleri, model ile veri arasındaki uyumun derecesini gösterir. Model ile veri arasında mükemmel bir uyum olduğunda bu istatistik değerleri 1 olmaktadır. Bunun yanı sıra; uyum-dışı istatistiği, beklenmeyen uç değerlere karşı uyum-İçi istatistiğine göre çok daha duyarlıdır. Uyum-İçi istatistiğinin 1’in üzerinde olması beklenenden daha yüksek, 1’den düşük olması ise beklenenden daha düşük bir varyansın gözlemlendiğine işaret eder (Güler, 2014; Hetherman, 2004). Alanyazında uyum istatistiklerine ilişkin kabul edilebilir değerlerin, 0.6 ile 1.5 (Lunz, Wright ve Linacre, 1990) ya da 0.5 ile 1.5 (Wright ve Linacre, 1994; Turner, 2003) arasında olabileceği belirtilmiştir. Buna göre, makale ölçüm raporunda yer alan uyum istatistikleri ortalamasının 1’e çok yakın olması, model-veri uyumunun oldukça yüksek olduğuna işaret etmektedir. Tablo 2’deki ayırma oranı ve güvenilirlik değerleri incelendiğinde; bu değerlerin sırasıyla 2.65 ve 0.88 olduğu gözlenmektedir. Buradaki güvenilirlik değeri, Cronbach alfa güvenilirlik değeri gibi 0 ile 1 arasında bir değer alır ve benzer şekilde yorumlanır. Elde edilen değerler oldukça yüksek çıkması, makalelerin belirlenen ölçütler açısından yeteri derecede ayırt edilebildiğini göstermektedir.

Makalelerin değerlendirilmesinde belirlenen ölçütlere ilişkin ölçüm raporu sonuçları Tablo 3’te verilmiştir.

Tablo 3. Değerlendirme Ölçütlerine İlişkin Ölçüm Raporu

Ölçütler	Ölçüm	Std. Hata	Uyum İçi	Uyum Dışı
Ö16	0.72	0.07	1.39	1.17
Ö14	0.46	0.06	1.16	1.24
Ö5	0.37	0.06	0.85	0.96
Ö7	0.25	0.06	1.28	1.31
Ö9	-0.07	0.06	1.10	0.73
Ö15	-0.34	0.06	1.19	1.18
Ö8	-0.74	0.07	0.93	0.87
Ö11	-0.76	0.07	1.06	1.20
Ö6	-0.87	0.08	0.73	0.75
Ö4	-0.99	0.08	0.61	0.87
Ö2	-1.12	0.09	0.55	0.80
Ö17	-1.24	0.10	0.79	0.80
Ö13	-1.35	0.11	0.95	1.07
Ö12	-1.47	0.12	1.04	1.11
Ö10	-1.53	0.12	1.49	1.16
Ö1	-1.55	0.12	0.50	0.92
Ö3	-1.58	0.13	0.68	0.93
Ortalama	-0.69	0.09	0.96	0.91
Standart Sapma	0.78	0.03	0.29	0.25
Güvenirlilik = 0.99	Ayırma Oranı = 8.69	Ki-kare = 1465.6	sd = 16	p = 0.00

Tablo 3'te yer alan ölçütlere ilişkin ölçüm raporu incelendiğinde, ölçütlerin güçlük düzeylerine göre en zor 0.72 logit değer ile Ö16 (*Ölçüt geçerliği analizi sonucu elde edilen değerler uygundur.*) ve en kolay -3.14 logit değer ile Ö3 (*Ölçeğin amacına uygun bir hedef kitle belirlenmiştir.*) ölçütünün olduğu gözlenmektedir. Bu ölçütlere ilişkin uyum-İçi ve uyum-dışı istatistikleri sırasıyla 0.50 ile 1.49 ve 0.80 ile 1.24 arasında değişen değerler almaktadır. Uyum istatistikleri için istenilen değer aralığının 0.5 ile 1.5 olup bu aralık dışında kalan istatistikler zayıf uyuma işaret etmektedir (Turner, 2003; Wright ve Linacre, 1994). Buna göre, model ile veri uyumunu olumsuz etkileyen bir ölçütün bulunmadığı söylenebilir. Ölçütlere ilişkin güvenirlilik değerinin 0.99 ile oldukça yüksek bir değer elde edilmesiyle de değerlendirmede kullanılan ölçütler, makalelerin belirlenen ölçütlere sahip olma düzeylerine göre ayırmada güvenilir bir şekilde işlev göstermişlerdir şeklinde yorumlanabilir.

Makaleleri belirlenen ölçütlere göre değerlendiren puanlayıcılara ilişkin ölçüm raporu sonuçları Tablo 4'te verilmiştir.

Tablo 4. Puanlayıcı Ölçüm Raporu

Puanlayıcılar	Ölçüm	Std. Hata	Uyum İçi	Uyum Dışı
P4	0.35	0.03	1.00	1.06
P3	0.15	0.04	0.86	0.75
P2	0.01	0.04	1.07	0.94
P1	-0.51	0.04	1.19	0.88
Ortalama	0.00	0.04	1.03	0.91
Standart Sapma	0.37	0.00	0.14	0.13
Güvenirlilik = 0.99	Ayırma indeksi = 9.77	Ki-kare = 260.4	sd = 3	p = 0.00

Tablo 4'te, puanlayıcıların ölçüm raporuna göre güvenirlilik ve ayırma indeksi değerlerinin sırasıyla; 0.99 ve 9.77 olduğu gözlenmektedir. Oldukça yüksek olan güvenirlilik değeri, puanlayıcıların puanlamalarının benzerliğinin değil farklılığının güvenirlilik değeridir (İlhan, 2016). Buna ek olarak, yüksek olan ayırma indeksi de puanlayıcıların puanlamaları arasındaki farklılığın düzeyini ifade

eder. Puanlayıcılar tam bir tutarlılık ile puanlama yaptıklarında ayırma indeksinin alacağı değer 0 olmaktadır. Böylece, bu iki değer göz önüne alındığında, çalışmada yer alan puanlayıcıların puanlamadaki katılık ve cömertlik düzeylerinin farklılık gösterdiği yorumu yapılabilir. Tablo 4’teki ölçüm değerleri incelendiğinde, P4 numaralı puanlayıcının en katı ve P1 numaralı puanlayıcının ise en cömert puanlama yaptıkları gözlenmektedir. Ancak, gözlenen uyum-içi ve uyum-dışı istatistiklerinin kabul edilebilir aralık olan 0.5 - 1.5 değerleri arasında olması, puanlayıcıların puanlarının modelle uyumluluğu olduğuna işaret eder (Wright ve Linacre, 1994). Buna göre, model-veri uyumunu olumsuz etkileyen bir puanlayıcının olmadığı söylenebilir.

Makale değerlendirme aşamasında kullanılan değerlendirme formunda yer alan ölçütlerin ölçüm raporu sonuçları Tablo 5’te verilmiştir.

Tablo 5. Değerlendirme Formu Ölçütlerinin Ölçüm Raporu

Puanlama düzeyleri	Frekans	%	Yığılmalı %	Ortalama Ölçüm	Beklenen Ölçüm	Uyum Dışı
0	823	21	21	-0.39	-0.37	0.70
1	193	5	26	0.11	0.09	0.70
2	445	11	38	0.74	0.62	1.10
3	2413	62	100	1.10	1.12	1.20

Tablo 5’te, değerlendirme formunda bulunan ölçütlerin frekans ve yüzde dağılımları incelendiğinde 0 - 3 arasında yapılan değerlendirmede “1” düzeyinin (bilgi var ama hiç açıklanmamış) en az kullanıldığı, bunu sırasıyla “2” (bilgi var ve kısmen açıklanmış) ve “0” (bilgi hiç yok) düzeylerinin takip ettiği gözlenmektedir. Değerlendirme formunda yer alan “3” düzeyinin (bilgi var ve gerekli şekilde açıklanmış) ise en çok kullanılan düzey olduğu belirlenmiştir. Düzeylerin işlevini yerine getirdiğinin söylenebilmesi için puanlama düzeylerindeki dağılımın dengeli olması ve her düzeyde en az 10 katılımcının bulunmuş olması ve uyum-dışı istatistiklerinin de 0.5 ile 1.5 aralığında değerler almış olması beklenir (İlhan, 2016; Linacre, 2014). Buna göre, düzeylere ilişkin frekans değerleri incelendiğinde tüm düzeyler için istenilen değerlere ulaşılabilirdiği ve uyum-dışı istatistiklerinin de istenilen aralıkta yer aldığı gözlenmektedir. Elde edilen bu sonuçlar, değerlendirme formunda yer alan ölçütlerin karşılanması beklenen varsayımları sağladığını, bir sorun olmadığını işaret etmektedir.

Makalelerin Belirlenen Ölçütleri Karşılama Düzeylerinin İncelenmesi

Verilerin ÇYRM ile analizi sonucunda makale yüzeyine ilişkin elde edilen logit ölçüm değerleri yukarıda sunulmuştur. Ancak Linacre (2014)’nin de ifade ettiği üzere, logit değerler ile rapor edilen makalelerin belirlenen ölçütleri karşılama düzeyleri tüm araştırmacılar ve okuyucular için anlaşılır olmayabilir. Bu nedenle, logit birim ile ifade edilen makale düzeyleri, değerlendirme formunda yer alan ölçüm düzeylerine dönüştürülmelidir. Tablo 6’da, alanyazında yer alan bu dönüştürme işleminin basamakları ayrıntılı olarak tanımlanmaya çalışılmıştır (İlhan, 2016; Linacre, 2014).

Tablo 6. Logit Birim ile İfade Edilen Makalelerin Düzeylerinin Değerlendirme Formundaki Düzeylere Dönüştürülmesi Süreci

Dönüşüm adımları	İşlemler
1. adım: En düşük logit değer ile en yüksek logit değer arasındaki fark hesaplanır (Bknz Şekil 1).	$1.13 - (-1.01) = 2.14$
2.adım: Değerlendirme formundaki değerlerin ranj değeri belirlenir.	$3 - 0 = 3$
3. adım: 2. adımda bulunan bu ranj değeri 1. adımda bulunan değere bölünür.	$3 / 2.14 = 1.402$
4. adım: 3. adımda bulunan değer, makalelere ilişkin elde edilen en düşük logit ölçümü ile çarpılır.	$-1.01 * 1.402 = -1.416$
5. adım: 4. adımda ulaşılan değeri değerlendirme formunun en küçük düzeyine (sıfıra) eşitleyecek sabit belirlenir.	Bu değer 1.416'dır.
6. adım: Her bir makale için elde edilen logit ölçümü 3.adımda elde edilen değer ile çarpılır ve bu değer ile 5. adımdaki değer toplanır.	$(\text{Logit puan} * 1.402) + 1.416$

Tablo 6'da yer alan süreç sonrası, incelenen makalelerden sadece birinin (% 1.7) değerlendirilme düzeyinin 3 olduğu gözlenmiştir. Bunu sırasıyla, 2.58 ile 2.20 arasında değerlendirilen 5 makale (% 8.8), 1.91 ile 1.01 arasında değerlendirilen 41 makale (% 72) izlemektedir. En düşük değerlendirme düzeyi olan 0.95 ile 0 arasında ise 10 makalenin (% 17.5) yer aldığı belirlenmiştir. Tablo 6'da yer alan adımlar, makale değerlendirme formunda yer alan değerlendirme ölçütleri için tekrarlanmış ve ölçüt düzeylerinin KTK'daki madde güçlük düzeyleri gibi 0 ile 1 arasındaki değerleri belirlenmiştir. Ancak burada KTK'dan farklı olarak 1'e en yakın ölçüt en zor, 0'a en yakın ölçüt ise en kolay ölçüt olarak yorumlanmaktadır. Ölçütlere ilişkin elde edilen bu değerler Tablo 7'de verilmiştir.

Tablo 7. Ölçütlerin Logit Değerlerinin [0, 1] Aralığındaki Değerleri

No	Ölçüt ifadeleri	Dönüştürülmüş değerler
1	Ölçülmesi amaçlanan değişken açıkça belirtilmiştir.	0.02
2	Yeni bir ölçeğe gereksinim duyulmasının nedenleri gerekçeleriyle belirtilmiştir.	0.21
3	Ölçeğin amacına uygun bir hedef kitle belirlenmiştir.	0.01
4	Madde havuzu uygun yöntemler kullanılarak oluşturulmuştur (Literatür tarama, Uzman görüşü alma, Kompozisyon yazdırma vb...)	0.26
5	Deneme uygulaması için üretilen madde sayısı, nihai ölçekte bulunması öngörülen madde sayısından daha fazla (en az 2-3 katı) olacak şekildedir.	0.85
6	Madde havuzu oluşturulduktan sonra, alan uzmanları ve ölçme ve değerlendirme uzmanlarından oluşan bir grup uzman tarafından maddeler gözden geçirilmiştir.	0.32
7	Oluşturulan madde havuzuna ön deneme uygulaması yapılmıştır.	0.80
8	Deneme uygulamasının örnekleme, ölçeğin amacı doğrultusunda ölçülecek özelliğe sahip olan bireylerin tümünü temsil edici niteliktedir.	0.37
9	Örnekleme büyüklüğünün belirlenmesinde ölçek ve madde analizinde yapılacak olan analizler dikkate alınmıştır.	0.66
10	Elde edilen veriler faktör analizine uygundur.	0.03
11	Belirlenen faktörlerin öz değerleri 1'den büyüktür.	0.36
12	Ölçeğin açıklanan varyansı %40'tan yüksektir (tek faktörlü ölçek için %30'dan büyük).	0.06
13	Seçilen maddelerin faktör yük değerleri 0.30'dan büyüktür.	0.11
14	Doğrulamalı faktör analizi için toplanan veriler ile ilgili betimsel bilgiler verilmiştir.	0.89
15	DFA analizi sonucunda elde edilen model veri uyumu indeksleri uygun değerlerdir.	0.54
16	Ölçüt geçerliği analizi sonucunda elde edilen değerler uygundur.	1.00
17	Ölçeğin güvenilirlik analizi sonucunda elde edilen değerler ölçeğin tamamı ve tüm alt boyutları için uygundur.	0.16

Tablo 7’de yer alan ölçütlerin [0, 1] aralığındaki değerleri incelendiğinde, değerlendiriciler tarafından makalelerin en büyük çoğunluğu tarafından istenilen düzeyde olduğu belirtilen ölçütün 3. ölçüt (Ölçeğin amacına uygun bir hedef kitle belirlenmiştir.) olduğu, bunu sırasıyla 1. (Ölçülmesi amaçlanan değişken açıkça belirtilmiştir.) ve 10. (Elde edilen veriler faktör analizine uygundur.) ölçütlerinin izlediği gözlenmiştir. Bunun yanı sıra, değerlendiriciler tarafından makalelerin büyük bir çoğunluğu tarafından istenilen düzeyin karşılanmadığı belirtilen ölçütün ise 16. ölçüt (Ölçüt geçerliği analizi sonucunda elde edilen değerler uygundur.) olduğu, bunu sırasıyla 14. (doğrulayıcı faktör analizi için toplanan veriler ile ilgili betimsel bilgiler verilmiştir.) ve 5. (Deneme uygulaması için üretilen madde sayısı, nihai ölçekte bulunması öngörülen madde sayısından daha fazla (en az 2-3 katı) olacak şekildedir.) ölçütlerin izlediği belirlenmiştir.

Çalışmada son olarak, incelenen 57 makalede, geliştirilen ölçeklere yer verilme durumu ile puanlamaya ilişkin bilginin verilme durumları gözden geçirilmiştir. Bu bağlamda makalelerin 27 tanesinde (%47,4) geliştirilen ölçekler makale içeriğinde verilirken 30 makalede (%52,6) bu durum gözlenmemiştir. Geliştirilen ölçeklerin diğer araştırmacılar tarafından kullanıldığında puanlamanın nasıl yapılacağına dair bilginin verilme durumu ise makalelerin %56,1’inde (32 makale) gözlenirken % 43,9’unda (25 makale) puanlama ile ilgili bir bilgiye rastlanmamıştır.

SONUÇLAR ve TARTIŞMA

Bu çalışmada, Türkiye’deki eğitim alanında SSCI dizininde yer alan üç dergide 2010-2015 yılları arasında yayınlanan 57 ölçek geliştirme makalesi, ÇYRM ile analiz edilerek incelenmiştir. Çalışma sonucu ulaşılan bulgularda, makalelerin güvenilir bir şekilde ayırt edilebildiği, belirlenen ölçütlerin güvenilir olduğu ve ölçüt kategorilerinin uygun ve yeterli ölçüm yapabildiği gözlenmiştir. Çalışmada yer alan dört puanlayıcının, makalelere verdikleri puanların, katılık-cömertlik düzeylerinin birbirinden farklı olduğu ancak model-veri uyumunun sağlandığı belirlenmiştir. Alan yazın incelendiğinde, farklı puanlayıcıların yer aldığı çalışmalarda, puanlayıcıların puanlama öncesi bilgilendirilme sürecinden, farklı eğitim süreçlerinden geçmiş olmaları ve farklı eğitim düzeylerinde olmalarından kaynaklanan puanlayıcılar arası farklılıkların olabileceği vurgulanmaktadır (Baştürk ve İşikoğlu, 2008; Mulqueen, Baker ve Dismuskes, 2000; Semerci, 2012; Semerci, 2011). Ayrıca, ÇYRM analizinde puanlayıcılar arası farklılığın olması bir sınırlılık olmamakta; bu farklılıklar kontrol altında tutularak birey ve maddelere (bu çalışma için makaleler ve ölçütlere) ilişkin istatistikler bu şekilde hesaplanmaktadır (Abu Kassim, 2007).

Analizler sonucu elde edilen bulgulara göre, çalışmada kullanılan formdaki 17 ölçütün tümü dikkate alındığında 57 makaleden sadece bir tanesinde, belirlenen bütün ölçütlere ilişkin bilginin bulunduğu ve gerekli açıklamaları kapsadığı belirlenmiştir. İncelenen makalelerden beşinde (% 8.8), ölçütlere ilişkin bilginin bulunduğu ve kısmen açıklama yapıldığı, 41’inde (% 72) sadece ölçütlere ilişkin bilginin yer aldığı ancak her hangi bir açıklama yapılmadığı ve 10’unda (% 17.5) ise ölçütlerle ilgili hiç bir bilginin yer almadığı belirlenmiştir. Alan yazın incelendiğinde, Türkiye’deki ölçek geliştirme makalelerinin incelendiği diğer çalışmalarda da benzer sonuçlara rastlanmaktadır. Acar Güvendir ve Özer Özkan (2015,) çalışmalarında toplam 26 ölçek geliştirme makalesi incelemişlerdir ve çalışmalarında yer alan “ölçek geliştirmede izlenmesi gereken 22 aşamadan” sadece sekizinin (%36.4), bu makalelerin %75’i ve daha fazlası tarafından sağlanabildiği belirlenmiştir.

Çalışmada kullanılan makale değerlendirme ölçütleri detaylı incelendiğinde, üç ölçütün incelenen makalelerin çoğunda oldukça yüksek düzeyde olduğu sonucuna ulaşılmıştır. Çalışmalarda kullanılan ölçeklerin çok farklı hedef kitleleri olmakta ve bu sebeple aynı değişken için dahi farklı hedef kitleleri için ölçeklerde yer alan ifadelerde farklılaşmaktadır. Bu sebeple ölçek geliştirme çalışması yürüten bir araştırmacının önceliklerinden birinin amacına uygun hedef kitlesini belirlemek olması gerekmektedir. Bu gereklilikten dolayı da incelenen çalışmalarda bu ölçütün karşılanma düzeyi oldukça yüksek çıkmıştır. Bunun yanı sıra, ölçek geliştirme çalışması yürüten araştırmacıların, araştırmalarına başlamadan önce, ölçme aracı geliştireceği değişken ile ilgili yeterli bilgiye sahip olması gerekmektedir. Bu sebeple yürüttüğü çalışmada da bu değişken ile ilgili bilgi vermesi kaçınılmazdır. Bu durumun yansımaları, incelenen makalelerde ölçülmesi amaçlanan değişkenin

açıklanma oranının oldukça yüksek çıkması bulgusu gözlenmektedir. Son olarak, incelenen 57 makalenin verilerinin faktör analizine uygunluğunun sağlanması ölçütü oldukça yüksek bir oranda gözlenmiştir. Ölçek geliştirme çalışmalarında araştırmacıların neredeyse tamamının topladıkları verilere faktör analizi yaptıkları gözlenmektedir. Toplanan veriler, faktör analizi yapmaya uygun değilse, çalışmanın ilerlemesi ve dolayısıyla bir yayına dönüşmesi de mümkün olmayacaktır. Bu ölçütün karşılanma oranının yüksek olmasının gerekçesi bu şekilde açıklanabilir.

Çalışmada kullanılan 17 makale değerlendirme ölçütünden yukarıda belirtildiği üzere oldukça yüksek oranlarda karşılanan ölçütlerin yanı sıra değerlendiriciler tarafından istenilen düzeyde karşılanamayan ölçütlere de rastlanmıştır. Bunların ilki ölçüt geçerliği ile ilgili olan ölçüttür. Bu değerlendirme ölçütünün karşılanma oranının düşük olmasının sebebi, araştırmacıların ölçek geliştirme esnasında böyle bir adımdan haberdar olmamalarından çok başka nedenlere dayandırılabilir. Öncelikle ölçek geliştirme çalışması yürüten araştırmacının, ölçüt geçerliği ile ilgili veri toplayabilmesi için öncelikle ölçek geliştirdiği değişken ile ilişkili bir başka ölçek olması gerekmektedir. Böyle bir ölçeğin bulunması her zaman olası olmayabilir. Bunun yanı sıra ölçüt bulunsa dahi ölçüt geçerliğinde kullanılacak ölçeğin kullanılarak yeniden veri toplanması zaman, maliyet vb. sebeplerden ötürü tercih edilmemiş olabilir. İstenilen düzeyde karşılanamayan bir diğer ölçüt ise DFA analizi için toplanan veriler ile ilgili betimsel bilginin verilme durumudur. İncelenen çalışmalarda da görüldüğü üzere, araştırmacıların büyük bir çoğunluğu topladıkları veriler üzerinde sadece açıklayıcı faktör analizi (AFA) yapmakta ancak doğrulayıcı faktör analizi yürütmemektedirler. Bu durumun ilk gerekçesi, araştırmacının böyle bir gerekliliğin farkında olmamasından kaynaklanabilir. Çünkü alanyazındaki pek çok çalışmada sadece AFA yapılmakta ve araştırmacılar bu çalışmaları örnek olarak çalışmalarını yürütmüş olabilirler. Bir diğer gerekçe ise AFA ve DFA'nın farklı örneklemelere uygulanması gerekliliğidir. Geliştirilmeye çalışılan ölçeğin uygulanacağı uygun hedef kitesinden bireylere ulaşılması bir diğer ifadeyle yeterli sayıda kişiden veri toplanması oldukça zahmetli bir süreçtir. AFA'nın yanı sıra DFA'da yaparak bu süreci daha fazla zorlaştırmamak adına araştırmacılar, bu adımı atlamış olabilirler. Bir diğer gerekçe ise AFA'ya kıyasla DFA'nın araştırmacılar tarafından çokta sık kullanılmayan bilgisayar programlarında yapılması olabilir. Araştırmacılar kendilerine yabancı ve bu sebeple daha zor gelen bu adımı atlıyor olabilirler. Son olarak, incelenen makalelerde karşılanma oranı düşük olan ölçüt, deneme uygulamasında üretilen madde sayısının nihai ölçektekinin 2-3 katı olması gerekliliğidir. Normal şartlar altında herhangi bir davranışı yoklamaya yönelik madde yazmak zahmetli ve detaylı bir iştir. Aynı davranışını ölçen birden fazla madde yazma süreci ise daha da zordur. Bu sebeple araştırmacılar her zaman bir davranışla ilgili birden fazla madde yazmakta zorlanabilirler. Bu durumda da deneme formundaki madde sayısının nihai ölçekte bulunması amaçlanan madde sayısının 2-3 katı olması gerekliliği karşılanamayabilir. Bu durumun bir diğer gerekçesi ise deneme formundaki madde sayısındaki artış, ölçeğin cevaplanma süresini ve dolayısıyla cevaplayıcıların konsantrasyonlarını etkileyeceği için araştırmacılar kasıtlı olarak daha az sayıda maddeden oluşan deneme formları geliştirmiş olabilir. Çalışma kapsamında son olarak makalelerde bulunma durumu incelenen iki madde yer almaktadır. Bunlardan ilki geliştirilen ölçeğin çalışmalarda verilme durumudur. Elde edilen sonuçlarda, geliştirilen ölçeğin verilme durumu yaklaşık %50 civarında gözlenmiştir. Geriye kalan makalelerde verilmeme nedeni, makalelerdeki sayfa sınırından ölçeklerin makale kapsamında yer alamaması olabileceği gibi ölçeği geliştiren çalışmacıların, geliştirdikleri ölçeği kullanan bireylerden anında haberdar olma istekleri ve kendilerinden izin alındıktan sonra ölçeği diğer araştırmacılara gönderme istekleri olabilir. Son olarak incelenen ikinci madde ise geliştirilen ölçeklerle ilgili puanlama bilgisinin verilme durumudur. İncelenen makalelerin yaklaşık yarısında (%56,1) puanlama ile ilgili bilgi bulunmaktadır. Puanlama bilgisi verilmeyen makalelerdeki bu bilgi eksikliği, geliştirilen ölçekleri çalışmalarında kullanacak diğer araştırmacıların elde ettikleri puanların yorumlamalarını güçleştirecektir. Bu sebeple, çalışmalarda ölçeklerden alınacak en yüksek ve en düşük puanların yanı sıra gerekiyorsa belirli aralıkların tanımlanması yoluyla ölçeğin hem alt boyutları hem de bütünü itibarıyla alınacak puanların yorumu ile ilgili bilgi verilmesi gerekmektedir.

