5.

EĞİTİMDE VE PSİKOLOJİDE ÖLÇME VE DEĞERLENDİRME KONGRESİ

1-3 Eylül 2016
Akdeniz Üniversitesi Eğitim Fakültesi

V. ULUSAL EĞİTİMDE VE PSİKOLOJİDE ÖLÇME VE DEĞERLENDİRME KONGRESİ

(ULUSLARARASI KATILIMLI)

1-3 Eylül 2016

Akdeniz Üniversitesi

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

Ekim, 2016

SUNU

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinin (Uluslararası Katılımlı) 5.'si 1-3 Eylül 2016 tarihlerinde Akdeniz Üniversitesi Eğitim Fakültesi koordinatörlüğünde Eğitim Fakültesinde gerçekleştirilmiştir. "Ölçme ve Değerlendirme – Psikometri" alanındaki akademik çalışmaların ve uygulamaların paylaşılmasını amaçlayan kongremize değişik üniversitelerden 370 akademisyen 126 bildiri ile katılmışlardır. Kongre öncesinde "R, Gerçek Yaşam Durumlarına Dayalı (Otantik) Açık Uçlu Madde Yazma, Puanlama ve Geri Bildirim Verme ile Tek Boyutlu ve Çok Boyutlu Bilgisayar Ortamında Bireyselleştirilmiş Testler"konularında değerli araştırmacılar Uzm.Eren Halil ÖZBERK ve Arş. Gör. Elif Bengi ÜNSAL ÖZBERK**,** Yrd. Doç. Dr.Ömer KUTLU ve Arş. Gör. Seval Kula KARTAL ile Yrd. Doç. Dr. Burhanettin Özdemir tarafından çalıştaylar gerçekleştirilmiştir. Çağırılı konuşmacılarımız ise: Prof. Dr. Anthony J. Nitko "Geniş Ölçekli Uluslararası Sınavlar", Doç Dr. Murat Kayri "Sinir Ağları Ölçme İlişkisi" ve Yrd. Doç. Dr. Kenan Demir "Müzik ve Dansla Ölçme Öğretimi" konusundaki çalışmaları ile kongreye renk katmışlardır. Kongre takvimi çerçevesinde toplanan bildiriler "Kongre Bilim Kurulu" tarafından değerlendirilmiş olup kabul edilen çalışmalar 5 Ağustos 2016 tarihinde duyurulmuş, 22 Ağustos 2016 tarihinde ise kongre programımız yayımlanmıştır. Kongrede sunulan bildirilerin geniş özetleri araştırmacıların son düzeltmeleri yapmalarını müteakiben EPODDER sitesinde yayınlanacaktır.

Kongremiz birçok kişi ve kuruluşun katkı ve çabasının birleşmesi sayesinde organize edilmiştir. Kongremizin gerçekleşmesinde bize talep ettiğimiz her türlü desteği sağlayan üniversite rektörlüğümüze, yine bize farklı boyutlarda katkı sağlayan Anı Yayınevi, Nobel Yayınevi, Pegem Akademi, Pearson, Ankara Üniversitesi Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi, Ankara Üniversitesi Sınav Yönetim Merkezi ve Kurukahveci Mehmet Efendi'yeteşekkür ederiz. Ayrıca kongremizin düzenlenmesinde emeği geçen "Kongre Düzenleme Kurulu" üyelerine ve "Bilim Kurulu" üyelerine sonsuz teşekkürlerimizi sunuyoruz.

Doç. Dr. Bayram BIÇAK

Kongre Düzenleme Kurulu Adına

Kongre Düzenleme Kurulu Onursal Başkanı

Prof. Dr. Mustafa ÜNAL, Akdeniz Üniversitesi Rektörü

Kongre Düzenleme Kurulu Başkanı

Doç. Dr. Bayram BIÇAK

Kongre Eş Başkanı

Prof. Dr. Hülya KELECİOĞLU, EPODDER Başkanı/ Hacettepe Üniversitesi Eğitim Fakültesi

Prof. Dr. Beykan ÇİZEL, Akdeniz Üniversitesi Rektör Yardımcısı

Kongre Onur Kurulu

Prof. Dr. Mehmet Fuat TURGUT (Eşi: Perihan TURGUT)

Prof. Dr. İlhan AKHUN

Prof. Dr. Yaşar BAYKUL

Prof. Dr. Süleyman Çetin ÖZOĞLU

Dr. Fethi TOKER

Prof. Dr. Ali BAYKAL

Prof. Dr. Nizamettin KOÇ

Prof. Dr. Durmuş Ali ÖZÇELİK (Eşi: Şükran ÖZÇELİK)

Doç. Dr. Halil TEKİN (Eşi: Hatice TEKİN)

Kongre Düzenleme Kurulu*

Prof. Dr. Beykan ÇİZEL
Prof. Dr. Hülya KELECİOĞLU
Prof. Dr. Şener BÜYÜKÖZTÜRK
Prof. Dr. Şerafettin KARAKAYA
Doç. Dr. Bayram BIÇAK
Doç. Dr. Mehmet CANBULAT
Doç. Dr. Tülin ACAR
Yrd. Doç. Dr. Emrah GÜL
Yrd. Doç. Dr. Güçlü ŞEKERCİOĞLU
Yrd. Doç. Dr. Hakan KOĞAR
Yrd. Doç. Dr. Ömer KUTLU
Yrd. Doç. Dr. Süleyman KARATAŞ
Dr. Kübra ATALAY KABASAKAL
Dr. Melek Gülşah ŞAHİN
Arş. Gör. Ahmet SU
Arş. Gör. Emine Ela KÖK
Arş. Gör. Ertunç UKŞUL
Arş. Gör. Gizem UYUMAZ
Arş. Gör. Merve AYVALLI
Arş. Gör. Merve TÜRK
Arş. Gör. Neslihan GÖK
Okt. Ahmet KÜTÜK

*Düzenleme kurulu unvan ve alfabetik sıra esas alınarak listelenmiştir.

Bilim Kurulu*

Prof. Dr. Çetin SEMERCİ
Prof. Dr. Ezel TAVŞANCIL
Prof. Dr. Hülya KELECİOĞLU
Prof. Dr. Mehtap ÇAKAN
Prof. Dr. Nükhet DEMİRTAŞLI
Prof. Dr. Ramazan BAŞTÜRK
Prof. Dr. Selahattin GELBAL
Prof. Dr. Şener BÜYÜKÖZTÜK
Doç. Dr. Bayram BIÇAK
Doç. Dr. Bayram ÇETİN
Doç. Dr. Burcu ATAR
Doç. Dr. Cem Oktay GÜZELLER
Doç. Dr. Devrim ALICI
Doç. Dr. Duygu ANIL
Doç. Dr. Hakan ATILGAN
Doç. Dr. Hakan Yavuz ATAR
Doç. Dr. Halil YURDUGÜL
Doç. Dr. İrfan YURDABAKAN
Doç. Dr. İsmail KARAKAYA
Doç. Dr. Mediha KORKMAZ
Doç. Dr. Neşe GÜLER
Doç. Dr. Nilüfer KAHRAMAN
Doç. Dr. Nuri DOĞAN
Doç. Dr. Ömay ÇOKLUK
Doç. Dr. Şeref TAN
Doç. Dr. Tahsin Oğuz BAŞOKÇU
Doç. Dr. Tuncay ÖĞRETMEN
Doç. Dr. Tülin ACAR
Doç. Dr. Zekeriya NARTGÜN
Yrd. Doç. Dr. Betül KURNAZ
Yrd. Doç. Dr. Bilge GÖK
Yrd. Doç. Dr. Celal Deha DOĞAN
Yrd. Doç. Dr. Deniz GÜLLEROĞLU
Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN
Yrd. Doç. Dr. Devrim ERDEM KEKLİK
Yrd. Doç. Dr. Emine ÖNEN
Yrd. Doç. Dr. Emre ÇETİN
Yrd. Doç. Dr. Ergül DEMİR
Yrd. Doç. Dr. Fatih KEZER
Yrd. Doç. Dr. Funda NALBANTOĞLU YILMAZ
Yrd. Doç. Dr. Güçlü ŞEKERCİOĞLU

Yrd. Doç. Dr. Hakan KOĞAR
Yrd. Doç. Dr. Güler YAVUZ
Yrd. Doç. Dr. Gülşen TAŞDELEN TEKER
Yrd. Doç. Dr. Hatice KUMANDAŞ
Yrd. Doç. Dr. İ. Alper KÖSE
Yrd. Doç. Dr. Kaan Zülfikar DENİZ
Yrd. Doç. Dr. Meltem ACAR GÜVENDİR
Yrd. Doç. Dr. Murat AKYILDIZ
Yrd. Doç. Dr. Neşe Öztürk GÜBEŞ
Yrd. Doç. Dr. Ömer KUTLU
Yrd. Doç. Dr. Önder SÜNBÜL
Yrd. Doç. Dr. Özen YILDIRIM
Yrd. Doç. Dr. Recep Serkan ARIK
Yrd. Doç. Dr. Safiye BİLİCAN
Yrd. Doç. Dr. Sevda ÇETİN
Yrd. Doç. Dr. Şeyma UYAR
Yrd. Doç. Dr. Sevilay KİLMEN
Yrd. Doç. Dr. Derya Çakıcı ESER

*Kongrede hakemlik yapanlardan oluşturulmuştur.

İçindekiler

SUNU iv

Kongre Düzenleme Kurulu vi

Bilim Kurulu vii

AÇILIŞ KONUŞMASI I xviii

AÇILIŞ KONUŞMASI II xix

AÇILIŞ KONUŞMASI III xxi

BİLDİRİ ÖZETLERİ 1

Kitapçık Türünün Test Maddelerinin Yanlılık Eğilimlerine Olan Etkisinin Değişen Madde Fonksiyonu Yöntemleriyle İncelenmesi 2

Metin ODABAS, Lokman AKBAY 2

Öğrenci Başarısının Belirlenmesinde Kullanılan Farklı Ölçme Araçlarından Elde Edilen Puanların Genellenebilirlik Kuramı ile Değerlendirilmesi 4

Muammer YÜKSEL, Gülşah BAŞOL 4

Çok Boyutlu Madde Tepki Kuramındaki Değişik Faktörlerin Yetenek Parametresi Kestirimi Doğruluğuna Etkisi 6

Levent YAKAR 6

Eğitimde Ölçme ve Değerlendirme Dersinin Uygulamadaki Yeri 8

Meltem ACAR GÜVENDİR, Yeşim ÖZER ÖZKAN 8

Likert Tipi Ölçeklerde Olumsuz Madde ve Kategori Sayısına İlişkin Rasch Modeli ile Bir İnceleme . 10

Mustafa İLHAN, Neşe GÜLER 10

Türkiye'de Eğitim Alanında Yayımlanan Ölçek Geliştirme Çalışmalarının Çok Yüzeyli Rasch Modeli İle İncelenmesi 12

Gülden KAYA UYANIK, Neşe GÜLER, Gülşen TAŞDELEN TEKER, Süleyman DEMİR 12

Klasik Test Kuramına Göre Sayısal ve Sözel Alanlar İçin Puanlama Güvenirliğinin Kayıp Veri Kapsamında İncelenmesi 14

Sibel ADA, İsmail KARAKAYA 14

Türkçe Konuşmaya İlişkin Kaygı Ölçeği Geliştirilmesi 16

Nesrin SEEF, Sakine GÖÇER ŞAHİN 16

Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Uygulanabilirliği: Bir Post-Hoc Simülasyon Çalışması* 18

Eren Can AYBEK, R. Nükhet DEMİRTAŞLI 18

SBS 2011 Türkçe Alt Testindeki Maddelerin Cinsiyet Değişen Madde Fonksiyonu Açısından İncelenmesi 20

Emre TOPRAK, Levent YAKAR 20

Matematik, Fizik, Kimya ve Biyoloji Öğretim Programlarının Öğretmen Görüşlerine Göre İncelenmesi 22

Abdullah Faruk KILIÇ, Cem Oktay GÜZELLER 22

Güvenirlik Katsayılarının Kayıp Veri Varlığında İncelenmesi 24

Çiğdem AKIN ARIKAN, Sümeyra SOYSAL, Hatice İNAL 24

Öğretmen Adaylarının Hazırladıkları Tanılayıcı Dallanmış Ağaçların Çok Yüzeyli Rasch Modeliyle Analizi 26

Funda NALBANTOĞLU YILMAZ 26

Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Madde Havuzu Özelliklerinin Test Uzunluğu ve Sınıflama Doğruluğu Üzerindeki Etkisi 28

Ceylan GÜNDEĞER, Nuri DOĞAN 28

İkili Puanlanan Ölçme Sonuçlarındaki Kayıp Veri Sorununun Çözümünde Kullanılan Çoklu Değer Atama Yöntemlerinin Karşılaştırılması 30

Duygu Gizem ERTOPRAK 30

Tek-Biçimli-Olmayan Değişen Madde Fonksiyonunun Belirlenmesinde Lojistik Regresyonun Performansı 32

Devrim ERDEM-KEKLİK 32

Eğitim, Ekonomi ve İnsan Arasındaki Sinerji 34

Tülin ACAR 34

Kategori Sayısının Psikometrik Özellikler Üzerine Etkisinin Parametrik Olmayan Madde Tepki Kuramı Modeline (Mokken Model'e) Göre İncelenmesi 36

Asiye ŞENGÜL AVŞAR 36

Öğrencilerin Matematik Kaygısı Üzerinde Etkili Olan Değişkenlerin Belirlenmesi: PISA 2012 Türkiye Örneği 38

Nermin KIBRISLIOĞLU UYSAL 38

Temel Eğitimden Ortaöğretime Geçiş Asıl Sınav ve Mazeret Sınavı Matematik Alt Testlerinin Madde Tepki Kuramına Göre Eşitlenmesi 40

Emrah GÜL, Çilem DOĞAN GÜL,Ömay ÇOKLUK, Mustafa ÖZKAN 40

Görme Engelli Üniversite Öğrencilerinin Sınıf İçi Ölçme-Değerlendirme Süreçlerinin İncelenmesi . 42

Ömer KAMIŞ, Ergül DEMİR 42

Çok Boyutlu Testlerde Alt Boyut ve Toplam Puanların Kestirimine İlişkin bir Karşılaştırma 44

Eren Halil ÖZBERK, Selahattin GELBAL, Terry ACKERMAN 44

Merkezi Ortak Sınav Matematik Alt Testinde Değişen Madde Fonksiyonunun Görme Engeli Durumuna Göre İncelenmesi 46

Hatun Betül OZARKAN, Emre KUCAM, Ergül DEMİR 46

Türkiye ve Amerika'da Engelli Öğrenciler için Yapılan Geniş Ölçekli Sınavların Yasal Sorumluluklar ve Uygulama Yöntemleri Açısından İncelenmesi 48

Gökhan AKSU, Derya ÇOBANOĞLU AKTAN, Mehmet Taha ESER 48

TIMMS 2011 Türkiye Sonuçlarına Göre 8. Sınıf Öğrencilerinin Akıl Yürütme Becerileri ile Duyuşsal Özelliklerinin İlişkisi 50

Neslihan Tuğçe ŞİMŞEK, Ergül DEMİR 50

Faktör Puanı Belirleme Yöntemlerinin Karşılaştırılması Üzerinde Bir Çalışma 52

Müge ULUMAN, Deha DOĞAN 52

Türkiye'deki Öğretmenlerin Karşılaştıkları Mesleki Sorunların İkili Karşılaştırma Yöntemi ile Ölçeklenmesi 54

Yasemin Duygu ESEN, Filiz İSNAÇ, Ergül DEMİR 54

PISA 2012 Matematik Okuryazarlığı Puanlarının Öz-Yeterliğe Dayalı Ölçüt Geçerliği 56

Seda DEMİR, Nuri DOĞAN 56

Farklı Kayıp Veri Baş Etme Tekniklerinin Klasik Test Kuramı ve Madde Tepki Kuramı İle Elde Edilen Madde Parametreleri Üzerindeki Etkisinin İncelenmesi 58

Ufuk AKBAŞ, Şeyhmus AYDOĞDU, Şener BÜYÜKÖZTÜRK 58

Genellenebilirlik Kuramındaki Karar Çalışmalarının İncelenmesi 60

Ömer KAMIŞ, C. Deha DOĞAN 60

GSE (Global Scale Of Englısh) Temelli İngilizce Yeterlik Sınavı: Anadolu Üniversitesi Örneği 62

Belgin AYDIN, Meral Melek ÜNVER 62

Kendi Aracını Getir Uygulaması: Özel Arı Okulları Örneği 64

Pınar Nuhoğlu KİBAR, Buket AKKOYUNLU, Seda DEMİR, Ebru ÇEVİK 64

Kayıp Veriler Yerine Yaklaşık Değer Atamak için Kullanılan Gelişmiş Yöntemlerin Farklı Koşullar Altında Karşılaştırılarak İncelenmesi 66

Elif Kübra DEMİR, Sait ÇÜM, Selahattin GELBAL, Tarık KIŞLA 66

Veri Zarflama Analizi ile Dumlupınar Üniversitesi Bünyesindeki Okulların Verimliliklerinin İncelenmesi 68

R. Serkan ARIK, Buket EREN JANSSEN 68

Dezavantajlı Yüksek Başarılı Öğrencilerin Okul ve Öğrenci Düzeyinde Matematik Başarılarını Etkileyen Değişkenlerin İncelenmesi 70

Elif Bengi ÜNSAL ÖZBERK, Leyla YILMAZ FINDIK, Eren Halil ÖZBERK 70

Performans Görevleri Hakkında Öğretmen Ve Öğrenci Görüşleri: Meslek Lisesi Örneği 72

Ayşegül ALTUN, Hülya KELECİOĞLU 72

Yapay Test Eşitleme Fonksiyonlarını Oluşturmada Kullanılan Farklı Ağırlık Sistemlerinin Karşılaştırılması 74

Serdar ÇAĞLAK 74

Test Eşitlemede Çok Boyutluluğun Eş Zamanlı ve Ayrı Kalibrasyona Etkisi 76

Neşe ÖZTÜRK GÜBEŞ 76

PISA 2012 Türkiye Uygulamasında Kararlılık, Problem Çözmeye Açıklık ve Akademik Başarı Arasındaki İlişkilerin Belirlenmesi 78

Ömer KUTLU, Seval KULA KARTAL, Neslihan Tuğçe ŞİMŞEK 78

Öğretmenlerin Öğretim Becerileri Algısı Ölçeğinin Geliştirilmesi 80

Ali SABANCI, Kemal KAYIKÇI, İzzet ÖZDEMİR 80

Öğretmenlerin Yansıtıcı Düşünme Ölçeğinin Geliştirilmesi 82

Kemal KAYIKÇI, İlhan GÜNBAYI, İzzet ÖZDEMİR 82

Öğretmenlik Mesleği Sorun Algısı Ölçeği (ÖMSAÖ)'nin Geçerlik Güvenirliğinin Klasik Test Kuramına ve Madde Tepki Kuramına Dayalı Olarak Karşılaştırılması Üzerine Bir Çalışma 84

Erol KARACA 84

Cronbach Alfa ve Mcdonalds ω Değerlerinin Gerçek Verilerde Karşılaştırılması 86

Sibel AYDOĞAN, Nermin KIBRISLIOĞLU UYSAL, Nuri DOĞAN 86

Türkiye Cumhuriyeti İnkılâp Tarihi ve Atatürkçülük Ders Kitabı Ölçme ve Değerlendirme Sorularının İncelenmesi 88

Ümmügülsüm CANDEĞER 88

Hong Kong, Finlandiya, Türkiye ve Tayland Ülkelerine Göre PISA 2012 Matematik Alt Testinin Eşitlenmesinde Grup Değişmezliğinin İncelenmesi 90

Hatice İNAL, Duygu ANIL 90

Kayıp Verilerin Varlığında DMF Analizlerinin Yürütülmesi 92

Duygu Gizem ERTOPRAK, Didem ÖZDOĞAN 92

İlkokul 4. Sınıf için Hazırlanan Araştırma Becerileri Öğretim Programının Etkililiğinin İncelenmesi 94

Betül POLAT DEMİR, Ömer KUTLU 94

Öğrencilerin Eğitim Bilimleri Enstitüsü Lisansüstü Programlarına Kabul Durumlarının Yordanması 96

İlhan KOYUNCU, Nuri DOĞAN, Pelin GÖKDEMİR, Müjgen KAHVECİ 96

Farklı Yöntemlerle Ve Birden Çok Puanlayıcı Tarafından Puanlanan Açık Uçlu Matematik Sınavının Çok Yüzeyli Rasch Modeli İle İncelenmesi 98

Hakan ATILGAN Pelin DAMLAR DEMİRCİ 98

Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Madde Havuzu Teşhir Oranı, Madde Havuzu Büyüklüğü ve Örneklem Büyüklüğü Arasındaki İlişki Üzerine Bir Araştırma 100

Alper ŞAHİN 100

Rehberlik ve Araştırma Merkezlerinde Psikolojik Test Uygulama Süreçleri 102

Bengü BÖRKAN, Şeyda ÇETİNTAŞ, Betül GÜLCAN, Merve ÖZCAN, Gizem ÖZTEMÜR, Ali VAROL, Osman YILMAZ 102

Çoktan Seçmeli Test ile Açık Uçlu Sınavın Psikometrik Açıdan Karşılaştırılması 103

Hatice GÜRDİL EGE, Duygu ANIL 103

Derecelendirilmiş Ölçekler İçin Ölçek Değişmezliği Varsayımına Ampirik Bir Bakış: Bireyler-İçi Ve Bireyler-Arası Farklılıklar İçin Soru Sınıflamaları Ve Cevaplayıcı Profilleri 105

Nilüfer KAHRAMAN, Derya ÇAVDAR, Esra SÖZER 105

Teog Başarı Puanı, Özgüven, Özyeterlik Ve Motivasyon Düzeyi Arasındaki İlişkinin İncelenmesi ..107

H. Gonca USTA 107

Eğitimde Ölçme ve Değerlendirme Kongrelerinde (EPOD) Sunulan Bildirilerin Doküman Analizi Yöntemi ile İncelenmesi.....109

Mahmut Sami KOYUNCU, Mehmet ŞATA, İsmail KARAKAYA.....109

Akademik Yazma Beceri Sinavinin Çok Bağlamlı Rasch Modeliyle Değerlendirilmesi.....111

Bengü BORKAN, Aylin ÜNALDI.....111

Hesap Verebilir Eğitim Sisteminde Bir Paydaş Olarak Öğretmenlerin Temel Eğitimden Orta Öğretime Geçiş (Teog) Uygulamasına Yönelik Görüşleri.....112

R. Nükhet DEMİRTAŞLI.....112

Çok Kategorili Puanlanan Maddelerde Mantel Test ve Olabilirlik Oran Testi'nin Madde İşlev Farklılığı Belirlenme Gücünün İncelenmesi.....114

Safiye BİLİCAN DEMİR , R. Nükhet DEMİRTAŞLI.....114

Bağıl Değerlendirme mi yoksa Mutlak Değerlendirme mi? (Türkiye'den bir Üniversite Örneği) 116

Erkan Hasan ATALMIŞ, Alpaslan ALKIŞ.....116

Denk Olmayan Gruplarda Ortak Madde Deseni Kullanılarak Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması.....118

Bilge GÖK, Hülya KELECİOĞLU, Kübra ATALAY KABASAKAL.....118

Uzaktan Eğitim Yoluyla Gerçekleştirilen Ölçme Ve Değerlendirme Dersine İlişkin Öğrenci Görüşleri 120

Fatma Betül KURNAZ.....120

Öğrencilerin PISA Başarısındaki Gizil Sınıfların ve Bazı Değişkenlerle Gizil Sınıfların İlişkilerinin Belirlenmesi.....122