KAYNAKÇA

- Abu Kassim, N. L. (2007, June). Exploring rater judging behaviour using the many-facet Rasch model. *Paper Presented in the Second Biennial International Conference on Teaching and Learning of English in Asia: Exploring New Frontiers (TELiA2)*, Holiday Villa Beach & Spa Resort, Langkawi. Faculty of Communication and Modern Languages, Universiti Utara Malaysia.
- Acar Güvendir, M. ve Özer Özkan, Y. (2015). Türkiye’deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarılama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, 14(52), 23-33.
- Anshel, M. H., Weatherby, N. L., Kang M. ve Watson, T. (2009). Rasch Calibration Of A Unidimensional Perfectionism Inventory For Sport. *Psychology of Sport and Exercise*, 10(2009), 210-216.
- Barrett, G. V. (1972). Research models of the future for industrial and organizational psychology. *Personnel Psychology*, 25, 1-17
- Baştürk, R. ve Işıkoğlu, N. (2008). Okul Öncesi Eğitim Kurumlarının İşlevsel Kalitelerinin Çok-Yüzeyle Rasch Ölçme Modeli İle Analizi. *Kuram ve Uygulamada Eğitim Bilimleri*, 8(1), 7-32.
- Boztunç Öztürk, N., Eroğlu, G. ve Kelecioğlu, H. (2014). Eğitim alanında yapılan ölçek uyarılama makalelerinin incelenmesi. *Eğitim ve Bilim*, 40(178), 123-137.
- Brinthaup, T. M. ve Kang, M. (2012). Many-Faceted Rasch Calibraton: An Example Using the Self-Talk Scale. *Assessment*, 21(2), 241-249.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
- Cohen R. J. & Swerdlik M.E. (2010). *Psychological testing and assessment*. Boston: McGraw-Hill Companies.
- Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). *The experience of work*. London: Academic Press.
- Çüm, S. ve Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarılama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.
- Delice, A. ve Ergene, Ö. (2015). Investigation of scale development and adaptation studies: An example of mathematics education articles. *Karaelmas Journal of Educational Sciences* 3, 60-75
- DeVellis, R.F. (2003). *Scale development: Theory and applications*. Newbury Park: Sage Publications, Inc.
- Edenborough R. (1999). *Using psychometrics: a practical guide to testing and assessment*. London: Kogan Page.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarılama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-1: Temel kavramlar ve işlemler*. Ankara: Pegem Akademi.
- Güler, N. (2014). Analysis Of Open-Ended Statistics Questions With Many Facet Rasch Model. *Eurasian Journal of Educational Research*, 55, 73-90.
- Güler, N. ve Gelbal, S. (2010). A Study Based On Classical Test Theory And Many Facet Rasch Model. *Eurasian Journal of Educational Research*, 38, 108-125.
- Hetherman, S. C. (2004). *An application of multi faceted Rasch measurement to monitor effectiveness of the written composition in English in the new york city department of education*. Unpublished phd. dissertation. Teacher College, Colombia University, Colombia.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988.
- Hinkin, T. R., Tracey, J. B. & Enz, C.A. (1997). Scale construction: developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research*, 21(1), 100-120.
- İlhan, M. (2016). Açık uçlu sorularla yapılan ölçmelerde klasik test kuramı ve çok yüzeyle Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(2), 346-368. Doi: 10.16986/HUJE.2016015182
- Karasar, N. (1998). *Araştırmalarda rapor hazırlama yöntemi*. Ankara: Pars Matbaacılık Sanayi.
- Kim, Y., Park, I. & Kang, M. (2012). Examining rater effects of the TGMD-2 on children with intellectual disability. *Adapted Physical Activity Quarterly*, 29, 346-365.
- Lee, M. Peterson, J. J. ve Dixon, A. (2010). Rasch Calibration Of Physical Activity Self-Efficacy And Social Support Scale For Persons With İntellectual Disabilities. *Research in Developmental Disabilities*, 31(2010), 903-913.
- Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. [Available online at: <http://www.winsteps.com/a/facets-manual.pdf>], Retrieved on July 13, 2015.

- Linacre, J. M. (2007). *A User's Guide to FACETS: Rasch Model Computer Programs*. Chicago, IL.
- Linacre, J. M. (1989). *Many-facet Rasch Measurement*. Yayınlanmamış doktora tezi. University of Chicago, Chicago, IL.
- Lunz, M. E., Wright, B. D. & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345.
- Mulqueen C., Baker D., & Dismukes, R. K. (2000, April). Using multifacet Rasch analysis to examine the effectiveness of rater training. *Presented at the 15th Annual Conference for the Society for Industrial and Organizational Psychology (SIOP)*. New Orleans
- Murphy K.R. & Davidshofer C.O. (2005). *Psychological testing: Principles and applications*. New Jersey: Pearson Education International.
- Randall, J. & Engelhard, G. Jr. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement*, 46(1), 1-18.
- Revesz, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. *Language Learning*, 62(1), 93-132.
- Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, 19, 385-417.
- Semerçi, Ç. (2011). Doktora yeterlikler çerçevesinde öğretim üyesi, akran ve öz değerlendirmelerin rasch ölçme modeliyle analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(2), 164-171.
- Semerçi, Ç. (2012). Öğrencilerin BÖTE Bölümüne İlişkin Görüşlerinin Rasch Ölçme Modeline Göre Değerlendirilmesi (Fırat Üniversitesi Örneği). *E-Journal of World Science Academy, NWSA-Education Sciences*, 1CO542, 7(2), 777-784.
- Smith, V. E. & Kulikowich, M. J. (2004). An application of generalizability theory and many facet Rasch measurement using a complex problem-solving skills assessment. *Educational and Psychological Measurement*, 64, 617-639.
- Stone, E. (1978). *Research methods in organizational behavior*. Glenview, IL: Scott, Foresman.
- Sudweeks, R.R., Reeve, S., & Bradshaw, W.S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239-261.
- Tavşancıl, E., Güler, G. ve Ayan, C. (2014). *2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreci açısından incelenmesi*. IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran, Hacettepe Üniversitesi, Ankara.
- Tezbaşaran, Ata. (2008). *Likert tipi ölçek geliştirme kılavuzu*. Üçüncü baskı, Türk Psikologlar Derneği Yayını, Ankara.
- Turner, J. (2003). *Examining on art portfolio assessment using a many facet Rasch measurement model*. Unpublished phd. dissertation. Boston College, Boston.
- Wright, B. D. ve Linacre, J. M. (1992). Combining and splitting categories. *Rasch Measurement Transactions*, 6(3) 233-235.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement: Transactions of the Rasch Measurement SIG*, 8(3), 370.
- Yıldırım, A. ve Simsek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.

EXTENDED ABSTRACT

Introduction

The main reason for using scales as measurement instruments is the intention to measure the variables that are regarded that they exist in education and psychology but which cannot be directly observed. Such variables that are considered to exist in individuals are called structures and those structures cause visible behaviors in people. Scales aim to determine the extent to which individuals have the structures by using rubrics.

The generalizability, functionality and firmness of the findings obtained by applying the scales to individuals are related with the reliability and validity of those instruments. The results obtained from studies using the same scale or from studies using different scales developed to measure the same properties, for instance, can be inconsistent. It is thought that this situation may basically stem from the fact that measurement tools yielding reliable and valid results have not been used.

Therefore, the quality of those tools used to measure the directly unobservable variables is very important.

An increase has been observed in scale development and adaptation studies especially in recent years. One reason of the increase is the pressure about doing research that academicians feel over them. This, in turn, results in low quality studies. On examining the studies performed, it is observed that there are a number of problems and inadequacies in the studies performed or in the process of conducting the studies.

This study aims to demonstrate the extent to which the studies on scale development published in scientific journals appearing in SSCI in the period between 2010 and 2015 took the necessary steps in the “process of scale development”. For this purpose, the data collected were primarily analyzed for each facet in terms of reliability by using many-faceted Rasch Model, situations appearing to be problematic were revised, and thus the data obtained were put to analysis. In this way, the levels and reliability of the facet of raters assessing the articles, the criteria used in rating and the articles considered were identified. Then, an attempt was made in this study to describe the strengths and weaknesses of the articles on “scale development” in terms of the criteria set. It is hoped that this study will contribute significantly to the literature by analyzing all the facets through many-faceted Rasch model and by determining the place of the articles considered and the criteria set in equal interval scale in addition to offering data supportive of similar studies evaluating scale development articles. This study also aims to set an example demonstrating that many-faceted Rasch model can be used in similar situations of measurement.

Method

This study analyses the articles concerning scale development and published in the journals of education surveyed in SSCI (Education and Science, Journal of Hacettepe University Faculty of Education, Journal of Educational Sciences in Theory and Practice) in the period between 2010 and 2015. Studies developing scales only for measuring typical behaviors and studies reporting the process of developing such scales have been included in this study. Thus, 57 articles in total whose full texts are accessible in three journals of education constitute the sample of this study.

This study reviews the references published in relation to the steps to be followed in and necessity of scale development as well as the studies analyzing scale development articles. A “Scale Development Article Investigation Form” has been developed by the researchers considering the general tendency in the literature. The form contains items, and the first 17 items are the criteria that can be scored between 0 and 3. Scoring in those criteria is described as “(0) no information; (1) information available but not explained; (2) information available and partly explained; (3) information available and explained in an appropriate way. The other two items are yes-no questions asking whether or not the scale developed was available in the studies and whether or not information on scoring the scale is available.

Results and Discussion

The findings demonstrated that the articles had been discriminated in a reliable way, the criteria set were reliable and the categories of the criteria were able to measure appropriately and adequately. It was found that the four raters available in the study differed in their levels of strictness and generosity in rating the articles but model-data fit had been attained. The findings also showed that only one out of 57 articles included information on the criteria set and offered the necessary explanation considering the 17 criteria in the form used in this study. Accordingly, five of all the studies (8.8%) contained information about the criteria set and offered partial explanations, 41 of them (72%) contained information on the criteria set but offered no explanations on them, and 10 of the studies (17.5%) contained no information on the criteria set. The 14th and 16th criteria which are related to descriptive information about the data collected for DFA and criterion validity respectively did not meet adequately. Finally, the criterion met at low levels in the articles considered was the necessity that the number of items created in the trial application should be two or two times more than in the final scale.

Kayıp Veriyle Baş Etme Yöntemlerinin Model Veri Uyumu Ve Madde Model Uyumuna Etkisi*

The Effect Of Missing Data Tecniques On Model Fit And Item Model Fit

Duygu KOÇAK** Ömay ÇOKLUK BÖKEOĞLU***

Öz

Bu çalışmanın amacı, kayıp veri baş etme yöntemlerinin Madde Tepki Kuramı 1 parametrelili lojistik modelinde model veri uyumuna ve madde model uyumuna olan etkisini incelemektir. Bu amaç doğrultusunda örneklem büyüklüğünün 500, 1000 ve 1500 olarak manipüle edildiği, madde sayısının 20 olarak sabitlendiği IPLM'e uyumlu veri setleri üretilmiştir. Üretilen verilerde madde güçlüğü -2 ile +2 arasında sınırlandırılmış, madde ayırt ediciliği 1.5 olarak sabitlenmiştir. Üretilen eksiksiz veri setleri üzerinde tamamen rastgele kayıp ve rastgele kayıp koşulları altında %5, %10 ve %15 oranlarında silme işlemi gerçekleştirilmiştir. Tamamen rastgele kayıp mekanizması, veri setindeki toplam hücre sayısı arasından rastgele değerler silinerek oluşturulmuştur. Liste bazında silme yöntemi için belirlenen oranda rastgele birim (kişi) silinmiştir. Rastgele kayıp mekanizması, veri setine üç düzeyi olan bir değişken tanımlanıp 1. düzeyden %20, 2. düzeyden %30 ve 3. düzeyden %50 oranında olacak şekilde hücrelerin rastgele silinmesi ile oluşturulmuştur. Oluşturulan kayıp veriler liste bazında silme, regresyonla atama ve beklenti maksimizasyon algoritması yöntemleriyle giderilmiştir. Model veri uyumunun kestirilmesinde $-2 \log \lambda$, AIC ve BIC değerlendirme kriterlerinden, madde model uyumunun kestirilmesinde χ^2 istatistiğinden faydalanılmıştır. Eksiksiz veri setlerinden elde edilen değerler, kayıp veri baş etme yöntemleriyle tamamlanan veri setlerinden elde edilen kestirimler için referans olarak kullanılmıştır. İncelemeler sonucunda, beklenti maksimizasyon algoritması yönteminin rastgele kayıp mekanizmasında iyi, tamamen rastgele kayıp mekanizmasında kısmen iyi performans sergilediği sonucuna ulaşılmıştır. Regresyonla atama yönteminin de belirli koşullar altında iyi performans sergilediği ancak liste bazında silme yönteminin performansının düşük olduğu sonucuna ulaşılmıştır. Tüm kayıp veri mekanizmalarında kayıp veri oranı arttıkça, kayıp veri baş etme yöntemlerinin performansı da düşmektedir. Tüm mekanizmalarda ve koşullarda kullanılabilecek, en iyi sonuçları veren tek bir yöntemin varlığından söz etmenin mümkün olmadığı belirlenmiştir.

Anahtar Kelimeler: kayıp veri, kayıp veri baş etme yöntemleri, model veri uyumu, madde model uyumu, madde tepki kuramı.

Abstract

The purpose of this study was to examine the effects of missing data handling techniques on model data fit and item model fit in the one parameter logistic Item Response Theory Model. For this purpose, data sets with sample sizes of 500, 1000, and 1500 and with 20 items that fit to one parameter logistic model were generated. Item difficulty values of the items in the generated data sets ranged from -2 to +2 and item discrimination was fixed as 1.5. The generated complete data sets were exposed to deletion at %5, %10, and %15 under missing at complete random (MCR) and missing at random (MR) conditions. Missing at complete random mechanism was obtained as a result of random values deleted among the total number of cells in the data set. A particular percentage of random units (individuals) were deleted for listwise deletion method. Missing at random mechanism was reached as a result of random deletion of cells pursuant to defining a three level variable in the data set at the following percentages: 20% at Level 1, 30% at Level 2 and 50% at Level 3. The generated missing data were resolved using listwise deletion method (LM), regression imputation, and expectation maximization algorithm (EMA). $-2 \log \lambda$, AIC, and BIC evaluation criteria were used for model data fit

*Bu çalışma, birinci yazarın Doç. Dr. Ömay ÇOKLUK BÖKEOĞLU danışmanlığında tamamlanan doktora tezinden türetilmiştir.

**Yrd. Doç. Dr., Alanya Alaaddin Keykubat Üniversitesi, Eğitim Fakültesi, Antalya-Türkiye, e-posta: duygu.kocak@alanya.edu.tr

***Doç. Dr., Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara-Türkiye, e-posta: cokluk@aducation.ankara.edu.tr

estimation and χ^2 statistics were used for item model fit estimation. Values obtained from the complete data sets were taken as reference for predictions in the data sets that were completed with the effect of missing data handling techniques. As a result of the examinations, it was concluded that expectation maximization algorithm had good performance in missing at random mechanism but partially good in missing at complete random mechanism. It was also seen that regression imputation had good performance under certain conditions but the performance of listwise deletion method was poor. In all missing data mechanisms, the performance of the effect of missing data handling techniques declines as missing data increase. It is certain that a single method to give best results in all mechanisms and under any conditions is unlikely to be assumed.

Keywords: missing data, missing data techniques, model fit, item model fit, item response theory.

GİRİŞ

Eğitim ve psikoloji alanında ölçme konusu olan özellikler çoğunlukla doğrudan gözlenemeyen, gizil (örtük) değişkenlerdir. Bireylerin ölçülmek istenen bu değişkenlerle ilgili olduğu düşünülen ve gözlenebilen değişkenlere verdikleri tepkiler aracılığıyla, gizil özelliklere dair çıkarımlar yapılmaktadır. Bireylerin gizil özelliklerini gözlenen değişkenler ile açıklamak için farklı test kuramları geliştirilmiştir ve söz konusu özellikler, bu kuramlara dayalı olarak geliştirilen ölçme araçları ile ölçülmektedir.

Psikolojik ölçme tarihindeki kuramlar ve bunların gelişimleri incelendiğinde test geliştirmede, geliştirilen testlerin puanlanmasında ve analizinde yaygın olarak Klasik Test Kuramı'nın (KTK) kullanıldığı görülmektedir. KTK'nın çeşitli sınırlılıkları nedeniyle Madde Tepki Kuramı (MTK), Genellenabilirlik Kuramı (GK), Konjenerik Test Kuramı (KonTK) gibi farklı kuramlar da giderek daha fazla tercih edilmeye ve kullanılmaya başlanmıştır (Reise, Ainsworth ve Haviland, 2005).

Klasik Test Kuramı'nda doğrudan gözlenemeyen iki farklı puan tanımlanmaktadır. Bunlar, gerçek puan ve hata puanıdır. Ölçülmesi hedeflenen gizil özelliğe ilişkin gerçek değeri ifade eden gerçek puan tanımlanır ve temel olarak ölçülen özelliğin gerçek değerine ulaşmak amaçlanır. Fakat ölçme sürecini etkileyen hatalar nedeniyle bu değer doğrudan kestirilememektedir. Bu nedenle kuramda gerçek puan, çeşitli varsayımlar ışığında gözlenen puanlar aracılığıyla kestirilir. Kuramda gerçek ile gözlenen puan farkı hata puanı olarak tanımlanır (Baykul, 2000; Crocker ve Algina, 1986; Embretson, 1999; Lord, 1980; Turgut, 1992; Wainer ve Thissen, 2001; Wilson, 2005).

KTK'da parametreler madde ve test olmak üzere iki başlıkta ele alınmaktadır. Madde güçlük indeksi, madde ayırt edicilik indeksi, madde standart sapması, madde varyansı ve madde güvenilirliği madde parametreleri olarak gruplandırılır (Cronbach, 1990). KTK'da madde güçlüğü ve madde ayırt ediciliği gibi parametreler verilerin toplandığı gruptan gruba farklılık göstermektedir bir diğer ifadeyle gruba bağlı olarak farklı değerler alabilmektedir (Lord ve Novick, 1968). Uygulama yapılan grup değiştiğinde, madde parametrelerinin de değişmesi, farklı gruplara ilişkin test puanlarının karşılaştırılmasında önemli bir sınırlılıktır. Bu durum KTK'da farklı testleri alan kişilerin karşılaştırılmasını güçleştirmektedir (Hambleton, Swaminathan ve Rogers, 1991). Bir diğer önemli sınırlılık ise testin bütünü için tek bir hata değerinin kestirilmesi ve testi alan bireylerin tümü için hata varyansının eşit kabul edilmesidir (Hambleton, Swaminathan ve Rogers 1991). Madde parametrelerinde olduğu gibi güvenilirlik gibi test parametreleri de uygulama yapılan gruba bağlı olarak farklı değerler alabilmektedir. Dolayısıyla bir grupta güvenilir olan test, başka bir grupta aynı durumu gösteremeyebilir yani güvenilir olmayabilir (Crocker ve Algina, 1986). Diğer bir sınırlılığı da testi alan kişilerin testte yer alan bir maddeye nasıl tepki vereceğine ilişkin kestirim imkânı sunamamasıdır (Hambleton ve Swaminathan, 1989).

Klasik Test Kuramı'nın yukarıda değinilen temel bazı sınırlılıklarına alternatif olarak geliştirilen önemli kuramlardan biri de MTK'dır. Kuram testi alan kişinin yetenek düzeyi ile maddelere verdiği yanıtlar arasındaki ilişkinin matematiksel fonksiyonlar ile açıklanabileceği iddiasındadır (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1989). Buna bağlı olarak da bireylerin yetenek düzeyleri testten bağımsız olarak maddelere verilen yanıtlardan hareketle kestirilmektedir. Çünkü MTK, farklı yetenek düzeylerindeki bireylerin her bir maddede nasıl tepki vereceklerine dair matematiksel modelleme sunmaktadır (Crocker ve Algina, 1986).

Madde Tepki Kuramı, test ve madde parametrelerinin gruptan, yetenek düzeyi kestiriminin ise testten bağımsız olduğu iddiasındadır. Bununla birlikte büyük örneklem aracılığıyla gruptan bağımsız şekilde kestirilen madde parametrelerinin değişmezlik özelliği ile farklı gruplar için ölçme sonuçlarının karşılaştırılmasına olanak sağlamaktadır. MTK'nın KTK'dan bir diğer üstünlüğü ise, KTK'da tek bir güvenilirlik katsayısı hesaplanırken MTK'da her bir yetenek düzeyi için test ve madde bilgi fonksiyonları aracılığıyla ayrı standart hata kestirimi yapılabilmesidir (Adams, 2005; Çelen, 2008; Embretson ve Reise, 2000; Yurdugül, 2006).

Madde Tepki Kuramı'nda yetenek (θ), maddelere verilen yanıtlar arasındaki kovaryansa karşılık gelen kesiksiz ve başat yapı yani özellik olarak ifade edilmektedir (Reeve, 2002). θ , lojit birimi ile ifade edilir ve kuramsal olarak $-\infty$ ile $+\infty$ arasında değer alır. Madde puanının, θ yetenek düzeyi üzerindeki regresyonu "Madde Karakteristik Fonksiyonu (MKF)", fonksiyonun grafiği ise "Madde Karakteristik Eğrisi (MKE)" olarak adlandırılır (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1989; Lord ve Novick, 1968). Madde Karakteristik Eğrisi'nin şeklini ise, maddeye ait üç parametre (güçlük, ayırt edicilik ve şans) belirlemektedir.

Madde Tepki Kuramı'nda testte yer alan her bir maddenin taşıdığı bilgi "Madde Bilgi Fonksiyonu (MBF)" ile hesaplanır (Yu, 2013). Madde bilgi fonksiyonları ve madde bilgi fonksiyonlarından elde edilen test bilgi fonksiyonları, madde ve testleri tanımlamada, test maddelerini seçmede ve testleri karşılaştırmada kullanılan önemli göstergelerdir. Test geliştirme sürecinde MBF'den faydalanılması MKE'nin veriye uyumlu olmasına bağlıdır (Hambleton ve diğ., 1991). MKE ile test verisi yeterince uyumlu değilse, bu durum bilgi fonksiyonundan yapılacak kestirimlerin hatalı olabileceğini gösterir. Bunun önüne geçebilmek adına varsayımlar test edilmeli ve sağlandığından emin olunmalıdır.

Madde Bilgi Fonksiyonu, maddenin ölçülen özelliğe ilişkin ne kadar bilgi verdiğini ortaya koyan bir fonksiyondur. Bir madde tarafından sağlanan bilgi, maddenin bireyin yetenek düzeyinin tahmin edilmesine sağladığı katkı olarak yorumlanmaktadır. Madde bilgisi, maddeye ilişkin hata varyansı ile ters orantılıdır (Reid, Kolakowsky-Hayner, Lewis ve Armstrong, 2007). TBF, uygulanan testin her bir yetenek düzeyinde verdiği bilgiyi gösterir ve MTK'da güvenilirliğin kestirilmesinde, standart hatanın hesaplanmasında temel alınır (DeMars, 2010). Bu nedenle testin geliştirilmesi sürecinde önemli bir yere sahiptir. Standart hata daha öncede ifade edildiği gibi her madde için, her yetenek düzeyi için ve her birey için kestirilebilmektedir ve özellikle testin uygulandığı kişilerin yetenek düzeyine uyumlu maddelerin testte yer almasıyla ilişkilidir (Hambleton ve diğ., 1991). MTK'nın, KTK'nın sınırlılıklarına karşılık olarak sağladığı birçok avantaj bulunmaktadır; ancak MTK'nın yukarıda bahsedilen avantajları yalnızca model veri uyumu kabul edilebilir derecede sağlandığında geçerlidir. Model-veri uyumu düşük olduğunda madde ve yetenek parametrelerinin değişmezliği sağlanamayacağından, avantajlarından da faydalanılamayacaktır (Hambleton ve diğ., 1991; Orlando ve Thissen, 2000). Bu nedenle parametrelerin kestiriminde hangi modelin kullanılacağına belirlenmesi gerekir. Bu belirleme model veri uyumunun değerlendirilmesi, dolayısıyla uyum indekslerinin incelenmesi ile mümkündür (Orlando ve Thissen, 2000).

Model veri uyumunun değerlendirilmesinde belirli bir MTK modelinin veriye ve maddelerin modele ne kadar uyumlu olduğunu ortaya koymak amaçlanır. Bir modelin bir veri setine kesin olarak uyumlu veya uygun olup olmadığı yargısına varmayı sağlayan mutlak bir yöntem henüz geliştirilememiştir. Model veri uyumunda ayrı ayrı maddelerin model ile uyumu ve modelin veri ile uyumu incelenebilir. Genellikle MTK uygulayıcıları, tüm maddelere karşılık gelen model uyumu yerine, her bir maddenin model ile uyumuna odaklanmıştır (Demars, 2010).

Bir modelde, model veri uyumu zayıf ise, madde istatistikleri ve madde bilgi fonksiyonları da yanıltıcı olacaktır (Ackerman, 2005; Chang, 1992; Hambleton ve diğ., 1991). Başka bir deyişle, MBF ve dolayısıyla TBF'nin ve standart hatanın yanıltıcı olması, model veri uyumunun sağlanamaması ile açıklanabilir. Bu nedenle güvenilirliğin sağlanması ve elde edilecek bilgilerin doğruluğu için, model ve madde uyumunun sağlandığı en doğru modelin seçilmesi gerekmektedir. Görüldüğü üzere model veri uyumu ölçmenin geçerliği ve güvenilirliği açısından büyük önem taşımaktadır.

Madde Tepki Kuramı'nda model veri uyumu aracılığıyla kullanılacak modele karar verirken göz önünde bulundurulmuş temel ölçüt birey hakkında en çok bilgiyi en az hata ile kestirebilecek modeli

seçmektir. Böylece bireyin verdiği yanıtlardan hareketle, ölçülmeye çalışılan gizil özelliği hakkında daha yüksek doğrulukla kestirimler yapılabilir. Testlerle ölçülmeye çalışılan gizil değişkenlere ulaşılabilmesi, bireylerin kendilerine testle yöneltilen maddelere yanıt vermesi ile mümkündür. Bu bağlamda yanıtlayıcıların kendilerine yöneltilen maddeleri herhangi bir nedenle yanıtlamaktan kaçınması, boş bırakması ya da atlaması, aslında bu tür bir çıkarımın yapılabilmesinin önündeki en önemli engeli oluşturur (Hohensinn ve Kubinger, 2011).

Araştırmalarda kayıp veri sorununun bulunması durumunda, tüm veriler ile uyumlu en doğru modelin seçildiğinin düşünülmesi doğru olmayacaktır ya da bir başka deyişle, veriler için doğru modelin seçiminde sorunlar yaşanacaktır. Yanıt örüntüsündeki eksiklikler, yeteneğin kestirilmesinin önünde de önemli bir engel oluşturacaktır. Ancak araştırmalarda her zaman eksiksiz veri setleri elde edebilmek ya da başka bir deyişle kayıp verilerin önüne geçebilmek çok da mümkün değildir. Hohensinn ve Kubinger (2011), kayıp değerlerin birçok çalışma kapsamında karşılaşılan ve çalışma uzman kişilerce ne kadar dikkatli planlanırsa planlansın tam olarak önüne geçilmesi mümkün olmayan bir problem olduğunu belirtmektedirler.

Kayıp veriler araştırma sürecinde önemli bir sorundur; çünkü standart istatistiksel yöntemlerin uygulandığı veri setleri, satırları gözlemlerden, sütunları ise değişkenlerden oluşan matrislerdir ve bir değişkene ilişkin bir gözlemin eksik olması, söz konusu gözlemi temsil eden hücrenin boş kalmasına neden olur. Bu da analiz sürecini engelleyecek / etkileyecek bir problem ile karşılaşılması anlamına gelir (Little ve Rubin, 1987). Bununla birlikte yanıtlamama ya da boş bırakmaya bağlı olarak ortaya çıkan kayıp veriler, veri setinin daralmasına ve yapılacak kestirimlerin zayıflamasına yol açar. Kayıp veriler içeren veri seti üzerinde istatistiksel analizler yapılamadığı gibi, yanıt veren ve vermeyen bireyler arasında da, çoğunlukla sistematik olan farklılıklardan dolayı, bir yanlılık oluşmasına neden olabilir (Rubin, 1987).

Rubin (1987), “yanıtlamama oranının” dikkat edilmesi gereken bir nokta olduğunu vurgulamaktadır. Çünkü yanıtlamama oranı arttıkça, sistematik hatanın yani yanlılığın ortaya çıkma olasılığının da artacağı yönünde görüşler bulunmaktadır (Baraldi ve Enders, 2010; Enders ve Bandalos, 2001; Hohensinn ve Kubinger, 2011). Buna karşın yanıtlamama oranı ile yanıtlamama yanlılığı arasında doğrudan bir ilişki bulunmadığı yönünde görüşler de mevcuttur. Bu farklı görüşler kayıp verilere ilişkin tartışmaları, kayıp verilerin ihmal edilebilir olup olmadığının yoklanması noktasına taşımıştır. Kayıp verilerin ihmal edilebilir olması, kayıp verilerin seçkisiz olarak oluştuğu, yani herhangi bir örüntü/yanlılık içermediği, dolayısıyla veri dağılımında bir sapma ya da farklılık oluşmayacağı, verinin çok değişkenli olması durumunda, çok değişkenli normallik sağlanırsa, her bir değişkene ilişkin kayıp veri oluşma durumunun eşit olacağı gibi anlamlara gelmektedir. Rubin (1976), kayıp verilerin herhangi bir örüntüye sahip olmadığını, rastgele dağıldığının kanıtlanması durumunda, eksiksiz veri seti ile kayıp veri içeren veri setinden yapılacak kestirimler arasında manidar bir fark olmayacağını ifade etmektedir.

Alanyazında, kayıp verileri ve kayıp veri baş etme yöntemlerini konu alan çok sayıda araştırma mevcuttur (Chen, Wang ve Chen, 2011; Çokluk ve Kayri, 2011; Demir, 2013; Demir ve Parlak, 2012; Doğanay Erdoğan, 2012; Enders ve Bandalos, 2001; Enders, 2004; Fiona ve diğ., 2006; Furlow ve diğ., 2007; Köse, 2014; Sijtsma ve van der Ark., 2003; Van Ginkel, 2007). Kayıp verilerin neredeyse her çalışmada karşılaşılan bir sorun olması ve önceden kestirilemediği için önlenemiyor olması, konuya ilişkin araştırmaların her geçen gün artmasına ve farklı koşullar altında etkilerinin neler olacağını incelenmesine yol açmaktadır.

Araştırmanın Amacı

Bu çalışmada kayıp verilerle baş etmede kullanılan yöntemlerin MTK modellerinden 1PLM’de model veri uyumu ve madde model uyumu üzerindeki etkilerinin farklı örneklem büyüklüğü ve kayıp veri koşulları dikkate alınarak incelenmesi amaçlanmaktadır. Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1.Farklı örneklem büyüklüğü ve kayıp veri oranlarında tamamen rastgele kayıp mekanizmasında liste bazında silme, regresyonla atama ve beklenti maksimizasyon algoritması ile atama yöntemlerinin

a. model veri uyumu

b. madde model uyumu üzerindeki etkileri nelerdir?

2. Farklı örneklem büyüklüğü ve kayıp veri oranlarında rastgele kayıp mekanizmasında liste bazında silme, regresyonla atama ve beklenti maksimizasyon algoritması ile atama yöntemlerinin

a. model veri uyumu

b. madde model uyumu üzerindeki etkileri nelerdir?

YÖNTEM

Bu bölümde araştırmanın modeli, verilerin üretilmesi, istenen koşullara uygun kayıplar içeren veri setlerinin elde edilmesi, kayıp veri içeren veri setlerinin farklı kayıp veri baş etme yöntemleri ile tamamlanması ve verilerin analizi ile ilgili açıklamalara yer verilmiştir.

Araştırmanın Modeli

Araştırma, farklı kayıp veri baş etme yöntemlerinin Madde Tepki Kuramı bir parametrelili lojistik modelinde model veri uyumuna ve standart hataya olan etkisini yapay (simulatif) veri setleri kullanarak ortaya koymayı amaçlayan bir simülasyon çalışmasıdır. Aynı zamanda var olan kuramsal bilgiye yenilerini katmayı amaçladığından, (Karasar, 2007) temel araştırma niteliği taşıdığı ifade edilebilir.

Verilerin Üretilmesi

Bu çalışmada verilerin üretilmesinde, üretilen eksiksiz veri setlerinde kayıp veri mekanizmalarına uygun ve istenen oranlarda kayıp veri oluşturulmasında R programı kullanılmıştır (R Development Core Team, 2011). Araştırmada veri üretimi, Madde Tepki Kuramı tek boyutlu lojistik modellerinden 1PLM temel alınarak gerçekleştirilmiştir. Üretilen veri seti için madde sayısı 20, kişi sayısı 500, 1000 ve 1500 ve her maddeye ait yanıt kategori sayısı iki (1-0 puanlama) olarak belirlenmiştir. Schafer (1997), tekrar ve kestirim sayısının 100'den az olduğu durumlarda kestirimlerin birleştirilmesi ile elde edilecek değerlerin güvenilir olmayacağını belirtmiştir. Bu nedenle üretilen verilerde, yapılacak atama sayısında ve parametre kestirimlerinde 100 Monte Carlo (MC) tekrarı yapılmıştır.

Madde sayısının belirlenmesi

Aiken'in (1995), iki kategorili puanlanan maddelerden oluşan testlerde 20 maddeden daha az madde bulunması durumunda, testle ölçülmek istenen özelliğin süreklilik gösterme olasılığının düşeceği yönündeki görüşü dikkate alınarak, üretilen veri setlerinin 20 madde ile sınırlandırılmasına karar verilmiştir.

Örneklem büyüklüğünün belirlenmesi

Alanyazında örneklem büyüklüğünü (n) dikkate alan araştırmalar incelendiğinde genellikle, 250 (Goldman ve Raju, 1986), 500 (Baker, 1998; Gao ve Chen, 2005; Goldman ve Raju,1986; Hulin, Lissak ve Drasgow, 1982; Thissen ve Wainer, 1982), 1000 (Goldman ve Raju, 1986; Lord, 1968; Hulin, Lissak ve Drasgow, 1982; Thissen ve Wainer, 1982; Yen, 1987) ve 2000 (Gao ve Chen, 2005;

Hulin, Lissak ve Drasgow, 1982) kişilik örneklemelerin kullanıldığı belirlenmiştir. Bu araştırmaların bulguları ve Madde Tepki Kuramı'nın varsayımları göz önünde bulundurularak, bu çalışmada örneklem büyüklüklerinin 500,1000 ve 1500 kişi ile sınırlandırılmasına karar verilmiştir.

Yetenek düzeyinin belirlenmesi

Veriler üretilirken örneklem büyüklüğü koşulunda belirlenen kişi sayıları göz önünde bulundurularak, kişilere ait yetenek (θ) parametrelerini içeren vektör minimum -3 logit ve maksimum +3 logit olacak şekilde sınırlandırılmıştır. Veriler, belirlenen bu değerler arasında $(\theta_{maksimum}-\theta_{minimum})/(\text{kişi sayısı}-1)$ birimlik artışlarla normal dağılım gösterecek şekilde oluşturulmuştur.

Madde parametrelerinin belirlenmesi

Veri üretiminde temel alınan modelde olması gereken parametreler güçlük (b) ve ayırt edicilik (a) parametreleridir. Madde güçlükleri (b), minimum -2 logit, maksimum +2 logit olacak şekilde sınırlandırılmıştır. Madde ayırt edicilik indeksinin (a) tanımlanan normal aralığı (0, 2)'dir (Hambleton, Swaminathan ve Rogers, 1991). Hambleton'a (1989) göre, 1.35 ile 1.69 aralığı yüksek ayırt ediciliğe işaret etmektedir buna göre üretilen verilerde madde ayırt ediciliği 1.5 olarak sabitlenmiştir.

Kayıp Verilerin Oluşturulması

Alanyazın incelendiğinde kayıp verilerle ilgili yapılan çalışmalarda en çok %5, %10 ve %15 (Chen, Wang, Chen, 2005; Doğanay Erdoğan 2012; Enders ve Bandalos 2001; Fiona ve diğ., 2006; Furlow ve diğ., 2007; Sijtsma ve van der Ark, 2003) kayıp veri oranlarının kullanıldığı belirlenmiştir. Bu çalışmaların bulguları göz önünde bulundurularak, üretilen verilerde Tamamen Rastgele Kayıp (TRK) ve Rastgele Kayıp (RK) mekanizmalarını sağlayan %5, %10 ve %15 oranlarında kayıp veri oluşturulmuştur.

Veri silme işlemi için her bir kayıp veri mekanizmasının belirlenen koşullarını sağlayacak kodlar R programında yazılmış ve kayıp verilerin oluşturulması sağlanmıştır. TRK mekanizmasında kayıplar oluşturulurken 500, 1000 ve 1500 kişilik örneklemelerde herhangi bir değişkene bağlı olmaksızın, örneklem büyüklüğünden belirlenen oranlarda gözenek silinmiştir. RK, kayıp veri mekanizmasını oluşturabilmek için öncelikle ölçülen özellikten farklı ama ölçülen bireylere ait, üç düzeyi olan sıralama ölçeği düzeyinde bir başka değişken rastgele olarak tanımlanmıştır. Daha sonra 1. 2. ve 3. düzeylerde sırasıyla %20, %30 ve %50 oranlarında kayıp veri oluşturulmuştur.

Elde edilen eksikli veri setleri, kayıp veri baş etme yöntemlerinden atamaya dayalı yöntemler ile veri tamamlama aşamasında tamamlanmıştır. Bu yöntemler basit atama yöntemlerinden olan Regresyonla Atama (RA), model tabanlı atama yöntemlerinden Beklenti Maksimizasyon Algoritması ile Atama (BMA) ve silmeye dayalı yöntemlerden biri olan Liste bazında Silme (LBS) yöntemleridir. LBS yönteminde veri seti, veri atama ile değil, eksik gözeneği olan kişileri silme yoluyla eksiksiz hale getirilir. Bu nedenle LBS yöntemi için farklı bir kayıp veri mekanizması kodu yazılmıştır. TRK'da 500, 1000 ve 1500 kişilik örneklemelerde herhangi bir değişkene bağlı olmadan örneklem büyüklüğünden belirlenen oranlarda kişi silinmiştir. RK mekanizmasını oluşturmak için ise kişilere rastgele olarak, üç düzeyi olan sıralama ölçeği düzeyinde bir başka değişken tanımlanmıştır. Daha sonra 1. 2. ve 3. düzeylerden sırasıyla %20, %30 ve %50 oranlarında kayıp veri oluşturulmuştur.

Eksik Verilerin Tamamlanması

Beklenti maksimizasyon algoritması ve regresyonla atama yöntemleri ile veri ataması IBM SPSS 21.00 paket programı ile yapılmıştır. Liste bazında silme yöntemi için R programı kullanılmıştır.

Liste bazında silme yönteminde belirlenen senaryolara (koşullara) uygun olacak şekilde kişi silme işlemi gerçekleştirilmiştir.

Verilerin Analizi

Araştırmada öncelikle üretilen eksiksiz veri setlerinden model veri uyumu kestirilmiştir. Ardından tamamlanmış veri setlerinde aynı kestirimler yapılmış ve eksiksiz veri setinden elde edilen değerler referans kabul edilerek, bu değerlerle karşılaştırılmıştır. Kestirimlerin hesaplanmasında R programı içerisinde yer alan {lrm} ve {dplyr} paketlerinden yararlanılmıştır.

Model veri uyumu

En iyi modeli seçmek için kullanılan yöntemlerden biri $-2 \log \lambda$ fark istatistiğine yani G^2 'ye göre karar vermektir (Dais, 2006; Morren, Gelisen ve Vermut, 2011). G^2 'nin istatistiksel olarak manidar olup olmadığına, söz konusu iki modelin serbestlik dereceleri arasındaki farkın kritik değer ile karşılaştırılması yoluyla karar verilir.

Serbestlik derecesinin yüksek olduğu koşullarda G^2 istatistiklerinin χ^2 dağılımı göstermeyebileceği belirtilmektedir (Collins ve Lanza, 2010). Bu durumlar için alternatif yol olarak AIC ve BIC bilgi kriterlerinin kullanılması genel olarak kabul görmüştür (Bauer ve Curran, 2003; Dias, 2006; Lin, 2006;2012; Nylund ve diğ., 2007; Vrieze, 2012; Yang ve Yang, 2007). Buna bağlı olarak bu çalışmada model veri uyumunun belirlenmesinde $-2 \log \lambda$ ile birlikte AIC ve BIC yöntemleri de kullanılmıştır.

Model veri uyumunun kestirilmesinde R programında yer alan {lrm} paketindeki fonksiyonlardan faydalanılarak $-2 \log \lambda$, AIC ve BIC değerleri kestirilmiştir. Aşağıda eksiksiz veri setlerinden kestirilen model veri uyumu değerleri sunulmuştur.

Tablo 1. Eksiksiz Veri Setinden Kestirilen Bir Parametrelili Lojistik Modelde $-2 \log \lambda$, AIC ve BIC Kriterlerine Göre Model Veri Uyumu

Bilgi Kriterleri	Örneklem Büyüklüğü		
	500	1000	1500
$-2 \log \lambda$	-4575.22	-8982.30	-13779.42
AIC	9192.45	18006.61	27600.85
BIC	9280.96	18109.68	27712.43

Eksiksiz veri setinin ve kayıp veri baş etme yöntemleri ile tamamlanmış veri setlerinin 1PLM'e uyumlarını gösteren $-2 \log \lambda$ değerleri hesaplanmış ve eksiksiz veri seti ile aralarındaki farkın manidarlığın belirlenmiştir. Böylece hangi kayıp veri koşulunda, hangi kayıp veri baş etme yönteminin eksiksiz veri setine daha benzer sonuçlar ürettiği belirlenmeye çalışılmıştır.

Tamamlanmış veri setlerinden elde edilen AIC ve BIC değerleri bulgular bölümünde eksiksiz veri setinden elde edilen değerler referans alınarak yorumlanmıştır. AIC değeri doğrudan yorumlanamaz; çünkü bir modele ilişkin model veri uyumunun iyi ya da kötü olduğuna ilişkin bilgi sunmaz, karşılaştırılan modellerden hangisinin veriye daha uygun olduğuna dair karşılaştırma yapma olanağı sunar. Buna göre, daha küçük AIC ve BIC değerine sahip modelin veriye ya da daha küçük AIC değerine sahip veri setinin modele daha uyumlu olduğu sonucuna varılır (Blozis ve diğ. 2007, Duncan ve diğ., 2006). Bu doğrultuda eksiksiz veri seti ve kayıp veri baş etme yöntemleri ile tamamlanmış veri setlerinden elde edilen AIC ve BIC değerleri karşılaştırılırken, kayıp veri baş etme yöntemlerinin performansı betimsel olarak yorumlanmıştır.

Madde model uyumu

Embretson ve Reise (2000), yalnız model veri uyumunun belirlenmesinin yeterli olmayacağını bunun yanı sıra madde model uyumunun da belirlenmesi gerektiğini belirtmiştir. Orlando ve Thissen (2000), bir maddenin model ile uyumunun belirlenmesinde bir grafik ya da χ^2 benzeri bir istatistik kullanılarak, gözlenen değerler ile beklenen (model yardımıyla kestirilen) değerlerin karşılaştırılmasını önermektedir. Bu nedenle bu çalışmada da madde model uyumunun belirlenmesinde χ^2 madde-uyum yaklaşımı kullanılmıştır. Bu yaklaşıma göre, modelde her bir madde için gözlenen ve beklenen değerler, farklı yetenek düzeylerindeki bireyler ve gruplar için karşılaştırılarak χ^2 dağılımı gösteren olabirlik (likelihood) oran istatistiği elde edilir ve bu istatistik için serbestlik derecesi, cevaplayıcılar için oluşturulan yetenek aralıkları sayısına eşittir (Mislevy ve Bock, 1990). Eksiksiz veri seti ve kayıp veri baş etme yöntemleri ile tamamlanan veri setlerinde madde model uyumu hesaplanmış ve eksiksiz veri seti referans alınarak tamamlanan veri setlerindeki uyumlu madde sayıları karşılaştırılmıştır.

Tablo 2. Eksiksiz Veri Setlerinde Kestirilen Madde Model Uyumu

Madde No	Örneklem Büyüklüğü					
	500		1000		1500	
	χ^2	P	χ^2	p	χ^2	p
M1	14.602	0.406	10.094	0.814	18.738	0.095
M2	18.925	0.168	15.252	0.433	9.387	0.670
M3	16.549	0.281	23.728	0.070	16.208	0.182
M4	8.373	0.869	17.199	0.307	12.024	0.444
M5	11.115	0.677	18.830	0.222	23.400	0.065
M6	12.606	0.558	21.085	0.134	23.241	0.126
M7	17.422	0.234	33.536	0.004	33.687	0.011
M8	11.605	0.638	26.549	0.033	22.789	0.070
M9	16.845	0.265	14.092	0.519	20.796	0.053
M10	23.987	0.046	14.500	0.488	19.967	0.068
M11	7.647	0.907	16.586	0.344	28.788	0.004
M12	10.611	0.716	24.414	0.058	15.794	0.201
M13	20.327	0.120	31.305	0.008	24.366	0.018
M14	18.154	0.200	18.884	0.219	39.570	0.000
M15	8.518	0.861	30.995	0.009	15.356	0.223
M16	6.161	0.962	19.946	0.174	23.006	0.028
M17	13.799	0.465	21.659	0.117	9.846	0.629
M18	20.957	0.103	28.457	0.019	24.337	0.058
M19	16.180	0.303	28.957	0.016	13.688	0.321
M20	13.608	0.479	24.459	0.058	20.566	0.089

Tablo 2’de yer alan değerler tamamlanmış veri setlerinden kestirilecek madde model uyumu için referans olarak kabul edilmiş ve yorumlamalar da bu değerlerle karşılaştırılarak yapılmıştır.

BULGULAR

Bu bölümde araştırmanın amaçları doğrultusunda elde edilen bulgulara yer verilmiştir. Araştırmada 500, 1000 ve 1500 kişilik 1PLM model ile uyumlu verilerde, %5, %10 ve %15 oranında TRK ve RK veri mekanizmalarında LBS, RA ve BMA yöntemlerinin model veri uyumuna, madde model uyumuna etkisi araştırılmıştır. Elde edilen bulgular TRK ve RK mekanizmaları için sırasıyla sunulmuştur. Tablo 3’te Tamamen Rastgele Kayıp mekanizmasında liste bazında silme yöntemi ile tamamlanan veri setlerinden elde edilen G^2 olabirlik oranlarına ilişkin bulgular sunulmuştur.

Tablo 3. Tamamen Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranı

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4347.43	227.83*	-8524.2	458.12*	-13059.80	719.74*
%10	-4123.78	451.48*	-8065.191	917.12*	-12418.96	1360.58*
%15	-3897.29	677.97*	-7650.436	1331.88*	-11716.71	2062.83*

* $p(\chi^2_{sd=19} > 30,144) < 0.05$

Tablo 3 incelendiğinde veri setinde TRK olduğu durumda LBS yöntemi ile tamamlanan veri setinde model veri uyumu $-2 \log \lambda$ ile sınılandığında tüm örneklem büyüklüklerinde (500, 1000 ve 1500) kayıp veri oranı arttıkça bu değerin düştüğü görülmektedir. Her bir $-2 \log \lambda$ sütununda ilk olarak eksiksiz veri setinden elde edilen $-2 \log \lambda$ değeri verilmiş olup sütunda aşağı doğru inildikçe kayıp veri miktarlarında da artış görülmektedir. Ancak buna karşılık $-2 \log \lambda$ değeri ise azalmaktadır. Dolayısıyla eksiksiz veri setinden elde edilen $-2 \log \lambda$ değeri ile, tamamlanan veri setlerinden elde edilen $-2 \log \lambda$ arasındaki fark yani G^2 değeri de giderek artmaktadır. χ^2 dağılımının tablo değeriyle karşılaştırıldığında eksiksiz veri seti ve LBS yöntemi ile tamamlanan veri setlerinden kestirilen model veri uyumları arasında manidar farklılık olduğu görülmektedir. TRK mekanizmasında LBS yöntemi ile tüm örneklem büyüklüklerinde, kayıp veri oranı arttıkça eksiksiz veri setinden sapmanın giderek artması, LBS yönteminden kaynaklı olarak örneklemin küçülmesi ile açıklanabilir. Çünkü LBS yönteminin TRK mekanizmasında kullanılması durumunda, kayıp verili kişiler silindiğinden, veri setinde kalan gözlemler eldeki verinin rastgele bir alt örneklemini oluşturmaktadır (Allison, 2003; Enders, 2010; McKnight ve diğ., 2007; Rosenthal ve Rosnow, 2008) ve örneklemin küçülmesi χ^2 istatistiğine dayalı uyum iyiliğini artırmaktadır (Drasgow, Levine, Tsien, Williams ve Mead, 1995; Bock, 1997). Madde Tepki Kuram'ında model parametrelerinin kestiriminde ve model veri uyumu üzerinde örneklem büyüklüğü önemli bir etkidir (Bolt ve Lall, 2003; Seungho-Yang, 2007). Hambleton ve diğerleri (1991), $-2 \log \lambda$ 'in χ^2 dağılımı gösterdiğini ve bu istatistiğin örneklem büyüklüğüne karşı oldukça hassas olduğunu, büyük örneklem söz konusu olduğunda model veri uyumunun neredeyse tüm modellerde sağlanmadığını vurgulamaktadır. Tablo 4'te TRK mekanizmasında LBS silme yöntemi ile tamamlanan veri setlerinden elde edilen AIC değerleri sunulmuştur.

Tablo 4. Tamamen Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	ΔAIC	AIC	ΔAIC	AIC	ΔAIC
	9190.54		18004.66		27599.07	
%5	8736.86	453.67	17090.40	914.26	26161.59	1437.48
%10	8289.57	900.96	16172.38	1832.22	24879.91	2719.16
%15	7836.58	1353.95	15342.87	2661.79	23475.42	4123.65

Tablo 4 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, AIC değerinde düşüş yaşandığı görülmektedir. Dolayısıyla kayıp veri oranı arttıkça, LBS yöntemi, AIC kriteri açısından eksiksiz veri setinden uzaklaşmış ve veriyi modele daha uyumlu hale getirmiştir. Tablo 5'te TRK mekanizmasında LBS yöntemi ile tamamlanan veri setlerinden elde edilen BIC değerleri sunulmuştur.

Tablo 5. Tamamen Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	ΔBIC	BIC	ΔBIC	BIC	ΔBIC
	9274.84		18102.81		27705.34	
%5	8824.29	450.54	17192.39	910.42	26272.09	1433.25
%10	8375.86	898.97	16273.23	1829.58	24989.28	2716.06
%15	7921.67	1353.16	15442.52	2660.29	23583.58	4121.76

Tablo 5 incelendiğinde, eksiksiz veri seti ve tamamlanan veri setlerinden elde edilen değerler karşılaştırıldığında, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, BIC değerinin düştüğü ve eksiksiz veri setinden uzaklaştığı görülmektedir. Bu uzaklaşma veri setinin 1PLM'e daha uyumlu hale gelmesine yol açmıştır. Model veri uyumunun değerlendirilmesinde kullanılan AIC ve BIC kriterlerinin formülleri incelendiğinde ($AIC = -2 \log \lambda + 2k$, $BIC = -2 \log \lambda + k \log(n)$) k sabit terim dahil parametre sayısı ve n gözlem sayısı (Akaike, 1973;1974; Hurvichve Tsai, 1989; Siungiura, 1978) olmak üzere, her iki kriterin de $-2 \log \lambda$ 'den türetildiği görülmektedir. Dolayısıyla örneklemin küçülmesinin χ^2 istatistiğine dayalı uyum iyiliğini artırdığı (Bock, 1997; Drasgow ve diğ., 1995) ve $-2 \log \lambda$ 'in χ^2 dağılımı gösterdiği göz önünde bulundurulduğunda, AIC ve BIC kriterlerinin de örneklemden etkileneceği, buna bağlı olarak LBS yönteminin örneklemini daraltmasıyla model veri uyumunu artırdığı sonucuna varılabilir. Kankaras ve diğerleri (2011) AIC ve BIC değerlerinin örneklem büyüklüğünden etkilendiğini belirtmektedir. Dolayısıyla LBS yönteminin kullanılması örneklemin küçültmesi sonucu AIC, BIC ve $-2 \log \lambda$ değerlerinde bir küçülmeye ve buna bağlı olarak da model veri uyumunda artışa neden olmaktadır. Tablo 6'da TRK mekanizmasında RA ile tamamlanan veri setlerinden elde edilen G^2 olabilirlik oranları sunulmuştur.

Tablo 6. Tamamen Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranı

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4505.32	69.945*	-8818.56	163.76*	-13490.39	289.15*
%10	-4392.78	182.48*	-8626.59	355.27*	-13259.96	519.58*
%15	-4290.10	285.16*	-8441.24	541.07*	-12929.28	850.26*

* $p(\chi^2_{sd=19} > 30,144) < 0.05$

Tablo 6 incelendiğinde, RA yöntemi ile tamamlanan veri setlerinin TRK mekanizması olduğu durumda tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, eksiksiz veri setinden kestirilen $-2 \log \lambda$ değerinden uzaklaştığı görülmektedir. G^2 oranının manidarlığı test edildiğinde ise, RA yöntemi ile tamamlanan veri setlerinin 1PLM'e eksiksiz veri setinden daha uyumlu olduğu görülmektedir. RA yöntemiyle tamamlanan veri setlerinde TRK olması durumunda, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, eksiksiz veri setinden kestirilen $-2 \log \lambda$ değerinden daha da uzaklaştığı ifade edilebilir. Tablo 7'de TRK mekanizmasında RA yöntemi ile tamamlanan veri setlerinden elde edilen AIC değerleri sunulmuştur.

Tablo 7. Tamamen Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	Δ AIC	AIC	Δ AIC	AIC	Δ AIC
	9190.54		18004.66		27599.07	
%5	9050.64	139.89	17677.12	327.54	27020.78	578.29
%10	8825.57	364.97	17293.19	711.47	26559.91	1039.16
%15	8620.21	570.32	16922.48	1082.12	25898.56	1700.51

Tablo 7 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, AIC değerinin azaldığı görülmektedir. Buna bağlı olarak RA yönteminin TRK mekanizmasına sahip kayıp veri olduğu durumda, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça eksiksiz veri setinden uzaklaştığı ve IPLM'e daha uyumlu bir hale dönüştürdüğü belirtilebilir. Tablo 8'de TRK mekanizmasında RA yöntemi ile tamamlanan veri setlerinden elde edilen BIC değerleri sunulmuştur.

Tablo 8. Tamamen Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	Δ BIC	BIC	Δ BIC	BIC	Δ BIC
	9274.84		18102.81		27705.34	
%5	9134.94	139.89	17775.28	327.53	27127.05	578.29
%10	8909.86	364.97	17391.35	711.46	26666.18	1039.16
%15	8704.50	570.33	17020.64	1082.17	26004.83	1700.51

Tablo 8 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, BIC değerinin düştüğü görülmektedir. Dolayısıyla Δ BIC değerleri de göz önünde bulundurulduğunda, TRK mekanizmasında kayıp veri oranı arttıkça, RA yönteminin BIC değeri bakımından eksiksiz veri setinden uzaklaştığı ve IPLM için model veri uyumunu artırdığı görülmektedir. TRK durumunda model veri uyumunda RA yöntemi kullanılarak tamamlanan veri setlerinden elde edilen değerlerin eksiksiz veri setinden küçük sapmalar gösterdiği görülmektedir. $-2 \log \lambda$, AIC ve BIC değerleri bakımından sapmaların olması, alanyazın ile de uyum göstermektedir. TRK mekanizması olduğu durumda RA yöntemi, bazı durumlarda tam veri setlerinden elde edilen değerlerden farklı değerler alabilmektedir (Roth ve Switzer, 1995; Suraphee, Raksmanee, Busaba, Chaisorn ve Nakornthai, 2006). Tablo 9'da TRK mekanizmasında BMA yöntemi ile tamamlanan veri setlerinden kestirilen model veri uyumu kriteri olan $-2 \log \lambda$ değerleri sunulmuştur.