Seher YALÇIN.....122

Öğrencilerin Sınav Kavramına Yönelik Metaforlarının İncelenmesi.....124

Duygu KOÇAK, Çilem DOĞAN-GÜL, Emrah GÜL, Ömay ÇOKLUK.....124

Bir Makalenin Atıf Sayısını Yordayan Değişkenlerin İncelenmesi.....126

Ercan ÇOBAN, Okan DEMİR, Volkan ALKAN, Ezel TAVŞANCIL.....126

Lisans Düzeyi İçin Hazırlanmış Ölçme ve Değerlendirme Kitaplarının İçerik Açısından İncelenmesi 128

Rana SALMANER, Sena GENCAN ACAR, İsmail KARAKAYA.....128

Bir Avuç Mutluluk: Ölçme Ve Değerlendirme Dersinde Bir Öğretim Tasarımı.....130

Kenan DEMİR.....130

PISA 2012 Matematik Alt Testlerinden Seçilen İki Kitapçığın Gözlenen Puan Eşitleme Yöntemleri İle Eşitlenmesi.....136

Sami PEKTAŞ, Mustafa KILINÇ.....136

Kopya Belirleme İndekslerinin Madde Ayırt Edicilik ve Güçlük Düzeylerine Göre Performanslarının İncelenmesi.....138

Önder SÜNBÜL, Seha YORMAZ.....138

Kayıp Veri Giderme Yöntemlerinin Değişen Madde Fonksiyonu Saptama Teknikleri Üzerindeki İşleyişinin İncelenmesi 140

Hüseyin SELVİ, Devrim ALICI 140

Farklı Puanlama Anahtarlarının Güvenirliğinin Genellenebilirlik Kuramı İle İncelenmesi 142

Mehtap AKTAŞ, N. Bilge BAŞUSTA, Devrim ALICI 142

PISA 2012 Türkiye Örnekleminde Okul Özelliklerinin Matematik Okuryazarlığına Etkisi 144

Melek Gülşah ŞAHİN, Yıldız YILDIRIM, Elif SEZER 144

Test Geliştirme Sürecinde Madde Seçimi İçin Hiyerarşik Puanlayıcı Model: Sinyal Algılama Puanlayıcı Modelinin Uyarlanması 146

T. Oğuz BAŞOKÇU, Elçin MANAP, Duygu GÜNGÖR 146

Çoktan Seçmeli Testlerde Seçenek Ağırlıklandırma Yöntemlerinin Testin Faktör Yapısına Etkisinin İncelenmesi 148

Başak ERDEM, Levent ERTUNA, Nuri DOĞAN 148

Thurstone ve Bradley-Terry Ölçekleme Yöntemlerinin Karşılaştırılması 150

Osman TAT, Duygu ANIL 150

CART, CHAID ve Lojistik Regresyon Analizi Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması 152

Bayram ÇETİN, Murat KAYRİ, Görkem CEYHAN 152

Amerika'da Görme Engelli Öğrenciler için Uygulanan Bireyselleştirilmiş Bilgisayar Testlerinin Yapısal Açısından İncelenmesine İlişkin Bir Derleme Çalışması 154

Derya ÇOBANOĞLU AKTAN, Mehmet Taha ESER,Gökhan AKSU 154

Farkı Sayıda Değişen Yargıcı Kararlarının Farklı Madde Türlerine Göre Tutarlılığının İncelenmesi . 156

Fazilet TAŞDEMİR, D. Bahar ŞAHİN 156

Eğitim Araştırmalarında Güç Analizi Ve Bir Uygulama 158

Hikmet ŞEVGİN, Bayram ÇETİN 158

Örtük Markov Modellerinde Örtük Statü Sayısının Belirlenmesi 160

Duygu GÜNGÖR 160

Ortalama Kovaryans Yapılarının Değişmezliği ve Örtük Sınıf Analizi Yöntemleri ile Ölçme Değişmezliğinin İncelenmesi 162

Alperen YANDI, İ. Alper KÖSE, Ömür UYSAL 162

Bifaktör Yapının Tek Boyutlu, Çok Boyutlu ve İkinci Dereceden Yapılarla Karşılaştırılması 164

Murat Doğan ŞAHİN, Murat AKYILDIZ 164

DinasoruS Değerlendirme Sistemi Yardımıyla Bilgisayar Ortamında Dinamik Olarak Özgün Soru Maddeleri Üretilmesi 166

Erol KARAKIRIK , Mustafa DOĞAN 166

Anadolu Üniversitesi İngilizce Yeterlilik Sınavının Bireyselleştirilmiş Bilgisayarlı Test (CAT) Olarak Uygulanması 168

Murat AKYILDIZ, Murat Doğan ŞAHİN 168

PISA 2006 Öğrenci Anketinde Yer Alan Tutum Maddelerine Uygulanan Faktör Analizi, Çok Boyutlu Ölçekleme ve Kümeleme Analizi Sonuçlarının Karşılaştırılması.....170

Seda KOZAN , N. Bilge BAŞUSTA 170

Yordayıcı Değişkenlerin Belirlenmesinde Chaıd Analizi İle Yol Analizi Karşılaştırılması 172

Çiğdem REYHANLIOĞLU KEÇEOĞLU, Emine Gülen ULUSOY, Didem ÖZDOĞAN.....172

Bireye Uyarlanmış Bilgisayarlı Test Ortamında Maksimum Fisher Madde Seçme Yönteminin İncelenmesi.....174

Mevra KAÇAR, Sema SULAK.....174

Okul Öncesi Eğitimde Ölçek Uyarlamaya Dayalı Çalışmaların İncelenmesi 176

Esra SÖZER, Seyhan SARITAŞ, İsmail KARAKAYA 176

Çok Değişkenli Regresyon, Yapay Sinir Ağları ve Bulanık Mantık Tahminleme Modellerinin Karşılaştırılması.....178

Cem Oktay GÜZELLER, Rasim ÖNDER, Caner DURSUN.....178

Faktör Analizi Kullanılan Yayınlamış Bilimsel Makalelerde Örneklem Özellikleri 179

Hakan ATILGAN, Burak AYDIN, Sungur GÜREL, Mehmet KAPLAN 179

PISA Sınavlarında Sorulan Soruların Math Taksonomi Çerçevesinde Analizi.....181

H. Sevgi MORALI, Hakan KARADUMAN 181

Ölçme Araçlarının Psikometrik Özellikler Açısından İncelenmesi 183

Esra Sözer, Derya Çavdar, Nilüfer Kahraman 183

Madde Parametre Kaymasının Test Eşitleme Üzerindeki Etkisi.....185

Kadriye Belgin DEMİRUS, İbrahim UYSAL 185

Bir Bilişsel Tanı Modeli Uygulaması: Niteliklerin Belirlenmesi ve Geçerlik Analizi.....187

Mehmet KAPLAN, Hakan ATILGAN, Sungur GÜREL, Burak AYDIN 187

Farklı Kategori Sayısı ve Yetenek Dağılımı Durumlarında Kategoriler Arası Psikolojik Uzaklıkların İncelenmesi.....189

Gizem UYUMAZ, Gözde SIRGANCI, Ömay ÇOKLUK.....189

Küçük Örneklemlerde Ortak Maddeli Eşdeğer Olmayan Gruplar Deseni Altında Farklı Eşitleme Yaklaşımlarının Karşılaştırılması.....191

Vildan BAĞCI, Ayşenur ERDEMİR, Derya ÇAVDAR 191

Performans Takip Sınavında (PTS) Kullanılan Alt Testlerin Psikometrik Özelliklerinin İncelenmesi193

Dilek GÜR, Özge GÖÇER, Evrim YALÇIN.....193

Many small or one large study? - Detection of effects and a prevention to the file drawer problem 195

Can GÜRER.....195

Yönlendirmeye ve Biçimlendirmeye Dayalı Çok Aşamalı Bilgisayarlı Testlerde Yetenek Kestirimlerinin Karşılaştırılması.....196

Murat Doğan ŞAHİN, Murat AKYILDIZ 196

Madde Tepki Kuramı'na Dayalı Madde Uyum İndekslerinin I. Tip Hata ve Güç Oranlarının Çeşitli Değişkenler Açısından İncelenmesi 198

Seçil ÖMÜR SÜNBÜL, Semih AŞİRET 198

Antalya İli Performans Takip Sınav Puanları İle Yüksek Öğretime Geçiş Sınav Puanları Arasındaki İlişkinin İncelenmesi 200

Şerife ZEYBEKOĞLU, Şevki Yetkin ODABAŞI 200

Çapraz Yüklerin Sıfırdan Farklı Olduğu Durumlarda Maksimum Olabilirlik ve Bayesçi Doğrulayıcı Faktör Analizi Sonuçlarının Karşılaştırılması 202

Derya ÇAVDAR, Ergün Cihat ÇORBACI, Vildan BAĞCI, Ayşenur ERDEMİR 202

Likert Tipli Envanterler Nasıl Puanlanmalı?.....204

Sungur GÜREL, Hakan ATILGAN, Burak AYDIN, Mehmet KAPLAN 204

PISA 2012 Türkiye Örneklemi Yüksek ve Düşük Performans Gösteren Okulların KNN Analizi ve Kümeleme Analizi ile İncelenmesi 206

Halime YILIDIRIM, Sinem ŞENFERAH, Menekşe UYSAL, Şener BÜYÜKÖZTÜRK 206

Farklı Örneklem Büyüklüğü ve Dağılımı Koşullarında WLS ve Robust WLS Yöntemlerinin Karşılaştırılması.....208

Halime YILIDIRIM, Menekşe UYSAL, Sinem ŞENFERAH, Şener BÜYÜKÖZTÜRK 208

Madde Seti Büyüklüğü ve Eğim Parametresinin CAT Test Uzunluğuna Etkisi 210

Ahmet Salih ŞİMŞEK, Ezel TAVŞANCIL.....210

Hiyerarşik Nitelik Yapı Seçiminde Model Seçme Metotlarının Kullanılabilirliği.....212

Lokman AKBAY.....212

Performans Takip Sınavının (PTS) Psikometrik Özelliklerinin Rasch Analizi ile İncelenmesi.....214

Hakan KOĞAR 214

Madde Takımları İçeren Testlerde Tek ve Çok Boyutlu Madde Tepki Kuramı Modellerinden Elde Edilen Madde ve Yetenek Parametrelerinin Karşılaştırılması.....216

Esin YILMAZ KOĞAR, Hülya KELECİOĞLU 216

Çok Boyutlu İş Motivasyonu Ölçeğinin Türk Kültürüne Uyarlanması.....218

Aydın ÇİVİLİDAĞ, Güçlü ŞEKERCİOĞLU 218

Kağıt-Kalem Testi ve Web Temelli Test Olarak Uygulanan Ölçme Araçlarının Ölçme Değişmezliği ve Evren Ayrışıklığı Bakımından Karşılaştırılması 220

Güçlü ŞEKERCİOĞLU 220

Tutum Ölçeği Geliştirmede Genel Tutum İfadelerinin Kullanımı Üzerine Bir Araştırma.....222

Nuri DOĞAN, Yusuf KARA 222

Açımlayıcı Faktör Analizi ve Hiyerarşik Kümeleme Analizi Yöntemlerinin Farklı Örneklem Büyüklüklerine Göre Karşılaştırılması 224

Murat KAYRİ, Fuat ELKONCA, Görkem CEYHAN 224

Bileşke Puan Güvenirlik Hesaplama Yöntemlerinin İki Ve Üç Boyutlu Yapılarda Karşılaştırılması ..226

Nuri DOĞAN, Yusuf KARA, Seda DEMİR, Buket EREN JANSSEN 226

Çok Düzeyli Modellerle PISA 2012 Matematik Testinde Değişen Madde Fonksiyonu ve Potansiyel Kaynak Olarak Öğretmen Davranışı Özelliklerinin İncelenmesi 228

Özen YILDIRIM, Ömer KUTLU, Hatice KUMANDAŞ 228

PISA 2012 Matematik Okuryazarlığı Testinde Farklı Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması 230

Şeyma UYAR, Burcu AKSEKİOĞLU, Neşe ÖZTÜRK GÜBEŞ 230

Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Örneklem Büyüklüğünün Ölçme Kesinliğine Etkisinin İncelenmesi 232

Hülya KELECİOĞLU, Begüm ÖZTEMUR, Melike ÖZER TAYMUR, Şule ÖTKEN 232

Pısa 2012'deki Matematiğe Yönelik Duyuşsal Özelliklerin Bölge, Okul Türü Ve Cinsiyete Göre Sınıflama Doğruluğunun İncelenmesi 236

Emrah BÜYÜKATAK, Duygu ANIL 236

8. Sınıf Öğrencilerinin Temel Eğitimden Ortaöğretime Geçiş Sınavı (TEOG) Kavramı Algıları Metaforları 238

Şemsettin ERCAN, Devran BAYADUR, Mürşet ÇAKMAK 238

Gelişimin Ölçülmesinde Örtük Gelişim Modelleme ve Aşamalı Doğrusal Modellemenin Model-Veri Uyumlarının Örneklem Büyüklüğüne Göre Karşılaştırılması 239

Gürkan CÜVİTOĞLU, Hakan ATILGAN 239

Hakem Katılığının Yapılandırılmış-Yanıtlı Maddelerde Yetenek ve Parametre Kestirimine Etkisi 241

T. Oğuz BAŞOKÇU, Sevcan GÖKÇE 241

BTM' de Ankor Test Deseni İle Gözlenen Puanlara Dayalı Lokal Eşitleme Çalışması 243

T. Oğuz BAŞOKÇU, Arzu CANPOLAT 243

Öğretmen Adaylarının Akranları ve Öğretim Elemanları Tarafından Değerlendirilmelerinden Elde Edilen Puanların Güvenirliğinin Karşılaştırılması 245

Gizem UYUMAZ, Nimet AKBEN 245

Genel Öğretmen Yeterliliklerinin İkili Karşılaştırmalar Yöntemi ile Ölçeklenmesi 247

Haydar KARAMAN, Atilla ÖZDEMİR, Rasim ÖNDER, Duygu ANIL 247

Doğrulayıcı Faktör Analizinde Farklı Parametre Kestirim Yöntemi, Örneklem Büyüklüğü ve Modelin Hatalı Kurulmasının Uyum İndekslerine Etkisi 249

Ayfer SAYIN 249

Performans Takip Sınavında (PTS) Kullanılan Alt Testlerin Psikometrik Özelliklerinin Resmi ve Özel Okullara Göre İncelenmesi 251

Bayram BIÇAK, Dilek GÜR, Evrim YALÇIN, Özge GÖÇER 251

Tahmin Parametresinin Yetenek ile Düzenlenmesi 253

Fatih ORÇAN 253

ÖLÇME VE DEĞERLENDİRME KONGRESİ'NİN ARDINDAN 255

AÇILIŞ KONUŞMASI I

Saygıdeğer akademisyenler, değerli basın mensupları, sevgili öğrenciler ve kıymetli misafirler,

Ülkemizin zor zamanlardan geçtiği bu dönemde, bu tür bilimsel etkinliklerin yapılmasını önemsiyor ve değerli buluyorum. Ülke olarak bir an önce normale dönmemiz ve tekrar eğitimin önemi hakkında konuşur hale gelmemiz gerektiğini düşünüyorum. Bu nedenle başta Kongre Düzeleme Kurulu Başkanı Doç. Dr. Bayram Bıçak hocam olmak üzere tüm kongre düzenleme kurulu üyelerine, gönderdikleri bildirilerle ve katılımlarıyla kongreye katkı sağlayan herkese çok teşekkür ediyorum.

Ben de Tıp Fakültesi Tıp Eğitimi Anabilim Dalı'nda öğretim üyesiyim. Eğitimde ölçme ve değerlendirmenin en önemli anahtar kavramlar arasında olduğunu düşünüyorum. Eğitimde hedefleri doğru koymak ve konulan bu hedefleri uygun ölçme araçları kullanarak ölçmek, başarıya ulaşmada temel kriterlerdendir. Bu açıdan Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'ni çok önemsiyor ve tüm katılımcılara katkılarından dolayı teşekkür ediyorum.

Prof. Dr. Erol GÜRPINAR

Akdeniz Üniversitesi Rektör Yardımcısı

AÇILIŞ KONUŞMASI II

Sayın rektörüm, saygıdeğer onur kurulu üyeleri, kıymetli akademisyenler, değerli basın mensupları sevgili öğrenciler ve kıymetli misafirler.

5.sini gerçekleştirmek üzere bir araya geldiğimiz EĞİTİMDE VE PSİKOLOJİDE ÖLÇME VE DEĞERLENDİRME KONGREMİZE hoş geldiniz, şeref verdiniz.

Kongremiz siz değerli hocalarımızın bir araya gelmesine vesile olmuştur. Sağlık sorunları ya da başka nedenlerle aramıza katılamayan ancak kalbi burada olan tüm hocalarımızı saygı ile anarken, yakın geçmişte hayatını kaybeden Prof. Dr. Mehmet Fuat Turgut, Prof. Dr. Ata Tezbaşaran, ve Prof. Dr. Şefik Uysal Hocalarımıza, hala alandaki sıcaklığını koruyan Prof. Dr. Halil Tekin hocamıza Allahtan rahmet diliyorum. Mekanları cennet olsun.

Bu kongreyi en son yapıldığı yer olan Hacettepe Üniversitesinde devraldığımızda çok heyecanlıydık. Akdeniz Üniversitesi olarak biz ne yapabiliriz üzerine başka ne koyabiliriz diye planlarken şu son bir yıl içerisinde yaşayacaklarımızdan habersizdik. İçinde bulunduğumuz coğrafyanın zorluğunun bizi bu kadar etkileleyebileceğini hiç hesaba katmadık. Geçen 15 Temmuz gecesi yaşanan alçak darbe girişimi ve başkaca yapılan terör saldırıları nedeniyle İnanın şu geçen bir ay içinde kongremizi ertelemeyi defalarca düşündük, konuştuk ancak, ertelemenin ya da iptalin bu terör gruplarının amacına hizmet edeceğini düşünerek normal proğramımıza sadık kaldık. Bu vesile ile akla uymayan her türlü oluşum ve cemaate ve terörle insanları sindireceğini sanan canavar ruhlulara bu ülkeyi yem etmeyeceğimizi bildirmek istiyorum.

Bu zorlu süreçte bizi yalnız bırakmayan değerli hocam, danışmanım alanımızın önemli hocalarından biri olan sayın Prof. Dr. Anthony J. Nitko yaşadığı ailevi sorunlara, eminim ülkemizle ilgili aman ha uyarılarına rağmen gelmekten çekinmemiş bu önemli çalışmamızda bizi yalnız bırakmamıştır. Kendisine katılımından dolayı yürekten teşekkür ediyorum. Hoşgeldiniz sayın hocam.

Değerli misafirler,

Kongreler ve benzeri toplantılar belirli bir konu alanında yapılan bilimsel araştırmaların paylaşıldığı, araştırmacıların duygu ve düşüncelerini birbirleriyle paylaştığı yeni araştırmalara ufuk açan çok önemli bilimsel toplantılardır. Aile duygusu bu toplantılarla canlı kalmakta, yeni üyeler ait oldukları ailenin büyüklüğünü bu toplantılarla idrak etmektedirler. Bu tür toplantılara katılmaktan imtina eden, buralarda görülmekten çekinen arkadaşlarımız çok önemli çalışmalara imza atmış olsalar da camia tarafından yeterince hatırlanmamakta, hakettikleri değeri görmemektedirler. Bu nedenle bu tür toplantılarda yer

almak, sorumluluk almak kongrelerimize sahip çıkmak, daha da büyütmek bizler açısından önemli bir sorumluluktur.

Değerli misafirler, ölçme değerlendirme camiası olarak ilgi alanımız eğitim sürecinin en hayati boyutunu ele almakta, karar verme sürecinin üzerine odaklanmaktadır. Doğru karar verme için dogru ölçme yapma kaçınılmaz bir zorunluluktur. Ne yazık ki birçok disiplin gibi ölçme ve değerlendirme disiplini de araç olmaktan çıkarılıp amaç haline getirilmiş, yıllardır üzerinden rant sağlanan bir yapı haline getirilmiştir. Sonuçları itibarı ile sosyal ve bireysel açıdan çok önemli doğurguları olan bu disiplinin Mutfağında görev alan akademisyenleri ve uygulayıcıları olarak bu alanı amaç olmaktan çıkarıp doğru kararlara hizmet eden araçlara dönüştürmek zorundayız. Yaptığımız çalışmaların önemini kavramakla beraber herşeyin üzerinde ve değişmez tabulara dönüştürmekten vaz gecmeliyiz, vaz geçirmeliyiz.

Kongremiz 161 sözlü bildirisi, alandan ve dışardan 400 civarında katılımcısı ile muhteşem bir çıkış yaşamıştır. Siz değerli konuklarımızı bir buçuk gün sürecek oturumlarda birdirinden değerli çalışmaları izlemeye, gördüğünüz eksik ya da yanlışları kırıcı olmadan eleştirmeye, yapıcı eleştirilerinizle alana olana olan ilgiyi ve katılımı teşvik etmeye davet ediyoruz. katılımınızla şeref verdiğinizi Kısa sürse de Antalya şehrimizi tanıyarak keyif almanızı diliyor, bizi kırmayıp geldiğiniz için teşekkürlerimi arz ediyorum.

Doç. Dr. Bayram BIÇAK

Kongre Düzenleme Kurulu Başkanı

AÇILIŞ KONUŞMASI III

5. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresine hoş geldiniz.

İlki 2008 yılında Ankara Üniversitesi tarafından düzenlenen ve iki yılda bir tekrarlanan Kongremizin bugün Akdeniz Üniversitesi'nde beşincisini düzenlemekteyiz.

Kongremiz 2014 yılına kadar ulusal, 2014 yılından itibaren uluslararası katılımlı olarak düzenlenmektedir. 2013 yılında kurulan EPODDER de kurulduğu günden bu yana kongre düzenleme kuruluna destek vermektedir. Kongrenin önemli yükü ve katkısı Akdeniz Üniversitesi'ndeki arkadaşlarımızda olmakla birlikte EPODDER kongreye, sponsor bularak, kongre düzenlemeleri konusunda önerilerde bulunarak katkı getirmeye çalışmıştır. Ayrıca kongrede sunulan bildirilerin tam metinleri, bildiri sahiplerinin talepleri doğrultusunda ve hakem sürecinden geçtikten sonra dergimiz EPOD'da yayınlanabilecektir.

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongreleri, derneğimizin güçlenmesiyle ve öncelikle bilim alanımızdan daha fazla akademisyenin hem derneğimize hem de kongreye katılımıyla uluslararası platformda kendine yer bulabilecektir. Türkiye'de, Psikometri/Ölçme ve Değerlendirme bilim alanları son yıllarda ele aldığı konularla hızlı bir gelişme göstermiştir. Siz değerli akademisyenler sayesinde yalnızca Türk eğitiminin sorunları değil dünyada ele alınan kuramsal konular da kongrelerde ele alınmıştır. Alanda yetişen genç akademisyenlerin de bu konularda standartları yüksek önemli çalışmalar yapmaları bilim alanımızın geleceği açısından güven vericidir.

Psikometri/Ölçme ve Değerlendirme bilim alanları 1850'li yıllardan itibaren ortaya koyduğu kuramlarla ve yaptığı bilimsel araştırmalarla insan davranışlarını anlama ve açıklama konusunda insanlığa önemli bilgiler sunmuştur, sunmaya da devam etmektedir. Ancak sürece bakıldığında insanoğlunun yarattığı teknolojiye dayalı bilginin çok hızla değişerek geliştiği görülmektedir. Bugün mühendislik, tıp, kimya vb. konularda gelinen yetkin düzeye ne yazık ki insan davranışlarında ulaşılamamıştır. İnsanlık hâlâ açlık, hastalık, çarpık gelir dağılımı, eğitim yetersizliği, savaş, çevre kirliliği vb. konularda çaresiz kalmakta insan ve toplum bilimlerinin bulgularını bu sorunların çözümü için yeterince seferber edememektedir.