Tablo 9. Tamamen Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranı

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4498.01	77.26*	-8806.507	175.82*	-13485.65	293.89*
%10	-4391.44	183.83*	-8613.895	184.13*	-1368.43	239.74*
%15	-4269.46	305.81*	-8410.51	571.81*	-12894.70	884.84*

* $p(\chi^2_{sd=19} > 30.144) < 0.05$

Tablo 9'da sunulan G^2 değerlerinin manidarlık testleri edildiğinde tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, model uyumunun iyileştiği gözlenmiştir. BMA yönteminin TRK mekanizmasında model veri uyumunda $-2 \log \lambda$ kriteri açısından uyumu artırıcı bir performans

sergilediği görülmektedir. Tablo 10'da TRK mekanizmasında BMA yöntemi ile tamamlanan veri setlerinden kestirilen AIC değerleri sunulmuştur.

Tablo 10. Tamamen Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	Δ AIC	AIC	Δ AIC	AIC	Δ AIC
	9190.54		18004.66		27599.07	
%5	9036.11	154.42	17654.61	350.05	27015.34	583.73
%10	8822.88	367.65	17273.01	731.65	26528.14	1070.93
%15	8582.22	608.31	16871.29	1133.37	25849.71	1749.36

Tablo 10 incelendiğinde, BMA yöntemiyle tamamlanan veri setlerinin daha düşük AIC değeri verdiği görülmektedir. Tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, AIC değerinin eksiksiz veri setinden kestirilen değerden daha da saptığı ve model veri uyumunun arttığı görülmektedir. TRK mekanizması olduğu durumda BMA yönteminin AIC kriteri bakımından uyumu artırıcı atamalar yaptığı sonucuna ulaşılmaktadır. Tablo 11'de TRK mekanizmasında BMA yöntemi ile tamamlanan veri setlerinden elde edilen BIC değerleri sunulmuştur.

Tablo 11. Tamamen Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	Δ BIC	BIC	Δ BIC	BIC	Δ BIC
	9274.84		18102.81		27705.34	
%5	9120.41	154.42	17752.77	350.04	27121.61	583.73
%10	8907.17	367.66	17371.17	731.64	26634.41	1070.93
%15	8666.51	608.32	16969.44	1133.37	25955.98	1749.36

Tablo 11 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, BIC değerinin düştüğü, dolayısıyla da model veri uyumunun BIC kriteri açısından arttığı görülmektedir. Kayıp veri oranı sabit tutulduğunda, örneklem büyüklüğü arttıkça eksiksiz veri setinden kestirilen BIC değeri ile BMA yöntemi ile tamamlanmış veri setlerinden elde edilen BIC değerleri arasındaki farkın giderek arttığı gözlenmektedir. TRK koşulunda BMA yöntemi ile yapılan atamalar model veri uyumu bakımından eksiksiz veri setine yakın kestirimler sunmaktadır. Enders (2010) ve Schafer ve Graham (2002) TRK mekanizmasında BMA yöntemi ile yansız parametre tahminleri elde edilebildiğini belirtmektedir. Bu çalışmada elde edilen bulgular da bunu desteklemektedir. Tablo 12'de TRK mekanizmasında 1PLM ile uyum gösteren madde sayıları sunulmuştur.

Tablo 12. Tamamen Rastgele Kayıp Mekanizmasında 1PLM ile Uyum Gösteren Madde Sayısı

	n=500			n=1000				n=1500				
	Eksiksiz Veri Seti	%5	%10	%15	Eksiksiz Veri Seti	%5	%10	%15	Eksiksiz Veri Seti	%5	%10	%15
LBS		20	19	20		15	14	17		17	16	17
RA	19	16	15	16	14	14	13	12	15	13	10	9
BMA		17	17	12		15	11	9		8	6	6

Tablo 12 incelendiğinde, örneklem büyüklüğü n=500 olan eksiksiz veri setinde 1PLM'e uyum gösteren madde sayısı 19, n=1000 örneklem büyüklüğünde 14 ve n=1500 örneklem büyüklüğünde 15'tir. LBS yöntemi ile tamamlanan veri setlerinde tüm örneklem büyüklüklerinde ve kayıp veri oranlarında LBS yöntemi ile eksiksiz hale getirilen verilerle uyumlu olan madde sayısının, eksiksiz veri setinden daha fazla olduğu görülmektedir. Madde Tepki Kuramı'nda model parametrelerinin kestiriminde ve model veri uyumu üzerinde "örneklem büyüklüğü" önemli bir etkidir (Bolt ve Lall, 2003; Seungho-Yang, 2007). Madde model uyumunun belirlenmesinde kullanılan χ^2 istatistiğinin de örneklem büyüklüğünden etkilendiği düşünülürse (Hambleton ve diğ.,1991), LBS yönteminin kullanılmasının örnekleme küçültürken χ^2 'nin manidar çıkmasına yol açtığı ve dolayısıyla uyumlu madde sayısını artırdığı düşünülebilir.

Regresyonla atama yöntemi ile tamamlanan veri setinde model ile uyum sağlayan madde sayısı incelendiğinde, tüm örneklem büyüklüklerinde, genel olarak eksiksiz veri setinden daha az sayıda maddenin uyumu olduğu görülmektedir. Örneklem büyüdükçe ve kayıp veri oranı arttıkça, uyumlu madde sayısı giderek azalmaktadır. Alanyazında RA yönteminin, kayıp veri mekanizmasının TRK olduğu durumlarda yansız parametre tahminleri sunduğu vurgulanmaktadır (Baraldi ve Enders, 2010; McKnight ve diğ., 2007). Ancak elde edilen bulgular bunu desteklememektedir. Buna karşılık bu bulgular, alanyazında RA yönteminin, bazı durumlarda tam veri setlerinden elde edilen değerlerden farklı değerler alabileceği bilgisiyle örtüşmektedir (Roth ve Switzer, 1995; Suraphee ve diğ., 2006).

Beklenti maksimizasyon algoritması yöntemi ile tamamlanan veri setlerindeki model ile uyumlu madde sayısı incelendiğinde, tüm örneklem büyüklüklerinde eksiksiz veri setinden daha az sayıda maddenin uyumlu olduğu ve kayıp veri oranı arttıkça, uyumlu madde sayısının da azaldığı görülmektedir. Enders (2010) ve Schafer ve Graham (2002), TRK koşullarında, BMA yönteminin yansız parametre tahminleri sağlayabildiğini belirtmektedir. Ancak elde edilen sonuçlar bunu desteklememektedir. En çok olabilirlik temelli atama yöntemleri, özellikle RK varsayımının sağlandığı durumlarda daha iyi tahmin değerleri vermektedir (Allison, 2003; Baraldi ve Enders, 2010). Buna bağlı olarak BMA yönteminin TRK mekanizmasında daha düşük bir performans sergilemesi olması bir sonuç gibi görünmektedir. Tablo 13'te RK mekanizmasında LBS yöntemi kullanılarak eksiksiz hale getirilen veri setlerinden kestirilen $-2 \log \lambda$ değerleri sunulmuştur.

Tablo 13. Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranı

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4361.74	213.52*	-8579.01	403.30*	-13164.56	614.98*
%10	-4141.48	433.78*	-8122.06	860.26*	-12500.73	1278.81*
%15	-3915.42	659.84*	-7685.31	1297.00*	-11812.62	1966.92*

*p ($\chi^2_{sd=19} > 30.144$) < 0.05

Tablo 13 incelendiğinde, G^2 değerleri χ^2 dağılımının tablo değeri ile karşılaştırıldığında, tüm örneklem büyüklüklerinde ve kayıp veri oranlarında LBS yöntemi ile eksiksiz hale getirilen veri setlerinin 1PLM ile daha uyumlu olduğu görülmektedir. Tablo 14'te RK mekanizmasında LBS yöntemi ile tamamlanan veri setlerinden elde edilen AIC değerleri sunulmuştur.

Tablo 14. Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	ΔAIC	AIC	ΔAIC	AIC	ΔAIC
	9190.54		18004.66		27599.07	
%5	8765.48	425.05	17200.03	804.63	26371.12	1227.95
%10	8324.97	865.65	16286.12	1718.54	25043.45	2555.62
%15	7872.84	1317.69	15412.63	2592.03	23667.24	3931.83

Tablo 14 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça AIC değerinin eksiksiz veri setinden uzaklaşarak düştüğü, dolayısıyla model veri uyumunun arttığı görülmektedir. RK koşulu altında LBS yönteminin kullanılması IPLM’de model veri uyumunda AIC kriteri açısından model veri uyumunu artırmaktadır. Tablo 15’te RK mekanizmasında LBS yöntemi ile tamamlanan veri setlerinden elde edilen BIC değerleri sunulmuştur.

Tablo 15. Rastgele Kayıp Mekanizmasında Liste Bazında Silme Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	ΔBIC	BIC	ΔBIC	BIC	ΔBIC
	9274.84		18102.81		27705.34	
%5	8852.91	421.92	17302.01	800.80	26481.62	1223.72
%10	8411.26	863.57	16386.97	1715.84	25152.82	2552.52
%15	7957.94	1316.89	15512.31	2590.50	23775.40	3929.94

Tablo 15 incelendiğinde tüm örneklem büyüklüklerinde ve kayıp veri oranlarında, LBS yöntemi ile tamamlanan veri setlerinin, eksiksiz veri setinden daha düşük BIC değerlerine sahip oldukları görülmektedir. Tüm örneklem büyüklüklerinde LBS yöntemi ile tamamlanan veri setlerindeki kayıp veri oranı arttıkça, eksiksiz veri setinden uzaklaştıkları ve IPLM’e uyumun arttığı görülmektedir. Bu durum da LBS yöntemine bağlı olarak örneklemin küçülmesi ve böylelikle χ^2 istatistiğine dayalı uyum iyiliğinin artması ile ilişkilendirilebilir (Bock, 1997; Drasgow ve diğ., 1995). Tablo 16’da RA yöntemi ile tamamlanan veri setlerinden elde edilen $-2 \log \lambda$ değerleri arasındaki farklara ilişkin G^2 olabilirlik oranları sunulmuştur.

Tablo 16. Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranları

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4574.09	1.17	-8965.79	16.52	-13761.04	18.50
%10	-4553.29	21.97	-8946.36	29.95	-13716.70	62.84
%15	-4562.16	13.10	-8932.99	49.32	-13695.28	84.26

* $p(\chi^2_{df=19} > 30,144) < 0.05$

Tablo 16 incelendiğinde, G^2 oranının manidarlığı test edildiğinde örneklem büyüklüğü n=500 iken RA yönteminin tüm kayıp veri oranlarında eksiksiz veri seti ile benzer sonuçlar verdiği görülmektedir. RA yöntemi küçük örneklerde tüm kayıp veri oranlarında $-2 \log \lambda$ kriteri açısından eksiksiz veri seti ile karşılaştırıldığında, aradaki farkların manidar olmadığı görülmektedir. Örneklem büyüklüğü n=1000 olduğunda %5 ve %10 kayıp veri oranlarında, örneklem büyüklüğü

n=1500 iken, %5 kayıp veri oranında RA yöntemi eksiksiz veri seti ile benzer sonuçlar üretmektedir. Örneklem büyük ve kayıp veri miktarı yüksek olduğunda, RA yöntemi ile tamamlanan veri setlerinde eksiksiz veri setinden önemli sapmalar meydana gelmektedir. Tablo 17’de RA yöntemi ile tamamlanan veri setlerinden kestirilen 1PLM’e uyumu gösteren AIC değerleri sunulmuştur.

Tablo 17. Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	Δ AIC	AIC	Δ AIC	AIC	Δ AIC
	9190.54		18004.66		27599.07	
%5	9188.19	2.34	17971.59	33.07	27562.08	36.99
%10	9146.59	43.94	17938.72	65.94	27473.39	125.68
%15	9132.45	58.08	17905.99	98.67	27430.57	168.5

Tablo 17 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça AIC değerlerinin düştüğü ve eksiksiz veri setinden uzaklaşarak 1PLM’e daha uyumlu hale geldiği görülmektedir. Tablo 18’de RA yöntemi ile tamamlanan veri setlerinden kestirilen BIC değerleri sunulmuştur.

Tablo 18. Rastgele Kayıp Mekanizmasında Regresyonla Atama Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	Δ BIC	BIC	Δ BIC	BIC	Δ BIC
	9274.84		18102.81		27705.34	
%5	9272.48	2.35	18069.75	33.06	27668.34	37.00
%10	9230.88	43.95	18036.88	65.93	27579.66	125.68
%15	9248.62	26.21	18004.14	98.67	27536.83	168.51

Tablo 18 incelendiğinde, tüm örneklem büyüklüklerinde, RA yöntemi ile tamamlanan veri setlerinde kayıp veri oranı arttıkça, BIC kriterinin düştüğü ve dolayısıyla model veri uyumunun arttığı görülmektedir. %5 kayıp oranında eksiksiz veri setine en yakın kestirimlerin elde edilmiş olması, RA yönteminin iyi performans sergilediğini göstermektedir. Allison (2002) ve Van Ginkel (2007) de, kayıp veri oranı %5’in üzerine çıktığında kayıp veri baş etme yöntemlerinin performansının düşeceğini belirtmektedir. Burada da elde edilen bulgu, bu durumu desteklemektedir. RK kayıp veri mekanizmasında RA yöntemi yansız parametre tahminleri sunmaktadır (Baraldi ve Enders, 2010; McKnight ve diğ., 2007; Roth ve Switzer, 1995; Suraphee ve diğ., 2006). Elde edilen sonuç alanyazın ile uyumludur. Tablo 19’da BMA ile tamamlanan veri setlerinden kestirilen $-2 \log \lambda$ değerleri sunulmuştur.

Tablo 19. Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen G^2 Olabilirlik Oranı

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2	$-2 \log \lambda$	G^2
	-4575.27		-8982.32		-13779.54	
%5	-4573.19	2.07	-8962.41	19.90	-13760.38	19.16
%10	-4557.04	22.22	-8947.24	35.07*	-13714.20	65.34*
%15	-4553.94	21.32	-8929.07	53.24*	-13690.30	89.24*

* $p(\chi^2_{sd=19} > 30.144) < 0.05$

Tablo 19 incelendiğinde, BMA ile tamamlanan veri setleri ve eksiksiz veri seti arasındaki $-2 \log \lambda$ değeri farkını gösteren G^2 olabilirlik oranının manidarlığı test edildiğinde, kayıp veri oranı %5 iken, $-2 \log \lambda$ farkının manidar olmadığı görülmektedir. Örneklem büyüklüğü $n=500$ olduğunda tüm kayıp veri oranlarında eksiksiz veri seti ile farkın manidar olmadığı, dolayısıyla iyi performans sergilediği görülmektedir. Örneklem büyüklüğü $n=1000$ ve $n=1500$ olduğunda kayıp veri oranı %10 ve üzerine çıkınca, tamamlanan veri setleri ile eksiksiz veri seti arasında $-2 \log \lambda$ farklarının manidar olduğu görülmektedir. BMA yöntemi kayıp veri oranı düşük olduğunda iyi performans sergilemektedir. Allison (2002) ve Van Ginkel (2007) kayıp veri oranı %5'in üzerine çıktığında kayıp veri baş etme yöntemlerinin performansının düşebileceğini belirtmektedir. BMA yönteminin %5 veri kayıplarının üzerinde düşük performans sergilemesi alanyazın ile uyumlu görünmektedir. Tablo 20'de BMA yöntemi ile tamamlanan veri setlerinden kestirilen AIC değerleri sunulmuştur.

Tablo 20. Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen AIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	AIC	Δ AIC	AIC	Δ AIC	AIC	Δ AIC
	9190.54		18004.66		27599.07	
%5	9186.39	4.14	17964.84	39.82	27560.76	38.31
%10	9154.09	36.44	17934.49	70.17	27468.40	130.67
%15	9147.89	42.64	17898.15	106.51	27420.61	178.46

Tablo 20 incelendiğinde, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, AIC değerinin düştüğü görülmektedir. Kayıp veri oranı arttıkça BMA yöntemi, AIC kriteri bakımından model veri uyumunu artırıcı bir performans sergilemektedir. Kayıp veri oranı sabit tutulup örneklem büyüklüğü artırıldığında ise, eksiksiz veri setinden kestirilen AIC değerinden uzaklaşmanın arttığı görülmektedir. Tablo 21'de BMA yöntemi ile tamamlanan veri setlerinden kestirilen BIC değerleri sunulmuştur.

Tablo 21. Rastgele Kayıp Mekanizmasında Beklenti Maksimizasyon Algoritması Yöntemi ile Tamamlanan Veri Setlerinden Elde Edilen BIC Değerleri

Eksiksiz Veri Seti	n=500		n=1000		n=1500	
	BIC	Δ BIC	BIC	Δ BIC	BIC	Δ BIC
	9274.84		18102.81		27705.34	
%5	9270.68	4.15	18062.99	39.82	27667.03	38.31
%10	9238.42	36.41	18032.64	70.17	27574.67	130.67
%15	9232.19	42.64	17996.31	106.5	27526.87	178.47

Tablo 21 incelendiğinde, tüm örneklerde kayıp veri oranı arttıkça, BIC değerinin azaldığı görülmektedir. Eksiksiz veri setinden elde edilen BIC değeri ile kıyaslandığında, BMA yönteminin veri setini IPLM'e daha uyumlu bir forma dönüştürdüğü görülmektedir. Kayıp veri oranı sabit tutulup örneklem büyüklüğü artırıldığında, eksiksiz veri setinden kestirilen BIC değeri ile olan farkın da arttığı görülmektedir. Bir diğer ifade ile örneklem büyüklüğü arttıkça, BMA yöntemi BIC değerinde sapma miktarını artırma eğilimi göstermektedir. Bu durum BMA'nın küçük örneklerde daha iyi performans sergilediği sonucuna götürmektedir. Tablo 22'de RK mekanizmasında model ile uyum gösteren madde sayısı sunulmuştur.

Tablo 22. Rastgele Kayıp Mekanizmasında Model ile Uyum Gösteren Madde Sayısı

	n=500			n=1000				n=1500				
	Eksiksiz Veri Seti	%5	%10	%15	Eksiksiz Veri Seti	%5	%10	%15	Eksiksiz Veri Seti	%5	%10	%15
LBS		19	20	18		17	16	19		8	12	9
RA		18	18	14		16	12	14		15	16	7
BMA	19	19	17	12	14	15	12	12	15	15	14	12

Tablo 22 incelendiğinde, eksiksiz veri setinde IPLM ile uyum gösteren madde sayılarının örneklem büyüklüğü n=500 olduğunda 19, n=1000 olduğunda 14 ve n=1500 olduğunda ise 15 olduğu görülmektedir. Kayıp veriler LBS yöntemi ile tamamlandığında, örneklem büyüklüğü n=500 olduğunda %5 kayıpta 19, %20 kayıpta 20 ve %15 kayıpta 18 madde uyumlu olacak şekilde etki gösterdiği görülmektedir. Örneklem büyüklüğü n=1000 olduğunda eksiksiz veri setinden daha fazla sayıda madde uyumlu görünürken, örneklem büyüklüğü n=1500 olduğunda, eksiksiz veri setinden daha az sayıda maddeyi uyumlu görmektedir. Bu sonuçlara bağlı olarak RK mekanizmasında LBS yönteminin, eksiksiz veri seti ile uyumlu bir performans göstermediği ve tutarlı olarak “düşük” ya da “yüksek” kestirimlere neden olduğu şeklinde bir genellemeye varılamayacağı görülmektedir. Madde Tepki Kuramı’nda model parametrelerinin kestiriminde ve model veri uyumunda örneklem büyüklüğü önemli bir etkidir (Bolt ve Lall, 2003; Seungho-Yang, 2007). Madde model uyumunun belirlenmesinde kullanılan χ^2 istatistiğinin örneklem büyüklüğünden etkilendiği göz önünde bulundurulduğunda (Hambleton ve diğ.,1991), LBS yönteminin kullanılmasının örneklemini küçülterek χ^2 ’nin manidar hale gelmesine yol açması beklenir. Ancak buradaki temel sorun kaybın TRK olmamasıdır. LBS yönteminin TRK koşulunda kullanılması durumunda, veri setinde kalan gözlemler eldeki verinin rastgele bir alt örneklemini olacağından, parametre kestirimlerinde yanlılığa yol açmayacaktır. Ancak TRK şartının sağlanmadığı durumlarda, LBS yönteminin kullanılmasıyla elde edilen alt örneklem, tüm gözlemlerin rastgele bir alt örneklemini olmayacağından, parametre kestirimleri yanlılık içerecektir (Allison, 2003; Enders, 2010; McKnight ve diğ., 2007; Rosenthal ve Rosnow, 2008;). LBS yönteminin kullanılması, yanıtlarında kayıp bulunan ve yanıtlarında hiç kayıp bulunmayan kişiler arasında sistematik farklılığın olduğu ölçüde yanlılığa da neden olabilecek ve buna bağlı olarak da kayıp veri içeren kişilerin veri setinden çıkarılmasıyla veri setinde kalan kişilerin evreni temsiliyeti düşecektir (Montalto ve Sung, 1996; Schafer, 1997). Dolayısıyla veri setinde RK mekanizmasında kayıp olduğu durumda, LBS yöntemi eksiksiz veri setini yansıtır bir performans sergilememektedir.

Regresyonla atama yöntemi ile kestirilen madde model uyumunda, uyumlu madde sayıları incelendiğinde bu yöntemin genel olarak eksiksiz veri setinden daha az sayıda maddeyi uyumlu gösterecek atamalar yaptığı görülmektedir. tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça kestirilen değerlerin eksiksiz veri setinden uzaklaştığı bir diğer ifadeyle yöntemin performansı düşmektedir. Alanyazında RA yöntemi, bazı durumlarda eksiksiz veri setlerinden elde edilen değerlerden farklı değerler olsa da, bu değerlerin genel olarak eksiksiz veri setlerinden elde edilen değerlere çok benzer ya da aynı olacağı belirtilmektedir (Roth ve Switzer, 1995; Suraphee ve diğ., 2006). Elde edilen sonuç da bu bilgi ile tutarlıdır.

Beklenti maksimizasyon algoritması ile atama yapıldığında, tüm örneklem büyüklüklerinde kayıp veri oranı arttıkça, uyumlu madde sayısının azaldığı görülmektedir. Özellikle kayıp veri oranı %15 olduğunda model ile uyumlu madde sayısında önemli azalışların olduğu göze çarpmaktadır. Bu nedenle rastgele kayıp mekanizmasına sahip kayıp veri ile karşılaşıldığı durumda kayıp veri oranı yüksek (%15) ise bu yöntemi kullanmaktan kaçınılması önerilebilir. BMA yöntemi özellikle RK varsayımının sağlandığı durumlarda daha iyi tahmin değerleri üretmektedir (Allison, 2003; Baraldi ve Enders, 2010). Ancak elde edilen bulgular madde model açısından bunu desteklememektedir.

SONUÇLAR ve TARTIŞMA

Bu çalışmadan elde edilen bulgular, $-2 \log \lambda$ kriterine bakılarak model veri uyumuna karar vermek istendiğinde, kayıp veri baş etme yöntemlerinden LBS yönteminin diğer yöntemlere göre eksiksiz veri setinden daha uzak kestirime yol açtığı, BMA yönteminin özellikle örneklem büyüklüğü $n=500$ olduğunda tüm kayıp veri oranlarında eksiksiz veri setinden kestirilen $-2 \log \lambda$ değerine istatistiksel olarak en yakın sonuçları verecek atamalar yaptığı sonucuna götürmektedir.

Model veri uyumunun belirlenmesinde kullanılan AIC ve BIC kriterleri bakımından kayıp veri baş etme yöntemlerinin performansları karşılaştırıldığında; LBS yönteminin diğer yöntemlere göre, eksiksiz veri setinden kestirilen AIC ve BIC değerlerinden daha büyük sapmalara neden olduğu göze çarpmaktadır. AIC ve BIC kriterleri bakımından eksiksiz veri setinden yakın kestirimlerin RA ve BMA yöntemleri ile tamamlanan veri setlerinden elde edildiği görülmektedir. Ancak örneklem büyüklüğü ve kayıp veri oranı arttıkça, bu yöntemlerdeki sapmanın da artış gösterdiği söylenebilir. Bu durumda LBS, RA ve BMA yöntemlerinin tümünün örnekleme heterojenleştirecek performans sergiledikleri belirlenmiştir. χ^2 'nin hesaplanmasında, evren ortalaması ile bağımsız değişken ortalaması arasındaki farkın, evren varyansına oranlanması söz konusudur (Chao, 1968). Buradaki oranlama nedeniyle varyans ile χ^2 ters orantılıdır buna bağlı olarak da varyansın büyümesi χ^2 'yi düşürürken, varyansın azalması χ^2 'yi artıracaktır. Dolayısıyla LBS, BMA ve RA yöntemleri, varyansı artırıcı, bir başka deyişle grubu heterojenleştirici performanslar sergiledikleri saptanmıştır.

Tamamen rastgele kayıp koşulunda, RA ve BMA yöntemleri ile tamamlanan veri setlerinde ise birbirine yakın sonuçlar elde edilmiştir. Model veri uyumu AIC kriteri açısından ele alındığında, $-2 \log \lambda$ 'ya benzer olarak, eksiksiz veri setine en uzak kestirimlerin LBS yöntemi ile tamamlanan veri setlerinden elde edildiği belirlenmiştir. BMA ve RA yöntemleri ile tamamlanan veri setlerinden ise birbirine yakın sonuçlar elde edilmiştir. Model veri uyumu BIC kriteri açısından ele alındığında ise AIC ve $-2 \log \lambda$ 'de olduğu gibi, eksiksiz veri setine en uzak kestirimin yine LBS yöntemi ile tamamlanan veri setlerinden elde edildiği, RA, ÇA ve BMA yöntemleri ile tamamlanan veri setlerinde ise, birbirine yakın kestirimler elde edildiği belirlenmiştir.

Liste bazında silme yöntemi, tüm kayıp veri mekanizmalarında $-2 \log \lambda$ kriteri açısından benzer performanslar göstermiştir. Eksiksiz veri seti ile tamamlanan veri seti arasındaki $-2 \log \lambda$ farkını gösteren G^2 değerleri incelendiğinde, bu oranın tüm kayıp veri mekanizmalarında birbirine çok yakın değerlerde olduğu görülmektedir. Örneğin; örneklem büyüklüğü $n=1000$, kayıp veri oranı %10 olduğunda G^2 olabilirlik oranı TRK kayıp koşulunda yaklaşık 917, RK koşulunda yaklaşık 860, ROK koşulunda ise yaklaşık 899'dur. Bu değerlerin manidarlığı test edildiğinde, tüm kayıp veri mekanizmalarında eksiksiz veri seti ile arasındaki farklılığın manidar olduğu görülmektedir. Dolayısıyla LBS yöntemi, 1PLM'de model veri uyumu için $-2 \log \lambda$ bakımından isabetsiz kestirimler yapılmasına yol açmaktadır. TRK ve RK koşullarında diğer yöntemlerle kıyaslandığında ise eksiksiz veri setinden en uzak kestirimleri sunduğu ve bu nedenle de en düşük performansı sergilediği yönünde karara varıldığı söylenebilir.

Liste bazında silme yönteminin model veri uyumu üzerindeki etkisini AIC ve BIC kriterleri bakımından değerlendirildiğinde, $-2 \log \lambda$ 'de olduğu gibi, tüm kayıp veri mekanizmalarında benzer performansı sergilediği görülmektedir. Eksiksiz veri setinden uzaklaşma miktarının diğer kayıp veri baş etme yöntemlerine göre fazla olmasının, LBS yönteminin model veri uyumunda da AIC ve BIC kriterleri bakımından düşük performans sergilemesine neden olduğu belirtilebilir. LBS yönteminin performansı diğer kayıp veri baş etme yöntemleri ile karşılaştırıldığında ise, en düşük performansı TRK ve RK mekanizmalarında sergilediği ifade edilebilir.

Regresyonla atama yönteminin 1PLM'de model veri uyumu üzerindeki performansı incelendiğinde, $-2 \log \lambda$ kriteri açısından en iyi performansı RK mekanizmasında gösterdiği saptanmıştır. Eksiksiz veri seti ile tamamlanan veri seti arasındaki $-2 \log \lambda$ farkını gösteren G^2 değerleri incelendiğinde, bu oranın tüm kayıp veri mekanizmalarında birbirinden farklı değerlerde olduğu görülmektedir. RA yöntemi 1PLM'de model veri uyumu için $-2 \log \lambda$ bakımından RK mekanizmasında isabetli kestirimler yapılmasını sağlamaktadır. RK ve TRK mekanizması olduğu durumda, RA ve diğer yöntemlerin $-2 \log \lambda$ üzerindeki performansları karşılaştırıldığında, RK durumunda RA yönteminin,

TRK koşulunda ise BMA yönteminin daha iyi performans sunduğu sonucuna ulaşılmıştır. Burada performansın iyi olmasından kastedilen, eksiksiz veri setine en yakın kestirimlerin elde edilmiş olmasıdır. RA yönteminin model veri uyumu üzerindeki etkisi, AIC ve BIC kriterleri bakımından değerlendirildiğinde, $-2 \log \lambda$ ile benzer bir tablo ile karşılaşılmıştır. Eksiksiz veri setinden uzaklaşma miktarının kayıp veri baş etme yöntemlerinin performansını gösterdiği göz önünde bulundurulduğunda, RA yöntemi IPLM’de model veri uyumu için AIC ve BIC bakımından RK mekanizmasında oldukça isabetli kestirimler yaparken, TRK mekanizmasında performansının düşük olduğu belirlenmiştir. RK ve TRK mekanizmalarının olduğu durumlarda, RA ve diğer yöntemlerin AIC ve BIC üzerindeki performansları karşılaştırıldığında, RK durumunda RA yönteminin oldukça iyi ve BMA ile benzer performans sergilediği, TRK koşulunda ise BMA yöntemleri ile benzer sonuçlar sunduğu söylenebilir.

Beklenti maksimizasyon algoritması ile atama yönteminin IPLM’de model veri uyumu üzerindeki performansı incelendiğinde, $-2 \log \lambda$ kriteri açısından en iyi performansı RK mekanizmasında gösterdiği sonucuna ulaşılmıştır. Eksiksiz veri seti ile tamamlanan veri seti arasındaki $-2 \log \lambda$ farkını gösteren G^2 değerleri incelendiğinde, bu oranın RK mekanizmasında düşük, TRK mekanizmasında ise daha yüksek olduğu görülmüştür. BMA yöntemi IPLM’de model veri uyumu için $-2 \log \lambda$ bakımından RK mekanizmasında bazı koşullarda isabetli kestirimler sunmaktadır. Eksiksiz veri setinden sapmaların yönüne bakıldığında ise TRK ve RK mekanizmalarında model veri uyumunu artırma yönünde performans gösterirken ROK mekanizmasında model veri uyumunu düşürme yönünde performans sergilediği belirlenmiştir. RK mekanizması olduğu durumda BMA ve diğer yöntemlerin $-2 \log \lambda$ üzerindeki performansları incelendiğinde, RA ile benzer kestirime yol açtığı görülmüştür. RK mekanizmasında özellikle örneklem büyüklüğü $n=500$ ve kayıp veri oranı %5 olduğunda en iyi performansı gösteren yöntem olarak ön plana çıkmaktadır. TRK durumunda da RA yöntemleri ile benzer ve iyi sonuçlar sunduğu belirlenmiştir. BMA ile atama yönteminin model veri uyumu üzerindeki etkisi AIC ve BIC kriterleri bakımından değerlendirildiğinde, diğer yöntemlerde olduğu gibi $-2 \log \lambda$ ’de karşılaşılan sonuç ile karşılaşılmaktadır. BMA yöntemi IPLM’de model veri uyumu için AIC ve BIC bakımından RK mekanizmasında oldukça isabetli kestirimler sunmuştur. TRK mekanizmasında eksiksiz veri setine en yakın kestirimleri veren yöntem olmamakla birlikte, eksiksiz veri setine yakın kestirimler üretmiştir. RK ve TRK mekanizması olduğu durumda BMA ve diğer yöntemlerin AIC ve BIC üzerindeki performansları incelendiğinde, her iki durumda da BMA yönteminin oldukça iyi ve RA yöntemine benzer performans sunduğu ifade edilebilir.

Model veri uyumu söz konusu olduğunda TRK mekanizmasında LBS, BMA yöntemlerinin iyi, RA yönteminin kısmen iyi, RK mekanizmasında tüm yöntemlerin iyi performans sergilediği saptanmıştır. Buna ek olarak tüm kayıp veri mekanizmalarında, tüm kayıp veri baş etme yöntemlerinin, kayıp veri oranı arttıkça, model veri uyumu kestirimleri bakımından eksiksiz veri setinden uzaklaştığı ve performanslarının düştüğü vurgulanabilir.

Liste bazında silme yöntem ile tamamlanan kayıp verilerin ardından TRK ve RK mekanizmalarında IPLM ile uyum gösteren madde sayıları, eksiksiz veri seti ile uyumu referans alınarak değerlendirildiğinde, TRK mekanizmasında olması gerekenden fazla sayıda maddeyi model ile uyumlu gösterme eğilimi taşıdığı görülmüştür. LBS yönteminin hiçbir kayıp veri mekanizmasında eksiksiz veri seti ile uyumlu sonuçlar üretmediği, dolayısıyla LBS yönteminin IPLM’de madde model uyumu kestirimi söz konusu olduğunda kayıp veri baş etme yöntemi olarak tercih edilmemesi gerektiği sonucuna ulaşılmıştır.

Regresyonla atama yöntemi, madde model uyumu söz konusu olduğunda TRK mekanizması altında genel olarak daha az sayıda maddeyi uyumlu gösterecek yönde performans sergilediği belirtilebilir. TRK mekanizmasında madde sayısındaki azalma kayıp veri oranı ile ilişkilidir. Bir diğer deyişle kayıp veri miktarı arttıkça, uyumlu madde sayısı azalmaktadır. Bu durum RA yönteminin kayıp veri oranından etkilendiğini göstermektedir. RK mekanizması genel olarak daha az sayıda maddeyi uyumlu gösterme eğiliminde olmakla birlikte, kayıp veri oranı %15 olduğunda, tüm örneklem büyüklüklerinde uyumlu madde sayısında önemli düşümlere yol açacak atamalar yaptığı belirtilebilir.

Beklenti maksimizasyon algoritması yönteminin TRK ve RK mekanizmalarında IPLM’de madde model uyumuna etkisi değerlendirildiğinde, RK mekanizmasında tüm örneklem büyüklüklerinde ve

%5 kayıp oranında, eksiksiz veri setine en yakın kestirimleri sunmuştur. Dolayısıyla veri setinde RK mekanizması olduğunda, kayıp veri oranı düşük ise kullanılabilir bir yöntemdir. Kayıp veri oranı arttıkça, uyumlu madde sayısı azalmakta ve dolayısıyla BMA yönteminin RK mekanizmasında performansı düşmektedir. TRK mekanizmasında performansı incelediğinde ise, genel olarak tüm örneklem büyüklüklerinde eksiksiz veri setinden daha az sayıda maddeyi uyumlu kılmaktadır. Kayıp veri oranı %15 olduğunda performansında önemli düşüşler yaşanmaktadır. Dolayısıyla TRK mekanizmasında kayıp veri oranı yüksek olduğunda, BMA'nın kullanılmaması gereken bir kayıp veri baş etme yöntemi olduğu belirtilebilir.

Veri setinde TRK mekanizmasında kayıp veriler olduğunda MTK 1 PLM'e göre modelleme yapılmak istendiğinde, model veri uyumunun en iyi şekilde tahminlenebilmesi için; LBS, RA ve BMA yöntemlerinin kullanılması önerilebilir. Veri setinde RK mekanizmasında kayıp veriler olduğunda ve MTK 1PLM'e göre modelleme yapılmak istendiğinde ise, model veri uyumunun en iyi şekilde tahmin edilebilmesi için, BMA ve RA yöntemlerinin kullanımı önerilebilir. Veri setinde TRK mekanizmasına sahip kayıp veri olması durumunda 1PLM'de kayıp veri baş etme yöntemlerinin performansları göz önünde bulundurulduğunda, madde model uyumu için eğer örneklem küçükse (örneğin n=500) BMA yöntemi tercih edilebilir. Veri setinde RK mekanizmasına sahip kayıp veri olması durumunda 1PLM'de kayıp veri baş etme yöntemlerinin performansları göz önünde bulundurulduğunda madde model uyumu söz konusu ise, BMA yöntemi düşük kayıp veri oranları için kullanılabilir bir yöntemdir.

KAYNAKÇA

- Ackerman, T.A. (2005). Multidimensional item response theory modeling. In A. Maydeu-Olivares & J.J. McArdle (Eds.). *Contemporary psychometrics* (3–26) Mahwah, NJ: Lawrence Erlbaum Associates.
- Ackerman, T.A., Gierl, M.J., & Walker, C.M. (2003). *Using multidimensional item response theory to evaluate educational and psychological tests*. *Educational Measurement: Issues and Practice*, 22(3), 37-51.
- Adams, R. (2005). *Reliability as a measurement design effect*. *Studies in Educational Evaluation*, 31, 162–172. DOI: 10.1016/j.stueduc.2005.05.008
- Aiken, L. R. (1995). *Psychological testing and assessment*. Boston: Allyn and Bacon
- Allison, P. (2002). *Missing Data*. Thousand Oaks, CA:Sage.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 545-557
- Baker, F. B. (1998). An investigation of the item parameter recovery of a Gibbs sampling procedure. *Applied Psychological Measurement*, 22, 153-169.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology*, 48, 5–37.
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulanması*. Ankara: ÖSYM.
- Blozis, S. A., Conger K. J., & Harring, J. R. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development: Special Issue on Longitudinal Modeling of Developmental Processes*, 31, 340-346.
- Bock, R. D. (1997). A brief history of Item Response Theory. *Educational Measurement: Issues and Practice*, 16, 21-33.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and non-compensatory multidimensional Item Response models using Markov Chain Monte Carlo. *Applied Psychological Measurement*, 27(6), 395-414. Web: <http://apm.sagepub.com> adresinden 02.10.2015'de alınmıştır.
- Cao, Y., & Poh, K. L. (2006). An accurate and robust missing value estimation for microarray data: Least absolute deviation imputation, *Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA'06)*, 157-161.
- Chang, Y. W. (1992). A comparison of unidimensional and multidimensional IRT approaches to test information in a test battery (Unpublished doctoral dissertation, University of Minnesota). <http://files.eric.ed.gov/fulltext/ED344940.pdf> adresinden 20.01.2015'de alınmıştır.
- Chen, S. F., Wang, S., & Chen, Y. C. (2011). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications*, 39, 4026–4031.

- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Newyork: Holt, Rinehart and Winston
- Cronbach, J. L. (1990). *Essentials of psychological testing* (5. Ed). Harper Collins Publishers, Inc
- Çelen, Ü. (2008). Klasik Test Kuramı ve Madde Tepki Kuramına dayalı olarak geliştirilen iki testin psikometrik özelliklerinin karşılaştırılması. *İlköğretim Online*, 7(3), 758-768.
- Çokluk, Ö. ve Kayrı, M. (2011). Kayıp değerlere yaklaşık değer atama yöntemlerinin ölçme araçlarının geçerlik ve güvenilirliği üzerindeki etkisi. *Kuram ve Uygulamada Eğitim Bilimleri*, 11(1), 289-309.
- DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. London: Oxford Press.
- Demir, E. (2013). *Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi* (Yayımlanmamış doktora tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara). <http://acikarsiv.ankara.edu.tr/> adresinden 08.10.2016 tarihinde alınmıştır.
- Demir, E. ve Parlak, B. (2012). Türkiye’de eğitim araştırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(1), 230-241
- Dias, J. G. (2006). *Latent class analysis and model selection*. In M. R. Kruse, C. Borgelt, A. Nürberger, & W. Gaul, *From data and information analysis to knowledge engineering* (pp. 95-102). Berlin: Springer-Verlag.
- Drasgow, F., Levine, M., Tsien, S., Williams, B., & Mead, A. (1995). Fitting polytomous Item Response Theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-165. doi:10.1177/014662169501900203
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446. Doi:10.1037/0012-1649.43.6.1428
- Doğanay Erdoğan, B., (2012). *Çoklu atama yöntemlerinin Rasch modelleri için performansının benzetim çalışması ile incelenmesi* (Yayımlanmamış doktora tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara). <http://acikarsiv.ankara.edu.tr> adresinden 13.01.2015 tarihinde alınmıştır.
- Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, 64, 407-433.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Enders, C. K. (2010). *Applied missing data analysis*. (1st ed.). New York: The Guilford Publications, Inc.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*. 8(3), 430-457.
- Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*. 64(3), 419-436.
- Fiona, M. S., Heather, S., Hude, G., & William G. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology* 6:57. DOI: 10.1186/1471-2288-6-57
- Furlow C. F., Fouladi R. T., Gagne P., & Whittaker T.vA. (2007). A Monte Carlo study of the impact of missing data and differential item functioning on theta estimates from two polytomous Rasch family models. *Journal of Applied Measurement*. 7(8),388-403
- Gao, F., & Chen, L. (2005). Bayesian or Non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education*, 18(4), 351-380.
- Goldman, S. H., & Raju, N. S. (1986). Recovery of one- and two-parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement*, 46(1),11-21.
- Hambleton R. K., Swaminathan H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.
- Hambleton, R. K. (1989). Principles and salected applications of Item Response Theory. R. L. Linn (Ed), *Educational Measurement* (pp.147-200). Washington, DC: American Council of Education.
- Hohensinn, C., & Kubinger, K. D. (2011). On the impact of missing values on the item fit and the model validness of the Rasch model. *Psychological Test and Assessment Modeling*, 53(3), 380-393.
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Hurvich, C., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297-307

- Kankaras, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items. A comparison of factor analytic, Item Response Theory and Latent Class Approaches. *Sociological Methods & Research*, 40(2), 279-310.
- Karasar, N. (2007). *Bilimsel araştırma yöntemi: Kavramlar, ilkeler, teknikler* (17. Baskı). Ankara: Nobel.
- Köse, İ. A. (2014). The effect of missing data handling methods on goodness of fit indices in confirmatory factor analysis. *Educational Research and Reviews*, 9(8), 208-215. Doi:10.5897/ERR2014.1709
- Lin, T. H. (2006). A comparison of model selection indices for Nested Latent Class Models. *Monte Carlo Methods and Applications*, 12(3-4), 239-259.
- Lin, T. H. (2012). Model selection information criteria in Latent Class Models with missing data and contingency question. *Journal of Statistical Computation and Simulation*, 84(1),159-170 doi:10.1080/00949655.2012.698621.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- Lord, F. M. (1968). An Analysis of the Verbal Scholastic Aptitude Test Using Birnbaum's Three-parameter Logistic Model. *Educational and Psychological Measurement*, 28, 989-1020.
- Lord, F. (1980). *Applications of Item Response Theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York: The Guilford Publications, Inc.
- Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3- item analysis and test scoring with binary logistic models* (2nd Ed). Mooresville: Scientific Software Inc.
- Montalto, C. P., & Sung, J. (1996). Multiple imputation in the 1992 survey of consumer finances. *Financial Counseling and Planning*, 7, 133 – 141.
- Morren, M., Gelissen, J., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: A restricted Latent Class factor analysis. *Sociological Methodology*, 1(1), 13-47. Doi: 10.1111/j.1467-9531.2011.01238
- Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in Latent Class analysis and Growth Mixture Modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous Item Response Theory models. *Applied Psychological Measurement*, 24(1), 24-50
- R Development Core Team (2011). R: A language and environment for statistical computing, a foundation for statistical computing. Vienna, Austria, ISBN 3-900051-0-70, Erişim:[<http://www.R-project.org>].
- Reeve, B. B. (2002). *An introduction to modern measurement theory*. <http://appliedresearch.cancer.gov/areas/cognitive/immmt.pdf> adresinden 21.12.2015 tarihinde alınmıştır.
- Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, A. N., & Armstrong, A. J. (2007). Modern psychometric methodology: Applications of Item Response Theory. *Rehabilitation Counselling Bulletin*, 50(3), 177-178.
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item Response Theory. Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14(2), 95-101.
- Rosenthal, R., & Rosnow, R. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd Ed). Boston : McGraw-Hill.
- Roth, P. L., & Switzer, F. S. (1995). A Monte Carlo analysis of missing data techniques in a HRM setting. *Journal of Management*, 21(5), 1003-1023.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/Crc.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147 – 177.
- Seungho Yang, M. A. (2007). *A comparison of unidimensional and multidimensional Rasch models using parameter estimates and fit indices when assumption of unidimensionality is violated* (Unpublished Doctoral Dissertation, The Ohio State University). <https://etd.ohiolink.edu> adresinden alınmıştır.
- Sijtsma, K. & van der Ark, L. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38(4), 505-528.
- Suraphee, S., Raksmanee, C., Busaba, J., Chaisorn, C., & Nakornthai, W. (2006). A comparison of estimation methods for missing data in Multiple Linear Regression with two independent variables. *Thailand Statistician*, 4, 13-26.
- Thissen, D., & Wainer, H. (1982). Some standart errors in Item Response Theory. *Psychometrika*, 47, 397-412.

- Turgut, F. (1992). *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Yargı.
- Van Ginkel, J. R. (2007). *Multiple imputation for incomplete test, questionnaire, and survey data*. (Unpublished Doctoral Dissertation, Tilburg University). <https://pure.uvt.nl> adresinden 21.09.2016 tarihinde alınmıştır.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion and the Bayesian Information Criterion. *Psychological Methods*, 17(2), 228-243.
- Wainer, H., & Thissen, D. (2001). True score theory: The traditional method. D. Thissen, & H. Wainer (Ed). *Test scoring*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New Jersey: Lawrence Erlbaum Associates.
- Yang, C., & Yang, C. (2007). Separating Latent Classes by information criteria. *Journal of Classification*, 24, 183-203.
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, 52, 275-291.
- Yu, H. C. (2013). A simple guide to the Item Response Theory (IRT) and Rasch Modeling. <http://www.creative-wisdom.com/computer/sas/IRT.pdf> adresinden 18.09.2016 tarihinde alınmıştır.
- Yurdugül, H. (2006). Paralel, eşdeğer ve konjenetik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15-37.

EXTENDED ABSTRACT

Introduction

The measured properties in education and psychology are mostly latent variables that do not allow direct observation. Various test theories have been developed to account for latent individual properties with the help of observable variables and those properties have been measured with instruments that have been developed based on the theories. One of the most significant theories developed is Item Response Theory (IRT). IRT is based on responses of test takers to items. It grounds on expressing the relationship between ability levels of individuals associated with the measured property and responses of test takers with the help of a mathematical function (Embretson and Reise, 2000; Hambleton and Swaminathan, 1989).

Item Response Theory claims the following: tests and item parameters must be independent from the group under application and individual levels of ability (θ) must be independent from the measurement instrument applied. Nevertheless, when the chosen model does not fit to the data or when there is low model data fit, the benefits offered by the theory will not be utilized in terms of items and ability parameters (Hambleton et al., 1991; Orlando and Thissen, 2000). If model data fit in a model is poor, item statistics and item information functions will be misleading (Ackerman, 2005; Chang, 1992; Hambleton et al., 1991).

In IRT, the main criterion to be considered when deciding a model to be employed through model data fit is choosing the model to present maximum individual data with minimum error in estimation. Thus, most probable estimations of latent features under measurement are likely with the help of individual responses. However, reaching latent variables that are attempted to be measured with tests is likely with individual responses to test items. In this context, when individuals avoid responses to items, skip items or give no response for any reason, the greatest handicap in implications will appear (Hohensinn and Kubinger, 2011).

In case of missing data issues in research, it is not appropriate to think that the best fitting model to all data has been decided or in other words, problems will arise in the process of choosing the accurate model for data. Deficiencies in response patterns will considerably hinder ability estimations. The study aims at examining the effects of missing data handling techniques on model data fit and item model fit of IRT one parameter logistic model, taking different sample sizes and missing data conditions.

Method

The study is a simulation that attempts to reveal the effect of various missing data handling techniques on model data fit in Item Response Theory one parameter logistic model, using artificial (simulative) data sets. At the same time, it could be suggested that it has the quality of fundamental research as it aims at contributing new findings to the available theoretical knowledge (Karasar, 2007).

In data generation, missing data generation R program was employed at a desired level in accordance with missing data mechanisms in the generated complete data sets (R Development Core Team, 2011). The number of items for the generated data set was 20 and the number of response category for each item was two (1-0 scoring). The number of individuals was 500, 1000 and 1500 and the data were generated based on IRT one parameter logistic model. Item difficulty in the generated data (b) was restricted from minimum -2 logit to maximum +2 logit and item discrimination was fixed as 1.5.

In the generated data sets, missing data at 5%, 10% and 15% were obtained to provide missing at complete random (MCR) and missing at random (MR) mechanisms. Missing data were generated in MCR mechanism whereas pores at pre-defined ratios were deleted in the samples of 500, 1000 and 1500 individuals, independently from any variables. In order to generate MR mechanism, another variable was randomly defined at a three level sequencing scale which was different from the measured feature, but still belonged to the measured individuals. Then, missing data were generated respectively at Level 1 (20%), Level 2 (30%), and Level 3 (50%). The generated missing data were resolved, using expectation maximization algorithm, regression imputation, and listwise deletion method (LM).

In the study, model data fit was estimated first in the generated complete data sets. Then, the same predictions were applied in the complete data sets and resulting values obtained from the complete data set were taken as reference and compared to the above mentioned values. $-2 \log \lambda$, AIC, and BIC values were used to evaluate the model data fit. χ^2 item fit approach was used to decide item model fit.

Results and Discussion

It is seen that data sets which are completed with MR and EMA methods give the closest estimations to the complete data set in terms of $-2 \log \lambda$, AIC, and BIC criteria that are employed to decide model data fit. However, as sample sizes and missing data amount increase, deviation in these methods could be suggested as increasing. In this case, all LM, MR, and EMA methods were found to have performance to heterogenize the sample. Under missing at complete random (MCR) condition, similar results were obtained in data sets completed with missing at random (MR) and expectation maximization algorithm (EMA) methods. When compared to other methods, listwise deletion method under MCR and MR conditions could be decided to present the most improbable estimations from the complete data set and thus to have the lowest performance.

When there is missing data in MCR mechanism of the data set, and modelling according to IRT one parameter logistic method is desired, LM, MR and EMA methods were found eligible for the best model data fit estimation; when there are missing values in MR mechanism, and modelling according to IRT one parameter logistic method is desired, EMA and MR were found eligible for the best model data fit estimation. In case of missing data in data sets with MCR and MR mechanisms, EMA method could be preferred for item model fit if sample is small.

Lise Öğrencilerinin Meslek Seçimlerini Etkileyen Faktörlerin İkili Karşılaştırmalar Yöntemiyle İncelenmesi

Investigation of the Factors Affecting Occupation Choices of High School Students with Paired Comparison Method

Funda NALBANTOĞLU YILMAZ *

Öz

Araştırmanın amacı, lise öğrencilerinin mesleki seçimlerinde rol oynayan faktörlerin etki sıralarının ikili karşılaştırmalar yoluyla belirlenmesidir. Araştırma, eksiksiz bilgi toplanabilen ve içtenlikle bilgi verdiği düşünülen 355 lise öğrencisi ile yürütülmüştür. Öğrencilerin mesleki seçimlerinde rol oynayan faktörlerin belirlenmesinde alan yazın taraması yapılmış ve öğrencilere yazdırılan kompozisyonlar incelenmiştir. Bu doğrultuda öğrencilerin mesleki seçimlerini etkileyen 9 faktör belirlenmiştir. Bu etkiler ikişerli olarak 36 maddede öğrencilere sunulmuş ve öğrencilerden her bir ikili gruptan hangisinin mesleki seçimlerinde daha etkili olduğunu belirtmeleri istenmiştir. Öğrencilerin mesleki seçimlerinde rol oynayan faktörlerin etki sıralarının belirlenmesinde ikili karşılaştırmalarla ölçekleme kullanılmıştır. Ölçekleme çalışması sonucunda, 11. sınıflar hariç öğrencilerin meslek seçimlerini bir mesleğe duyulan ilginin en fazla etkilediği görülmektedir. Aynı durum cinsiyete bağlı incelendiğinde de kız ve erkek öğrencilerin meslek seçimlerini en fazla mesleğe duydukları ilginin etkilediği sonucuna ulaşılmıştır. Bununla birlikte öğrencilerin üniversite giriş sınavından alacakları puanlarında meslek seçimlerini büyük oranda etkilediği belirlenmiştir. Diğer taraftan, öğrencilerin mesleki seçimlerini en az çevrenin (arkadaş, öğretmen vb.) etkilediği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Meslek seçimi, ikili karşılaştırmalar, ölçekleme

Abstract

The aim of the research is to determine effect sequences of factors that play a role in the professional choices of high school students. The study was conducted from 355 high school students who can give thorough and sincere information. The compositions written by students were examined and the field literature survey was done to determine the factors that play a role in the professional selection of the students. In this respect, 9 factors that affect the professional selection of the students have been determined. These factors were presented to students in pairs. And then students were asked to indicate which pair is more effective in their professional selection. Paired comparison scaling was used to determine the order of influence of the factors that affect the professional preferences of the students. According to the results of the analysis, it seems that the most important effect of professional selection of students (except 11th grade students) is the interest in a profession. Also the professional selection of girls and boys are mostly influenced by their interests. At the same time, it is observed that students' university entrance examination scores substantially affect their professional selection. On the other hand, the preferences of the students are influenced by at least the environment.

Keywords: Choice of profession, paired comparison, scaling

GİRİŞ

Meslek ve kariyer seçimi öğrencilerin hayatlarındaki dönüm noktalarından biridir. Meslek, yararlı mal veya hizmet üretmek ve karşılığında bir gelir elde etmek için yapılan, belirli bir eğitimle kazanılan ve kuralları toplumca belirlenmiş faaliyetler bütünüdür (Kuzgun, 2014). Meslek, bireyin etrafta saygı görmesine, toplumda yer edinmesine olanak sağlar (Kuzgun, 2000). Melek seçimi ise,

* Yrd. Doç. Dr., Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, Türkiye, fundan@nevsehir.edu.tr

“bireyin meslekleri çeşitli yönleri ile değerlendirip ihtiyaç ve beklentileri açısından istenilen yönleri çok, istenilmeyen yönleri az olan birine yönelmeye karar vermesi” olarak tanımlanabilir (Arsoy, 2010, s. 171). Böyle bir kararda bireyin kişisel seçimleri önemlidir.

Meslek seçimi bireyin yaşantısında çok önemli bir yere sahiptir. Meslek bireyin yaşamında sürdürdüğü geçici bir iş değil, sürekli bir uğraştır. Meslek seçimi bireyin yaşantısına yön verir. Meslek seçimleri bireyin yaşantısında kiminle evleneceği, çalışma biçimi, günlük yaşantı biçimi, yaşam tarzı, dünya görüşü gibi çoğu durumu etkiler.

Meslek seçiminde birey, meslekler arasında ayırım yaparak, ilgi, yetenek ve istekleri doğrultusunda seçeceği mesleği belirler. Bireyin meslek seçiminde ve bireyleri kendilerine uygun bir mesleğe yönlendirmede okullarda bulunan rehberlik ve psikolojik danışma birimleri görevlidir. Ayrıca okullarda yürütülen mesleki rehberlik ve kariyer danışmanlığı programları mesleğe yönlendirmede önemli bir yere sahiptir. Mesleki rehberlik ve kariyer danışmanlığının yetersiz olduğu noktada bireyler korkular, kaygılar, yanlış bilgiler, çevre etkisi gibi birçok faktör nedeniyle yanlış meslek seçimi yapabilmektedir.