Tüm dünyayı özellikle 2000'li yıllardan itibaren etkileyen bir diğer olumsuz gelişme ise terördür. Terör başta Ortadoğu ve Türkiye olmak üzere bugün tüm dünyayı ciddi biçimde tehdit etmektedir. Hepimizin tanık olduğu gibi 2015 yılından sonra dünyada ve özellikle ülkemizde terör saldırıları giderek artmıştır. Canlı bomba eylemleri insan yaşamını konserde, futbol maçında, markette, sokakta, gösteride, düğünde kısaca yaşamın her alanında insanlığı tehdit eden bir noktaya gelmiştir.

Ülkemiz 15 Temmuzda toplum yaşamını derinden etkileyen hain bir terör örgütünün darbe girişimiyle farklı boyuta taşınmıştır. 15 Temmuzda kalkışılan darbe girişimi başarılı olsaydı Türkiye'nin içine düşeceği durum ne olurdu? sorusu her Türk vatandaşının ciddi olarak düşünmesine neden olmuştur.

Darbe girişimini yapan FETÖ örgütlenmesini eğitim açısından ele aldığımızda çok çarpıcı ve dramatik durumları görmekteyiz. Özellikle darbe girişimi sonrası tüm olumsuz yönleriyle iyice deşifre olan FETÖ yapısı, eleştirel ve sorgulayıcı düşüncenin, eğitim

sisteminin temel taşı olmasının gerekliliğini ortaya koymuştur. Hain örgütün yıllarca eğitim sisteminin içinde devleti hedef alarak örgütlenmesi, bir lidere sorgulamadan körü körüne bağlı, dogmatik anlayışa sahip insanlar ve bu şekilde teröristler yetiştirmesi eğitimin bir toplumu dönüştürmede ve biçimlendirmede ne kadar önemli olduğunu ortaya koymaktadır.

Eğitim bir toplumu dönüştürebilmede en önemli ve en etkili güçtür. Burada üzerinde önemle durulması gereken nokta bu gücün nasıl yönlendirildiğidir. Mustafa Kemal Atatürk eğitimin bu dönüştürücü gücünü, "Eğitimdir ki, bir milleti ya özgür, bağımsız, şanlı, yüksek bir topluluk halinde yaşatır; ya da esaret ve sefalete terk eder" sözleri ile vurgulamıştır. Bir an önce ülke olarak bu durumu idrak etmeli, eğitim ve sınav sistemlerini yap-boz tahtasına çevirmekten vazgeçmeli, güçlü, kendini yenileyebilen çağdaş, laik ve demokratik eğitim sistemini kendimize şiar edinmeli ve temellerini sağlamlaştırarak yerleştirmeliyiz.

15 Temmuz sonrasında açığa çıkan gerçekler laik, demokratik ve sorgulayan beyinler yetiştiren bir eğitim sisteminin bu ülkenin bugünleri ve geleceği için ne kadar önemli olduğunu çok acı bir şekilde bizlere göstermiştir. Türkiye Cumhuriyeti'nin kurucusu Mustafa Kemal Atatürk, "Ben manevi miras olarak hiçbir ayet, hiçbir dogma, hiçbir donmuş, kalıplaşmış kural bırakmıyorum. Benim manevi mirasım ilim ve akıldır. Benden sonra, akıl ve ilmin kılavuzluğunu kabul edenler benim manevi mirasçım olurlar." sözleri ile de akılcı, bilimsel ve özgür düşüncenin önemine vurgu yapmıştır.

EPODDER olarak bizler de Atatürk'ün eğitim alanında yaktığı meşalenin savunucusu ve uygulayıcısıyız. Beşincisini yaptığımız kongremizin de temel felsefesi budur.

Tüm bu olumsuz gelişmeler, eğitimin insan davranışları üzerindeki olumlu etkisini arttırma konusunda daha fazla çalışma yapmamız gerektiğini ortaya koymaktadır. Psikometri/Ölçme ve Değerlendirme bilim alanlarında bilimsel niteliği yüksek, daha fazla sayıda çalışmalar yapmaya özen göstermeliyiz. Elde ettiğimiz bulguları eğitimdeki karar vericilerle ve paydaşlarla paylaşmalı ve onları önlemler almaya davet etmeli, hatta ikna etmeliyiz. Türk eğitim sistemi hakkında alınan kararların arkasında bilimsel araştırmaların olup olmadığını sorgulamalı, kararların araştırmalara dayalı olarak alınmasında ısrarcı olmalıyız. Psikometri/Ölçme ve Değerlendirme bilim alanlarının yalnızca olguları betimleyici ve açıklayıcı değil, toplumların değişmesinde ve gelişmesinde rol oynayan itici bir güce sahip olduğunu da görmeliyiz.

Sözlerime burada son verirken Kongrenin Türkiye'nin en önemli turizm merkezlerinden olan Antalya'da bulunan Akdeniz Üniversite'sinde yapılmasını sağlayan ve önemli destekler veren Akdeniz Üniversitesi Rektörlüğüne ve Akdeniz Üniversitesi Eğitim Fakültesi Dekanlığına; ayrıca kongreye çeşitli konularda destek sağlayan Ankara Üniversitesi Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi'ne, Ankara Üniversitesi Sınav Yönetim Merkezi'ne, Pearson, Pegem Akademi, Nobel Yayıncılık, Anı Yayıncılık ve Expo 2016'ya teşekkür eder, verimli, dinamik ve yapıcı bir kongre süreci geçirmeyi dilerim.

Yaşasın, Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongremiz...

Saygılarımla...

Prof. Dr. Hülya KELECİOĞLU

EPODDER Başkanı

BİLDİRİ ÖZETLERİ

Bildiri No: 5

Kitapçık Türünün Test Maddelerinin Yanlılık Eğilimlerine Olan Etkisinin Değişen Madde Fonksiyonu Yöntemleriyle İncelenmesi

Metin ODABAS, Lokman AKBAY

Özet

Bu çalışmada Temel Eğitimden Ortaöğretime Geçiş (TEOG) 2013 Kasım Matematik testinde farklı kitapçıklarda cinsiyete göre değişen madde fonksiyonu (DMF) analizleri yapılmıştır. DMF analizleri üç farklı ölçme kuramına göre geliştirilmiş olan DMF belirleme yöntemlerine göre yürütülmüştür. Ankara'da 2013 Kasım TEOG sınavına giren öğrencilerin A ve B kitapçıklarına göre oluşturduğu evrenlerden elde edilen 1000 kişilik 100'er veri setinin her birisi DINA-DMF, Raju'nun alan ölçüleri ve Mantel-Haenszel DMF belirleme yöntemleriyle cinsiyete göre DMF analizleri yapılmıştır. Bu çalışmanın ana amaçlarından birisi farklı kitapçık türlerinde aynı maddelerin cinsiyete göre DMF gösterme eğilimlerini incelemek kaydıyla kitapçık türlerinin maddelerin psikometrik özelliklerini ne ölçüde değiştirdiği incelemektir. Ayrıca, üç farklı DMF analiz yöntemi kullanımı farklı test kuramlarına dayalı yöntemlerin DMF belirlemedeki tutarlılık ve benzerliklerinin görülmesine imkân sağlamıştır.

Anahtar Kelimeler

Değişen madde fonksiyonu, DINA model, raju'nun alan ölçüleri, mantel–haenszel.

Paper No: 5

Investigation of Booklet Type Effect on Item's Bias Tendency with DIF Methods

Abstract

In this study, differential item functioning (DIF) analyses are performed in different booklets with respect to gender for a mathematics exam, which was administered in November 2013 and is referred to as Transition From Elementary to Secondary Education (TEOG: Temel Egitimden Ortaogretime Gecis). DIF analysis are performed through three different methods, each of which are representative of three different testing therioes. One hundred sample data sets are randomly selected for each of the A and B booklets of the test from the responses of all examinees took the test in Ankara. DINA-DIF, Raju's area measures, and Mantel-Haenszel DIF detection methods are used to detect gender DIF across the booklets. One of the main purpose of this study is to investigate DIF flagging tendency of the items in different booklets. The results provides evidence to argue that booklet effect has an impact on the psychometric properties of items that can make an item function differently across gender groups in one booklet but not in the other. Furthermore, employment of three representative DIF methods of the measurement theories provides us the opportunity to compare and contrast the results obtained by DIF methods with different underlying testing theories.

Key Words

Differential item functioning, DINA model, raju's area measures, mantel–haenszel

Bildiri No: 6

Öğrenci Başarısının Belirlenmesinde Kullanılan Farklı Ölçme Araçlarından Elde Edilen Puanların Genellenebilirlik Kuramı ile Değerlendirilmesi *

Muammer YÜKSEL**, Gülşah BAŞOL***

Özet

Araştırmanın amacı, aynı konu alanında farklı ölçme araçlarına ait puanların G kuramı analizleriyle karşılaştırılmasıdır. Araştırmacılar tarafından geliştirilen, ön uygulaması yapılan ölçme araçları Kars il merkezinde 48 öğrenciye birer hafta arayla uygulanmıştır. Uygulanan klasik sınav üç farklı puanlayıcıtarafından puanlanmıştır. Sonuçlara göreklasik sınav için karışan hata miktarı çoktan seçmeliye göre daha yüksek düzeydedir. Çoktan seçmeli testlerin ortalamasının daha yüksek olduğu, varyansmiktarı ise iki sınav için paralellik gösterdiği görülmüştür. Ayrıca klasik sınav için puanlayıcılar arası puanlama farkları tespit edilmeye çalışılmış, karar çalışması yapılarak her iki ölçme aracı için G ve Phi katsayılarındaki farklılıklar incelenmiştir. Sonuçların iki ölçme aracı için paralellik gösterdiği görülmüştür.

Anahtar Kelimeler

Genellenebilirlik Kuramı, Klasik Sınav, Başarı.

---

*Bu çalışma, ilk yazarın aynı isimle Doç. Dr. Gülşah BAŞOL danışmalığında hazırlamış olduğu yüksek lisans tez çalışmasının bir kısmından oluşmaktadır.

**Ayşe Temizel Ortaokulunda Soma/MANİSA Elektronik posta: muammer8845@hotmail.com.

***Gaziosmanpaşa Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Taşlıçiftlik Yerleşkesi, 60100 Tokat. Elektronik posta: gulsah.basol@gop.edu.tr & gulsahbasol@gmail.com.

Paper No:6

Comparison of Scores Obtained From Different Measurement Tools Used in the Determination of Student Achievement

Abstract

This is a generalizability study, measuring students' performance through different measurement tools and comparing the finding though G theory and K in order to identify the errors associated with the raters and items to improve the future applications. The sample consisted of 48 randomly selected 8th graders, registered to Atatürk Secondary School, located in the center of Kars province during 2013-2014 academic years. Two different types of exams (a multiple choice test and an essay) prepared by the researcher were applied to students. Three raters graded the students' exams. Essays were scored by three different raters and G and K analyses were performed on the results. According to the findings, the error rate was higher for the essays in comparison to multiple choice test. The mean score was higher for the multiple choice test, on the other hand the variances were found to be similar. For the essays, there were no differences among the scores given by different coders. Decision studies indicated that the findings were in line for both type of the measurements according to G and Phi coefficients. The results indicated that increasing the number of items was resulted in an increase in the G and Phi coefficients. According to the findings, the Person Factor was the main source of the variation in the data for both measurement types. As the findings indicated consistent result among the coders, we suggest increasing the number of coders rather than increasing the number of questions for the essay. Considering the excessive work, associated with grading an essay with more question, when available, it can be better to increase the number of raters.

KeyWords

GeneralizabilityTheory, Essay, Success.

Bildiri No: 7

Çok Boyutlu Madde Tepki Kuramındaki Değişik Faktörlerin Yetenek Parametresi Kestirimi Doğruluğuna Etkisi

Levent YAKAR*

Özet

Bu çalışmanın amacı değişik faktörlerin çok boyutlu madde tepki kuramı (ÇBMTK) yetenek parametresi kestirim doğruluğuna etkisini incelemektir. Bu amaçla 100 tekrarla yapılan simülasyon çalışması, telafi edici ve telafi edici olmayan 1PL, 2PL ve 3PL ÇBMTK modellerine etki eden madde ayırt edicilik (düşük, orta ve yüksek), madde yapısı (çoğunlukla karmaşık, eşit düzeyde ve çoğunlukla basit), test uzunluğu (15,30 ve 60) ve yetenek parametreleri arası korelasyon (0 ve 0,6) faktörleri çaprazlanarak 54 koşul üzerinden gerçekleştirilmiştir. Çalışmada 2 boyut üzerinde durulmuş ve madde parametresi kestirim doğruluğundan etkilenmemek adına birey sayısı yüksek tutularak 5000 bireylik veri setleri oluşturulmuştur. Analiz sonucunda elde edilen yetenek parametreleri ile gerçek yetenek parametrelerinin mutlak farkı, RMSEA değeri, kestirimden elde edilen yetenek parametrelerinin ortalaması ve standart hataları hesaplanmıştır. Elde edilen sonuçlara göre telafi edici model kestirimleri telafi edici olmayan model kestirimlerine göre kısmen daha başarılıdır. 1PL ve 2PL modelleri arasında fark yok denecek kadar az iken, 3PL modelden elde edilen parametreler 1PL ve 2PL'ye göre gerçek değerlerden daha uzaktır. Daha fazla maddenin tek boyuta yük vermesi kestirim doğruluğunu artırmaktadır. Beklendiği üzere madde kalitesi artıkça kestirim doğruluğu da artmaktadır. Aynı şekilde test uzunluğunun artması kestirim doğruluğunu artırmaktadır. Yetenekler arası korelasyon artışı da kestirim doğruluğunu artışını beraberinde getirmektedir. ÇBMTK doğası gereği çok sayıda parametre kestirimini gerektirdiğinden, tek boyutlu MTK ile kıyaslandığında parametrelerin kestirim doğruluğu açısından daha zayıf kalabilir. Ancak testin ÇBMTK gerektirmesi durumunda tek boyutlu MTK uygulaması eksik ve çok daha yanlış sonuçlar verebilir. Birden fazla yeteneğin aynı anda ölçülmek istenmesi durumunda elde ettiğimiz sonuçlar ışığında testin hazırlanması ile birey yeteneklerinin daha doğru bir düzeyde elde edilmesi mümkündür.

Anahtar Kelimeler

Çok Boyutlu Madde Tepki Kuramı, Yetenek Parametresi, Madde Parametreleri

* Hacettepe Üni. Eğitim Fak. Eğitim Bilimleri Böl. Tel: 0312 297 85 50

Paper No: 7

Effect of Different Factors in Multidimensional Item Response Theory to Ability Score Estimation Accuracy

Abstract

This study aims investigate of effect of different factors in multidimensional item response theory (MIRT) to ability parameter estimation accuracy. For this purpose, 100 replications performed simulation study for compensatory and non-compensatory 1PL, 2PL and 3PL MIRT models. Investigated factors are item discriminations (low, medium and high), item structures (mostly complex, same rate and mostly simple), test length (15, 30 and 60) and the correlation between the ability parameters (0 and 0.6), totally 54 conditions crossed. The study has focused on two dimensions. For eliminate precision of item parameters estimation, sample size was specified huge enough, n=5.000. End of analysis, absolute difference of true and estimated ability parameters, RMSEA, estimated ability parameters' average and standard errors were obtained and reported. In results, compensatory models are slightly more successful on accuracy than non-compensatory models. While 1PL and 2PL accuracy are almost identical, 3PL model estimations is poorer than 1PL and 2PL results. Increasing rate of simple structured items provide closer estimation to true abilty score. As expected, the higher item discriminations have more accurate estimations. Likewise, increasing test length cause of increasing estimation accuracy. Also increasing correlation between ability scores, increases accuracy. Because of the nature MIRT requires a large number of parameter estimation, may be weaker in terms of parameter estimation accuracy compared with unidimensional IRT. However, although if the test fits to MIRT, application of unidimensional IRT on test can cause incomplete and incorrect results. In case of measuring multiple abilities in a test, preparation of the test in consideration of these results, can be obtained more accurate ability scores.

Key Words

Multidimensional Item Response Theory, Ability Parameter, Item Parameters

Bildiri No: 10

Eğitimde Ölçme ve Değerlendirme Dersinin Uygulamadaki Yeri

Meltem ACAR GÜVENDİR*, Yeşim ÖZER ÖZKAN**

Özet

Bu araştırmanın amacı yükseköğretimde verilen ölçme ve değerlendirme dersinin uygulamadaki yerini belirleyebilmektir. Bu amaç doğrultusunda Trakya Üniversitesi Eğitim Fakültesi İngiliz Dili Eğitimi Anabilim Dalı'ndan mezun olan ve en fazla bir senedir öğretmenlik yapan toplam 30 öğretmen çalışma grubunu oluşturmaktadır. Çalışma grubuna en fazla bir sene öğretmenlik yapan kişilerin seçilmesinin nedeni üniversitede ölçme ve değerlendirme dersini yakın zamanda almış olmalarından dolayı bu dersin uygulamadaki yerini daha iyi değerlendirebilecekleri düşüncesidir. Öğretmenlerle internet üzerinden görüşmeler yapılmıştır. Bu yönüyle araştırma nitel araştırma desenlerinden durum çalışması niteliğindedir. Veriler araştırmacılar tarafından beş öğretmenle yapılan ön görüşmeler sonucunda hazırlanan görüşme formu ile toplanmıştır. Görüşme formu; "Üniversitede almış olduğunuz ölçme ve değerlendirme dersine ilişkin hatırladıklarınız nelerdir?", "Üniversitede ölçme ve değerlendirme dersinde öğrendiğiniz konulardan hangisi/hangileri sınıf içi ölçme değerlendirme uygulamalarınızda en çok katkı sağlamaktadır? Neden?", "Üniversitede ölçme ve değerlendirme dersinde öğrenmiş olduğunuz fakat uygulamada hiç kullanmadığınız konu/konular nelerdir? Neden?", "Öğrencilerinizin ders/konu/ünite sonlarındaki öğrenme eksikliklerini değerlendirmek amacıyla hangi değerlendirme yöntem/teknik/ araçlarını kullanıyorsunuz?," "Sınav hazırlarken üniversitede aldığınız eğitimin size katkı sağladığını düşünüyor musunuz? Neden?", "Sınav hazırlarken en çok hangi konuda zorlanıyorsunuz? Neden?", "Sınav hazırlarken en çok tercih ettiğiniz soru tipi hangisidir? Neden?" ve "Sınav sonuçlarını nasıl değerlendiriyorsunuz?" olmak üzere sekiz soru içermektedir. Araştırmanın veri analizinde betimsel analiz kullanılmıştır. Elde edilen sonuçlara göre öğretmenler üniversitede aldıkları ölçme ve değerlendirme dersinin özellikle sınav hazırlama aşamasında faydalı olduğunu düşünmekteler ve bu dersten en çok sınav hazırlamada dikkat edilecek durumlar ve soru türleri konularında faydalanmaktadırlar. Öğretmenler en çok öğrencilerin seviyelerine yönelik soru hazırlamada sıkıntı yaşadıklarını belirtirken, madde analizi ve test istatistikleri konularını ihtiyaç duymadıkları için kullanmadıklarını belirtmişlerdir.

Anahtar Kelimeler

Ölçme ve değerlendirme, yükseköğretim, İngilizce öğretmeni, uygulama, teori.

*Trakya Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:meltemacar@gmail.com Tel: +902842120808.

**Gaziantep Üniversitesi Gaziantep Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:yozer80@gmail.com Tel: +903423604372.

Paper No: 10

Practicality of Measurement and Evaluation Course in Education

Abstract

The purpose of the study is to determine practicality of measurement and evaluation course in education. The study group includes 30 teachers who graduated from English Language Teaching Department at Trakya University, Faculty of Education and have been teachers for one year. As the teachers in the study group had recently finished university and successfully completed measurement and evaluation course, they were considered to have a higher potential to evaluate and compare this course and its functionality in the occupational life of the teachers. The teachers were interviewed on Internet. Thus, this research is a case study making it a qualitative research. Data were obtained by an interview form which was prepared considering pre-interview results of five teachers. The interview form contains eight questions: "What do you remember about measurement and evaluation course that you took at university?", "Which topic/s of measurement and evaluation course has/have been contributing to your class at most? Why?", "Which topic/s of measurement and evaluation course have you not been using in your class? Why?", "Which assessment methods/technics/tools have you been using for measuring your students' lack of knowledge? Why?", "Do you think that measurement and evaluation course contributes to your exam preparation? Why?", "What do you find most challenging while you are preparing an exam? Why?", "Which question type do you prefer at most when you prepare an exam? Why?", and "How do you evaluate exams' scores? Descriptive analysis was used for data analysis. According to the results, the teachers thought that measurement and evaluation course, which they took at the college, has contributed to them while they have prepared exams. They have also been making use of this course in terms of question types and key points of exam preparation. The teachers found it difficult to prepare questions according to students' level and they have not been using test and item statistics as they have not needed them.

Key Words

Measurement and Evaluation, Higher Education, English Teacher, Practice, and Theory

Bildiri No: 12

Likert Tipi Ölçeklerde Olumsuz Madde ve Kategori Sayısına İlişkinRasch Modeli ile Bir İnceleme

Mustafa İLHAN*, Neşe GÜLER**

Özet

Bu araştırmada Likert tipi ölçeklerde olumsuz madde ve kategori sayısı sorunu ele alınmıştır. Araştırma, biri 197 ve diğeri 115 katılımcıdan oluşan iki ayrı çalışma grubu üzerinde yürütülmüştür. Çalışmada veri toplama aracı olarak Olumsuz Değerlendirilme Korkusu Ölçeği (ODKÖ) ile Oxford Mutluluk Ölçeği-Kısa Formu (OMÖ-K) kullanılmıştır. Birinci çalışma grubundaki katılımcılara ODKÖ, ikinci çalışma grubundaki katılımcılara OMÖ-K üç, beş ve yedili dereceleme ile uygulanmıştır. Araştırma kapsamında toplanan veriler Rasch modeline göre analiz edilmiştir. Analiz sonucunda; ODKÖ ile OMÖ-K'deki olumlu maddelerde gözlenen ve beklenen test karakteristik eğrilerinin büyük ölçüde örtüştüğü, her üç kategori sayısının da etkin bir biçimde çalıştığı ve ölçek kategorileri arasındaki farkların katılımcılar tarafından başarılı bir biçimde ayırt edildiği belirlenmiştir. Öte yandan olumsuz maddelerde gözlenen ile beklenen test karakteristik eğrileri arasında önemli farklılıklar olduğu ve ölçek kategorilerinin etkin bir biçimde çalışmadığı saptanmıştır. Üç, beş ve yedili derecelendirmeden hangisi kullanılırsa kullanılsın, ODKÖ ile OMÖ-K'deki olumsuz maddelerde kategorilerin katılımcılar tarafından ayırt edilemediği tespit edilmiştir. Bu tespitin ardından olumsuz maddeler veri dosyasından çıkarılarak analiz tekrarlanmıştır. Analiz sonuçlarına göre, ODKÖ kullanılarak elde edilen ölçümlerde birey yüzeyi için rapor edilen güvenirlik katsayıları üçlü derecelemede .71, beşli derecelemede .82 ve yedili derecelemede .71 olarak bulunmuştur. Üçlü, beşli ve yedili derecelemede madde yüzeyi için rapor edilen güvenirlik katsayıları ise sırasıyla, .94, .95 ve .93 olarak hesaplanmıştır. OMÖ-K kullanılarak elde edilen ölçümler analiz edildiğinde, birey yüzeyine ait güvenirlik katsayıları üçlü derecelemede .70, beşli derecelemede .76 ve yedili derecelemede .70 olarak saptanmıştır. Madde yüzeyine ait güvenirlik katsayıları ise üçlü, beşli ve yedili derecelemede sırasıyla; .96, .97 ve .97 şeklinde belirlenmiştir. Bu sonuçlara göre, ölçekteki kategori sayısının madde yüzeyine ilişkin güvenirlikte bir farka yol açmadığı söylenebilir. Birey yüzeyi için rapor edilen güvenirlik değerleri ise Likert tipi ölçeklerde beş kategorili bir derecelemenin Türk kültürü için daha uygun olacağına işaret etmektedir.