Meslek seçiminde bireyin kişisel nitelikleri ile meslek özellikleri arasındaki bağ önem taşımaktadır. Yazıcı (2012), bireyin meslek seçimini etkileyen iki tür kaynaktan bahsetmiştir. Bunlardan ilki; işsizlik, iş alanlarının çeşitlenmesi, çalışma koşullarının değişmesi, aile, okul ve çevre etkileri gibi bir mesleğe karar verme sürecini etkileyen dışsal kaynaklardır. İkincisi ise, yetenekler, ilgiler, kişilik özellikleri ve benlik kavramı gibi meslek seçiminde etkili olan içsel kaynaklardır. Yeşilyaprak (2013), bireyi bir mesleği seçmeye götüren süreci etkileyen faktörleri bireyin ailesi, ilişki kurduğu kişiler, eğitimi, ihtiyaçları, ilgileri, yetenekleri, değerler sistemi, benlik kavramı, kişilik özellikleri, yaşadığı çevrenin özellikleri, mevcut iş alanları, iş koşulları ve arz talep durumu olarak belirtmiştir. Arsoy (2010) ise, bireyin meslek seçme kararının toplumun mesleklere attığı değerlerden, bireyin ailenin sosyo-ekonomik ve sosyo-kültürel düzeyinden, bireyin ilgi ve yeteneklerinden, bireyin benliğine uygun bir uğraşı bulma arzusundan, iş olanaklarından, aile ve çevre etkisinden etkilendiğini belirtmiştir. Super (1957), bireyin meslek seçimini etkileyen faktörleri biyolojik (cinsiyet, fiziksel özellikler vb.), sosyolojik (aile, çevre beklentisi vb.), kişisel ve psikolojik (ilgi ve yetenekler, ihtiyaçlar), ekonomik (ailenin ekonomik düzeyi, insan gücü ihtiyacı), politik (iş bulma olanakları, yetiştirme fırsatları vb.) ve şansla ilgili faktörler (iş yerlerini etkileyen beklenmedik olaylar vb.) olarak sınıflamıştır (akt., Arsoy, 2010).

Bir ülkenin kalkınmasında insan gücü önemli kaynaklardan biridir. İnsan gücünün etkili ve verimli bir biçimde kullanılabilmesi, bireylerin ilgi ve yeteneklerine uygun meslek alanlarına yerleşmelerine oldukça bağlıdır (Arsoy, 2010). Bireyin meslek seçimi ile ilgili doğru bir karara varması önemlidir. Çünkü bireyin kendini ifade yolu olarak görülebilen meslek seçimi, istek, ilgi ve ihtiyaçları en iyi şekilde karşılamaya yönelik gerçekleştirildiği takdirde, birey kendini gerçekleştirebilecek ve sağlıklı benlik algısı oluşturabilecektir (Aydın, 2011). Böylece birey psikolojik gereksinimlerini karşılayabilecek, ruhen sağlıklı ve mutlu olabilecektir. Bununla birlikte meslek seçimi, bireyin gelecekteki başarı ve başarısızlığını, işinden hoşlanıp hoşlanmayacağını hatta yaşam biçimini de etkileyebilmektedir (Kuzgun, 1982). Fakat meslek belirleme sürecinde bireyin kişisel bazı düşünceleri seçimlerini etkileyebilir. Hatta bireyin geleceğini planlama aşamasında seçimini etkileyen faktörlerin önem sırası bireyin farklı meslekleri seçmesini, farklı mesleklere yönelmesini de sağlamaktadır. Bununla birlikte, günümüzde iş ve meslek sayısının çok fazla olduğu, buna rağmen işsizlik oranının fazlalığı düşünüldüğünde meslek seçmenin, mesleki tercihleri etkileyen faktörleri belirlemenin ve mesleki rehberlik çalışmalarının ne kadar önemli olduğu anlaşılmaktadır. Literatürde öğrencilerin meslek seçimini konu alan çalışmalar bulunmaktadır. Bu çalışmalardan bir kısmı (Alkan, 2014; Çiftçi, Bülbül, Bayar-Muluk, Çamur-Duyan ve Yılmaz, 2011; Dinç, 2008; Genç, Kaya ve Genç, 2007; Özyürek ve Atıcı, 2002; Pekmaya ve Çolak, 2013; Sarıkaya ve Khorshid, 2009; Şirin, Öztürk, Bezci, Çakar ve Çoban, 2008) üniversite öğrencilerinin meslek seçimini konu almakta iken bir kısmı (Kalkan, 2014; Kars, Arslan, Erik, Avcı, Bucaktepe, Celepkolu ve Şahin, 2014) lise öğrencilerinin meslek seçiminde karşılaştıkları sorunları ve meslek liselerini tercih etmede etkili olan faktörleri ele almaktadır. Literatürdeki sınırlı sayıda çalışma (Kıyak, 2006; Vurucu, 2010;

Yelken, 2008) ise lise öğrencilerinin meslek seçimlerini etkileyen faktörleri ve lise öğrencilerinin meslek seçimi yaparken temel aldığı kriterleri belirlemeyi içermektedir. Bu nedenle, lise öğrencilerinin meslek seçimlerinde (tercihlerinde) etkili olan faktörlerin ve bu faktörlerin öğrencilerin hayatlarındaki önem sıralarının belirlenmesi iş gücünün kullanımı, ülke kalkınması ve öğrencilere yönelik yapılacak mesleki rehberlik çalışmaları açısından önemli görülmektedir.

Araştırmanın Amacı

Araştırmada lise öğrencilerinin meslek seçimlerinde rol oynayan faktörlerin etki düzeylerinin ikili karşılaştırma yoluyla ölçeklenmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıdaki sorulara cevap aranmıştır.

1. Öğrencilerin meslek seçimlerini etkileyen faktörlerin ölçek değerlerine göre sıralaması nasıldır?
2. Cinsiyete göre öğrencilerin meslek seçimlerini etkileyen faktörlerin ölçek değerlerine göre sıralaması nasıldır?
3. Sınıf düzeylerine göre öğrencilerin meslek seçimlerini etkileyen faktörlerin ölçek değerlerine göre sıralaması nasıldır?

YÖNTEM

Araştırma Türü

Araştırmada, lise öğrencilerinin mesleki tercihlerini etkileyen faktörlerin etki sıralarının ikili karşılaştırmalar yöntemiyle incelenmesi amaçlanmıştır. Araştırmada öğrencilere ait var olan durum belirlenmeye çalışıldığından betimseldir.

Çalışma Grubu

Araştırma, 2016-2017 eğitim-öğretim yılı bahar döneminde Nevşehir il merkezindeki 5 farklı lisenin 11 ve 12. sınıflarında okuyan ve eksizsiz bilgi toplanabilen 355 öğrenci ile gerçekleştirilmiştir. Araştırmaya katılan öğrencilerin %28.2'si (n=100) 11. sınıf, %71.8'i (n=255) 12. sınıf öğrencisidir. Ayrıca öğrencilerin %65.9'u (n=234) kız, %34.1'i ise (n=121) erkektir. Lise yılları (15-18 yaşları), özellikle 11 ve 12. sınıf, bireyin kendisi ve meslekler hakkında oluşturduğu algılara dayanarak mesleki karar verme sürecini en yoğun yaşadığı dönemdir (Arsoy, 2010; Yeşilyaprak, 2013). Bu nedenle çalışma grubuna lise öğrenimine devam eden 11 ve 12. sınıf öğrencileri seçilmiştir.

Veri toplama Aracı

Araştırmada, öğrencilerin meslek seçimini etkileyen faktörlerin ikili karşılaştırmalar yöntemiyle ölçeklenmesi amacıyla araştırmacı tarafından bir veri toplama aracı geliştirilmiştir. Veri toplama aracı hazırlanmadan önce mesleki tercihler ve meslek seçimi ile ilgili alan yazın taraması yapılmıştır (Sarıkaya ve Khorshid, 2009; Dinç, 2008; Pekkaya ve Çolak, 2013; Kalkan, 2014; Kıyak, 2006; Özen, 2011; Kars ve diğerleri, 2014; Özyürek ve Atıcı, 2002; Varol, 1990). Yapılan alan yazın taramasında, çoğunlukla bir meslek grubuna ait bireylerin o mesleği seçme nedenlerinin araştırıldığı, sınırlı sayıda çalışmada ise lise öğrencilerinin meslek seçimi ve mesleki tercihlerine odaklanıldığı belirlenmiştir. Alan yazın taramasından sonra hedef gruba benzer bir grupta, öğrencilerin meslek tercihlerini yaparken neleri göz önünde bulundurdıkları ile ilgili kompozisyon yazdırılmıştır. Öğrencilerin mesleki tercihlerini etkileyen faktörlerin belirlendiği çalışmalardan, Arsoy (2010), Kıyak (2006), Yazıcı (2012) ve Yeşilyaprak'ın (2013) tanımladığı meslek seçimini etkileyen faktörlerden ve kompozisyon verilerinden hareketle lise öğrencilerinin mesleki tercihlerinde etkili olan 9 faktör belirlenmiştir. Bu faktörler; mesleğin maddi olarak kazanç durumu (A), iş bulma

kolaylığı (B), bireysel yetenekler (C), mesleğe duyulan ilgi (D), mesleğin toplumsal saygınlığı (E), aile etkisi/beklentisi (F), çevre etkisi (arkadaş vb.) (G), üniversite giriş sınavından alınacak puan (H) ve mesleğe ait eğitimin kolay (I) olmasıdır. Belirlenen bu faktörlerin her biri birbiriyle ikili gruplar halinde eşleştirilmiş ve 36 uyarıcı içeren veri toplama aracı hazırlanmıştır. Veri toplama aracının anlaşılabilirliği ise iki ölçme ve değerlendirme uzmanından alınan görüşler doğrultusunda sağlanmıştır.

Verilerin Toplanması

Veriler araştırmacı tarafından öğrencilerin sınıf ortamına gidilerek toplanmıştır. Verilerin toplanmasından önce öğrencilere ölçme aracının uygulanma gerekçesi ve aracın doldurulması ile ilgili bilgiler verilmiştir. Verilerin toplanması yaklaşık olarak 20 dakika sürmüştür.

Verilerin toplanmasında araştırmacı tarafından hazırlanan ölçme aracı kullanılmıştır. Hazırlanan ölçme aracında, öğrencilerin mesleki tercihlerine yön veren 9 faktör ikili olarak birbiriyle karşılaştırılarak 36 maddede sunulmuştur. Öğrencilerden, kendilerine sunulan her bir ikili gruptaki faktörlerden bir tanesini meslek tercihlerini etkileme durumlarına göre seçmeleri istenmiştir.

Verilerin Analizi

Öğrencilerin mesleki tercihlerini etkileyen faktörlerin etki sırasını belirlemede ikili karşılaştırmalar yoluyla ölçekleme kullanılmıştır. İkili karşılaştırmalarla ölçekleme, uyarıcıların cevaplayıcılara ikişerli olarak sunulduğu, kişinin uyarıcıları algılama ve ayırt etme olarak tepkide bulunduğu, böylece ayırt etme yargılarının belirlendiği bir yöntemdir (Turgut ve Baykul, 1992).

Araştırmada öğrencilere ikişerli olarak 36 maddede meslek seçimlerinde rol oynayan 9 faktör sunulmuştur. Öğrencilerden kendilerine ikili olarak sunulan uyarıcılardan birini meslek seçimlerindeki etkisine göre seçmeleri istenmiştir. Öğrencilerin bir grup uyarıcı arasından karar verme ve alternatifler arasından seçme olanaklarının olduğu birçok durumda, uyarıcılardan hangisinin daha çok etki ettiği ve hangisinin daha çok tercih edildiğini belirlemede ikili karşılaştırmalarla ölçekleme kullanılabilir.

Verilerin analizinde öncelikle, öğrencilere ikili halde sunulan mesleki tercihlerine yön veren 9 faktöre ait tercih frekansları belirlenmiştir. Bu frekanslardan hareketle 9x9'luk frekans matrisi oluşturulmuştur. Elde edilen frekans matrisi toplam öğrenci sayısına bölünerek oranlar matrisi elde edilmiştir. Oranlar matrisinden hareketle birim normal sapmalar matrisi hesaplanmıştır. Birim normal sapmalar matrisinin sütun toplamalarının ortalamalarından öğrencilerin mesleki tercihlerini etkileyen faktörlere ait ölçek değerleri belirlenmiştir. Belirlenen ölçek değerlerinin iç tutarlılığı için ortalama hata (0,021) hesaplanmıştır. Bununla birlikte ölçek değerlerinin verilere uygun olup olmadığı ki-kare testi ile belirlenmiştir (Ki-Kare=137,76; sd=28). Elde edilen ki-kare değeri 28 serbestlik derecesi, 0,05 anlamlılık düzeyindeki kritik ki-kare değeri ile karşılaştırılmıştır. Bu karşılaştırmada hesaplanan ki-kare değerinin tablo değerinden büyük olması nedeniyle V. Hal denklemi yerine III. Hal denklemi ile ölçekleme yapılmıştır. Ölçekleme işlemi araştırma soruları doğrultusunda cinsiyet ve sınıf değişkenleri için ayrı ayrı tekrar edilmiştir. Gerekli hesaplamalar da Microsoft Office Excel programı kullanılmıştır.

BULGULAR

Öğrencilerin mesleki tercihlerine yön veren faktörlerin etki derecelerini belirlemek için yapılan ikili karşılaştırmalarla ölçekleme işlem basamakları aşağıda verilmiştir.

Öğrencilerin mesleki tercihlerini etkileyen 9 faktör ikişerli gruplar halinde 36 uyarıcı çifti olarak öğrencilere sunulmuş ve öğrencilerden her bir ikişerli gruptan mesleki tercihlerini daha fazla etkileyen bir uyarıcıyı seçmeleri istenmiştir. Öğrencilerin tercihlerine göre öncelikle, her bir faktörün

diğerine göre tercih edilme frekansları ($U_j > U_k$) bulunmuş ve 9x9'luk frekans matrisi oluşturulmuştur. Oluşturulan frekans matrisi Tablo 1'de verilmiştir.

Tablo 1'de verilen frekans matrisinde k. satır ile j. sütuna karşılık gelen eleman (U_{jk}); öğrencilerin meslek seçimlerinde sütundaki uyarıcıyı satırdaki uyarıcıya göre daha etkili bir faktör olarak gören katılımcı sayısını yansıtmaktadır. Örneğin; U_{AB} elemanına karşılık gelen 135, A uyarıcısının B uyarıcısına göre bireyin meslek seçimini daha fazla etkilediğini ($U_A > U_B$) göstermektedir. Buna göre frekans tablosunun sütun toplamları, U_j uyarıcısını U_k uyarıcısından daha etkili bulanların sayısını göstermektedir. Böylece her bir uyarıcıya ait frekans tablosundaki sütun toplamları, o uyarıcının diğerlerine göre meslek seçimlerinde daha etkili yargılarına ait frekans toplamlarını göstermekte ve uyarıcıların ölçek değerlerine göre sıralamaları hakkında ipucu verebilmektedir. Örneğin; frekans tablosunda her bir uyarıcıya ait sütun frekanslarına ve frekans toplamlarına bakıldığında, D uyarıcısının öğrencilerin meslek seçimlerini diğer uyarıcılara göre daha fazla etkilediği, G uyarıcısının ise diğer uyarıcılara göre öğrencilerin mesleki tercihlerini daha az etkilediğini düşündürmektedir.

Tablo 1. Lise Öğrencilerinin Meslek Seçimlerini Etkileyen Faktörlere Ait Frekans Matrisi

Uyarıcı (U) U_k	U_j								
	A	B	C	D	E	F	G	H	I
A		220	250	289	211	139	98	249	149
B	135		194	252	190	117	79	197	125
C	105	161		245	168	104	84	200	158
D	66	103	110		153	102	70	175	120
E	144	165	187	202		132	93	223	153
F	216	238	251	253	223		105	253	177
G	257	276	271	285	262	250		274	227
H	106	158	155	180	132	102	81		129
I	206	230	197	235	202	178	128	226	
Toplam	1235	1551	1615	1941	1541	1124	738	1797	1238

Tablo 1'de verilen frekans matrisinin düzenlenmesinde bir hata yapıp yapılmadığını kontrol etmek amacıyla önerilen $(N.K.(K-1))/2$ formülü ile tercih frekansları toplamı hesaplanmış ve bu değer matrisin satır ve sütun toplamları ile karşılaştırılmıştır. Bu karşılaştırma sonucunda tercih frekansları toplamı ve matrisin satır ve sütun toplamlarının (12780) eşit olduğu görülmüştür.

Tablo 1'de verilen frekans matrisi katılımcı sayısına bölünerek oranlar matrisi oluşturulmuştur. Oranlar matrisi Tablo 2'de verilmiştir.

Tablo 2. Lise Öğrencilerinin Meslek Seçimlerini Etkileyen Faktörlere Ait Oranlar Matrisi (P)

Uyarıcılar	A	B	C	D	E	F	G	H	I
A		0,620	0,704	0,814	0,594	0,392	0,276	0,701	0,420
B	0,380		0,546	0,710	0,535	0,330	0,223	0,555	0,352
C	0,296	0,454		0,690	0,473	0,293	0,237	0,563	0,445
D	0,186	0,290	0,310		0,431	0,287	0,197	0,493	0,338
E	0,406	0,465	0,527	0,569		0,372	0,262	0,628	0,431
F	0,608	0,670	0,707	0,713	0,628		0,296	0,713	0,499
G	0,724	0,777	0,763	0,803	0,738	0,704		0,772	0,639
H	0,299	0,445	0,437	0,507	0,372	0,287	0,228		0,363
I	0,580	0,648	0,555	0,662	0,569	0,501	0,361	0,637	
Toplam	3,479	4,369	4,549	5,468	4,341	3,166	2,079	5,062	3,487

Tablo 2’de yer alan oranlar matrisi, birim normal sapmalar matrisine dönüştürülmüş ve Tablo 3’te verilmiştir. Birim normal sapmalar matrisinden yararlanarak gözlemci varyansları kestirilmiştir. Bunun için, öncelikle oluşturulan birim normal sapmalar matrisinin sonunda her bir uyarıcıya ait birim normal sapma değerleri toplamı ($\sum z_{jk}$) ve birim normal sapmalar kareleri toplamı hesaplanmıştır. Ardından, Z matrisinin sütun elemanlarının standart kaymaları (V_j) hesaplanmıştır. Hesaplanan her bir standart kayma uyarıcı sayısı (K) ile çarpılarak KV_j ve $1/KV_j$ hesaplanmıştır. Daha sonra tabloda verilen hesaplama formülüyle KC sabiti hesaplanmıştır. Bu hesaplamalardan hareketle de ayırt etme yargılarının standart kaymaları “ $\sigma_j=(KC/KV_j)-1$ ” formülüyle bulunmuştur. Hesaplanan standart kaymaların karesi alınarak da her bir uyarıcıya ait ayırt etme yargılarının varyansları (σ_j^2) hesaplanmıştır. Bu işlem adımları Tablo 3’te verilmiştir.

Tablo 3. Birim Normal Sapmalar Matrisi (Z) ve Gözlemci Varyanslarının Kestirilmesi

Uyarıcılar	A	B	C	D	E	F	G	H	I
A		0,305	0,537	0,893	0,239	-0,275	-0,595	0,528	-0,203
B	-0,305		0,117	0,553	0,088	-0,441	-0,764	0,138	-0,380
C	-0,537	-0,117		0,496	-0,067	-0,545	-0,717	0,160	-0,138
D	-0,893	-0,553	-0,496		-0,174	-0,561	-0,852	-0,018	-0,418
E	-0,239	-0,088	0,067	0,174		-0,327	-0,637	0,327	-0,174
F	0,275	0,441	0,545	0,561	0,327		-0,537	0,561	-0,004
G	0,595	0,764	0,717	0,852	0,637	0,537		0,745	0,357
H	-0,528	-0,138	-0,160	0,018	-0,327	-0,561	-0,745		-0,349
I	0,203	0,380	0,138	0,418	0,174	0,004	-0,357	0,349	
$\sum z_{jk}$	-1,429	0,993	1,465	3,965	0,897	-2,170	-5,203	2,791	-1,308
$\sum z_{jk}^2$	1,985	1,361	1,408	2,595	0,750	1,592	3,553	1,423	0,659
$K\sum z_{jk}^2$	17,865	12,249	12,672	23,356	6,748	14,327	31,976	12,807	5,927
$(\sum z_{jk})^2$	2,042	0,986	2,146	15,718	0,805	4,711	27,070	7,790	1,711
KV_j	4,185	3,540	3,397	2,526	2,564	3,187	1,308	2,133	2,122
$1/KV_j$	0,239	0,282	0,294	0,396	0,390	0,314	0,764	0,469	0,471
KC	K.C=2K/ $\sum(1/K*V_j)$ =4,973								
σ_j	0,188	0,405	0,464	0,968	0,940	0,560	2,800	1,332	1,343
σ_j^2	0,035	0,164	0,215	0,938	0,883	0,314	7,841	1,773	1,805

Tablo 3’te hesaplanan ayırt etme yargılarının varyansları kullanılarak satır ve sütun toplamları ile varyans toplamları matrisi oluşturulmuştur. Oluşturulan varyans toplamları matrisi Tablo 4’te verilmiştir.

Tablo 4. Varyans Toplamları Matrisi

Uyarıcılar	A	B	C	D	E	F	G	H	I
	(0,035)	(0,164)	(0,215)	(0,938)	(0,883)	(0,314)	(7,841)	(1,773)	(1,805)
A (0,035)	0,071	0,199	0,250	0,973	0,918	0,349	7,877	1,808	1,840
B (0,164)	0,199	0,327	0,379	1,101	1,047	0,478	8,005	1,937	1,968
C (0,215)	0,250	0,379	0,430	1,153	1,098	0,529	8,056	1,988	2,020
D (0,938)	0,973	1,101	1,153	1,875	1,820	1,252	8,779	2,711	2,742
E (0,883)	0,918	1,047	1,098	1,820	1,766	1,197	8,724	2,656	2,688
F (0,314)	0,349	0,478	0,529	1,252	1,197	0,628	8,155	2,087	2,119
G (7,841)	7,877	8,005	8,056	8,779	8,724	8,155	15,683	9,614	9,646
H (1,773)	1,808	1,937	1,988	2,711	2,656	2,087	9,614	3,546	3,578
I (1,805)	1,840	1,968	2,020	2,742	2,688	2,119	9,646	3,578	3,609

Varyans toplamları matrisindeki her bir elemanın karekökü alınarak Tablo 5'te verilen varyans toplamlarının karekökü matrisi oluşturulmuştur.

Tablo 5. Varyans Toplamlarının Karekökü Matrisi

Uyarıcılar	A	B	C	D	E	F	G	H	I
A	0,266	0,446	0,500	0,986	0,958	0,591	2,807	1,345	1,357
B	0,446	0,572	0,615	1,049	1,023	0,691	2,829	1,392	1,403
C	0,500	0,615	0,656	1,074	1,048	0,727	2,838	1,410	1,421
D	0,986	1,049	1,074	1,369	1,349	1,119	2,963	1,646	1,656
E	0,958	1,023	1,048	1,349	1,329	1,094	2,954	1,630	1,639
F	0,591	0,691	0,727	1,119	1,094	0,792	2,856	1,445	1,456
G	2,807	2,829	2,838	2,963	2,954	2,856	3,960	3,101	3,106
H	1,345	1,392	1,410	1,646	1,630	1,445	3,101	1,883	1,891
I	1,357	1,403	1,421	1,656	1,639	1,456	3,106	1,891	1,900

Varyans toplamlarının karekökü matrisiyle birim normal sapmalar matrisinin elemanları birbiriyle çarpılarak S matrisi hesaplanmış ve Tablo 6'da gösterilmiştir. S matrisinin köşegenleri sıfırdır. Matrisin köşegen üstü ve altındaki değerler mutlak değerce eşittir. Bu nedenle matris toplamı ve sütun ortalamaları toplamı sıfıra eşittir (Turgut ve Baykul, 1992).

Tablo 6. S Matrisi

Uyarıcılar	A	B	C	D	E	F	G	H	I
A	-	0,136	0,269	0,881	0,229	-0,163	-1,669	0,711	-0,275
B	-0,136	-	0,072	0,580	0,090	-0,305	-2,161	0,192	-0,533
C	-0,269	-0,072	-	0,533	-0,070	-0,396	-2,036	0,225	-0,196
D	-0,881	-0,580	-0,533	-	-0,235	-0,628	-2,524	-0,029	-0,692
E	-0,229	-0,090	0,070	0,235	-	-0,358	-1,882	0,533	-0,285
F	0,163	0,305	0,396	0,628	0,358	-	-1,532	0,811	-0,005
G	1,669	2,161	2,036	2,524	1,882	1,532	-	2,310	1,109
H	-0,711	-0,192	-0,225	0,029	-0,533	-0,811	-2,310	-	-0,661
I	0,275	0,533	0,196	0,692	0,285	0,005	-1,109	0,661	-
$\sum z_{jk}$	-0,119	2,199	2,281	6,101	2,007	-1,123	-15,22	5,413	-1,538
S_j	-0,013	0,244	0,253	0,678	0,223	-0,125	-1,691	0,601	-0,171
S_c	1,678	1,935	1,944	2,369	1,914	1,566	0,000	2,292	1,520

Tablo 6'da verilen S matrisinde sütun toplamları ($\sum z_{jk}$) uyarıcı sayımına (9) bölünerek uyarıcılara ait ölçek değerleri (S_j) hesaplanmıştır. Hesaplanan ölçek değerlerinden en küçüğünün başlangıç noktası sıfıra kaydırılarak her bir faktöre ilişkin ölçek değerleri (S_c) elde edilmiştir. Öğrencilerin meslek seçimlerinde rol oynayan faktörlere ait ölçek değerleri ve etki sıraları Tablo 7'de verilmiştir.

Tablo 7. Öğrencilerin Mesleki Seçimlerinde Rol Oynayan Faktörlere Ait Ölçek Değerleri ve Etki Sıraları

Meslek Seçiminde Rol Oynayan Faktörler	Ölçek Değeri	Etki Sırası
A. Mesleğin maddi olarak kazanç durumu	1,678	6
B. İş bulma kolaylığı	1,935	4
C. Bireysel yetenekler	1,944	3
D. Mesleğe duyulan ilgi	2,369	1
E. Mesleğin toplumsal saygınlığı	1,914	5
F. Aile etkisi	1,566	7
G. Çevre etkisi (arkadaş vb.)	0	9
H. Üniversite giriş sınavından alınacak puan	2,292	2
I. Mesleğe ait eğitimin kolay olması	1,520	8

Tablo 7'ye göre, öğrencilerin meslek seçimlerini/tercihlerini en fazla bir mesleğe duydukları ilgi etkilemektedir. Bunu sırasıyla üniversite giriş sınavından alacakları puan, bireysel yetenekleri, iş bulma kolaylığı, mesleğin toplumsal saygınlığı, mesleğin maddi olarak kazanç durumu, aile etkisi, mesleğe ait eğitimin kolay olması ve son olarak çevre etkisi takip etmektedir.

Öğrencilerin cinsiyetlerine göre mesleki seçimlerini etkileyen faktörler ayrı ayrı ölçeklendiğinde ise elde edilen ölçek değerleri ve etki sıraları Tablo 8'de gösterilmiştir.