Anahtar Kelimeler

Likert tipi ölçek, olumsuz madde, kategori sayısı, Rasch modeli

* Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, İlköğretim Bölümü. İletişim: mustafailhan21@gmail.com, Tel: +90 505 516 2262

** Sakarya Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: gnguler@gmail.com, Tel: +90 505 505 3370

Paper No: 12

A Study with the Rasch Model Regarding the Reverse Directional Item and Number of Response Categories in Likert-type Scales

Abstract

This study addressed reverse directional item and the number of response categories problems in Likert-type scales. The study was conducted with two study groups, one of 197 students and the other of 115 participants. The Fear of Negative Evaluation Scale (FNES) and the Oxford Happiness Questionnaire-Short Form (OHQ-S) were used as data collection tools. The FNES was administered to the participants in the first study group while the OHQ-S was administered to the participants in the second group using three, five and seven-point rating scales. The data of the study were analyzed according to the Rasch model. The analysis found that the observed and expected test characteristic curves were largely overlapped, each of the three rating scales worked effectively, and the differences between response categories could be distinguished successfully by the participants in straightforward directional items. On the other hand it was determined that there were significant differences between the observed and expected test characteristic curves in reverse directional items. It was also found that no matter which one of these three, five and seven-point rating scales was used, the participants could not distinguish the response categories of the reverse directional items on the FNES and the OHQ-S. Afterwards, the reverse directional items were removed from the data file, and the analysis was repeated. The analysis found the reliability coefficients for the persons facet in the measurements obtained using the FNES to be .71 for the three-point rating scale, .82 for the five-point rating scale and .71 for the seven-point rating scale. The reliability coefficients reported for the item facet for the three, five and seven-point rating scales were found to be .94, .95 and .93, respectively. The analysis of measurements obtained using the OHQ-S showed that the reliability coefficients for the persons facet were .70 for the three-point rating scale, .76 for the five-point rating scale and .70 for the seven-point rating scale. The reliability coefficients reported for the item facet for the three, five and seven-point rating scales were found to be .96, .97 and .97, respectively. These results, indicate that the number of response categories on the scale does not make a difference in reliability for the item facet. The reliability values reported for the persons facet show that a five-point rating in Likert-type scales is more appropriate for Turkish culture.

Key Words

Likert-type scale, reverse directional items, number of response categories, Rasch model

Bildiri No: 14

Türkiye'de Eğitim Alanında Yayımlanan Ölçek Geliştirme Çalışmalarının Çok Yüzeyli Rasch Modeli İle İncelenmesi

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

Özet

Bu çalışmanın amacı Türkiye'de 2010-2015 yılları arasında SSCI dizininde taranan bilimsel dergilerde yayımlanmış ölçek geliştirme çalışmalarının, "ölçek geliştirme süreci"ndeki adımları ne derece sağlayabildiklerine ilişkin incelemektir. Araştırma kapsamında Türkiye'de eğitim alanında SSCI dizininde yer alan üç dergide 2010-2015 yılları arasında yayınlanan 57 ölçek geliştirme makalesi, Ölçme ve Değerlendirme alanında 4 uzman tarafından araştırmacıların geliştirdiği "Ölçek Geliştirme Makaleleri İnceleme Formu" ile puanlanmıştır. Ölçek "Geliştirme Makaleleri İnceleme Formu" alanyazında olan genel görüşler dikkate alınarak 20 maddeden oluşan bir form olarak düzenlenmiştir. Analizler çok yüzeyli Rasch modeli (ÇYRM) ile 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 düzeylerine ait bulgular ayrı ayrı elde edilmiştir. Analizler sonucunda, Logit cetvele göre çalışmada esas alınan ölçütleri karşılayabilme düzeyi en yüksek 14 numaralı makalenin (1.13) ve en düşük 27 (-1.01) 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 yüzeyli Rasch Modeli, eğitim bilimleri

Paper No: 14

The Investigation of Scale Development Studies Conducted In Educational Sciences Published In Turkey By Using Many Facet Rasch Model

Abstract

The aim of this study is to investigate the scale development studies which were conducted in Turkey from 2010 to 2015 and published in the journals indexed in 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 and indexed in SSCI were rated according to "Scale Development Article Investigation Form" by four measurement and evaluation expert. The 20-item "Scale Development Article Investigation Form" was composed according to the view of literature. The many facet Rasch model was used and analyses were done by using FACETS program developed. The analysis results of the three facets (the level of articles by means of verification of determined 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 (The values obtained from criterion related validity were appropriate.) was the most difficult one. On the other hand, the 3rd 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 4th rater was the strictest one and the first rater was more lenient one. Moreover, the developed scales were given in just 27 of 57 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) or investigated articles, there were no such information for 43,9 % of investigated articles.

Key Words

Scale development, many facet Rasch model, educational sciences

Bildiri No: 15

Klasik Test Kuramına Göre Sayısal ve Sözel Alanlar İçin Puanlama Güvenirliğinin Kayıp Veri Kapsamında İncelenmesi*

Sibel ADA**, İsmail KARAKAYA***

Özet

Çalışmanın amacı puanlama güvenirliğini belirlemek için kullanılan uzlaşma katsayılarını karşılaştırmak ve veri setindeki kayıp veri oranına göre bu katsayıları incelemektir. Sayısal ve sözel ders için hesaplanan uzlaşma katsayılarının incelenmesi çalışmanın bir diğer amacını oluşturmaktadır. Çalışma grubunda 71'i sekizinci sınıf ve 159'u dokuzuncu sınıf öğrencisi olmak üzere toplam 230 öğrenci bulunmaktadır. Çalışma kapsamında elde edilen veriler üzerinden uzlaşma katsayıları ve sınıf içi korelasyon değerleri kayıp veri oranının yüzde beş, on ve on beş olduğu durumlara göre hem iki puanlayıcı ve hem de üç puanlayıcı için ayrı ayrı hesaplanmıştır. İki puanlayıcı için hesaplanan uzlaşma katsayıları Cohen'in kappa, Gwet'in AC2, Scott'nin pi, Krippendorff'un alfa, Brennan-Prediger, uzlaşma yüzdesi; üç puanlayıcı için hesaplanan uzlaşma katsayıları Fleiss'nin kappa, Gwet'in AC2, Krippendorff'un alfa, Conger'in kappa, Brennan-Prediger ve uzlaşma yüzdesi şeklindedir. Toplam puan üzerinden ise sınıf içi korelasyon katsayısı hesaplanmıştır. Elde edilen güvenirlik katsayılarına ait değerleri karşılaştırılırken uzlaşma katsayıları için standart hata değerleri ve sınıf içi korelasyon katsayısı için hata varyansları dikkate alınmıştır. Uzlaşma yüzdesinin her zaman en yüksek güvenirlik katsayısı değerine sahip olduğu tespit edilmiştir. Uzlaşma katsayısından sonra genel olarak en yüksek katsayı değerine ve en düşük standart hata değerine sahip katsayı Gwet'in AC_2 katsayısıdır. Ayrıca Brennan-Prediger ile Gwet'in AC_2 ve Cohen'in/ Conger'in kappa, Scott'nin pi/Fleiss'nin kappa ile Krippendorff'un alfa değerlerinin genel olarak yakın değerlere sahip olduğu görülmüştür. Kayıp veri oranının yüzde on beşe kadar olduğu durumlarda hem uzlaşma katsayılarının hem de sınıf içi korelasyon değerlerinin iki yüz elli kişilik veri grubu için kayıp veri olmadığı durumla benzer sonuçlar verdiği görülmektedir. Ayrıca üç puanlayıcı için elde edilen güvenirlik katsayısı değerlerinin iki puanlayıcıya göre ve sayısal dersler için elde edilen güvenirlik katsayısı değerlerinin sözel derslerden elde edilene göre daha yüksek olduğu tespit edilmiştir.

Anahtar Kelimeler

Puanlama Güvenirliği, Puanlayıcılar Arası Güvenirlik Katsayısı, Sınıf İçi Korelasyon Katsayısı, Kayıp Değer

---

*Bu çalışma 19.11.2015 tarihinde Gazi Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Bilim Dalı'nda yüksek lisans tezi olarak kabul edilmiştir.

**Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü. İletişim: sibelada.sa@gmail.com

*** Gazi Üniversitesi, Gazi Fakültesi, Eğitim Bilimleri Bölümü. İletişim: ikarakaya@gazi.edu.trTel: +90 312 202 17 49

Paper No: 15

Examining Interrater Reliability Considering Missing Value For Numerical And Verbal Courses According To Classical Test Theory

Abstract

The purpose of this research was to compare agreement coefficient that was used to determine the interrater reliability and to examine this coefficient according to missing value ratio in data set. Examination of compromise agreement coefficients calculated for numerical and verbal courses was another purpose of this research. There were 230 students (71 in 8th grade and 159 in 9th grade) in study group. Agreement coefficients and intra-class correlation coefficient were calculated for both two and three raters according to missing value ratio consisting of five, ten and fifteen percent. The agreement coefficients calculated for two raters were as follows: Cohen's Kappa, Gwet's AC2, Scott's pi, Krippendorff alpha, Brennan-Prediger and Agreement Percentage. The agreement coefficients calculated for three raters were as follows: Fleiss' Kappa, Gwet's AC2, Conger's kappa, Krippendorff alpha, Brennan-Prediger and Agreement Percentage. Total scores were used for calculating intra-class correlation coefficient. While the values of reliability coefficients were compared, standard error values for agreement coefficient and error variance for intra-class correlation were taken into consideration. It was determined that Agreement Percentage had maximum reliability coefficient values. After Agreement Percentage coefficient, the coefficient of maximum reliability values and minimum standard error were usually Gwet's AC_2. It was also showed that values of Brennan-Prediger and Gwet's AC2, and values of Cohen's/Conger's kappa, Scott's pi/Fleiss' kappa and Krippendorff alpha had almost same values. In the case wheremissing value ratio was to up fifteen percent, values of both agreement coefficients and intra-class correlations had same results with no missing value for two hundred and fifty data sets. It was also determined that values of reliability coefficient for three raters were higher than two raters' and that the values of reliability coefficient for numerical courses were higher than verbal course.

Key Words

Inter-rater Reliability, Inter-rater Reliability Coefficient, Intra-class Reliability Coefficient, Missing Value

Bildiri No: 16

Türkçe Konuşmaya İlişkin Kaygı Ölçeği Geliştirilmesi

Nesrin SEEF*, Sakine GÖÇER ŞAHİN**

Özet

Yabancı uyruklu bireylerin her dönem arttığı Türkiye'de bireylerin Türkçe konuşmaya yönelik kaygılarının belirlenmesi ve buna yönelik eğitsel, sosyal tedbirlerin alınması onların Türkiye kültürüne daha kolay uyum sağlamasını sağlayacaktır. Bu çalışmanın amacı, Türkiye'de eğitim gören öğrenciler için Türkçe konuşmaya ilişkin kaygı ölçeği'ni (TKİKÖ) geliştirmektir. Bu amaçla Ankara'daki üç büyük üniversiteye bağlı Türkçe Öğretim Merkez (TÖMER)'lerine devam eden ve anadili 38 farklı dil olan 235 yabancı uyruklu öğrenciden veri toplanmıştır. Ölçeğin geçerliği, kapsam, yapı ve ölçüt dayanaklı geçerlik bağlamında incelenmiştir. Kapsam geçerliği için ölçme değerlendirme (2), psikolojik danışmanlık ve rehberlik (PDR) (3) ve TÖMER'de çalışan Türkçe öğretimi (2) uzmanlarından görüş alınmıştır. Yapı geçerliği için faktör analizinden faydalanılmıştır. Varimax döndürme tekniğinin kullanıldığı analiz sonunda 3 boyut ve 13 maddeli bir yapı elde edilmiştir. Bu üç boyutun birlikte toplam varyansın %51,545'ini açıkladığı belirlenmiştir. Elde edilen boyutlar yine PDR alanındaki uzmanların görüşleri doğrultusunda fizyolojik belirtiler, kaçınma ve kendine güven olarak adlandırılmıştır. Ölçüt dayanaklı geçerlik çalışması kapsamında öğrencilerin ölçekten aldıkları puanların çeşitli demografik özelliklere göre değişip değişmediği incelenmiştir. Buna göre geliştirilen ölçeğin, Türkiye'de daha uzun zaman geçirenler, daha önceden Türkçe bilenler ve anadili Türkçe ile aynı dil ailesinde olanlar dolayısıyla kaygı objesi ile ilgili daha çok deneyimi olan bireyler ile olmayanların diğer bir ifadeyle kaygısı az ve çok olan bireyleri ayırmada etkili olduğu söylenebilir. 13 maddelik ölçeğin güvenirliği ile her bir alt boyutun güvenirliği Cronbach alfa katsayısı ile hesaplanmıştır. Ölçeğin genel güvenirliği 0,808; boyutların ise sırasıyla 0,735; 0,755 ve 0,468 olarak hesaplanmıştır. Madde sayısı az olan 3.boyut dışındaki iki boyut ile ölçeğin geneli için güvenirliğin iyi düzeyde olduğu belirlenmiştir. Tüm bulgulara dayanarak Türkçe Konuşmaya İlişkin Kaygı Ölçeğinin yabancı uyruklu öğrencilerin bu kaygılarını güvenilir ve geçerli bir şekilde ölçtüğü söylenebilir.

Anahtar Kelimeler

Türkçe, yabancı dil, kaygı, geçerlik, güvenirlik

* Doktora öğrencisi. Hacettepe Üniversitesi, Eğitim Fakültesi Bölümü, Eğitimde Ölçme Değerlendirme Abd, Ankara. İletişim: n-seef@hotmail.com Tel: +90 312 2978550

** Arş. Gör. Hacettepe Üniversitesi, Eğitim Fakültesi Bölümü, Eğitimde Ölçme Değerlendirme Abd, Ankara. İletişim: sgocersahin@gmail.com Tel: +90 312 2978550

Paper No: 16

Developing Turkish Speaking Anxiety Scale

Abstract

Detection of anxiety towards speaking Turkish in Turkey where the number of foreign nationals increases day by day and to take related educational and social precautions will help those people adapt to Turkish culture. This study aims to develop a Turkish Speaking Anxiety Scale (TSAS) for students who receive education in Turkey. With this purpose, data were collected from 235 foreign national students who have 38 different mother tongues and receive education at Turkish Teaching Center (TÖMER) affiliated to three greatest universities of Ankara province. Opinions of experts in assessment and evaluation (2), psychological counseling and guidance (3) and Turkish teaching at TÖMER (2) were asked for content validity. Factor analysis was used for construct validity. Varimax rotation technique was used for the analysis; consequently the obtained construct was 13 items in 3 dimensions. It was found that all these three dimensions explain 51.545% of total variance. The obtained dimensions were named after physiological indications, avoiding and self-confidence in line with the opinions of experts in psychological counseling and guidance. The question whether the scores that students got from the scale differ by various demographic features or not was analyzed within the scope of criteria-based validity study. Accordingly, it can be said that the developed scale is effective to differentiate those who do not have experience with anxiety and those who have more experience due to characteristics of the participants who already spent longer time in Turkey, already knew Turkish and whose mother tongue is from the same language family with Turkish; in other words, this scale differentiates those who have less anxiety and those who have more. Cronbach alpha coefficient was used to measure the reliability of 13-item scale and each sub-dimension. The reliability of overall scale and sub-dimensions respectively was found 0.808, 0.735, 0.755 and 0.468. Reliability of the scale and two dimensions, except for the 3rd dimension having less items, was found in good level. According to the overall findings, it can be said that Turkish Speaking Anxiety Scale assesses the anxiety of foreign national students as a valid and reliable construct.

Key Words

Turkish, foreign language, anxiety, validity, reliability

Bildiri No: 17

Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Uygulanabilirliği: Bir Post-Hoc Simülasyon Çalışması*

Eren Can AYBEK[1**], R. Nükhet DEMİRTAŞLI[***]

Özet

Bu araştırma, okullarda öğrenci yönlendirme hizmetlerinde öğrencinin yeteneklerini, mesleki ilgi ve değerlerini tanımasına yardımcı olmak üzere sıklıkla kullanılan Kendini Değerlendirme Envanteri (KDE)'nin Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT) olarak uygulanabilirliğini araştırma amacıyla yapılmıştır. Kendini Değerlendirme Envanteri, bireyin algıladığı yeteneklerini, ilgilerini ve mesleki değerlerini ölçmeyi amaçlamaktadır. Envanterde üç boyut altında toplanmış 23 faktör ve toplam 230 madde yer almaktadır. Her faktör 10 maddeden oluşmaktadır. Çalışma kapsamında işbirliği faktörü için .63, değişiklik faktörü için .68; diğer faktörler içinse .71 ile .88 arasında iç tutarlılık katsayıları elde edilmiştir. Çalışmada Firestar yazılımı kullanılarak; 653 kız, 490 erkek ve cinsiyet bilgisini paylaşmayan 1 öğrenci ile birlikte toplam 1144 lise öğrencisinden toplanan gerçek verilere dayalı post-hoc simülasyonlar yapılmıştır. Simülasyonlarda, .30, .40 ve .50 standart hata durdurma kuralları; Genelleştirilmiş Kısmı Puan (GKPM) ve Kademeli Tepki (KTM) Modelleri; Fisher En Yüksek Bilgi (FEYB), Beklenen En Yüksek Bilgi (BEYB), Beklenen En Düşük Sonsal Varyans (BEDSV) ve Beklenen En Yüksek Sonsal Ağırlıklandırılmış Bilgi (BEYSAB) madde seçim yöntemleri kullanılmıştır. Simülasyonlar KDE'nin 23 faktörü için ayrı ayrı yapılmış ve simülasyon sonuçlarına göre madde tepki modeli olarak GKPM, madde seçim yöntemi olarak FEYB, durdurma kuralı olaraksa .40 standart hatanın BOBUT uygulaması için en uygun parametreler olduğuna karar verilmiştir. Ayrıca kağıt-kalem formu 230 madde olan Kendini Değerlendirme Envanteri, simülasyon sonuçlarına göre 113 madde ile sonlanmış ve kağıt-kalem formundan alınan puanlar ile simülasyonla kestirilen θ düzeyleri arasında .90 ile .96 arasında korelasyon katsayıları elde edilmiştir.

Anahtar Kelimeler

Bilgisayar Ortamında Bireye Uyarlanmış Test, Madde Tepki Kuramı, Kendini Değerlendirme Envanteri, Mesleki Rehberlik

---

*Bu çalışma "Kendini Değerlendirme Envanteri'nin Bilgisayar Ortamında Bireye Uyarlanmış Test Olarak Uygulanabilirliğinin Araştırılması" başlıklı doktora tezinden üretilmiştir.

**Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:erencan@aybek.net

***Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: rnukhet@yahoo.com

Paper No: 17

Applicability of the Self-Assessment Inventory as a Computerized Adaptive Test: A Post-Hoc Simulation Study

Abstract

The purpose of this study is investigating the applicability of the self-assessment inventory as a computerized adaptive test (CAT) form which developed to assess students' abilities, vocational interests and values. Self-Assessment Inventory aims to measure abilities, interests and vocational values of high school students. Inventory has 230 items under 23 factors and three dimensions. Each factor has 10 items. Only two factors have .63 and .68 internal consistency coefficient and other 21 factors have internal consistency coefficient between .71 to .88. Post-hoc simulations have been executed with data collected from 1144 (653 female, 490 male, 1 unknown) high school student. .30, .40 and .50 standard error stopping rules; Generalized Partial Credit (GPCM) and Graded Response (GRM) Models; Maximum Fisher Information (MFI), Maximum Expected Information (MEI), Minimum Expected Posterior Variance (MEPV) and Maximum Expected Posterior Weighted-Information (MEPWI) item selection methods are used in simulations. According to the simulation results, GPCM was chosen as the IRT Model, MFI as the item selection method and .40 standard error was chosen as the stopping rule for the live CAT form. Simulation study showed that mean 113 items has been used, while paper-pencil form has 230 items; and the correlation coefficients between simulation θ and the paper-pencil scores for 23 factors change .90 to .96.

Key Words

Computerized Adaptive Test, Item Response Theory, Self-Assessment Inventory, Vocational Counseling

Bildiri No: 18

SBS 2011 Türkçe Alt Testindeki Maddelerin Cinsiyet Değişen Madde Fonksiyonu Açısından İncelenmesi

Emre TOPRAK*, Levent YAKAR **

Özet

Bu araştırmada SBS Türkçe alt testinde yer alan 23 maddede cinsiyete göre DMF bulunma durumunun farklı DMF belirleme yöntemleriyle incelenmesi amaçlanmıştır. Ayrıca ikinci bir amaç olarak, DMF içeren maddelerin bulunması halinde bunların coğrafi bölgelere göre durumunun ne olduğunun ortaya konulması belirlenmiştir. Araştırma verileri 2011 yılı 8. sınıf SBS'nin 23 soruluk Türkçe alt testine aittir.Değişen Madde Fonksiyonunu (DMF) tespit etmek üzere verilerin analizinde klasik test teorisine dayalı olan Mantel-Haenszel ve Lojistik Regresyon yöntemleri kullanılırken; Madde Tepki Kuramına dayalı olan SIBTEST, Wald istatistiği ve Olabilirlik Oranı yöntemleri kullanılmıştır. Araştırmadan elde edilen bulgulara göre lojistik regresyon modeli hariç diğer 4 yöntem DMF içeren farklı maddeleri göstermektedir. SIBTEST ve Mantel-Haenszel yöntemleri sadece 9. ve 21. maddede DMF tespit ederken; Lojistik Regresyon yöntemiyle sadece 2. maddede DMF tespit edilmiştir. Madde tepki kuramına göre kestirim yapan SIBTEST sadece 9. ve 21. maddeler, Olabilirlik Oranı bu maddelere ek olarak 14. ve 18. maddelerde ve Wald İstatistiği ise 9. ve 21. madde dışında 14. maddede DMF tespit etmiştir. Lojistik Regresyon diğer modellerin tespit ettiği maddelerde sorun görmemesi ve DMF tespit ettiği 2. maddenin diğer yöntemler tarafından sorunsuz görülmesi nedeniyle yorumlarda kullanılmamıştır. DMF tespit edilen maddelerin 2'sinde kızlar lehine, 2'sinde ise erkekler lehine Değişen Madde Fonksiyonu gözlenmiştir.

Anahtar Kelimeler

Değişen madde fonksiyonu, SBS

* Erciyes Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: etoprak@erciyes.edu.tr
** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: leventyakar@hacettepe.edu.tr

Paper No: 18

Analysis of SBS 2011 Turkish Subtest Items In Terms Of Differential Item Functioning Based On Gender

Abstract

In this study SBS Turkish subtest was analyzed in terms of whether or not test items changes based on gender in a comparative way between different differential item functioning (DIF) determination techniques. And also test items were analyzed according to the geographical regions. Research data from the year 2011 belongs to the 8th grade SBS's 23 Turkish subtest. Mantel-Haenszel, Logistic Regression, SIBTEST, Wald statistic and Likelihood Ratio were used for analysis of DIF. SIBTEST and Mantel-Haenszel determined item 9 and 21; Logistic regression determined only item 2; SIBTEST determined item 9 and 21; Likelihood Ratio determined item 14 and 18 in addition to these items and Wald statistic determined item 14 except item 9 and 21. Logistic regression determined DIF only on item 2 but other methods didn't determine on item 2, so this item wasn't used for analysis. All methods, except the logistic regression were identified in item 9 and 21. Items detected DIF were found 2 in favor of girls and 2 in the men's favor.