Tablo 8. Cinsiyete Göre Öğrencilerin Mesleki Seçimlerini Etkileyen Faktörlere Ait Ölçek Değerleri ve Etki Sıraları

Meslek Seçiminde Rol Oynayan Faktörler	Kız		Erkek	
	Ölçek Değeri	Etki Sırası	Ölçek Değeri	Etki Sırası
A. Mesleğin maddi olarak kazanç durumu	1,527	6	1,082	5
B. İş bulma kolaylığı	1,960	3	1,053	6
C. Bireysel yetenekler	1,945	4	1,138	4
D. Mesleğe duyulan ilgi	2,462	1	1,398	1
E. Mesleğin toplumsal saygınlığı	1,858	5	1,316	3
F. Aile etkisi	1,511	7	0,774	8
G. Çevre etkisi	0	9	0	9
H. Üniversite giriş sınavından alınacak puan	2,367	2	1,345	2
I. Mesleğe ait eğitimin kolay olması	1,364	8	0,948	7

Tablo 8'e göre, kız ve erkek öğrencilerin mesleki seçimleri en fazla mesleğe duydukları ilgiden, en az ise çevre etkisinden etkilenmektedir. Cinsiyete göre öğrencilerin mesleki tercihlerini etkileyen ilk ve son etkinin değişmediği söylenebilir. Fakat bazı faktörlerin öğrencilerin seçimlerine olan etki sıralamaları cinsiyete göre farklılık göstermektedir. Örneğin, kız öğrenciler bir mesleği seçerken en çoktan en aza doğru sırasıyla; mesleğe duydukları ilgi, üniversite sınavından alınacak puan, mesleğe ait iş bulma kolaylığı, bireysel yetenekler, mesleğin toplumsal saygınlığı, kazanç durumu, aile etkisi, eğitimin kolay olması ve çevre etkisinden etkilenmektedir. Erkek öğrenciler ise bir mesleği seçerken en çoktan en aza doğru sırasıyla; mesleğe duydukları ilgi, üniversite giriş sınavından alınacak puan, toplumsal saygınlık, bireysel yetenekler, mesleğin kazanç durumu, iş bulma kolaylığı, eğitimin kolay olması, aile etkisi ve çevre etkisinden etkilenmektedir.

Öğrencilerin sınıf düzeylerine göre meslek seçimlerini etkileyen faktörler ayrı ayrı ölçeklendiğinde elde edilen ölçek değerleri ve etki sıraları Tablo 9'da gösterilmiştir.

Tablo 9. Sınıf Düzeyine Göre Öğrencilerin Meslek Seçimlerini Etkileyen Faktörlere Ait Ölçek Değerleri ve Etki Sıraları

Meslek Seçiminde Rol Oynayan Faktörler	11. Sınıf		12. Sınıf	
	Ölçek Değeri	Etki Sırası	Ölçek Değeri	Etki Sırası
A. Mesleğin maddi olarak kazanç durumu	1,969	8	1,244	6
B. İş bulma kolaylığı	2,008	6	1,577	3
C. Bireysel yetenekler	2,102	4	1,588	4
D. Mesleğe duyulan ilgi	2,443	2	2,049	1
E. Mesleğin toplumsal saygınlığı	2,240	3	1,476	5
F. Aile etkisi	2,004	7	1,054	7
G. Çevre etkisi	0	9	0	9
H. Üniversite giriş sınavından alınacak puan	2,465	1	1,991	2
I. Mesleğe ait eğitimin kolay olması	2,057	5	1,004	8

Tablo 9'a göre, 11. ve 12. sınıf öğrencilerinin meslek seçimlerini etkileyen faktörlerin sıralamaları arasında bazı farklılıklar bulunmaktadır. 11. sınıf öğrencilerinin mesleki seçimlerini en fazla üniversite giriş sınavından alacakları puan etkilerken, 12. sınıf öğrencilerinin mesleki seçimlerini en fazla mesleğe duydukları ilgi etkilemektedir. İş bulma kolaylığı, mesleğin maddi olarak kazanç durumu gibi faktörler 12. sınıfların tercihlerini 11. sınıflara göre daha ön sırada etkilemektedir. Bununla birlikte bazı faktörlerin öğrencilerin tercihlerine olan etki sıraları sınıf düzeyine göre benzerlik göstermektedir. Örneğin, meslek belirleme sürecinde önemli olan yetenekler öğrencilerin seçimlerini her iki sınıf düzeyinde de dördüncü olarak etkilemektedir. Ayrıca 11 ve 12. sınıf öğrencilerinin mesleki seçimleri en az çevreden (arkadaş vb.) etkilenmektedir.

SONUÇLAR ve TARTIŞMA

Meslek seçimi sadece ekonomik gereksinimlerin karşılanması için değil, psiko-sosyal gelişim ve doyum sağlamak için de sürdürülen bir uğraştır (Yazıcı, 2012). Bireyin psiko-sosyal gelişimini ve doyumunu sağlamasında doğru bir meslek seçmesi önemlidir. Meslek seçmenin zor bir karar olduğu düşünüldüğünde, öğrencilerin bu aşamada mesleki seçimlerini etkileyen faktörlerin etki sıralarının belirlenmesinin mesleki rehberlik açısından gerekli olduğu düşünülmektedir. Bu doğrultuda araştırmada, öğrencilerin meslek seçimlerini etkileyen faktörlerin etki sıraları ikili karşılaştırmalar yoluyla incelenmiştir.

Araştırmadan elde edilen bulgulara göre, öğrencilerin mesleki seçimlerini en fazla mesleğe duydukları ilginin etkilediği, bunu sırasıyla üniversite sınavından alınacak puanın, bireysel yeteneklerinin, iş bulma kolaylığının, mesleğin toplumsal saygınlığının, mesleğin maddi kazanç durumunun, aile etkisinin, mesleğe ait eğitimin kolay olmasının ve çevre etkisinin takip ettiği sonuçlarına ulaşılmıştır.

Sağlıklı bir meslek seçiminin ilgi ve yetenekler dikkate alınarak verilen bir karar olduğu bilinmektedir (Kuzgun, 2014). Araştırma sonuçları da öğrencilerin mesleki seçimlerindeki ilk üç etkinin bir mesleğe duyulan ilgi, üniversite sınavından alınacak puan ve bireysel yetenekler olduğunu göstermektedir. Bu doğrultuda öğrencilerin mesleki seçimlerinde sağlıklı kararlar alma eğiliminde olduğu, fakat üniversite giriş sınavından alınacak puanında meslek seçiminde öğrenciler için yadsınamaz bir gerçek olduğu söylenebilir. Oysaki öğrencilerin üniversite giriş sınavından alacakları puana başvurmadan önce alanların ve mesleklerin gerektirdiği yetenek düzeyi ile kendisinin hangi alanda ve ne düzeyde yeteneğe sahip olduğu konusunda bilgi sahibi olması gerekmektedir (Kutlu, 2006 aktaran Kıyak, 2006).

Öğrencilere sunulan dokuz faktörden mesleğe ait iş bulma kolaylığının, öğrencilerin meslek seçimlerini dördüncü sırada etkilediği belirlenmiştir. Kıyak (2006), lise son sınıf öğrencileri ile yaptığı çalışmada öğrencilerin meslek seçimlerini etkileyen ilk beş faktörü; iş güvencesi, insanlara faydalı olma, sosyal güvence, kolay iş bulabilme ve iyi bir maaş olarak belirlemiştir. Pekkaya ve Çolak (2013), iktisadi ve idari bilimler fakültesi öğrencilerinin meslek seçimini etkileyen faktörlerin

önem derecelerini belirledikleri çalışmalarında, öğrencilerin iş güvencesi, meslek kazançları ve kariyer imkânı kriterlerine çok önem verdiğini ve bu kriterlerin genç bireylerin kararında yaklaşık %64,35 etkili olduğu sonucuna ulaşmışlardır. Kars ve diğerleri (2014) tarafından yapılan ve lise son sınıf öğrencilerinin meslek seçiminde karşılaştıkları sorunların belirlendiği çalışmada da öğrencilerin mesleki tercihlerini ilk sırada iş olanaklarının, ikinci sırada ilgi ve yeteneklerine uygunluğun, üçüncü sırada da ekonomik getirinin etkilediği belirlenmiştir. Yapılan bu çalışmalar, iş bulma ve iş güvencesi gibi faktörlerin öğrencilerin meslek seçimlerini etkilediğini destekler niteliktedir. Bu doğrultuda Türkiye’de iş bulma konusunda yaşanan güçlüklerin ve işe girme çabasının, öğrencilerin mesleki seçimlerini etkilediği düşünülmektedir.

Eksiklik ve kusurlara karşı hoşgörüsüz bir ortamda yetişen bireyler yakın çevrenin ilgisinden ve sevgisinden yoksun kalmamak için kendi ilgi, isteklerini yadsıyıp başkalarının istediği biçimde davranmaya çalışmaktadır (Kuzgun, 2014). Araştırmada öğrencilerin mesleki seçimlerini en az çevrenin (arkadaş, öğretmen vb.) etkilediği sonucuna ulaşılmıştır. Bu doğrultuda öğrencilerin bir mesleği seçerken mesleki ilgilerini ön plana, çevre etkisini ise arka plana alması doğru bir yaklaşım olarak kabul edilebilmektedir.

Öğrencilerin mesleki seçimlerini etkileyen faktörlerin etki sırası cinsiyet değişkenine göre incelendiğinde, kız ve erkek öğrencilerin mesleki seçimlerini en fazla mesleğe duyulan ilginin, sonrasında ise üniversite sınavından alınacak puanın etkilediği belirlenmiştir. Bununla birlikte öğrencilerin mesleki seçimleri cinsiyete göre ilk iki etkiden sonra farklılaşmaktadır. Erkekler için bir mesleğin toplumsal saygınlığı meslek seçiminde daha önemli üçüncü etki olmakla birlikte, kızlar üçüncü sırada iş bulma kolaylığının mesleki seçimlerini daha çok etkilediğini belirtmiştir. Bununla birlikte kızlar bir mesleği seçerken iş bulma kolaylığını mesleğin kazanç durumuna göre daha ön sırada tutarken, mesleğin maddi olarak kazanç durumunu erkekler kızlara göre daha ön sırada tutmaktadır. Bu durum cinsiyet değişkeni bakımından toplumun bireylere yüklediği sorumluluğa bağlanabilir.

Kız ve erkeklerin meslek seçimlerini, kişisel yeteneklerinin dördüncü sırada etkilediği görülmüştür. Kız ve erkek öğrencilerin meslek seçimlerini kişisel yeteneklerinden önce başka değişkenlerin daha fazla etkilediği sonucuna ulaşılmıştır. Oysa doğru bir meslek seçme sürecinde bireysel yetenekler ve ilgi dikkat edilmesi gereken önemli etkenlerdendir. Bunun aksi bir durum öğrencileri mesleki seçim sürecinde gerçekçi olmayan tercihler yapmaya zorlamaktadır (Kuzgun, 2014).

Öğrencilerin meslek seçimlerini etkileyen faktörlerin etki sırası sınıf düzeyine göre incelendiğinde, 11. sınıflar için ilk sırada üniversite sınavından alacakları puanın etkisi, 12. sınıflar için de ilk sırada mesleğe duyulan ilginin olduğu belirlenmiştir. Sınıf düzeyi arttıkça öğrencilerin daha bilinçli meslek tercihi yaptığı söylenebilir. Her iki sınıf düzeyinde de öğrenciler, meslek seçme süreçlerinde çevre ve aile etkisinden son sıralarda etkilendiklerini belirtmiştir. Bu durum öğrencilerin meslek seçim sürecinde bir başkasının kararlarına göre seçimlerine yön verme eğiliminde olmadıklarına işaret etmektedir. Sınıf düzeyine göre öğrencilerin bir mesleği seçmelerini etkileyen faktörlerin etki sıraları arasında bazı farklılıklar bulunmaktadır. Örneğin, 12. sınıflar için iş bulma kolaylığı bir meslek seçme sürecinde daha ön sıralarda bulunurken, 11. sınıflarda bu etki öğrencilerin meslek seçimlerini daha az etkilemektedir. Bu durum öğrencilerin sınıf düzeyi arttıkça hayata ve toplumsal sorunlara daha gerçekçi bakmaları olarak yorumlanabilir. Çünkü liseden üniversiteye geçme döneminde bireyin, ilgi, yetenek ve değer gibi birey içi etkenlerden çalışma koşulları, mesleğe hazırlanma süresi ve kazanç durumu gibi çalışma hayatındaki gerçeklere odaklandığı bilinmektedir (Yeşilyaprak, 2013).

Sonuç olarak, 11. sınıflar hariç, bir mesleğe duyulan ilginin öğrencilerin meslek seçimlerini en fazla etkilediği görülmektedir. Bununla birlikte, öğrencilerin üniversite giriş sınavından alacakları puanların da mesleki seçimlerini büyük oranda etkilediği görülmektedir. Genel olarak meslek seçim sürecinde öğrencilerin ilgileri ön plandadır. Fakat yeteneklerinden önce, öğrencilerin meslek seçimlerinde üniversite giriş sınavından alacakları puanı daha çok düşündükleri gözlenmektedir. Bu durum, öğrencilerin bir meslek dalına yönelmesinde üniversite giriş sınavından alınan puan etkisinin çok fazla olduğu, öğrencilerin bireysel yeteneklerinden ziyade sınav puanına göre tercih

yapma eğiliminde olduğu, bir meslek dalına ait eğitimin alındığı üniversiteye yerleştirme sisteminin öğrencileri meslek seçmede çok etkilediği ve bu nedenle de hatalı yönlendirebileceği sonuçlarına ulaşılmıştır.

Meslek seçiminde daha önemli bir yere sahip olması gereken yeteneklerden önce öğrencilerin meslek seçimlerini başka faktörler etkilemektedir. Araştırmadan elde edilen sonuçlara göre genel olarak, bireysel yeteneklerin olması gerekenden daha az dikkate alındığı söylenebilir. Bu nedendir ki bireyler başka etkiler nedeniyle yeteneklerine uygun olmayan mesleklere yönelebilmektedir. Oysaki okullarda yürütülen mesleki rehberlik ve mesleğe yönlendirme sürecinde öğrencilerin özellikleri ile mesleklerin özellikleri incelenip, ikisi arasındaki ortak noktalar tespit edilerek öğrencilerin kendine uygun bir meslek seçmesine yardımcı olunmalıdır (Arsoy, 2010). Bu doğrultuda okullardaki mesleki rehberlik çalışmaları artırılmalıdır. Mesleki rehberlik sürecinde öğrencilerin ilgi ve yeteneklerini tanımasına daha fazla fırsat verilmelidir. Öğrencilerin okul öncesinden lise sonuna kadar ilgi ve yetenekleri belirlenmeli ve kişisel yeteneklerine göre yönlendirme yapılmalıdır.

KAYNAKÇA

- Alkan, N. (2014). Üniversite adaylarının bölüm tercihleri: Bir kariyer araştırma yöntemi olarak bölüm tanıtımları. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, 5(41), 61-74.
- Arsoy, S. (2010). Mesleki rehberlik. Esra İşmen Gazioğlu ve Şengül Mertol İlgar (Ed.), *Öğretmen ve öğretmen adayları için rehberlik içinde* (s. 169-204). Ankara: Pegem Akademi.
- Aydın, B. (2011). *Rehberlik*. Ankara: Pegem Akademi.
- Çiftçi, G. E., Bülbül, S. F., Bayar-Muluk, N., Çamur-Duyan, G. ve Yılmaz, A. (2011). Sağlık bilimleri fakültesini tercih eden öğrencilerin, üniversite ve meslek tercihlerinde etkili olan faktörler (Kırıkkale üniversitesi örneği). *J Kartal TR*, 22(3), 151-160.
- Dinç, E. (2008). Meslek seçiminde etkili olan faktörlerin incelenmesi: Meslek yüksekokulu muhasebe programı öğrencileri üzerine bir araştırma. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2(16), 90-106.
- Genç, G., Kaya, A. ve Genç, M. (2007). İnönü üniversitesi tıp fakültesi öğrencilerinin meslek seçimini etkileyen faktörler. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 8(14), 49-63.
- Kalkan, Ö. K. (2014). Meslek liselerini tercih etmede etkili olduğu düşünülen faktörlerin ikili karşılaştırmalar yöntemiyle ölçeklenmesi. *Mesleki Bilimler Dergisi*, 3(1), 129-146.
- Kars, V., Arslan, N., Erik, L., Avcı, N., Bucaktepe, P. G., Celepkolu, T. ve Şahin, H. A. (2014). Lise son sınıf öğrencilerinin meslek seçiminde karşılaştığı sorunlar ve bu sorunların anksiyete ve depresyonla ilişkisi. *Dicle Tıp Dergisi*, 41(1), 187-190.
- Kıyak, S. (2006). *Genel lise öğrencilerinin meslek seçimi yaparken temel aldığı kriterler*. (Yüksek Lisans Tezi, Yeditepe Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Kuzgun, Y. (1982). *Mesleki rehberliğin bireylerin yetenek ve ilgilerini tanımalarına etkisi*. Ankara: Ankara Üniversitesi Eğitim Bilimleri Fakültesi.
- Kuzgun, Y. (2000). *Meslek danışmanlığı*. Ankara: Doğu Matbaacılık.
- Kuzgun, Y. (2014). İlköğretimde meslek gelişimi. Yıldız Kuzgun (Ed.), *İlköğretimde rehberlik içinde* (s.125-153). Ankara: Nobel Akademi.
- Özen, Y. (2011). Kişisel sorumluluk bağlamında kariyer seçimini etkileyen sosyal psikolojik faktörler. *Eğitim ve İnsani Bilimler Dergisi: Teori ve Uygulama*, 2(3), 81-96.
- Özyürek, R. ve Atıcı, M. (2002). Üniversite öğrencilerinin meslek seçimi kararlarında kendilerine yardım eden kaynakların belirlenmesi. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, 2(17), 33-42.
- Pekkaya, M. ve Çolak, N. (2013). Üniversite öğrencilerinin meslek seçimini etkileyen faktörlerin önem derecelerinin AHP ile belirlenmesi. *International Journal of Social Science*, 6(2), 797-818.
- Sarıkaya, T. ve Khorshid, L. (2009). Üniversite öğrencilerinin meslek seçimini etkileyen etmenlerin incelenmesi: Üniversite öğrencilerinin meslek seçimi. *Türk Eğitim Bilimleri Dergisi*, 7(2), 393-423.
- Şirin, A., Öztürk, R., Bezci, G., Çakar, G. ve Çoban, A. (2008). Hemşirelik öğrencilerinin meslek seçimi ve mesleği uygulamaya yönelik görüşler. *Dirim Tıp Gazetesi*, 83, 69-75.
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM.
- Varol, S. (1990). *Lise son sınıf öğrencilerinin kaygılarını etkileyen etmenler* (Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Sosyal Bilimler Enstitüsü, Samsun). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.

- Vurucu, F. (2010). *Meslek lisesi öğrencilerinin meslek seçimi yeterliliği ve meslek seçimini etkileyen faktörler* (Yüksek Lisans Tezi, Yeditepe Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yazıcı, H. (2012). Eğitsel ve mesleki rehberlik. Gürhan Can (Ed.), *Psikolojik danışma ve rehberlik içinde* (s. 81-125). Ankara: Pegem Akademi.
- Yelken K. (2008). *Orta öğretim son sınıf öğrencilerinin üniversite tercihlerini ve meslek seçimini etkileyen faktörler "Sakarya il merkezi örneği"* (Yayımlanmamış Yüksek Lisans Tezi, Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, Sakarya). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yeşilyaprak, B. (2013). *21. yüzyılda eğitimde rehberlik hizmetleri*. Ankara: Nobel Akademi.

EXTENDED ABSTRACT

Introduction

In the process of occupational identification, the individuals' personal preferences may influence their professional choice. The choices of the individual are effected by some factors. At the same time, the number of jobs and professions are very high nowadays, but when considering the high rate of unemployment, it is understood how important the vocational selection, the factors affecting professional preferences and the vocational guidance work are important. For these reasons, determining the factors that affect high school students' preferences in their career choices and the importance of these factors in their life are important in terms of the use of workforce and country development and vocational guidance work towards students. In this respect, the aim of the research is to determine effect sequences of factors that play a significant role in the professional choices of high school students. For this aim, the researchers tried to find answers to the following questions:

1. How do factors affecting the choice of students' occupations rank according to their scale values?
2. How do ranking of the factors affecting the choice of students' occupations change according to gender?
3. How do ranking of the factors affecting the choice of students' occupations change according to the class level?

Method

By examining of the literature and the composition data written by students, 9 factors were identified that were effective in the professional choices of high school students. These factors are financial status of the profession, ease of finding a job, individual abilities, interest in the profession, occupational social prestige, family factor, environmental factor, university entrance examination score and easy job training. 9 factors were presented to 355 high school students in pairs. Thus, students were asked to select one of the stimuli presented in pairs according to their influence in their professional choice. Paired comparison scaling (third state equation) was used to determine the order of influence of the factors affecting the professional preferences of the students. Paired comparison scaling is a method in which stimuli are presented as two-way responders and the stimuli of the person are reacted as perception, discrimination, so that the discrimination judgments are statistically determined. Microsoft Office Excel program is used for calculations.

Results and Discussion

According to the results of the analysis, it was determined that the students were mostly influenced by their personal interests in the selection of profession, then the scores and individual abilities they got from the university exam. It has been determined that the ease of finding occupation is the fourth most influential factor in the career choice of students. It is thought that the difficulties in finding a job in Turkey and the effort to get a job affect the professional choices of the students.

When choosing a profession, students' preferences are least influenced by environmental factors. It can be considered as a correct approach that students should take the environment in the background when choosing a profession.

When the factors affecting the professional selection of the students were examined according to the gender variable, it was determined that the occupational choices of the male and female students firstly influenced by the interests then the university exam' scores. It was seen that personal abilities affected the professional preferences of boys and girls in the fourth place. It was found that male and female students were mostly influenced by other variables on the professional choices than their personal abilities during the vocational selection process. However, in the process of choosing the true occupation, individual abilities and interests are important influences that need to be considered.

As a result, it seems that the most important effect of professional selection of students (except 11th grade students) is the interest in a profession. At the same time, it is observed that the students have a great influence of their professional choices from the scores they will get from university entrance examination. In general, students interests are at the forefront of the occupational selection process. But before their abilities, students are thinking about the scores they will get from university entrance examination in their vocational choices. This suggests that students tend to choose according to exam scores rather than their own abilities. For this reason, students can choose jobs that are not suited to their abilities. In the vocational guidance process, more opportunities should be given to identify the individual's interests and abilities. Students should obtain their interests and abilities from the pre-school to the end of high school and be guided according to their personal abilities.

Örtük Markov Model: Latent GOLD 5.1'e Genel Bakış*

Latent Markov Model: An Overview of Latent GOLD 5.1

Duygu GÜNGÖR ** Selva ÜLBE***

Öz

Hem eğitim hem de psikoloji alanında gelecekteki durumların geçmişe bağlı olduğu düşünüldüğünde zamana bağlı değişimin nasıl gerçekleştiği önemli bir araştırma alanıdır. Bu kapsamda örtük Markov analizi gözlenebilir değişkenler kullanılarak, kategorik örtük değişkenler hakkında ileriye yönelik tahminleme yapmak için başvurulan istatistiksel bir yöntemdir. Teknolojik gelişmelerin hızlanmasıyla örtük Markov modelleri çeşitli özgün yazılımlarla incelenebilmekte ve bu yazımlar arasından seç-tıkla özelliği ile Latent GOLD programı öne çıkmaktadır. Bu yazının amacı boylamsal kategorik verilerin analizinde kullanılan ve örtük Markov modelini test eden Latent GOLD 5.1 programını uygulamacılara tanıtmaktır. Bu kapsamda öncelikle örtük Markov modelin kısa bir açıklaması yapılmıştır. İkinci olarak Latent GOLD 5.1'in genel özelliklerinden bahsedilmiştir. Son olarak bu program yardımıyla analizlerin nasıl yapılacağına bir örnek göstermek için Monte Carlo simülasyonu kullanılarak üretilen bir veri seti üzerinden model testi gerçekleştirilmiş ve çıktıların yorumlanmasıyla ilgili bilgiler sunulmuştur.

Anahtar Kelimeler: Örtük Markov model, Latent GOLD 5.1, boylamsal veri

Abstract

In the field of both education and psychology, if future situations are thought to be related to the past, how time-based change is taking place is an important research area. In this context, latent Markov analysis is a statistical method applied to make prediction about categorical latent variables by using observed variables. With acceleration of technological developments, latent Markov models can be examined with various special software, among which Latent GOLD program with select-click feature stands out. The aim of this paper is to introduce the Latent GOLD 5.1 program which is used to analyze the longitudinal categorical data and to test the latent Markov models to users. In this context, firstly a brief explanation of latent Markov model was given. Secondly, the general features of Latent GOLD 5.1 were mentioned. Finally, to give an example of how the analyses would be carried out via this program, a model test was conducted on a data set generated from Monte Carlo simulation and the information about the interpretation of the outputs was provided.

Key Words: Latent Markov model, Latent GOLD 5.1, longitudinal data

GİRİŞ

Temellerini örtük sınıf analizinden alan örtük Markov modeller hem gözlenen hem de örtük değişkenin kategorik olduğu araştırma verilerinin boylamsal analizi için kullanılmaktadır. Psikoloji alanında özellikle madde kötüye kullanımı ile ilgili çalışmalar (Cosden, Larsen, Donahue ve Nylund-Gibson, 2015; Guo, Aveyard, Fielding ve Sutton, 2009; La Flair vd., 2013; Lanza ve Bray, 2010) ve eğitim alanı (Vermunt, Langeheine ve Bockenholt, 1999) örtük Markov modelin kullanıldığı uygulamalı alanlara örnek verilebilir. Ölçme modeli ve yapısal model olarak adlandırılan iki alt bölümden oluşan örtük Markov modellerinde üç temel parametre tahminlenmektedir: örtük statü olasılıkları, madde koşullu olasılıkları ve örtük geçiş olasılıkları (ayrıntılı bilgi için bkz: Collins ve Lanza, 2010; Lanza ve Bray, 2010; Vermunt, Tran ve Magidson, 2008). Bu parametrelerden ilk ikisi örtük sınıf analizindeki parametrelerle doğrudan ilişkilidir. Örtük sınıf analizinde örtük sınıf olasılığı olarak adlandırılan parametre, örtük Markov modellerinde sınıfın dinamik yapısını tanımlamak amacıyla genellikle örtük statü olasılığı kavramı ile adlandırılmaktadır (Collins ve Lanza, 2010). $a=$

* Bu çalışmanın bir kısmı TÜBİTAK 2219-Yurtdışı Doktora Sonrası Araştırma Bursu ile desteklenmiştir.