Key Words

Differential Item Functioning, SBS

Bildiri No: 19

Matematik, Fizik, Kimya ve Biyoloji Öğretim Programlarının Öğretmen Görüşlerine Göre İncelenmesi

Abdullah Faruk KILIÇ*, Cem Oktay GÜZELLER**

Özet

Bu araştırmanın amacı matematik, fizik, kimya ve biyoloji öğretim programlarının öğretmen görüşleri açısından değerlendirilmesidir. Araştırmanın evrenini Türkiye'de Anadolu, fen, sosyal bilimler, spor ve güzel sanatlar liselerinde görev yapan matematik, fizik, kimya ve biyoloji öğretmenleri oluşturmaktadır. Örneklem ise tabakalı ve küme örnekleme yöntemiyle seçilmiş 5767 öğretmenden oluşmuştur. Veri toplama aracı olarak araştırmacılar tarafından hazırlanan anket formu kullanılmıştır. Anket formunda 5'li likert tipinde 15 madde bulunmaktadır. Hazırlanan anket internet ortamında uygulanarak veriler toplanmıştır. Katılımda dürüstlük, içtenlik ve gönüllülük esas alınmıştır. Veri analizi için SPSS programı ile basit uyum analizi yapılmıştır.

Araştırma sonucunda biyoloji ve fizik öğretmenlerinin program tasarımının, programı anlamayı kolaylaştırıcı buldukları, yenilenen programın eski programa göre daha iyi tasarlandığını belirttiği, kimya öğretmenlerinin ise bu görüşe katılmadığı söylenebilir.

Biyoloji programındaki etkinlik örnekleri sayıca yeterli bulurken matematik programındakilerin yetersiz bulunduğu gözlenmiştir. Biyoloji ve fizik öğretmenleri programın içeriğindeki sınırların açık ve net bir şekilde belirlendiğini düşünürken, kimya öğretmenleri bu görüşe katılmamaktadır.

Fizik ve biyoloji öğretmenleri kazanımların öğrencinin günlük yaşantısıyla bağ kurmasına yardımcı olduğunu ve yaparak-yaşayarak öğrenmeyi temele alan bir felsefe üzerine kurulduğunu düşünürken matematik ve kimya öğretmenleri bu görüşe katılmamaktadır.

Fizik, kimya ve biyoloji öğretmenleri kazanımlarda güncel bilgilerin yer aldığını düşünmektedir. Matematik, fizik ve biyoloji öğretmenleri programlarda gereksiz bilgi ve fazla ayrıntının bulunmadığını düşünürken kimya öğretmenleri fazla ayrıntılı olduğunu düşünmektedir.

Biyoloji öğretmenleri programda kullanılan kavramların sınıf seviyesine uygun olduğunu düşünürken kimya öğretmenleri uygun olmadığını, fizik ve biyoloji öğretmenleri öğretim programının sınıf seviyesine uygun olduğunu, matematik ve kimya öğretmenleri ise uygun olmadığını düşünmektedir. Matematik ve biyoloji öğretmenleri ise programın ders saatleriyle uyumlu olduğunu düşünürken kimya ve fizik öğretmenleri programlarının içeriğinin belirtilen ders saatleriyle uyumlu olmadığını düşünmektedir.

Fizik ve biyoloji öğretmenleri programın konu sıralamasının bilinenden bilinmeyene doğru olduğunu düşünürken kimya öğretmenleri bu görüşe katılmamaktadır.

Fizik, kimya ve biyoloji öğretmenleri programın ölçme ve değerlendirme yaklaşımını benimsediği ve programların güncel ölçme ve değerlendirme uygulamalarına uygun olduğunu düşündüğü gözlenmiştir.

Anahtar Kelimeler

Matematik, fizik, kimya ve biyoloji öğretim programı, uyum analizi, öğretmen görüşü, ortaöğretim

*Hacettepe Üniversitesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. İletişim: abdullahfarukkilic@gmail.com

**Hacettepe Üniversitesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. İletişim: cguzeller@gmail.com

Paper No: 19

Evaluation by Teachers' Views of Mathematics, Physics, Chemistry and Biology Curriculum

Abstract

The aim of this study is to evaluate mathematics, physics, chemistry and biology curriculum in terms of teachers' opinion. The population of study consist of mathematics, physic, chemistry and biology teachers working at Anatolian high school, science high school, social science high school, sport high school and fine arts high school. The sample of study is 5767 teachers who was chosen by stratified sampling and cluster sampling method. A questionnaire developed by researchers was used as data collection tool. The questionnaire has 15 items which is 5 point Likert scale. Data was collected via online forms. Honesty, sincerity, and volunteerism was based for participation. Correspondence analysis was conduct via SPSS for data analysis.

As a result of research, according to the views of biology and physics teachers, it has been observed that design of the biology and physics curriculum is facilitator to understand the curriculum. Biology and physics teachers also mentioned renewed curriculum was designed better than old curriculum. But chemistry teachers don't agree with this view.

According to view biology teachers activity examples in biology curriculum are sufficient in number but mathematics teachers found that the number of examples insufficient in mathematics curriculum. The boundaries of contents of the curriculum defined clearly according to teacher of this subjects but chemistry teachers think opposite in terms of same indicator in chemistry curriculum.

According to views of the physics and biology teachers, objectives in curriculum are supporting in terms of linking with daily life of students and objectives were constructed on the basis of learning-by-doing. Mathematics and chemistry teachers shared their opinion in opposite way.

Objectives of physics, chemistry and biology curriculum have up-to-date information according to these subjects' teachers. Chemistry teachers think their curriculum is over-detailed and has some unnecessary information. But mathematics, physics and biology teacher don't think that their curriculum have over detailed information.

While biology teachers think that concepts used in the curriculum are appropriate to the grade level, chemistry teachers have opposite opinion about this subject. Physics and chemistry teachers think that their curriculum are appropriate to the grade level but mathematics and chemistry teachers don't think like them. Mathematics and biology teacher think that their curriculum are appropriate for weekly course hours, but physics and chemistry teachers think in opposite way.

Physics and biology teacher think that order of concepts in curriculum are from the known to the unknown on the other hand chemistry teachers don't agree with them. It is observed that chemistry, physics and biology teachers adopt the measurement and evaluation approaches and think that those curriculum are appropriate for contemporary measurement and evaluation methods.

Key Words

Mathematics, physics, chemistry and biology curriculum, correspondence analysis, teacher's views, secondary education.

Bildiri No: 20

Güvenirlik Katsayılarının Kayıp Veri Varlığında İncelenmesi

Çiğdem AKIN ARIKAN[*], Sümeyra SOYSAL[**],Hatice İNAL[***]

Özet

Bu çalışmanın amacı farklı örneklem büyüklüğü, kategori sayısı, kayıp veri oranı ve kayıp veri atama teknikleri koşulları altında güvenirlik katsayı kestirimlerinin incelenmesidir. Bu doğrultuda örneklem büyüklüğü 500, 1000 ve 5000 olacak şekilde 20 maddeden oluşan veri seti üretilmiştir. Üretilen tam veri setlerinde tümüyle seçkisiz kayıp koşulu altında %5, %10,%20 ve %30 oranlarında silme işlemi gerçekleştirilmiştir. Silme işleminden sonra kayıp veri atama yöntemlerinden 0, regresyon, ortalama ve çoklu atama yöntemleri ile değer ataması yapılmıştır. Güvernirlik katsayıları olarak ise Cronbach α, Standartlaştırılmış α, Armor'un θ, Guttman Lambda 4, Guttman Lambda 5, Guttman Lambda 6 ve McDonald'ın ω güvenirlik katsayıları kullanılmıştır ve elde edilen kestirimler tam veri üzerinden elde edilen kestirimler ile karşılaştırılmıştır. Çoklu atama ve regresyon atama tekniklerinin birbirine benzer ve genel olarak yüksek performans gösterdiği, 0 atama tekniğinin ise en kötü performans gösterdiği sonucuna ulaşılmıştır. Ayrıca koşullar arasında fark olup olmadığı belirlemek için ANOVA testi yapılmıştır. Yapılan ANOVA analizi yalnızca McDonald'ın ω güvenirlik katsayısının örneklem büyüklüğüne göre anlamlı farklılık göstermediği, diğer güvenirlik katsayılarının çalışmada ele alınan tüm değişkenlere göre anlamlı farklılık gösterdiği bulunmuştur.

Anahtar Kelimeler

Güvenirlik katsayıları, kayıp veri, kayıp veri atama yöntemleri

[*]Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim: akıncgdm@gmail.com Tel: +0(312) 297-85 71

[**]Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim: sumeyrasoysal@hotmail.com Tel: +0(312) 297-85 71

[***]Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim: hinal@hacettepe.edu.tr Tel: +0(312) 297-85 71

Paper No: 20

Investigation Reliability Coefficients On Presence Of Missing Value

Abstract

The purpose of this study is to invastigate reliability coefficients under different sample size, number of choise, missing data percent and different missing data techniques. Fort his purpose, the data sets were generated 20 number of items for sample size 500, 1000 and 5000 respectively. The full data sets were deleted under missing completely at random condition by five, ten, twenty and thirty percent. After deleting data sets, missing data techniques 0, mean, regresion and multiple imputation were carried out on incomplete data sets. Reliability coefficients which used this study were Cronbach α, standardised α, Armor'un θ, Guttman Lambda 4, Guttman Lambda 5, Guttman Lambda 6 and McDonald'ın ω, and the reliability estimations were compared with the full data sets of the reliability estimations. Results show that regresion and multiple imputation techniques had similar and generally high performance results, but 0 impuations had low performance. Afterwards, ANOVA test were conducted in order to find out whether there were any difference between conditions. According to ANOVA test, only McDonald'ın ω reliability coefficient had not have any difference under sample size, but the others had.

KeyWords

Reliability coefficients, missing data, missing data techniques

Bildiri No: 21

Öğretmen Adaylarının Hazırladıkları Tanılayıcı Dallanmış Ağaçların Çok Yüzeyli Rasch Modeliyle Analizi

Funda NALBANTOĞLU YILMAZ*

Özet

Araştırmanın amacı, öğretmen adaylarının grupça hazırladıkları tanılayıcı dallanmış ağaçların çok yüzeyli Rasch modeli ile incelenmesidir. Araştırmanın çalışma grubunu, 2015-2016 eğitim öğretim yılı bir devlet üniversitesindeki eğitim fakültesi sınıf öğretmenliği bölümü öğrencilerinden ölçme ve değerlendirme dersini almakta olan 24 öğretmen adayı oluşturmaktadır. Öğretmen adayları ikişerli olarak grup oluşturmuştur. Böylece araştırma 12 çalışma grubu üzerinden yürütülmüştür. Öğretmen adaylarının hazırladıkları tanılayıcı dallanmış ağaçların puanlanmasında bir öğretim üyesi ve her gruptan gönüllü olarak seçilen birer öğretmen adayı kullanılmıştır. Hazırlanan tanılayıcı dallanmış ağaçların puanlanmasında 11 ölçütten oluşan puanlama formu kullanılmıştır. Puanlama formunun hazırlanmasında literatür taraması yapılmış, tanılayıcı dallanmış ağaç özellikleri de dikkate alınarak ilgili beceri tanımlanmış ve üründeki gözlenebilir ölçütler belirlenmiştir. Belirlenen ölçütler uygunluk, gereklilik, anlaşılırlık bakımlarından dört ölçme ve değerlendirme uzmanına ve her bir çalışma grubundan gönüllü olarak seçilen ve tanılayıcı dallanmış ile ilgili eğitim almış 12 öğretmen adayına incelettirilmiştir. Alınan görüşler doğrultusunda benzer olduğu belirtilen bazı ifadeler düzeltilmiş/çıkartılmıştır. Uzman ve öğretmen adayı görüşlerine göre kapsam geçerlik oranları ve kapsam geçerlik indeksi hesaplanmıştır. Puanlamadan önce form sınıfa tanıtılmış ve bir örnek puanlama yapılmıştır. Araştırma sonucunda, sekizinci ve dokuzuncu grubun tanılayıcı dallanmış ağaç hazırlama bakımından en yüksek, yedinci grubun ise en düşük yetenek seviyesine sahip olduğu belirlenmiştir. Öğretmen adayları kendilerine verilen bir konuda tanılayıcı dallanmış ağaç hazırlarken önermeleri genelden özele doğru sıralamada, tek bir ifadenin doğruluğunu ya da yanlışlığını ölçen önerme hazırlamada ve anlaşılır önermeler yazmada zorlanmışlardır. Puanlayıcıların katılık/cömertlikleri açısından aralarında anlamlı farklılıklar olduğu belirlenmiştir. Bazı puanlayıcıların puanlamalarındaki farklılığın yeterli olmadığı, bazılarının ise puanlamalarında tutarsızlıklar olduğu sonucuna ulaşılmıştır. Bu doğrultuda farklı bir araştırmada elde edilen bu verilerle puanlayıcılara ait puanlayıcı etkisinin araştırılması önerilmektedir.

Anahtar Kelimeler

Çok yüzeyli Rasch ölçme modeli, Tanılayıcı dallanmış ağaç

*Nevşehir Hacı Bektaş Veli Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:fundan@nevsehir.edu.tr Tel: +90(0384)2281000-21034

Paper No: 21

Analyzing Teacher Candidates' Diagnostic Tree With Many Facet Rasch Measurement Model

Abstract

The aim of the study is to investigate the prospective teachers' diagnostic tree with many facet Rasch model. The study group is comprised of 24 teacher candidates taking the measurement and evaluation course in a state university during the 2015-2016 education year. Teacher candidates formed the group in pairs. Thus, the study is conducted over 12 groups. A tutor and teacher candidates selected as volunteer from each group are used for scoring of the diagnostic trees. The scoring of the diagnostic tree was used rating scale consisting of 11 criteria. In the preparation of rating scale made a literature review, defined relevant skills and set the criteria. And then specified criteria analyzed the four measurement and evaluation experts and 12 teacher candidates with respect to suitability, necessity, clarity. Some criteria corrected/removed by all accounts. The content validity and content validity index was calculated according to the views. And before scoring rating scale was introduced to study group and made an example of scoring. According to stufy results; It was determined that the eighth and ninth groups with highest ability level and the seventh group with the lowest ability level in terms of preparing diagnostic tree. It was concluded that the teacher candidates face the difficulties in rank the true-false questions from the general to the specific and preparation true-false questions. It was determined to be significant differences between raters lenient/severity. It was found that some raters inconsistency. In this regard it is recommended to investigate the rater effects in a different study.

Key Words

Many facet Rasch model, Diagnostic tree

Bildiri No: 22

Bireyselleştirilmiş Bilgisayarlı Sınıflama Testlerinde Madde Havuzu Özelliklerinin Test Uzunluğu ve Sınıflama Doğruluğu Üzerindeki Etkisi

Ceylan GÜNDEĞER, Nuri DOĞAN

Özet

Bu çalışmada bireyselleştirilmiş bilgisayarlı sınıflama testlerinde (BBST) madde havuzu özelliklerinden dağılım ve büyüklüklerin ortalama test uzunluğu ve ortalama sınıflama doğruluğu üzerindeki etkisi incelenmiştir. Bu amaçla, sivri ve basık dağılımlı 50, 100, 200 ve 300 maddelik madde havuzlarında; tesadüfi madde seçme yöntemi (TMSY), Maksimum Fisher Bilgisi (MFB) ve Kullback-Leibler Bilgisi (KLB) yöntemleri incelenmiştir. 1000 bireye ait yetenek parametreleri -3,3 aralığında N(0,1) olacak şekilde türetilmiştir. Sivri dağılıma sahip madde havuzlarındaki maddelerin a parametresi U[0,5; 2,0] aralığından; b parametresi N(1, 0,4) ve c parametresi N(0,15, 0,05) şeklinde; basık dağılıma sahip madde havuzlarındaki maddeler ise a parametresi U[0,5; 2,0] aralığından; b parametresi N(1, 1,5) ve c parametresi N(0,15, 0,05) şeklinde türetilmiştir. BBST simülasyonu R'da gerçekleştirilmiş, sonuçlar araştırmacı tarafından yazılan fonksiyonlarla elde edilmiştir. Çalışma sonucunda tüm madde havuzlarında ortalama test uzunluğu bakımından en yüksek değerin TMSY'ye ait olduğu; MFB ve KLB yöntemlerinin birbirine oldukça benzer çalıştıkları söylenebilir. Madde havuzu büyüklüğü arttıkça test uzunluklarının kısaldığı; sınıflama doğruluklarının azaldığı ancak tüm koşullarda 0,90 üstünde yüksek sınıflama doğruluğu elde edildiği görülmüştür. Ayrıca sivri dağılıma sahip madde havuzlarında test uzunluğunun kısaldığı ve test etkililiğinin arttığı; sınıflama doğruluklarının ise değişmediği görülmüştür. Bu sonuçlar dikkate alındığında, BBST'de çok sayıda maddeden oluşan sivri dağılıma sahip madde havuzları ile yüksek sınıflama doğruluğuna sahip daha kısa testlerin oluşturulabileceği söylenebilir.

Anahtar Kelimeler

Bireyselleştirilmiş Bilgisayarlı Sınıflama Testleri, Madde Havuzu Dağılımı, Madde Havuzu Büyüklüğü, Test Uzunluğu, Sınıflama Doğruluğu

Paper No: 22

The Effects of Item Pool Characteristics on Test Length and Classification Accuracy in Computerized Adaptive Classification Testings

Abstract

In this study it was investigated that the effects of distrubitions and sizes on avarage test length and avarage classification accuracy in computerized adaptive classification testings (CACT). For that purpose random item selection method (RISM), Maximum Fisher Informatiıon (MFI) and Kullback-Leibler Information (KLI) were studied in boared and peaked item pools with 50 items, 100 items, 200 items and 300 items, Thetas are derived from N(0,1). In peaked item pools items are simulated from U[0,5; 2,0] for a parameters, N(1, 0,4) for b parameters and N(0,15, 0,05) for c parameters; and in broad item pools items are simulated from U[0,5; 2,0] for a parameters, N(1, 1,5) for b parameters and N(0,15, 0,05) for c parameters. The simulation was performed in R and the results were obtained with the functions which are written by the researcher. The study results show that RISM has the maximum value with respect to avarage test length; and MFI and KLI perform similar. The more items in the pool, the shorter test length and fewer the classification accuracy but in all conditions classification accuracy has high rate above 90%. In addition, in peaked item pools it is seen that the avarage test lengths are getting shorter and the test effectiveness is getting higher; but the classification accuracies are not changing. In conclusion it can be said that with the peaked item pools with more items, BBST provides shorter tests and high classification accuracy.

Key Words

Computerized Adaptive Classification Testing, Item Pool Distrubition, Item Pool Size, Test Length, Classification Accuracy

Bildiri No: 23

İkili Puanlanan Ölçme Sonuçlarındaki Kayıp Veri Sorununun Çözümünde Kullanılan Çoklu Değer Atama Yöntemlerinin Karşılaştırılması

Duygu Gizem ERTOPRAK*

Özet

Eğitim alanında yapılan ölçme işlemlerinde, gözlenen verilerden hareketle gizil değişkenlere yönelik çıkarımların yapılması amaçlanır. Gözlenen verilerin önemi dikkate alındığında, veri toplama sürecinde sıklıkla karşılaşılan ve gizil değişkenlere ait kestirimlerin gücünü azaltan kayıp veri sorunu ile başa çıkılması gerekir. Çoklu değer atama (ÇDA), bir veri setinde bulunan kayıpların yerine geçecek uygun değerlere sahip birden çok tam veri seti oluşturma temeline dayanan ve farklı değer atama yöntemleri bulunan bir başa çıkma yöntemidir. Bu çalışmanın amacı, doğru ya da yanlış yanıta sahip ikili puanlanan maddelerin meydana getirdiği veri setlerindeki kayıp veri sorununun çözümünde kullanılan farklı ÇDA yöntemlerini karşılaştırmaktır. Bu amaç doğrultusunda, MTK'nın 3 pl modeline uygun olacak şekilde, 2000 bireyin 100 maddeye verdiği ikili puanlanmış yanıtları içeren tam bir veri seti türetilmiştir. Sonuçların genellenebilirliğini arttırmak amacıyla, her bir koşul altında türetilen veri setlerine ait 20 tekrar yapılmıştır. Sonraki aşamada, tam veri setlerinden; tamamen rastgele kayıp (TRK), rastgele kayıp (RK) ve rastgele olmayan kayıp (ROK) türleri altında, %10 ve %30 oranlarında veri silme işlemleri gerçekleştirilerek eksik veri setleri elde edilmiştir. Eksik veri setlerinin yeniden tam veri setlerine dönüştürülmesi işlemleri için zincirleme denklemler ile çok değişkenli atama (ZDÇDA) algoritması kapsamında uygulamaları gerçekleştirilen lojistik regresyon (LR), diskriminant fonksiyonu (DF) ve tahmine dayalı eşleştirme (TDE) adlı değer atama yöntemleri kullanılmıştır. Verilerin türetilmesi için R yazılımının "irtoys", parametre kestirimleri için "ltm", kayıp verilere yönelik atama işlemleri için ise "mice" paket programlarından yararlanılmıştır. ÇDA yöntemlerinin performansı, her bir simülasyon koşuluna ait kestirilen parametreler ile tam veri setlerine ait gerçek parametrelerin farklarını temele alarak hesaplanan BIAS ve RMSE değerleri üzerinden karşılaştırılmıştır. İncelenen tüm koşullar dikkate alındığında, en düşük $BIAS_{ort}$ değerine sahip yöntem LR (.002) iken en yüksek DF (.013); en düşük $RMSE_{ort}$ değerine sahip yöntem DF (.642) iken en yüksek LR (.657) çıkmıştır. Sonuçta, TDE yönteminin bu tip veri setleri üzerindeki kullanımının en az DF yöntemi kadar uygun olduğu söylenebilir.

Anahtar Kelimeler

Kayıp veri, Çoklu değer atama, Zincirleme denklemler ile çok değişkenli atama, Tahmine dayalı eşleştirme

*Amasya Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: duygugizemertoprak@gmail.com
Tel: +903582526230 (Dahili:3332)

Paper No: 23

Comparison of Multiple Imputation Methods Used to Handle Missing Data Problem in Binary Data Set

Abstract

The main aim is to make inferences about latent variables on the basis of observed data in educational measurement process. Missing data problem which reducing the power of estimation is needed to handle when the importance of observed data is taken into consideration. Multiple imputation (MI) generates independent multiple copies data sets which contain plausible values instead of missing values and it has different imputation methods. The main purpose of this study is to compare different multiple imputation methods used to handle missing data problem in binary data set which has correct or wrong answer. Moving on this purpose, a binary data set consists of 100 item response given by 2000 person based on IRT 3-parameter model was created. 20 replications were applied on data sets which created under every simulation condition to increase generalizability. Then, missing data sets were obtained by deleting 10% and 30% proportion data from complete data sets under missing completely at random (MCAR), missing at random (MAR), missing not at random (MNAR) mechanism. Logistic regression (LR), discriminant function (DF) and predictive mean matching (PMM) imputation methods were used to reproduce complete data sets from missing data sets by using multivariate imputation by chained equations (MICE) algorithm. R-software "irtoys" package was used to create data sets, "ltm" package was used to estimate ability and item parameters, "mice" package was used to impute value instead of missing data. The performance of different multiple imputation methods were compared with BIAS (bias) and RMSE (root mean squared error) values which calculated based upon the difference between true parameter values and estimated parameter values. While LR method had the lowest BIASmean value (.002), DF method had the largest (.013) and DF method had the lowest RMSEmean value (.642), LR method had the largest (.657) when all simulation conditions were considered. Consequently, it can be said that the use of PMM method to handle missing data problem on binary data set is suitable as DF method as.

Key Words

Missing data, Multiple imputation, Multivariate imputation by chained equations, Predictive mean matching

Bildiri No: 24

Tek-Biçimli-Olmayan Değişen Madde Fonksiyonunun Belirlenmesinde Lojistik Regresyonun Performansı

Devrim ERDEM-KEKLİK*

Özet

Bir testin geçerliğini tehdit eden en ciddi faktörlerden biri yanlılıktır. Yanlılık belirleme sürecinde öncelikle incelenmesi gereken maddenin değişen madde fonksiyonu (DMF) içerip içermediğidir. DMF, tek biçimli ve tek biçimli olmayan şekilde ortaya çıkmaktadır. Literatürde tek-biçimli DMF belirleme tekniklerinin performansını inceleyen pek çok araştırma olmasına rağmen, tek-biçimli-olmayan DMF araştırmaları nispeten daha azdır. Bu nedenle, bu araştırmanın amacı tek biçimli olmayan DMF belirlemede, gerçek test koşullarında ortaya çıkabilecek referans ve odak grup yetenek dağılımı ve örneklem büyüklüğünün çeşitli koşulları altında, iki kategorili maddelerde lojistik regresyon (LR) tekniğinin performansını incelemek şeklinde ortaya konmuştur.

Bu araştırmada, 20 maddelik iki kategorili benzetim verileri 2PL modele göre WinGen3 programı kullanılarak elde edilmiştir. İlk 17 madde DMF içermeyecek biçimde ve son üç madde tek-biçimli olmayan DMF içerecek biçimde üretilmiştir. Yetenek dağılımı, örneklem büyüklüğü ve örneklem büyüklüğü oranları manipüle edilen değişkenlerdir. Her bir koşul için 100 tekrar yapılmıştır.

Araştırmada ortaya çıkan bulgulara göre, odak ve referans grupların yetenek dağılımları eşit ve örneklem büyüklüğü oranları 1:1 iken LR I. Tip hata oranları nominal α (.05) düzeyindedir. Ancak, bu koşullarda istatistiksel güç gözlenen en düşük düzeyde seyretmiştir. Odak ve referans grup dağılımları eşit ancak örneklem büyüklüğü oranları 1:2 olduğunda LR I. Tip hata oranları nominal α düzeyinin üstüne çıkmıştır.Odak ve referans grup dağılımları eşit olmadığı koşullar altında ise LR I.Tip hata oranlarında ve istatistiksel güç değerlerinde büyüme (şişme) meydana geldiği gözlenmiştir.

Anahtar Kelimeler

Değişen madde fonksiyonu, Lojistik regresyon, Benzetim, I Tip hata oranı.

*Niğde Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: erdem_devrim@yahoo.com Tel:+903882254367

Paper No: 24

Type I Error Rate and Power of Logistic Regression in the Detection Non-uniform Differential Item Functioning

Abstract

Bias is one of the major issues that can threaten a test's validity. First stage of the bias detection procedure is the statistical detection of differential item functioning (DIF). There are two types of DIF: uniform and nonuniform DIF. Although performance of uniform DIF detection methods has been well studied, that of nonuniform DIF detection methods necessitates further examination. Thus the purpose of the current study was to investigate and compare the performance of the logistic regression procedure in detecting nonuniform differential item functioning (DIF) with dichotomous item responses under a variety of conditions using simulated data.

Two-parameter logistic item response model was used to generate for a 20-item test using WinGen3. The first 17 items were always free of DIF and the last three items were the studied items, and had nonuniform DIF. Factors manipulated were ability distributions differences between focal and reference groups, sample size (three levels; 600, 1200, 2400) and sample size ratio. Each condition was replicated 100 times.

Based on the findings, under all the conditions where groups were equal in ability and sample size ratio was 1:1, TypeI error rates of LR were at nominal α (.05) level. Meanwhile, under these conditions power of LR had the lowest level. When groups were equal in ability but sample size ratio was 1:2, TypeI error rates of LR exceeded the nominal level and took values between 0.65 and 0.103. When groups were unequal in ability and became more skewed distribution of focal group, inflated TypeI error rates were observed. Furthermore, TypeI error rates of LR procedure increased as sample size increased with the exception of the F:R=1:2 condition when the ability distributions were the same. Finally, results showed that increases in the ability differences between groups corresponded with more inflated TypeI error and power rates.

Key Words

Differential item functioning, Logistic regression, Simulation, Type I error rate.

Bildiri No: 26

Eğitim, Ekonomi ve İnsan Arasındaki Sinerji

Tülin ACAR

Özet

Bu araştırmanın amacı, katılımcıların eğitim uygulamalarına ilişkin yapmış oldukları sıralamalar arasındaki algılanan büyüklüğün belirlenmesidir. Katılımcılara göre acil çözülmesi, reform yapılması ya da tartışılması ve değerlendirilmesi gereken eğitim uygulamalarının önem derecesine göre sıralama işlemi "sıralama yargılarıyla ölçekleme yöntemi" ile analiz edilmiştir. Katılımcıların eğitim politikalarına ilişkin ilk üç sıralamasının kız çocuklarının okullaşması (s_j=0,999), öğretmen niteliği (s_j=0,918) ve müfredat programlarının içeriği (s_j=0,748) olduğu gözlenmiştir. Akıl düzleminde paydaşlarla yeterince tartışılmadan reform uygulamaktan vaz geçilmesi gerekmektedir.

Anahtar Kelimeler

Eğitim politikaları, Eğitim sorunları, Sıralama yargılarıyla ölçekleme

Paper No: 26

The Synergy among Education, Economy and Human Beings

Abstract

The aim of the study is to determine metric size in the scaling the participants with regard to educational system. Ranking was analyzed through the method of rank order judgment scaling. Participants answered are ranked as education of female student, the quality of teachers and content of curriculum respectively. It is suggested that reforms should be implemented without a through discussion among the stakeholders involed.

Key Words

Education policy, Educational issues, rank order judgment scaling

Bildiri No: 27

Kategori Sayısının Psikometrik Özellikler Üzerine Etkisinin Parametrik Olmayan Madde Tepki Kuramı Modeline (Mokken Model'e) Göre İncelenmesi

Asiye ŞENGÜL AVŞAR*

Özet

Araştırmanın amacı kategori sayısının psikometrik özellikler üzerindeki etkisinin parametrik olmayan madde tepki kuramı (POMTK) modeli ile belirlenmesidir. Belirlenen amaç doğrultusunda iki farklı test uzunluğu (10 ve 30 madde), üç farklı sayıda kategoriye (3, 5 ve 7) sahip maddeler olmak üzere, altı test koşulunda 20 tekrar yapılarak simülatif veriler üretilmiştir.

Kategori sayısının psikometrik özellikler üzerindeki etkisi POMTK modellerinden Mokken Homojenlik Modeli (MHM) ile araştırılmıştır. Yapılan araştırma temel araştırma olarak tasarlanmıştır. Verilerin üretiminde WinGen3 programı, MHM'ye göre yapılan analizlerde R 3.1.3 programı kullanılmıştır.

MHM'ye göre yapılan ölçekleme sonucunda kategori sayısının testlerin psikometrik özelliklerini etkilediği görülmüştür. Genel olarak hem kısa testlerde, hem de uzun testlerde kategori sayısı artıkça geçerlik ve güvenirlik kestirimlerinin daha yüksek yapıldığı görülmüştür. Ancak belirlenen test koşullarında en ideal koşula beş kategoriden oluşan maddelerden ulaşılmıştır.

Anahtar Kelimeler

Çok kategorili puanlanan maddeler, Kategori sayısı, Mokken Homojenlik Modeli

*Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:asiye.sengul@erdogan.edu.tr Tel: +464 5328454/2382

Paper No: 27

Analysis of Number of Category Effects on Psychometric Properties According to Nonparametric Item Response Theory Model (Mokken Model)

Abstract

The aim of the research is to analyze category effects on psychometric properties by nonparametric item response theory (NIRT) model. According to do this aim simulated data sets which are two different test lengths (10 and 30 items), three number of category (3, 5 and 7) generated by 20 replication in six test conditions.

Category effects on psychometric properties of polytomous items are searched by NIRT model which is Mokken Homogeneity Model (MHM). The research is designed as a basic research. In the generation of data WinGen3 program, for analysis made according to MHM R 3.1.3 program were used.

According to scaling MHM, it is seen that number of category affects psychometric properties of tests. In general both of short and long tests, as the number of category increased prediction of validity and reliability are increased. However, under the test conditions it is determined to be the most ideal condition is consist from five category.

Key Words

Polytomous items, Number of category, Mokken Homogeneity Model

Bildiri No: 28

Öğrencilerin Matematik Kaygısı Üzerinde Etkili Olan Değişkenlerin Belirlenmesi: PISA 2012 Türkiye Örneği

Nermin KIBRISLIOĞLU UYSAL*

Özet

Araştırmanın amacı PISA 2012 uygulaması Türkiye örnekleminde öğrencilerin matematik kaygıları üzerinde etkili olan faktörlerin belirlenmesidir.Araştırma kapsamında matematik kaygısıyla ilişkisi incelenen PISA 2012 öğrenci anketi B formunda yer alan, sosyoekonomik düzey, matematik öz yeterliği, matematik benlik algısı, başarısızlığı atfetme ve matematik ilgisi indeks değişkenleri ele alınmıştır. Araştırmada söz konusu değişkenlerin matematik kaygısı üzerinde etkili olup olmadığını belirlemek amacıyla CHAID sınıflama analizi kullanılmıştır. Araştırmanın sonuçları öğrencilerin matematik kaygıları üzerinde en etkili olan değişkenin benlik algısı değişkeni olduğunu ve onu sırasıyla başarısızlığı atfetme ve matematik ilgisi değişkenlerinin takip ettiğini göstermektedir. Sosyoekonomik düzey ve öz yeterlik değişkenleri ise elde edilen karar ağacında yer almadığı görülmektedir.

Anahtar Kelimeler

Matematik kaygısı, CHAID, PISA 2012

*Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: nkibrislioglu@hacettepe.edu.tr Tel: 05543900302

Paper No: 28

The Investigation of Variables Which Affect Students' Mathematics Anxiety: PISA 2012 Turkey Case

Abstract

The purpose of this study is to investigate factors affecting students' mathematics anxiety in the case of PISA 2012 Turkey sample. Socio-economic status, mathematics self-efficacy, mathematics self-concept, attribution to failure and mathematics interest index variables placed in student survey B form were taken into consideration. In order to determine whether these variables are influential on mathematics anxiety, CHAID classification analysis was conducted. The results of the study indicated that most influential variable on mathematics anxiety was mathematics self-concept and it continues as attribution to failure, and mathematics interest in order. The socio-economic status and mathematics self-efficacy variables were not placed on decision tree as an influential variable

Key Words

Mathematics anxiety, CHAID, PISA 2012

Bildiri No: 30

Temel Eğitimden Ortaöğretime Geçiş Asıl Sınav ve Mazeret Sınavı Matematik Alt TestlerininMadde Tepki Kuramına Göre Eşitlenmesi

Emrah GÜL[*], Çilem DOĞAN GÜL[**],Ömay ÇOKLUK[***], Mustafa ÖZKAN[****]

Özet

Bu araştırmanın amacını TEOG asıl sınav matematik alt testi ile TEOG mazeret sınavı matematik alt testlerinin Madde Tepki Kuramı'na (MTK) dayalı yöntemler kullanılarak eşitlenmesi ve hangi eşitleme yönteminin (ortalama-ortalama, ortalama-sigma, Haebara ve Stocking-Lord) daha uygun olduğuna karar verilmesi oluşturmaktadır. Burada amaçlanan en az hatalı sonuçlar üreten yöntem ve koşulların belirlenmesi olduğundan, bu yönüyle araştırmanın temel araştırma niteliği taşıdığı ifade edilebilir. Araştırmanın çalışma grubunun oluşturulması için öncelikle 2013-2014 eğitim-öğretim yılı I. yarıyılında TEOG kapsamında uygulanan merkezi ortak sınavlardan matematik mazeret sınavını alan öğrenci sayısı belirlenmiştir. Bu sayının 3392 olduğu, bunlar içerisinden A kitapçığını alan öğrenci sayısının ise 1747 olduğu saptanmıştır. Bu belirlemenin ardından çalışma grubunu mazeret sınavında A kitapçığını alan 1747 öğrenci ve buna denk olarak asıl sınav uygulamasında A kitapçığını alanlar arasından seçkisiz olarak belirlenen 1747 olmak üzere toplam 3494 öğrenci oluşturmuştur. Veriler MTK'ya dayalı olarak analiz edildiğinden öncelikle kuramın varsayımları R paketlerinden "polycor" ve "ltm" kullanılarak incelenmiştir. Testlerin eşitlenmesi için, madde ve birey parametreleri flexMIRT programı kullanılarak kestirilmiştir. Eşitleme işlemi R "equateIRT" paketi kullanılarak yapılmıştır. Araştırmadan elde edilen sonuçlar, en düşük eşitleme hatasının karakteristik eğri yöntemlerinden elde edildiğini ortaya koymaktadır. Bu yöntemlerden de en düşük eşitleme hatasının Stocking-Lord yönteminden elde edildiği belirtilebilir. Ortalama-Ortalama yöntemi ise en yüksek eşitleme hatası üreten yöntemdir. Bu araştırmadan elde edilen bulgular Stocking-Lord yönteminin TEOG asıl sınav ve mazeret sınavı kapsamında yer alan matematik alt testlerinin eşitlenmesinde daha uygun olduğunu göstermiştir. Her yıl testlerin farklı formlarının geliştirilerek kullanıldığı sınavlarda testlerin eşitliğinin sağlanması öğrenciler açısından çok önemlidir. Bu çalışmadan elde edilen bulgular Stocking-Lord yönteminin en az hata üreten yöntem olduğunu ortaya koymuş olsa da sınav uygulayıcılarına öncelikle testlerin eşitliklerini sınamaları ve bu sınamalar sırasında farklı eşitleme yöntemlerini kullanmaları önerilir. Böylelikle sınavların bireyler arasında herhangi bir yanlılığa neden olmasının da önüne geçilmiş olur.

Anahtar Kelimeler

TEOG, Test Eşitleme, Madde Tepki Kuramı, Mazeret Sınavı

[*]Hakkâri University, Faculty of Education, Department of Primary Education. Contact info: emrahgul@hakkari.edu.tr Phone: +904382121212

[**] Hakkâri University, Faculty of Education, Department of Educational Sciences. Contact info: cilemdogan@hakkari.edu.tr Phone: +904382121212

[***]Ankara University, Faculty of Educational Sciences, Department of Measurement and Evaluation. Contact info:cokluk@education.ankara.edu.tr Phone: +903123633350

[****]Ministry of National Education, General Directorate of Lifelong Learning. Contact info: mustafaozkan@meb.gov.tr Phone:+903124132111

Paper No: 30

Equating of Maths Subtests In Transition From Primary To Secondary Education (TEOG) And Make-Up Examination According To Item Response Theory

Abstract

The purpose of the study is to equating maths subtests of TEOG and make-up examination, using Item Response Theory (IRT) based methods and to decide which equating method (mean-mean, mean-sigma, Haebara and Stocking-Lord) is appropriate. It could be suggested that it is a pure research as it aims to explore the best methods and conditions with minimum error outcomes. First, the number of students taking the maths test of TEOG make-up exam, included in TEOG common examinations (2013-2014: Fall semester) in order to form the research group. The number was found 3392, and the number of those with Booklet A was found 1747. After number determination, the research group consisted of 3494 students in total, randomly chosen among the ones with booklet A in the make-up examination (1747 students) and the Booklet A owners in TEOG examination (1747 students). “polycor” and “ltm” (of R packets) were used to study the theoretical assumptions of IRT because the data were analysed based on IRT. For test equalization, item-person parameters were estimated, using flexMIRT programme. R “equateIRT” pack was employed for the equating. The research results have shown that the minimum number of equating errors is obtained from characteristic curve methods. It could also be suggested that the Stocking-Lord method gives the lowest equating error. The mean-mean method produces the highest equating error. The findings of the study have shown that, when compared to the others, the Stocking-Lord method is more appropriate to equate the maths subtests of TEOG and make-up examinations. It is critical for students to ensure test equating in examinations in which various test forms are designed and applied every year. Test appliers are advised to test equating and use different methods of equating during the testing process although the study has found the Stocking-Lord method with the lowest equating error. Thus, examinations are prevented from leading to any bias of individuals.

Key Words

TEOG, Test Equating, Item Response Theory, Excuse Exam

Bildiri No: 34

Görme Engelli Üniversite Öğrencilerinin Sınıf İçi Ölçme-Değerlendirme Süreçlerinin İncelenmesi

Ömer KAMIŞ*, Ergül DEMİR**

Özet

Bu araştırmanın amacı, 2015/2016 eğitim öğretim döneminde Ankara ilindeki üniversitelerin çeşitli fakülte ve bölümlerinde öğrenim gören görme engelli öğrencilerin sınıf içi ölçme-değerlendirme süreçlerini öğretim elemanı ve öğrenci görüşlerine göre incelemektir. Araştırma nitel araştırma yaklaşımıyla yürütülmüştür. Çalışma grubunun oluşturulmasında kartopu örnekleme yöntemi kullanılmıştır. Çalışma grubunu 17 görme engelli öğrenci ile 3 öğretim üyesi oluşturmaktadır. Araştırmanın verileri, araştırmacılar tarafından hazırlanan yarı yapılandırılmış görüşme formu ile toplanmıştır. Verilerin analizinde doküman analizi yöntemi kullanılmıştır. Araştırmanın bulgularına göre, görme engelli öğrenciler ölçme-değerlendirme süreçlerinde zaman, uygulama, soru ve materyal anlamında güçlükler yaşamaktadır. Görme engelli öğrencilerin öğretim elemanlarından ve üniversite yönetiminden çeşitli beklentileri vardır. Benzer şekilde öğretim üyeleri de görme engellilerin tüm süreçleri için ayrı bir planlama yapılması ve materyal sağlanması gerektiği yönünde görüş belirtmişlerdir.

Anahtar Kelimeler

Görme engelli öğrenci, sınıf içi ölçme-değerlendirme, nitel araştırma

*Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: okamis@ankara.edu.tr Tel: 0312 363 33 50 / 2042

**Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: erguldemir@ankara.edu.tr Tel: 0312 363 33 50 / 3122

Paper No: 34

Investigation of the Classroom Assessment and Evaluation Process of Visually Impaired University Students

Abstract

The aim of this research is examine the classroom assessment and evaluation process of visually impaired students at the studying from various faculties and departments of university in Ankara according to the students and instructors opinions. The research was conducted with qualitative research approach. To determine of study groups, purposively sampling and snowball sampling method was used. The study group composed of 17 visually impaired students and 3 instructors. The data of research were obtained from semi-structured interview form prepared by the researchers. Document analysis method was used to analyze data. According to the research findings, visually impaired students are experiencing difficulties in terms of time, application, question and material in assessment and evaluation process. There are several expectations of visually impaired students from the instructors and university management. Similarly, the instructors expressed their opinions that a separate planning of the whole process of visually impaired students and material should be provided.

Key Words

Visually impaired student, classroom assessment and evaluation, qualitative research

Bildiri No: 35

Çok Boyutlu Testlerde Alt Boyut ve Toplam Puanların Kestirimine İlişkin bir Karşılaştırma

Eren Halil ÖZBERK*, Selahattin GELBAL**, Terry ACKERMAN***

Özet

Bu çalışmanın amacı çok boyutlu yapılarda madde sayısına ve kalibrasyon yöntemlerine göre desenlemiş testlerde alt boyut ve toplam puanların kestirimlerini karşılaştırmaktır. Bu amaçla üç boyutlu basit ve karmaşık yapıda 30 ve 45 maddeden oluşan madde havuzları oluşturulmuştur. Boyutlar arasındaki korelasyon değerleri. 2, .5 ve .8 olarak alınmıştır. Madde bankasının kalibrasyonu için tek boyutlu ayrı kalibrasyon ile çok boyutlu Bock ve Aitkin's EM ile Metropolis-Hastings Robbins-Monro algoritmaları kullanılmak üzere üç farklı kalibrasyon yöntemi seçilmiştir. Madde parametreleri a ~ LN{0, 0.2), b ~ N(0, 1), ve c ~ Beta{6,16) dağılımlarına göre simule edilmiştir. Çok boyutlu test yapısı desenleri için gerçek yetenek parametreleri ise MVN(0, ∑). koşulunu sağlamak üzere çok değişkenli normal dağılımdan örneklenmiştir. Araştırmanın sonuçları incelendiğinde madde sayısı arttıkça puanlar daha az hata ile kestirilmiştir ve puanların tutarlığı artmıştır. Alt boyut puanlarının aksine, toplam puanlarda testin karmaşıklığı arttıkça korelasyonların arttığı belirlenmiştir. B alt boyut puanlarının test karmaşıklığına (özelliklerin karmaşıklaşmasına) ve çok boyutluluğa karşı daha hassas olduğu belirlenmiştir. Toplam puanlar ise sadece çok boyutluluğa karşı hassasiyet göstermiştir. Araştırmadan elde edilen diğer önemli bulgu da kalibrasyon yöntemlerine göre farklılıkların oluşmasıdır. MHRM yöntemi tüm koşullarda diğer iki kalibrasyon yönteminden daha kötü performans göstermiştir. Alt boyut puanlarına bakıldığında basit yapılı testlerde BAEM yöntemi tek boyutlu ayrı kalibrasyona göre kısmen daha iyi performans sergilemiştir. Testin karmaşıklığı ve çok boyutluluğu azaldığında tek boyutlu ayrı kalibrasyon BAEM yöntemine yakın hatta daha iyi sonuçlar vermiştir. Toplam puanlarda çok boyutluluk özellikle karmaşık yapılı testlerde çok az etkiye sahipken, basit yapılı testlerde önemli bir etkiye sahip olduğu görülmüştür.

Anahtar Kelimeler

Çok boyutlu testler, kalibrasyon, toplam puan

*Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. erenozberk@gmail.com
**Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. sgelbal@gmail.com
***University of North Carolina Greensboro, Educational Research and Methodology. taackerm@uncg.edu

Paper No: 35

A Comparison of Domain and the Overall Ability Score Estimations in Multidimensional Test Designs

Abstract

The purpose of this study is to compare domain and the overall ability estimation accuracy in with respect to MIRT test design and calibration strategies. A three-dimensional 30-item and 45-item item pool was simulated with simple and complex structure. Dimensions correlated at ρ = .2, .5, and .8. Three calibration strategies, separate unidimensional and two multidimensional (Bock and Aitkin's EM and Metropolis-Hastings Robbins-Monro algorithm) calibration were examined. Item parameters were generated from a ~ LN{0, 0.2), b ~ N(0, 1), and c ~ Beta{6,16) distributions. For multidimensional test structure design 1000×3 matrix of true ability parameters were randomly generated from the multivariate normal distribution MVN(0, $\sum$). As expected, longer test provides more accurate scores. Opposite to the domain scores, correlations increase in overall ability score when the complexity of test increases. The reason for that domain scores are more sensitive to test complexity (trait contamination) and multidimensionality. However, overall scores are only sensitive to multidimensionality. Another significant finding following previous statements was the inequality of calibration approaches. MHRM approach performed slightly worse than other calibration approaches for all the conditions. The results of domain scores show that, BAEM performs slightly better than SU in simple structured test design and for all other criteria. As the test gets more complex, and multidimensionality decreases, SU calibration perform better or similar than BAEM. For overall scores, even multidimensionality has a small effect on overall ability score recovery on complex structured tests, it has a considerable effect on simple structured tests.