** Yrd. Doç. Dr., Dokuz Eylül Üniversitesi, Edebiyat Fakültesi, İzmir-Türkiye, e-posta: duygu.gungor@deu.edu.tr

*** Araş. Gör., Dokuz Eylül Üniversitesi, Edebiyat Fakültesi, İzmir-Türkiye, e-posta: selva.ulbe@deu.edu.tr

1, 2,, A Zaman-1'deki örtük statüler olsun. Bu durumda Zaman-1'deki örtük statü olasılığı π_a^A olarak gösterilebilir. $\sum \pi_a^A = 1$ olacağından A-1 tane parametre tahminlenecektir (Collins ve Lanza, 2010). Koşullu olasılıklar ise faktör analizindeki faktör yüklerine benzemekte ve tıpkı örtük sınıf analizindeki gibi yorumlanmaktadır. Örtük sınıf analizinden farklı olarak, örtük Markov analizinde her z zamanı için farklı koşullu olasılıklar hesaplanabilmekte ve bu olasılıklar örtük statüler hakkında bilgi vermektedir. A, Zaman-1'deki örtük statüler, B, Zaman-2'deki örtük statüler, D, Y, G ise hem Zaman-1 hem de Zaman-2 için gözlenen değişkenler olmak üzere, D değişkeninin $i=1, 2, \dots, I$; Y değişkeninin $j=1, 2, \dots, J$ ve G değişkeninin $k=1, 2, \dots, K$ tane tepki kategorisi olsun. Bu durumda a örtük statüsünde olan bir bireyin/gözlemin Zaman-1'deki koşullu olasılıkları; $\pi_{ia}^{I/A}, \pi_{ja}^{J/A}, \pi_{ka}^{K/A}$, Zaman-2'deki koşullu olasılıkları ise $\pi_{ib}^{I/B}, \pi_{jb}^{J/B}, \pi_{kb}^{K/B}$ şeklinde olacaktır. Zaman-1'de A örtük statüsünde olan bir bireyin/gözlemin Zaman-2'de B örtük statüsünde olma koşullu olasılığı ise $\pi_{ba}^{B/A}$ şeklinde gösterilir ve bu koşullu olasılıklar Zaman-1'den Zaman-2'ye örtük statüler arasındaki örtük geçiş oranını temsil eder. Z zamanında alınan ölçümlerde bir kişi ancak bir örtük statüde yer alabilmektedir ve örtük statü olasılıklarının toplamı z zamanı için 1'e eşittir. Ancak z+1. zamanda kişiler, A örtük statüsünden B örtük statüsüne geçiş yapabilir ve statülerin yaygınlığı z. ve z+1. zamanda farklılaşabilir.

Son yıllarda artış gösteren örtük Markov model çalışmaları Mplus (Muthen ve Muthen, 1998-2007) ve Latent GOLD (Vermunt ve Magidson, 2013) gibi ticari programlarla ve PROC LTA (Lanza ve Collins, 2008) ve LMest (Bartolucci ve Pandolfi, 2016) gibi özgür yazılımlarla yürütülebilmektedir. Bu yazının amacı Latent GOLD 5.1 programı kullanılarak örtük Markov analizinin yapılması ve çıktıların yorumlanması konusunda uygulayıcılara bilgi sunmaktır. Bu programla pek çok olası ve karmaşık model test edilebilmekle birlikte, bu yazıda programı anlaşılır kılmak için birinci düzey örtük Markov model üzerine odaklanılmıştır.

Latent GOLD

Latent GOLD, Statistical Innovations Incorporated'e bağlı olarak Vermunt ve Magidson (2016) tarafından geliştirilmiş ticari bir istatistik paket programıdır. Program şu an için yalnızca Windows donanımlı bilgisayarlarda kullanılabilir. Basic, Advance ve Syntax versiyonları bulunan program, temel olarak örtük değişken modellerin parametrelerinin tahmin edilmesinde geniş alternatifler sunmaktadır. Bu üç farklı versiyon farklı şekilde ücretlendirilmektedir. Akademik kullanıcılar yıllık (345\$) ya da sınırsız (795\$) lisans seçeneklerini kullanabilirken her bir lisansın üç bilgisayar için kullanılmasına izin verilmektedir. Öğrencilere daha ekonomik çözümler sunulurken, herhangi bir akademik lisansa sahip olan kullanıcılara 2017 itibarıyla 25\$'lık sınıf kullanımı için lisanslama seçeneği de sunulmaktadır. <http://www.statisticalinnovations.com> sitesinden programa ait güncel bilgilere ulaşılabilir; programın deneme (demo) versiyonu ücretsiz olarak indirilip incelenebilir. Deneme versiyonu ile yalnızca programda bulunan örnek veriler üzerinden çalışılabilmektedir. Ayrıca aynı sitede programa ait kullanıcı kılavuzu, teknik bilgiler ve sentaks modülü kılavuzu olmak üzere üç ayrı kitap bulunmaktadır.

Latent Gold 5.0'dan daha önceki versiyonlarda sentaks modülü ile yapılabilen Markov analizleri yeni sürümle birlikte Latent GOLD 5.0 ve Latent GOLD 5.1 Advanced modülü ile de yürütülebilmektedir. Bu gelişmeyle birlikte modeller, seç-tıkla şeklinde kullanıcı dostu olarak da yürütülebilmektedir.

Veri Girişi

Markov analizlerinde özellikle veri girişinin doğru şekilde yapılması çok önemlidir. Klasik analizlerde boylamsal verilerin girişi her bir katılımcı için satır bazında yapılmakta iken örtük Markov analizlerinde ise verilerin sütun bazlı girilmesi gerekmektedir. Bu noktada boylamsal çalışmalarda fazlaca bulunan kayıp verilerin veri setinde mutlaka tanıtılması gerekmektedir. Veri seti

hazırlanırken her gözleme ve zaman dilimine ait değerler iki sütunda tanımlanmalıdır. Tablo 1'de üç gözlenen iki kategorili değişkenin bulunduğu ve dört zaman diliminde ölçüm alınmış örnek bir veri seti sunulmuş, olası kayıp değerler (99) gösterilmiştir.

Tablo 1. Örnek Veri Seti

Denek No	Zaman	Madde_1	Madde_2	Madde_3
1	1	0	1	1
1	2	0	0	1
1	3	99	99	99
1	4	99	0	1
2	1	1	1	1
2	2	1	1	1
2	3	1	1	1
2	4	0	0	1
.
.
n	4	.	.	.

Program, veri setini okurken her gözlem ve her zaman dilimi için verinin girilmiş olduğunu varsayarak modeli tahminlemektedir. Bu nedenle veri girişi uygun şekilde yapılmazsa sonuçlar yorumlanabilir olmayacaktır ya da yanlış şekilde karşımıza çıkacaktır (Vermunt ve Magidson, 2013). Veri setinde denek numaraları ile ölçüm alınan zaman tutarsız şekilde girilirse uygulayıcılar, boylamsal profilde ölçüm alınan zaman sayısından daha fazla zamanın tanımlı olması gibi bir sorunla karşılaşabilirler.

Simülasyon Verisinin Özellikleri

Bu yazının temel amacı, Latent GOLD programı ile örtük Markov modellerinin nasıl yürütülebileceğine ilişkin bir örnekle birlikte; kullanıcılara, program hakkında genel bir bilgi sunmaktır. Latent GOLD programının üstün özelliklerinden biri de çoklu simülasyon verilerini hızlı bir şekilde üretmeye ve tek bir komut ile aynı anda çok sayıda veri setini analiz etmeye olanak sağlamasıdır. Her ne kadar simülasyon çalışmalarının nasıl yapıldığı bu yazının kapsamında olmasa da kullanılan örnek veri, Latent GOLD programında Monte Carlo simülasyon yöntemi kullanılarak üretilmiştir. Veri seti oluşturulurken daha önce yapılmış çalışmalar (Collins ve Wugalter, 1992; Gudicha, Schmittmann, ve Vermunt, 2015; Vermunt, 2010) temel alınarak iki ayrı zaman dilimi, iki kategorili beş madde, 200 örneklem büyüklüğü kullanılarak 1000 replikasyon sonucu elde edilmiş simülasyon verileri için;

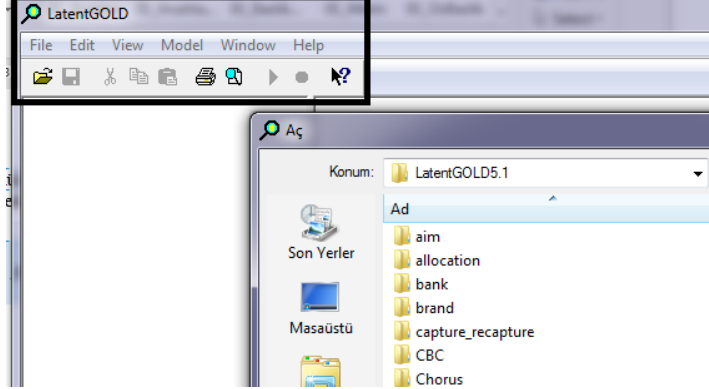
- eşit örtük statülerine sahip (.50, .50) iki örtük statü tanımlanmıştır.
- İki kategorili beş maddenin koşullu olasılıkları birinci örtük statüde olanlar için .90, .10 ve ikinci örtük statüde olanlar için .10, .90 olarak değişimlenmiştir.
- örtük geçiş matrisi (A^t) aşağıdaki gibidir;

$$A^{(t)} = \begin{bmatrix} .80 & .20 \\ .10 & .90 \end{bmatrix}$$

Örneğin bu veri seti, eğitim araştırmalarında sınıfın iki farklı yetenek düzeyinde (birinci örtük statü düşük yetenek düzeyi, ikinci örtük statü yüksek yetenek düzeyi gibi) olduğu durumlarda iyi bir eğitim sonucu yüksek yetenek düzeyindeki grup yeteneklerini korurken, diğer grubun da aynı yetenek düzeyine ulaşmayı başardığı durumlarda karşımıza çıkabilir.

Verinin Programa Tanıtılması

Latent GOLD programı .sav, .dat, .csv, .txt, .rsp ve .cho uzantılı veri setlerini okuyabilmektedir. Veri setinin kayıtlı olduğu klasör, programın dosya aç kısmı tıklanarak Şekil 1’de sunulduğu şekilde açılmaktadır.



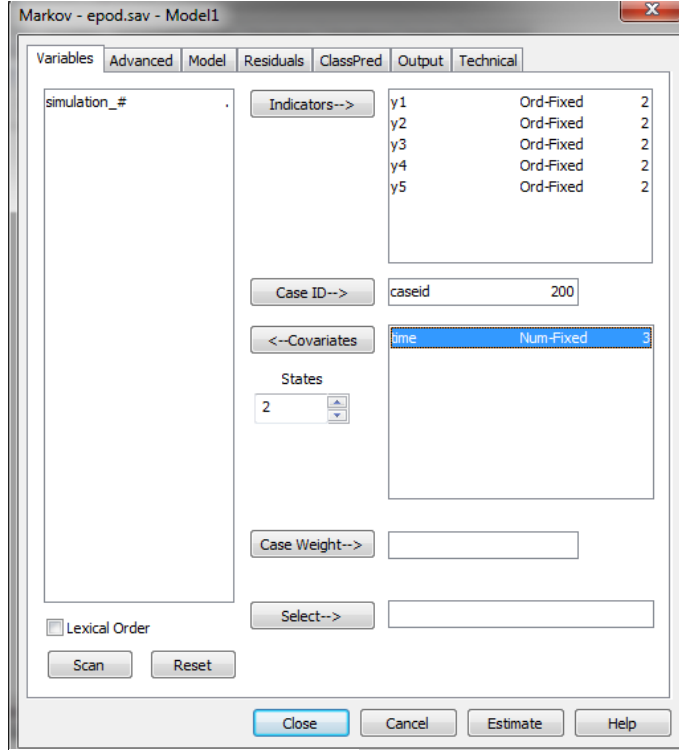
Şekil 1. Veri Setinin Tanıtılması

Programda çeşitli modeller test edilecekse uygun klasörlerin açılması önerilmektedir. Çünkü analiz sonuçları, veri setinin olduğu klasöre otomatik olarak kaydedilmektedir. Aynı klasörde aynı isimle analiz ya da çıktı dosyası kaydedilirse son işlem önceki işlemin üzerine yazılacak ve bu durum bulgu kayıplarına yol açacaktır.

Örtük Markov Modelin Tahminlemesi

Veri seti tanıtıldıktan sonra veri setinin adı -bu örnekte epod.sav- ve altında yer alan Model 1 yazısı veri setinin tanıtıldığı pencerede çıkacaktır. Model tanıtımı için iki alternatif bulunmaktadır. Birinci alternatifte Model 1 yazısının üzerine gelerek sağa tıkladığında yürütebilecek analizlerin adı çıkacaktır. İkincisinde ise pencerenin üstünde bulunan Model yazısı tıkladığında yine aynı analiz seçenekleri karşımıza çıkmaktadır. Burada “Markov” seçilirse gösterge değişkenlerinin, gözlemlerin tanımlanabileceği Caseid ve kovariate değişkenlerin sürüklenerek taşınabileceği pencere açılmaktadır. “Indicators” kısmına gösterge değişkenler sürüklenir. Gösterge değişkenler bu örnekteki gibi iki kategorili değilse ölçek düzeyi doğru şekilde tanıtılmalıdır. Ancak iki kategorili değişkenler için sınıflayıcı ve sıralayıcı seçimi parametre tahminlerini etkilememektedir. “Case ID” kısmına her gözleme ait etiket değeri, “Covariates” kısmına da olası ortak değişkenler (covariate) sürüklenebilir (Şekil 2). Bu değişkenler cinsiyet, grup, yaş gibi zamana göre sabit kalan değişkenler (time constant) olabileceği gibi iş durumu, madde kullanımı, başarı gibi her zaman diliminde değişebilecek değişkenler de olabilir (time varying covariates). Boylamsal çalışmalarda en önemli ortak değişken zamanın kendisidir. Zaman değişkeni bu kısma atılırsa geçiş olasılıkları her zaman dilimi için ayrı ayrı tahminlenirken (zaman-heterojen model) zaman değişkeni ortak değişken olarak kullanılmazsa geçiş olasılıkları her zaman diliminde eşit olacak şekilde hesaplanmaktadır (zaman-homojen model).

Model tanımlanırken belirlenmesi gereken en önemli noktalardan biri statü sayısıdır. Uygulamada her zaman dilimi için örtük statü sayısı ayrı ayrı inceleyebilir ya da araştırmacılar önsel olarak bu bilgiye sahip olabilirler. Bu örnekte önsel olarak statü sayısının iki olduğu bilindiği için “States” iki olarak belirlenmiştir. Ancak statü sayısı bilinmiyorsa farklı sayılardaki modeller iç içe test edilip veriye en iyi uyum sağlayan model seçilebilir.

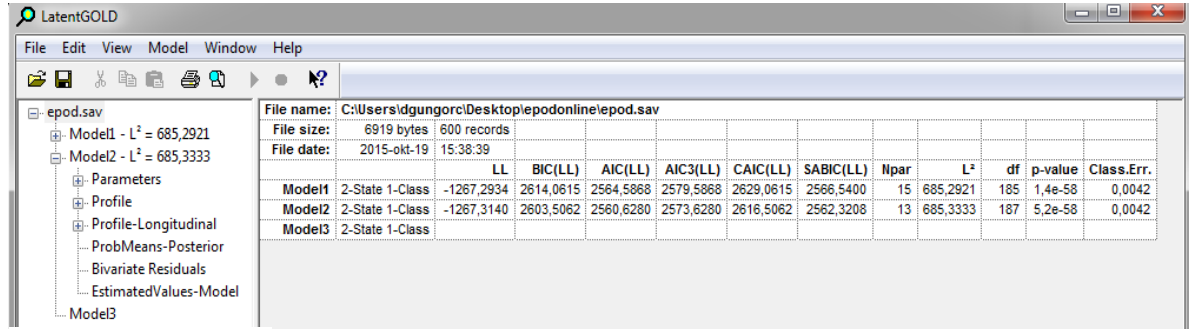


Şekil 2. Örtük Markov Model

Estimate sekmesine tıklandığında “WARNING: Time varying covariates have been removed from Model for Initial States: time” uyarısı çıkacaktır. Bu ilk zaman dilimi için zaman değişkeninin etkisinin tahminlenmeyeceği anlamına gelmektedir.

Model Seçimi

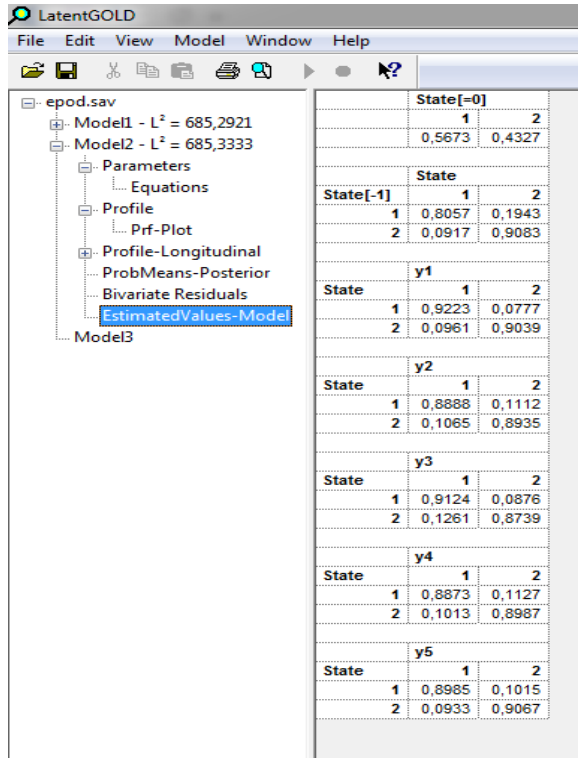
Örtük Markov modellerinde model seçim sürecinde, veriye uyum sağlayan en sade (parsimony) yani en az parametre tahminlenen model aranmaktadır. Bu koşulları sağlayan modeli bulmak için Pearson ki-kare (χ^2), en çok olabilirlik ki-kare oranı (likelihood ratio chi square) (L^2) gibi istatistikler ve Akaike bilgi kriteri (Akaike Information Criteria) (AIC), Bayesian bilgi kriteri (Bayesian Information Criteri) (BIC) gibi bilgi kriterleri kullanılabilir (McCutcheon, 2002). Bilgi kriterleri kullanılarak yapılan model seçim süreci diğer modeller arasından en iyisini seçmeye yöneliktir ve hangi bilgi kriteri kullanılırsa kullanılsın kriterin aldığı daha düşük değer daha iyi model anlamına gelmektedir. Alternatif modeller farklı örtük statü sayısına sahip modeller olabileceği gibi örnekte sunulduğu şekilde parametrelere kısıt konulmuş çeşitli modellerden birini seçmeye de yönelik olabilir. Örnekte Model-1 geçiş olasılıklarının her zaman dilimi için ayrı ayrı tahminlendiği zaman-heterojen model, Model-2 ise zaman-homojen modeldir. Şekil 3'te görülebileceği gibi tüm bilgi kriterleri Model 2'de daha düşük değeri almıştır. Bu durumda Model 2'nin veriye daha iyi uyum sağladığını söylemek mümkündür. Bu seçim işlemi yapıldıktan sonra veriye en iyi uyum sağlayan modele ait parametre değerleri raporlanıp yorumlanabilir.



Şekil 3. Model Seçimi

Örtük Markov Model Parametre Değerlerinin Yorumlanması

Seçilen modele ait parametre tahminleri, “Profile” ve “EstimatedValues-Model” sekmeleri tıklanarak incelenebilir. Şekil 4’te “EstimatedValues-Model” sekmesi tıklandığında açılan pencere gösterilmektedir.



Şekil 4. Parametre Tahminleri

State [=0] başlığının altında sunulan değerler, ölçüm alınan son zaman dilimindeki örtük statü olasılıklarını göstermektedir. Buradaki değerlere göre bir bireyin/gözlemin araştırmanın sonunda örtük statü-1’de olma olasılığının .57 ve örtük statü-2’de olma olasılığının .43 olduğu söylenebilir. Bir alttaki satırlarda sunulan olasılık değerleri ise örtük geçiş olasılıklarıdır (Seçilen model zaman-homojen model olduğu için t ve t-1 zamanlar arasındaki geçişin ölçüm alınan zaman sayısından bağımsız olarak aynı olduğu not edilmelidir). Bu matris şu şekilde yorumlanabilir; birinci zamanda örtük statü-1’de olan bireyin/gözlemin ikinci zamanda da örtük statü-1’de olma olasılığı .80’dir. Bu durumda örtük statü-1’den örtük statü-2’ye geçiş olasılığı da basitçe .20 olarak söylenebilir. Örtük statü-2’de gözlemlerin aynı statüde kalma olasılığı ise .90 ve örtük statü-1’e geçiş olasılıkları .10

olarak hesaplanmıştır. Son olarak sunulan olasılıklar ise madde tepki olasılıklarıdır ve faktör yükleri gibi yorumlanıp örtük statüler hakkında bilgi sunarlar.

SONUÇ

Zamana bağlı değişim hem eğitimciler hem de psikologlar için önemli bir araştırma alanıdır. Değişimin incelenmesinde araştırma sorusuna yetenek, bilgi düzeyi, klinik tanı gibi örtük değişkenler dahil olduğunda ve veri seti kategorik olduğunda tek alternatif örtük Markov model kullanmaktır. Teknolojik gelişmelerle birlikte hız kazanan örtük Markov çalışmalarında Latent GOLD 5.0 ve 5.1 Advanced modül programları seç-tıkla özelliği ile öne çıkmaktadır. Diğer programların sadece sentaks ile çalıştırılabildiği göz önünde bulundurulduğunda Latent GOLD'un daha kullanıcı dostu olduğunu söylemek mümkündür. Ancak araştırmacılar daha karmaşık modelleri test etmek isterse bu programlarda da sentaks yazmaları gerekmektedir. Latent GOLD büyük veri setlerini hızlı bir şekilde analiz edebilmektedir. Ancak benzetim çalışmaları yapan araştırmacıların örneğin 1000 tekar sayısı ve iç içe on model test etmek gibi bir araştırma desenleri varsa, analiz süresi parametre tahmin gücüne de bağlı olarak on ile kırk dakika arasında değişebilmektedir. Programın bir diğer avantajı kullanıcı kılavuzlarının pek çok konuda bilgiler içermesidir. Latent GOLD programın en büyük sınırlılığı ise şu an için yalnızca Windows işletim sistemi ile çalışılabilir olmasıdır.

KAYNAKÇA

- Bartolucci, F., & Pandolfi, S. (2016). LMest package, available at <https://cran.rproject.org/web/packages/LMest/index.html>
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New Jersey: John Wiley ve Sons.
- Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27(1), 131-157.
- Cosden, M., Larsen, J., Donahue, M. T., & Nylund-Gibson, K. (2015). Trauma symptoms for men and women in substance abuse treatment: A latent transition analysis. *Journal of Substance Abuse Treatment*, 50, 18-25.
- Gudicha, D., Schmittmann, V., & Vermunt, J. K. (2015). Power computation for likelihood ratio tests for the transition parameters in latent Markov models. *Structural Equation Modeling: A multidisciplinary Journal*, 23, 234-245. DOI: 10.1080/10705511.2015.1014040
- Guo, B., Aveyard, P., Fielding, A., & Sutton, S. (2009). Using latent class and latent transition analysis to examine the transtheoretical model staging algorithm and sequential stage transition in adolescent smoking. *Substance Use & Misuse*, 44(14), 2028-2042.
- La Flair, L. N., Reboussin, B. A., Storr, C. L., Letourneau, E., Green, K. M., Mojtabai, R., . . . Crum, R. M. (2013). Childhood abuse and neglect and transitions in stages of alcohol involvement among women: A latent transition analysis approach. *Drug and Alcohol Dependence*, 132: 491-498.
- Lanza, S. T., & Bray, B. C. (2010). Transitions in drug use among high-risk women: An application of latent class and latent transition analysis. *Advances and Applications in Statistical Sciences*, 3(2), 203-235.
- Lanza, S. T., & Collins, L. M. (2008). A new SAS procedure for latent transition analysis: Transition in dating and sexual behavior. *Developmental Psychology*, 44, 446-456.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis In J. A. Hagenaars, & A. L. McCutcheon (Ed.), *Applied latent class analysis* (pp. 56-89). New York: Cambridge University Press.
- Muthen, B., & Muthen, L. (2007). *Mplus user's guide fifth edition*. Los Angeles: Muthen ve Muthen.
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.
- Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 179-207.
- Vermunt, J. K., & Magidson, J. (2013). *LG-Syntax user's guide: Manual for Latent GOLD 5.0 Syntax Module*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., & Magidson, J. (2016). *Technical guide for latent GOLD 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In Menard, S. (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373-385). Burlington: Elsevier.

EXTENDED ABSTRACT

Introduction

Latent Markov models, based on their grounds of Latent Class Analyses, are used for longitudinal analysis of research data where both observed and latent data are categorical. Latent Markov models consist of two sub-parts, which are named as measurement model and structural model, and three parameters (i.e., latent status probabilities, item conditional probabilities, and latent transition probabilities) are predicted in these models. Latent status probabilities and item conditional probabilities in latent Markov models are directly associated with the parameters of Latent Class Analysis. The latent class probability parameter in Latent Class Analysis is usually named as latent status probability in Latent Markov models in order to define the dynamic structure of class. Furthermore, the conditional probabilities in Latent Markov models are interpreted as in latent class analysis. The Latent Markov model studies that have increased in recent years are conducted via number of commercial programs such as Mplus (L. Muthen ve Muthen, 1998-2007) and Latent GOLD (Vermunt ve Magidson, 2013) and via free software like PROC LTA (Lanza ve Collins, 2008) and LMest (Bartolucci and Pandolfi, 2016). The objective of this paper is to provide information about how Latent Markov Analysis is carried out and how its outcomes are interpreted by using Latent GOLD 5.1 program to researchers.

Latent GOLD is a commercial statistical package program developed by Vermunt and Magidson under Statistical Innovations Inc. Currently, the program is only available for Windows-equipped computers. With Basic, Advance and Syntax versions, the program offers a wide range of alternatives to predict the parameters of the models. Markov analysis, conducted with the syntax module in versions earlier than Latent Gold 5.0, can be now carried out by the user-friendly point-and-click interface versions, namely by Latent GOLD 5.0 and Latent GOLD 5.1 Advanced modules.

In Markov analyses, it is quite important to enter data correctly. In classical analyses, the entry of longitudinal data for each participant is done by raw by raw, but in latent Markov analysis; data must be entered column-by-column. At this point, it is absolutely necessary to enter the missing data which are likely to be found in the longitudinal studies. When the program reads the data set, it estimates the parameters of model assuming that the data has been entered for each observation and every time slice. For this reason, if the data entry is not done properly, the results cannot be interpreted or can be inaccurate (Vermunt ve Magidson, 2013). Latent GOLD program can read data sets with .sav, .dat, .csv, .txt, .rsp and .cho extensions.

While the model is being defined, one of the most crucial point that must be determined is the number of status. For each time occasion, status number can be examined separately, or the researchers may have a priori knowledge about the number of status. However, if the number of statuses is unknown, the different number of models can be tested, and then the best suited model can be selected. Also, case id and covariate variables are determined for presentation of the model. If the variable of time is considered as a covariate variable, the transition probabilities are always estimated separately for each slice (time-heterogeneous model); if not, the transition probabilities are calculated to be equal at all time occasions (time-homogeneous model).

During the model selection process, the most parsimonious model fit the data is sought in Latent Markov models. To find the best model, information criteria such as Pearson Chi-Square (χ^2), Likelihood Ratio Chi Square (L^2) and Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) Information criteria can be used (McCutcheon, 2002). Model selection process using information criteria is aimed at choosing the best model among other models, and whichever information criterion is used, the lower value, which the criterion holds, points at the better model.

Results

Time-dependent change is a crucial area of research for both educators and psychologists. When research question includes latent variables such as ability, knowledge level, clinical diagnosis, etc. and when the data set is categorical, the time-related change is examined by using Latent Markov Model. As compared to other programs that can only be run with syntax, Latent GOLD program with the point-click feature is more user-friendly for running Latent Markov analyses. However, researchers who want to test more complex models need to write syntax in this program. Furthermore, Latent GOLD is able to quickly analyze large dataset. Manuals of the program also provide users with useful information on a variety of topics. As for major limitation of Latent GOLD program, it can only work with Windows operating system.