Key Words

Multidimensional item response theory, calibration, composite score

Bildiri No: 36

Merkezi Ortak Sınav Matematik Alt Testinde Değişen Madde Fonksiyonunun Görme Engeli Durumuna Göre İncelenmesi

Hatun Betül OZARKAN*, Emre KUCAM**,Ergül DEMİR***

Özet

Bu çalışma, 2015-2016 Eğitim Öğretim yılı I.Dönem Merkezi Ortak Sınavı matematik alt testi maddeleri için değişen madde fonksiyonunun, bireylerin görme engel durumuna göre incelenmesi amacıyla yapılmıştır. Merkezi Ortak Sınavlar altı temel ders için 8.sınıfta öğretmen tarafından dönemsel olarak yapılan sınavlardan, iki yazılısı olan derslerden birincisi, üç yazılısı olan derslerden ikincisi olmak üzere yılda iki defa yapılmaktadır. Ortak sınavlardan alınan puanlar öğrencinin derse ait yazılı notlarından birisi yerine geçerek yılsonu başarı puanını etkilediği gibi ortaöğretim kurumlarına yerleştirmeye esas puana da %70 oranında katkı sağlamaktadır. Değişen madde fonksiyonu belirleme yöntemlerinden Mantel-Heanszel yöntemi kullanılmıştır. Çalışma, 583'ü hiçbir engel durumu olmayan, 282'si hafif düzeyde görme engeli olan ve 152'si hiç görmeyen olmak üzere 1017 öğrenci üzerinden yürütülmüştür. Matematik alt testini alan, hiç görmeyen ve hafif düzeyde görme engeli olan öğrencilerin tamamı çalışmaya dâhil edilmiştir. Hiçbir engel durumu olmayan öğrenci sayısının ise görme engel durumu olan öğrencilerin sayısına yakın olmasına dikkat edilmiştir. Matematik alt testinde dört seçenekli çoktan seçmeli yirmi madde bulunmaktadır. Verilerin tek boyutluluk varsayımını sağlayıp sağlamadığını kontrol etmek amacıyla tetrakorik korelasyona dayalı açımlayıcı faktör analizi uygulanmıştır. Analiz sonucu 14 madde tek boyutlu bir yapıyı temsil etmiştir. Bu 14 madde ile Mantel-Heanszel yöntemine dayalı değişen madde fonksiyonu belirleme analizleri yapılmıştır. Araştırmanın sonucunda, Matematik alt testi; a) görme engeli olan ve olmayan bireylere göre b) görme engeli olmayan ve hiç görmeyen bireylere göre c) görme engeli olmayan ve hafif düzeyde görme engeli olan bireylere göre d) hiç görmeyen ve hafif düzeyde görme engeli olan bireylere göre B ve C düzeyinde değişen madde fonksiyonu içermediği belirlenmiştir.

Anahtar Kelimeler

Değişen madde fonksiyonu, Mantel-Haenszel yöntemi, Merkezi Ortak sınav, Görme engel durumu

*Öğretmen. Milli eğitim Bakanlığı, Ankara, hbetulcan@gmail.com
**Ölçme Değerlendirme Uzmanı, Zafer Koleji, Ankara, emrekucam@gmail.com
***Yrd. Doç. Dr., Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara, erguldemir@ankara.edu.trTel:+90 312 3633350

Paper No: 36

An Investigetion of Differential Item Functioning According to the Visually Handicapped Situation for the Central Joint Exam Math Subtest

Abstract

The aim of this study is to investigate differential item functioning (DIF) according to the visually handicapped situation of the individual for the math subtest in 2015-2016 academic year 1st term Central Joint Exam. Mantel-Haenszel (MH) method was used to determine the DIF. Study was carried out with 1017 students, including 583 without handicap; 282 with a low level of visually handicapped and 152 with total visual handicapped. There are four choices in mathematics subtest 20 multiple-choice item. Tetrachoric correlation based on a factor analysis was performed in order to check that the data satisfies the one-dimensional assumption. At the end of the factor analysis, 14 items represented a one-dimensional structure. As a result of the research, math subtest a) according to individuals with visually handicapped and without handicapped b) according to individuals with total visual handicapped and without handicapped c) according to individuals with a low level of visually handicapped and without handicapped d) according to individuals with total visual handicapped and a low level of visually handicapped was not found DIF in any of the items at B and C level.

Key Words

Differential item functioning, Mantel-Haenszel method, Central Joint Exam, Visually handicapped situation

Bildiri No: 37

Türkiye ve Amerika'da Engelli Öğrenciler için Yapılan Geniş Ölçekli Sınavların Yasal Sorumluluklar ve Uygulama Yöntemleri Açısından İncelenmesi

Gökhan AKSU[*], Derya ÇOBANOĞLU AKTAN[**], Mehmet Taha ESER[***]

Özet

Bu çalışmada Türkiye ve Amerika'da engelli oldukları belirlenen öğrenciler için yapılan geniş ölçekli sınavların dayandıkları yasal sorumluluklar ve uygulama yöntemleri açısından karşılaştırılması amaçlanmıştır. Bu amaç kapsamında Türkiye'de ÖSYM tarafından yapılan sınavlar ile Amerika'da ETS tarafından gerçekleştirilen büyük ölçekli sınavlarda uygulanan yöntemlerin belirlenerek iki ülke arasındaki benzer ve farklı yönler ortaya çıkarılmıştır. Çalışma sonucunda Amerika'da uygulanan testin sunum şekli, zamanlamada yapılan değişiklikler, cevapların veriliş formatı ve sınava ilişkin genel düzenlemeler olmak üzere genel olarak dört başlık altında toplanan uyumsama çalışmalarının ülkemizde tam olarak bir karşılığı olmadığı belirlenmiştir. Türkiye'de ekstra süre verme ve sınavı özel bir ortamda gerçekleştirme dışında diğer uygulamaların hiçbirinin gerçekleşmediği belirlenmiştir. Elde edilen bulgulara göre Amerika'da 1970'li yıllarda ele alınan engellilere yönelik düzenlemelerin ülkemizde de biran önce yetkililer tarafından ele alınarak yasal düzenlemelerin yerine getirilmesi gerektiği düşünülmektedir.

Anahtar Kelimeler

Engelli öğrenci, Geniş Ölçekli sınav, Eşitlik, Uyumsama çalışmaları

[*] Adnan Menderes Üniversitesi, Aydın Meslek Yüksekokulu, AYDIN, gokhanaksu1983@hotmail.com
[**] Hacettepe Üniversitesi, Eğitim Fakültesi, ANKARA, dcaktan@hacettepe.edu.tr
[***] Türkiye Belediyeler Birliği, Eğitim Şube Müdürlüğü, tahaeser@gmail.com

Paper No: 37

Investigation of Legal Responsibilities and Practices of the Large Scale Exams for Students with Disabilities in Turkey and the US

Abstract

In this study, it is aimed to compare the legal responsibilities and methods of application for students determined to be disabled in Turkey and the United States had relied on large-scale assessments. Under this purpose, it revealed similarities and differences between the two countries conducted determining the methods applied in the examinations by ETS in the United States with large-scale examinations conducted by ÖSYM in Turkey. As a result, it has been observed that there hasn't been any unrequited collected under four headings including in presentation of the test performed in the US, changes made to the schedule and the general regulations on the answer format and exam answers in our country. It is determined that any of the other applications besides providing extra time and performing exams in a private setting isn't applied in Turkey. According to the findings, it is thought to be necessary to the fulfillment of legislation arrangements for the disable studenst considered by the authorities as soon as possible which was discussed in the 1970s in America

Key Words

Disability Studenst, Large Scale Assessments, Fairness, Accomadations

Bildiri No: 38

TIMMS 2011 Türkiye Sonuçlarına Göre 8. Sınıf Öğrencilerinin Akıl Yürütme Becerileri ile Duyuşsal Özelliklerinin İlişkisi

Neslihan Tuğçe ŞİMŞEK*, Ergül DEMİR**

Özet

Bu araştırmanın amacı, ilköğretim 8. sınıf öğrencilerinin bir üst düzey zihinsel süreç olan akıl yürütme becerilerini matematik öğrenme alanında yordayan duyuşsal özelliklerinin incelenmesidir. Ayrıca akıl yürütme becerilerini yordayan duyuşsal özelliklerin; öğrencilerin cinsiyetlerine, evlerinde internet bağlantısı, bilgisayar ve çalışma masası olup olmama durumlarına göre farklılık gösterip göstermediği incelenmiştir. Araştırma, ilişkisel tarama modelinde yürütülmüştür. Aynı zamanda TIMMS 2011 Türkiye uygulaması verileri üzerinden yürütülen bir ikincil analiz çalışmasıdır. TIMMS 2011 Türkiye evreni, Türkiye'de 2010-2011 eğitim ve öğretim yılında öğrenim görmekte olan 4. ve 8. sınıf öğrencileridir. Örneklem ise tabakalı seçkisiz yöntemle seçilmiş 6928 8. sınıf öğrencisini içermektedir. Bu öğrencilerden kayıp veri sebebiyle indeks puanları hesaplanamayan üç öğrenci çıkarılmış ve çalışma 6925 öğrenci üzerinden yürütülmüştür. Analizlerde adımsal çoklu doğrusal regresyon analizi ve bağımsız örneklem t testi kullanılmıştır. Araştırma sonuçlarına göre öğrenci duyuşsal özelliklerinden matematiğe yönelik özgüven ve matematikle ilgilenme, Türkiye'de 8. sınıf öğrencilerinin matematikte akıl yürütme becerilerini manidar düzeyde yordamaktadır. Matematiğe yönelik özgüvenlerinde öğrencilerin cinsiyetlerine göre manidar bir farklılık gözlenememiştir. Diğer taraftan evde internet bağlantısı, bilgisayar ve çalışma masası olup olmamasına göre manidar farklar gözlenmiştir. Matematikle ilgilenme düzeylerinde ise cinsiyete göre manidar düzeyde farklılık gözlenirken kendine ait bir çalışma masasına sahip olmaya, evde bilgisayar ve internet bağlantısı olup olmamasına göre manidar bir fark gözlenememiştir.

Anahtar Kelimeler

Akıl Yürütme, Matematik, TIMMS 2011

*Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı. İletişim:simsekneslihantugce@gmail.com Tel: +90 312 363 33 50

**Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı. İletişim:erguldemir@ankara.edu.tr Tel: +90 312 363 33 50

Paper No: 38

The Relationship between 8th Grade Students' Affective Features and Their Mathematical Reasoning Skills According to TIMMS 2011 Turkey Results

Abstract

The aim of this study is to investigate the affective variables that might predict the 8th grade students' mathematical reasoning skills in Turkey. Also, the significant differences between gender subgroups and situation for possession of internet connection and computer and study desk at home were examined. Study was designed as a relational survey model and also a secondary analysis of which TIMSS 2011 Turkey practice. TIMSS 2011 Turkey population consists of all 8th grade students who were being educated in 2010-2011 education terms. Sample consists of 6928 Turkish 8th grade students who were selected via stratified random method. Three students whose index values were not calculated due to the missing values are excluded from the sample. The study was conducted with 6925 students. Step-wise multiple linear regression and independent samples t test were used for analysis. According to the results, confidence with mathematics and mathematical engagement are the predictors of mathematical reasoning skills of the 8th grade students in Turkey. Confidence with mathematics does not indicate any significant differences between gender subgroups; however, it indicates significant differences with possessing a computer, a study table and a network connection. Mathematical engagement indicates a significant difference based on gender while there are no differences between groups based on possessing a computer, a study desk for self and a network connection.

Key Words

Reasoning, Mathematics, TIMMS 2011

Bildiri No: 39

Faktör Puanı Belirleme Yöntemlerinin Karşılaştırılması Üzerinde Bir Çalışma

Müge ULUMAN, Deha DOĞAN

Özet

Bu çalışmanın amaçlarından ilki, faktör analizinin sıklıkla kullanıldığı eğitim, psikoloji ve sosyal bilimlerde daha az değişkenle çalışmaya olanak sunan farklı faktör puanı hesaplama yöntemlerinin varlığına dikkat çekmektir. Bu yöntemlerin kullanımının araştırma sonuçlarını etkileyip etkilemediği önemli bir problemdir. Bunun yanı sıra, araştırmacının kullanacağı faktör analizi yönteminin avantajlarını ve dezavantajlarını bilmesi ve araştırmanın amacına en uygun yöntemi seçmesi deoldukça önemlidir. Bu bağlamda, araştırma kapsamında ele alınan iki faktörlü yapıya ait gerçek veri setinde rafine edilmiş ve rafine edilmemiş faktör puanı hesaplama yöntemleriyle faktör puanları elde edilmiştir. Bu çalışmada farklı faktör puanı hesaplama yöntemleri kullanılarak elde edilen faktör puanları arasındaki korelasyon katsayıları incelenmiş ve pozitif yönlü yüksek korelasyon katsayıları bulunmuştur. Rafine edilmiş ve rafine edilmemiş yöntemler için birinci ve ikinci faktör puanları arası manidar bir farklılık görülmüşken, yöntemler kendi içinde birinci ve ikinci faktör puanları arası manidar bir farklılık göstermemiştir.

Anahtar Kelimeler

Faktör puanları, rafine edilmiş yöntemler, rafine edilmemiş yöntemler

Paper No: 39

A Study About Comparison Of Factor Score Computation Methods

Abstract

The first aim of this study is to draw attention to the presence of different factor score computation methods that allow working with a smaller number of variables in areas of education, psychology and social sciences which often utilize factor analysis. It is an important problem whether the use of these methods affect results of research. It is also of high importance that the researcher knows advantages and disadvantages of the factor analysis method to be used and select the most appropriate method for the purpose of the research. In this context, factor scores were derived with refined and non-refined factor computation methods in the real data set investigated in the study. This study examined the correlation coefficients among the factor scores obtained using different factor calculation methods for each factor, and it was determined that there are high positive correlations among them. While the correlation between the first and second factor results in significant difference between non-refined and refined methods, the non-refined and refined methods do not pose a significant difference within themselves. Factor score calculation methods can influence the fitness of the selected transformation in the context of the study carried out. So, while calculating factor analysis, when refined methods are decided to be used, usage of orthogonal rotation is advised since it requires no correlation between sub-factors of the scale.

Key Words

Factor scores, non-refined methods,refined methods

Bildiri No: 41

Türkiye'deki Öğretmenlerin Karşılaştıkları Mesleki SorunlarınİkiliKarşılaştırma Yöntemi ile Ölçeklenmesi

Yasemin Duygu ESEN*, Filiz İSNAÇ**,Ergül DEMİR***

Özet

Bu çalışmada öğretmenlerin yaşadıkları mesleki sorunların belirlenmesi ve öğretmenlerin bu sorunları önemli görme düzeylerinin ikili karşılaştırma yöntemi ile ölçeklenmesi amaçlanmıştır. Çalışma tarama modelinde yürütülmüştür. Araştırmanın örneklemini Türkiye'de 2015-2016 eğitim ve öğretim yılında MEB'e bağlı devlet okullarının ilköğretim ve ortaöğretim kademelerinde görev yapmakta olan toplam 484 öğretmen oluşturmaktadır. Veri toplama aracı olarak araştırmacılar tarafından geliştirilmiş olan "Öğretmenlik Mesleği Mesleki Sorunlar Anketi" kullanılmıştır. Verilerin analizinde, Thurstone'nun karşılaştırmalı yargılar kanununun III. Hal denklemi ile ölçekleme yöntemi kullanılmıştır. Araştırma sonucunda elde edilen bulgulara göre öğretmenlerin mesleki sorunları; öğretmen yetiştirme sorunları ve öğretmen niteliği, özlük hakları ve ekonomik sorunlar, mesleki saygının azalması, MEB. politikaları ile ilgili sorunlar, sendikal faaliyetlerle ilgili sorunlar, iş yükünün fazlalığı, okuldaki idari sorunlar, fiziki koşullar ve alt yapı yetersizliği, veli ile ilgili yaşanan sorunlar, ve öğrenci ile ilgili sorunlar temaları altında toplanmıştır. Öğretmenlere göre en önemli sorun Milli Eğitim Bakanlığı politikalarıdır. Bunu sırası ile mesleki saygınlığın azalması, fiziki koşullar ve alt yapı yetersizliği, öğrenci ile yaşanan sorunlar, özlük hakları ve ekonomik sorunlar, okuldaki idari sorunlar, öğretmen yetiştirme sorunları ve öğretmen niteliği, veli ile yaşanan sorunlar, iş yükü, sendikal faaliyetlerle ilgili sorunlar takip etmiştir. Bu sorunlar, kıdem değişkenine göre kısmi farklılıklar göstermektedir.

Anahtar Kelimeler

Öğretmen Mesleki Sorunları, İkili Karşılaştırma Yöntemi, Ölçekleme

*M.E.B., İletişim: yaseminduyguesen@gmail.com, Tel: +90 272 448 30 07
**M.E.B., İletişim: filiz.temel30@gmail.com, Tel: +90 324 473 27 67
***Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü, İletişim: erguldemir@ankara.edu.tr, Tel: +90 312 363 33 50

Paper No: 41

Scaling the Teachers' Professional Problems in Turkey with the Paired Comparison Method

Abstract

In this study, teachers' professional problems were investigated and thelevelsfor considering them as significant were scaled with the paired comparison method. The study was carried out in general survey model. The study group consisted of 484 teachers working in public schools at primary and secondary level in Turkey. As a data collection tool, "The Teacher Professional problems Survey", which was developed by the researchers, was used. In analysis of data, the scaling method with the third conditional equation of Thurstone's law of comparative judgement was used. According to the results, the teachers' professional problems were observed as teacher training and the quality of teacher, employee rights and financial problems, reduction of professional reputation, the problems with MoNE policies, the problems with union activities, workload, the problems with administration in school, physical conditions and the lack of infrastructure, the problems with parents, the problems with students. According to teachers, the most significant problem is MoNE policies. This is followed respectively by reduction of professional reputation, physical conditions and the lack of infrastructure, the problems with students, employee rights and financial problems, the problems with administration in school, teacher training and the quality of teacher, the problems with parents, workload and the problems with union activities. When teachers' professional problems were analyzed seniority variable, there was little difference in scale values. These problems had difference according to the teachers' length of service.

KeyWords

Teachers' Professional Problems, PairedComparisonMethod, Scaling

Bildiri No: 42

PISA 2012 Matematik Okuryazarlığı PuanlarınınÖz-Yeterliğe Dayalı Ölçüt Geçerliği

Seda DEMİR*, Nuri DOĞAN**

Özet

Bu araştırmanın amacı matematik okuryazarlığı alanının ağırlıklı olduğu 2012 Uluslararası Öğrenci Değerlendirme Programı (PISA 2012) sonuçlarına göre öğrencilerin matematiğe yönelik öz-yeterlik puanları ile matematik okuryazarlığı puanlarından elde edilen %20'lik beşer grubun paralellik gösterip göstermediğini yani beklentiye uygun sonuçlar verip vermediği incelemektir. Bu amaç doğrultusunda beklenti tablolarından yararlanılmıştır. Beklenti tablosu, iki testten elde edilen sonuçlar arasındaki ilişkiyi gösteren iki yönlü bir tablodur. Dolayısıyla beklenti tablosu, yapılan kestirimin geçerliğine dair verileri görmek amaçlı kullanılabilmektedir. Buna göre yapılan bu çalışma bir geçerlik çalışması niteliğindedir. Bu çalışmada, ölçüt olarak belirlenmiş öz-yeterlik puanlarına göre oluşturulan gruplarla matematik okuryazarlığı puanlarına göre oluşturulan gruplar arasında yüksek eşleşme yüzdeleri elde edildiği takdirde, beklenti tabloları aracılığıyla hem öz-yeterliğin başarı üzerindeki etkisi kanıtlanmış olacak hem de ölçüt geçerlik düzeyi belirlenmiş olacaktır. Ölçüt geçerliğinin sağlıklı değerlendirilmesi, öz yeterliğin başarı üzerindeki etkisinin bilişsel başarı düzeyleri farklı olan ülkelerde aynı şekilde işleyip işlemediğinin saptanması için analizler farklı başarı düzeyine sahip üç ülkeye ait puanlar üzerinde yinelenmiştir. Araştırmanın çalışma grubunu PISA 2012 Matematik uygulamasına katılan Türkiye'den 3130 öğrenci, İsviçre'den 7246 öğrenci, Çin-Şangay'dan 3416 öğrenci oluşturmaktadır. Araştırma sonucunda, Türkiye için oluşturulan gruplarda elde edilen eşleşme yüzdelerinin beklenenin aksine oldukça düşük bulunması sonucu, Türkiye alt grup olarak düşünüldüğünde orta grup kabul edilebilecek İsviçre ve üst grup kabul edilebilecek Çin-Şangay için de aynı incelemeler gerçekleştirilmiştir. Sonuç olarak hem Türkiye, hem İsviçre hem de Çin-Şangay için oluşturulan gruplarda daha yüksek oranda olması beklenen eşleşme yüzdelerinin beklenenin çok altında olduğu, bunun yanı sıra üst gruptaki ülkeye ait eşleşme yüzdelerde herhangi bir düzenli artışın olmadığı görülmüştür. Dolayısıyla yapılan araştırmada, PISA 2012 uygulamasına katılan Türkiye, İsviçre ve Çin-Şangay örneklemleri için, genel anlamda kabul görülenin aksine öğrencilerin duyuşsal giriş özelliklerinin (matematiğe yönelik öz-yeterlik) bilişsel düzeydeki öğrenci başarıları (matematik başarısı) üzerinde anlamlı bir etkisinin olmadığı sonucuna ulaşılmıştır. Sonuç olarak PISA 2012 Türkiye, İsviçre ve Çin-Şangay verilerinin öz-yeterliğe dayalı ölçüt geçerliğini sağlayamadığı söylenebilir.

Anahtar Kelimeler

PISA 2012, öz-yeterlik, matematik okuryazarlığı

*Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Doktora Öğrencisi. İletişim: seda.demir@hacettepe.edu.tr Tel: +90 312 297 8550

**Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: nurid@hacettepe.edu.tr Tel: +90 312 297 8550

Paper No: 42

Criterion Validity Based On Self-Efficacy Of PISA 2012 Mathematics Literacy Scores

Abstract

The aim of this research is to investigate whether the students' self-efficacy scores about maths according to the results of the PISA 2012 with a focus on mathematics, and each five group consisting of 20% show parallelism, in other words, whether they propose results matching with the expectations. This study, which used to expectancy table, is in the form of validity check. In this research, if high levels of matching percentages are gained, both the effect of self-efficacy on success and criterion validity will be checked between the groups between created according to self-efficacy scores defined as measurement criteria and those created according to maths literacy scores through expectancy tables. For the purpose of making a healthy evaluation of the criteria validity, and determining whether the effect of self-efficacy on success works the same in the countries that have different cognitive success levels, the analysis were repeated over the scores that belong to three countries with different success levels (Turkey-lower group, Switzerland-average group, Shandhai-China—super group). Thus the population of this study consisted of 3130 students from Turkey, 7246 students from Switzerland and 3416 students from Shanghai-China who participated in the PISA 2012 Mathematics application. As a result, it was seen that the match is expected to be at a higher rate for Turkey, Switzerland, and also Shanghai-China couldn't be reached, and that the matching per cents in the groups created were quite lower than expected. Therefore, contrary to well accepted view, the conclusion that the affective input characteristics of the students (self-efficacy of math) in Turkey, Switzerland and Shanghai-China samplings who participated in the PISA 2012 application don't have any significant effect on their successes in cognitive level (math success) was reached. To sum up, it can be said that the data of PISA 2012 for Turkey, Switzerland and Shanghai-China couldn't provide criteria validity based on self-efficacy.

Key Words

PISA 2012, self-efficacy, math literacy

Bildiri No: 43

Farklı Kayıp Veri Baş Etme Tekniklerinin Klasik Test Kuramı ve Madde Tepki Kuramı İle Elde Edilen Madde Parametreleri Üzerindeki Etkisinin İncelenmesi

Ufuk AKBAŞ[*], Şeyhmus AYDOĞDU[**],Şener BÜYÜKÖZTÜRK[***]

Özet

Bu araştırmanın amacı, örneklem büyüklüğü ve kayıp veri oranının farklı koşulları altında, kayıp verilerle baş etmede kullanılabilecek üç farklı tekniğin klasik test kuramı ve klasik madde kuramı ile elde edilen madde parametreleri üzerindeki etkililiğini test etmektir. Bu amaç doğrultusunda, farklı örneklem büyüklüklerinde ve tek boyutlu madde tepki kuramı kapsamında iki parametreli modele uygun otuzar farklı yapay veri üretilmiştir. Kayıp veri teknikleri ile elde edilen değerlerin karşılaştırılabilmesi amacıyla her iki kuram için, eksiksiz veri setlerine ilişkin madde parametreleri hesaplanmıştır. Eksiksiz veri setlerinde, tümüyle seçkisiz kayıp koşulu altında %2, %5 ve %10 oranlarında silme işlemi gerçekleştirilmiş ve liste bazında silme, beklenti – maksimizasyon algoritması ve regresyonla değer atama teknikleri ile elde edilen yeni veri setlerine ait madde parametreleri hesaplanmıştır. KTK kapsamında, araştırmada yer verilen tüm örneklem büyüklükleri için liste bazında silme tekniği ile elde edilen güçlük ve ayırt edicilik değerlerinin kayıp veri oranına bağlı olarak sıfıra yaklaştığı, regresyonla değer atama ve beklenti – maksimizasyon algoritması ile ise eksiksiz veriye çok yakın değerlerin elde edildiği görülmüştür. MTK kapsamında elde edilen sonuçlar incelendiğinde ise liste bazında silme tekniğinin, %2'lik kayıp veri oranında bile model – veri uyumunu ciddi şekilde tehdit edeceği, regresyonla değer atama ve beklenti maksimizasyon algoritması ile eksiksiz veriden elde edilenlere çok yakın parametrelere ulaşılabileceği görülmüştür.

Anahtar Kelimeler

Kayıp veri, Klasik test kuramı, Madde tepki kuramı, Madde parametreleri

[*] Hasan Kalyoncu Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:ufuk.akbas@hku.edu.tr, Tel: +90(342)2118080

[**] Nevşehir Hacı Bektaş Veli Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:aydogduseyhmus@gmail.com Tel: +90(384)2281000

[***] Hasan Kalyoncu Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:sener.buyukozturk@hku.edu.tr, Tel: +90(342)2118080

Paper No: 43

The Impact Of Missing Data Techniques On Item Parameters Estimated According To Classical Test Theory And Item Response Theory

Abstract

The purpose of this study is to investigate the impact of three missing data tecniques, which will be applied under different sample size and missing ratio conditions, on item parameters obtained by classical test theory (CTT) and item response theory (IRT). For each condition thirty data sets were generated under 2PL. Item parameters for these sets are used as reference points for completed data sets. Assuming missing completely at random condition, data were deleted for %2, %5 and %10 ratio. Listwise deletion, regression imputation and expectation – maximization algorithm applied to these sets and item parameters were reestimated. Under CTT, listwise deletion reduced item parameters as missingness ratio increased; regression imputation and EM result were similar consistent with complete data. Under IRT, listwise deletion was a serious problem even for %2 missing raito and regression imputation and EM result were similar and accurate.

Key Words

Missing data, Classical test theory, Item response theory, Item parameters

Bildiri No: 45

Genellenebilirlik Kuramındaki Karar Çalışmalarının İncelenmesi

Ömer KAMIŞ*, C. Deha DOĞAN**

Özet

Bu araştırmanın amacı, Genellenebilirlik Kuramı karar çalışmalarından elde edilen G ve Phi katsayıları ile karar çalışmasındaki yüzeylerine ve koşullarına ilişkin gerçek durumlardan elde edilen G ve Phi katsayılarını karşılaştırmaktır. Araştırma temel araştırma türündedir. Çalışma grubunu 84 lisans öğrencisi ile 3 yüksek lisans öğrencisi ve 1 öğretim üyesi oluşturmuştur. Veri toplama aracı olarak araştırmacılar tarafından geliştirilen başarı testi kullanılmıştır. Verilerin analizinde Edu G 6.1-e programı kullanılmıştır. Bulgular incelendiğinde gerçek durumlarda elde edilen ve farklı karar çalışmaları sonucunda kestirilen G ve Phi katsayılarının birbirlerine yakın değerler almalarına rağmen farklılaştıkları görülmektedir. G kuramının kullanıldığı araştırmalarda yapılan karar çalışmalarında kestirilen G ve Phi katsayıları yorumlanırken, kestirilen bu katsayıların gerçek durumdaki değerleriyle aynı olmayabileceğinin bilinmesi ve yorumlamada bu duruma dikkat edilmesi gerekir.

Anahtar Kelimeler

Genellenebilirlik Kuramı, Karar çalışması, G Katsayısı, Phi Katsayısı.

*Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: okamis@ankara.edu.tr Tel: 0312 363 33 50 / 2042

**AnkaraÜniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: ddogan@ankara.edu.tr Tel: 0312 363 33 50 / 5211

Paper No: 45

Examination of Decision Studies in Generalizability Theory

Abstract

The aim of this research is to compare the G and Phi coefficients obtained by D studies in G theory and the actual cases for the same conditions of similar facets in D studies. This research is a basic research. In the study convenient sampling was used. The study group consist of 84 university students, 4 raters working at a state university. The relevant data was gathered from the achievement test in Research Methodology Courses developed by the researchers. Edu G 6.1-e programme was used to analyze the data. According to the research findings, although G and Phi coefficients obtained by D studies and actual cases close to each other, the coefficients are different. The findings revealed that there is no systematic relationship between G and Phi coefficients obtained by D studies and actual cases.

Key Words

Generalizability Theory, Decision study, G coefficent, Phi coefficent.

Bildiri No: 46

GSE (Global Scale Of Englısh) Temelli İngilizce Yeterlik Sınavı: Anadolu Üniversitesi Örneği

Belgin AYDIN*, Meral Melek ÜNVER**

Özet

Eğitim dili tamamen ya da kısmen İngilizce olan bir çok üniversitenin Yabancı Diller Yüksekokullarında öğrencilerin dil yeterliğini ölçmek amacıyla bir çok sınav yapılmaktadır. Ancak Türkiye genelinde aynı temellere oturtularak tasarlanmış bir yabancı dil yeterlik sınavı bulunmamaktadır. Türkiye'de en fazla öğrenci sayısına sahip 12 üniversitenin Yabancı Diller Yüksekokulu yöneticileri ve sınav birimi üyeleriyle görüşmeler yapılan bir projenin sonucunda, "dil yeterliği" kavramının kurumlar arasında farklılık gösterdiği saptanmıştır (Aydın, Akay, Polat, Geridönmez, 2016). Aynı proje sonuçlarına göre bu kurumlarda yürütülen programlarda Avrupa Ortak Dil çerçevesi (CEFR) (Council of Europe, 2001) ölçütlerinin baz alındığı ve B2 seviyesinin çıkış seviyesi olarak belirlendiği ifade edilse de B2 nin ne olduğuna dair kurumlar arası bir fikir birliği olmadığı görülmüştür. Bu çalışmada, Anadolu Üniversitesi Yabancı Diller Yüksekokulunda öğrencilerin lisans eğitimlerine başlayabilmeleri için yeterli dil becerilerine sahip olup olmadıklarını ölçmek için kullanılan ve Türkiye bağlamında ilk ve tek olan GSE (Global Scale of English) (Pearson, 2015) temelli İngilizce Yeterlik Sınavı tasarım süreci ve bu süreçte testin bileşenleri, madde türleri ve madde dağılımıyla ilgili verilen kararlar paylaşılacaktır. Buna ek olarak sınav biriminin çalışma prensipleri, iş akış şeması ve madde yazım sürecine de değinilecek, son olarak da testin çoktan seçmeli kısmına ait faktör analizi ve bu bağlamda yürütülen revizyon çalışmaları anlatılacaktır. Hangi parametrelerin göz önünde bulunduruluğuna bakmaksızın, düşük dil becerisine sahip öğrencilerin yabancı bir dilde verilen lisans eğitimlerine devam edebilmelerini sağlayacak ve bunu kısa bir süre içerisinde başaracak olan bir program geliştirmek ve uygulamak tüm kurumlar için büyük bir öneme sahiptir. Dolayısıyla, Türkiye'de yabancı dil ölçme-değerlendirme sürecini standartlaştırma adına öncü olacak bir deneyimin paylaşıldığı bu çalışmanın, diğer kurumlara da ilham vereceği öngörülmektedir.

Anahtar Kelimeler

İngilizce yeterlik sınavı, yabancı dilde ölçme değerlendirme, İngilizce test geliştirme, İngilizce Sınavı

Paper No: 46

*Anadolu Üniversitesi Yabancı Diller Yüksekokulu. İletişim: baydin@anadolu.edu.tr Tel: +90222-3350580/6001

** Anadolu Üniversitesi Yabancı Diller Yüksekokulu. İletişim: mmunver@anadolu.edu.tr Tel: +90222-3350580/6003.

English Proficiency Test Based On Global Scale Of English (GSE): The Case Of Anadolu University

Abstract

In Turkey, universities with English-medium instruction deploy many English proficiency tests via their School of Foreign Languages in order to assess language proficiency of their students. However, there is no standardized proficiency test based on the same principles throughout the country. A recent project conducted with the directors and testing unit members of the school of foreign languages in 12 most populated universities in Turkey revealed that the perception of "language proficiency" varies among the institutions (Aydın, Akay, Polat, Geridönmez, 2016). The results of the same project also indicated that although the programs were based on Common European Framework of Reference (CEFR) (Council of Europe, 2001) and B2 was determined as the exit level, there was no concensus on what B2 really meant to these institutions. This study particularly focuses on the Proficiency Test developed by the School of Foreign Languages of Anadolu University, which is the first and only Global Scale of English (GSE) (Pearson, 20015) based proficiency test in the Turkish context. To this end, the decisions made regarding the components of the test, the item types and numbers, the distribution of the items will be shared. In addition, the working principles of the testing unit, the work flow chart, and the item writing processes will be outlined. Finally, the focus will be on the factor analysis of the multiple-choice component of the test, and the revision processes we have been going through. No matter which parameters are taken into consideration, preparing and implementing curricula that will enable low proficient students to continue their undergraduate education in a foreign language in such a short time period is of great importance for all the institutions. Therefore, this presentation will give insights into the other institutions as a leading standardization process in language testing in Turkey.

Key Words

English Proficiency Test, Testing English, Test Design, Testing and Assessement in Language Teaching

Bildiri No: 47

Kendi Aracını Getir Uygulaması: Özel Arı Okulları Örneği

Pınar Nuhoğlu KİBAR*, Buket AKKOYUNLU**, Seda DEMİR***, Ebru ÇEVİK****

Özet

Bu çalışmada 2014-2015 eğitim - öğretim yılında ortaokul öğrencileriyle gerçekleştirilen infografik tasarım uygulaması, Kendi Aracını Getir (KAG) boyutuyla paylaşılmaktadır. KAG, öğrencilerin öğrenme ortamlarına kendi dijital araçlarını öğrenme sürecinde kullanmak amacıyla getirmeleri olarak tanımlanmaktadır (Johnson ve diğerleri, 2015). Öğrencilerin öğrenme sürecine kendi araçlarıyla katılmaları, bireyselleştirilmiş araçlarla çalışmalarını (Alberta Education, 2012) ve süreçte biçimlendirmeye dönük dönüt alabilmelerini sağlaması (Attewell, 2015) yönleriyle eğitsel bağlamda önemsenmektedir. Araştırma kapsamında gerçekleştirilen infografik tasarım uygulaması, öğrencilerin karmaşık bilgiyi yalın ve etkili bir biçimde görselleştirebilmelerini sağlamayı amaçlayan öğrenme stratejisi olarak ele alınmaktadır. İnfografik tasarım, öğrenme görevinin kapsadığı bilgi ile doğru orantılı olarak zamana yayılan bir uygulama olup, özellikle bu strateji ile yeni tanışan öğrenenler için biçimlendirmeye dönük değerlendirmenin yapıldığı bir süreç izlenmesini gerektirmektedir. Uygulamanın gerçekleştirildiği okul tarafından benimsenen KAG modeli doğrultusunda, infografik tasarım sürecinde öğrencilerin biçimlendirmeye dönük dönüt alabilmelerinin kolaylaştırılması ve bireyselleştirilmiş araçlarla çalışmalarının sağlanması amaçlanmıştır. Birbirini takip eden iki döngü ile yürütülen çalışmada ilk döngü bilgisayar laboratuvarında (Nuhoğlu Kibar & Akkoyunlu, 2014) ikinci döngü ise ilk döngü sonundaki bulgular doğrultusunda öğrencilerin kendilerine ait tabletlerle gerçekleştirilmiştir. İlk döngü sonunda alınan öğrenci ve öğretmen yansımaları doğrultusunda biçimlendirmeye dönük dönütün infografik tasarım sürecinin önemli bir boyutu olduğu belirlenmiştir. Öğrenciler ve öğretmenler zaman yetersizliğini vurgulamış, infografik tasarım süreci için okuldaki tasarım sürecinin daha uzun olması gerektiğini belirtmişlerdir. Tasarım sürecinde aktif rol alan öğretmenler, ilk döngüde infografik tasarım proje dosyalarının saklanması ve aktarımına yönelik yaşanan sorunlara dikkat çekmişlerdir. İnfografik tasarım uygulaması ikinci döngüsünde, ilk döngüde belirlenen kolaylaştırıcı bileşenler ve sorunlar göz önünde bulundurularak KAG modeli uygulanmıştır. İkinci döngü sonunda alınan öğrenci ve öğretmen yansımaları doğrultusunda öğrencilerin kendi araçlarıyla öğrenme sürecine katılmış olmalarının süreci kolaylaştırdığı belirlenmiştir. Çalışma kapsamında "Yönetim ve tercihler", "teknolojik destek ve kontrol", "kullanım yeri", "pedagoji ve öğretmen gelişimi" (Attewell, 2015) boyutlarına göre uygulanan KAG modeli açıklanmış, infografik tasarım süreci KAG uygulaması açısından irdelenmiştir.

Anahtar Kelimeler

Kendi Aracını Getir, Ortaokul öğrencileri, Biçimlendirmeye dönük dönüt, Bireyselleştirilmiş araç

* Arş. Gör., Hacettepe Üniversitesi,Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, pnuhoglu@hacettepe.edu.tr

**Prof. Dr., Hacettepe Üniversitesi,Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, buket@hacettepe.edu.tr

***Uzm., Özel Arı Okulları, Araştırma ve Geliştirme Birimi, seda_demir06@hotmail.com@gmail.com

****Uzm., Özel Arı Okulları, Araştırma ve Geliştirme Birimi, ebrukcevik@hotmail.com@gmail.com

Paper No: 47

Bring Your Own Device Implementation: Arı Private Schools Sample

Abstract

In this study, infographic design application implemented on secondary school students in 2014-2015 education term is shared in the scope of Bring Your Own Device (BYOD) model. BYOD is defined as the term that the students bring their own digital instruments to their learning habitat to be used in learning processes (Johnson et al, 2015). Bringing student's own instruments to their learning processes pedagogically pays much attention because of providing students to study with more individual tools (Alberta Education, 2012) and acquiring formative feed backs (Attewell, 2015). Infographic design application performed by this study is mentioned as a learning strategy aiming to provide students to visualize complex information in a lean and effective way. Infographic design is an application which extends over a period of time proportional to the information that is covered by the learning task and it necessitates an evaluation process in a formative manner to follow especially for the ones who get in contact with this strategy recently. It is intended to facilitate students to acquire formative feed backs and provide them to study with more individual instruments in the scope of BYOD model that have been adopted by the school of application. Two subsequent loops were performed; first one in the computer lab (Nuhoğlu Kibar & Akkoyunlu, 2014) and second one in the personal tablets of students towards the findings of the first loop. At the end of first loop; it was determined by the reflections of students and teachers that the formative feed back is a vital dimension of infographic design. Students and teachers emphasized the insufficiency of time and pointed out that the infographic design phase in the school should be longer. Teachers who had been taking an active role during the design phase have drawn attention to the problems encountered about keeping and transferring of infographic desing project documents during the first loop. BYOD model was applied in the second loop of infographic design application by taking into account of problems and facilitative elements determined during the first loop. It is specified at the end of the second loop by the student and teacher reflections that the learning process is simplified by the getting student's own learning tools involved to the process. In the scope of this study, BYOD model which is applied by "Management and Preferences", "Technological Support and Control", "Place of Use", "Pedagogy and Teacher Development" (Alberta Education, 2012) dimensions is explained and infographic desing process is examined in terms of BYOD model.

Key Words

Bring Your Own Device (BYOD), Secondary school students, Formative feedback, Individualized device

Bildiri No: 48

Kayıp Veriler Yerine Yaklaşık Değer Atamak için Kullanılan Gelişmiş Yöntemlerin Farklı Koşullar Altında Karşılaştırılarak İncelenmesi

Elif Kübra DEMİR, Sait ÇÜM,Selahattin GELBAL, Tarık KIŞLA[2]

Özet

Bu araştırmada, farklı oranlarda (%15 ve %25) ve yapılarda (TROK ve ROK) oluşturulan kayıp veriler yerine farklı yöntemlerle yaklaşık değer atanması sonucu elde edilen veri setlerinin tam veri setleriyle karşılaştırılarak incelenmesi amaçlanmıştır. Araştırma verilerinin toplandığı grup, PISA'ya (2012) Türkiye'den katılan 15 yaş grubundaki 4848 öğrenci arasından matematik özyeterliği anketine katılan ve eksiksiz bir şekilde yanıtlayan 3129 öğrencinin puanlarından oluşan veri seti üzerinde yürütülmüştür. Söz konusu veri seti içerisinden farklı yapılar oluşturulacak şekilde farklı oranlarda veri silinerek eksik veri setleri oluşturulmuştur. Bu veri setleri BM, BVA, ESE, MUA, MZMC ve RA olmak üzere altı farklı gelişmiş yaklaşık değer atama yöntemiyle tamamlanmıştır. Elde edilen veri setleri öğrenci toplam puanları üzerinden tam veri setiyle karşılaştırılmıştır. Araştırma kapsamında tüm kayıp veri atamalarında elde edilen öğrenci toplam puanları ile tam veriye ait toplam puanlar arasındaki korelasyonların oldukça yüksek olduğu görülmüştür. Benzer şekilde toplam puanlar göz önüne alındığında kayıp veri atama yöntemlerinin birbirleri ile olan ilişkileri de oldukça yüksek bulunmuştur. Tam veri seti ile yaklaşık değer atanmış veri setlerinden hesaplanan öğrenci toplam puanları arası farkların mutlak değer toplamları ve ortalamaları göz önünde bulundurulduğunda her üç koşul altında da en iyi çalışan yaklaşık değer atama yöntemlerinin MZMC ve BM olduğu sonucuna ulaşılmıştır.

Anahtar Kelimeler

Kayıp Veri, Değer Atama Yöntemleri

[22]Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:elif.kubra.demir@ege.edu.tr Tel:+902323113373

[2]Çerikli Atatürk Ortaokulu. İletişim: saitcum@hotmail.com Tel:+903186381023

[2]Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: gelbal @hacettepe.edu.tr Tel: +903122978550

[2] Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: tarik.kisla@ege.edu.tr Tel: +902323113145

Paper No: 48

The Examination of the Comparison of Advanced Methods Used for Value Imputation Instead of Missing Values under Various Conditions

Abstract

In this study, it is aimed to comparatively research of data sets obtained imputation for missing values that is formed by different ratios(%15 and %25) and in different structures (MCAR and MAR) with different methods. This study has been conducted on data set formed by points of 3129 students who participated in mathematics self-efficacy survey and answered it completely among 4848 students- age group of 15-who participated in PISA 2012 from Turkey. Missing data sets have been constituted by deleting data in different ratios to be constitute different structures in the data set. These data sets have been completed by six different nearby value imputation including EM, BIM, PSM, MCMC, MDIM, and RIM. Obtained data sets have been compared with full data sets by total points of students. In the scope of the research, correlation between obtained total score of the students and total points of actual data has been seen quite high. Similarly, when total points is considered, correlation of missing data imputation methods with each other have also been found quite high. Considering the difference between the totals and avarages of student scores calculated from the full data set and imputed data sets EM and MCMC is founded that the best missing data imputation methods under all conditions.

Key Words

Missing Values, Imputation Methods

Bildiri No: 49

Veri Zarflama Analizi ile Dumlupınar Üniversitesi Bünyesindeki Okulların Verimliliklerinin İncelenmesi

R. Serkan ARIK*, Buket ERENJANSSEN**

Özet

Bu çalışmada Dumlupınar üniversitesi bünyesinde yer alan okulların (fakülte, enstitü, yüksekokul) etkinliğinin veri zarflama analizi ile belirlenmesi amaçlanmaktadır. Analizlerde Dumlupınar Üniversitesi'nin verimliliği girdi değişkenleri; 2015-2016 eğitim-öğretim yılında okullarda kayıtlı olan toplam öğrenci sayıları, akademik olmayan personel sayıları, akademik personel sayıları ile çıktı değişkenleri; bir yıllık bir zaman aralığında mezun olan öğrenci sayıları, mezun olan öğrencilerin genel ortalamaları, okulu bırakan öğrenci sayılarına göre değerlendirilmiştir. Çalışmada 7 fakülte, 15 yüksekokul ve 3 enstitüye ait veriye dair analizler yapılmış ve bu okulların verimlilikleri incelenmiştir. Analizler sonucu toplam teknik etkinlikler belirlenmeye çalışıldığı için ölçeğe göre sabit getiri varsayımına dayanan CCR modeli tercih edilmiştir. Okulların hem girdi hem çıktı değişkenlerinin verimliliğe olan etkisi incelenmek istendiği için hem girdi hem çıktı odaklı olarak analizler yapılmıştır. Analizlerde DEA Solver LV yazılımı kullanılmıştır. Elde edilen sonuçlara göre analize alınan 25 okuldan 15'inin etkin olduğu belirlenmiş ve etkin olmayan okullar için yapılması gereken iyileştirmeler ifade edilmiştir.

Anahtar Kelimeler

Veri Zarflama Analizi, Verimlilik

* Dumlupınar Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:rserkan.arik@dpu.edu.tr Tel: +90 274 265 2031/4616

**Dumlupınar Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:buket.eren@dpu.edu.tr Tel: +90 274 265 2031/4567

Paper No: 49

Examining the Efficiency of the Schools Contained Within Dumlupınar University by Data Enveloping Analysis

Abstract

The aim of this research is to examine the efficiency of the schools (faculty, institute, academy) contained within Dumlupınar University by data envelopment analysis. The input variables are the total number of students registered on 2015-2016 education period, the number of non-academic personnel, the number of academic personnel. The output variables are the number of graduate students in one year period, the general grade point average of graduated students, the number of student who dropped out. The analysis had been done with 7 faculties, 15 academies and 3 institutes and the efficiencies of these schools were examined. Because the total efficiencies were tried to be determined, the CCR model which relied on fixed values according to scale was used. Both input and output analysis had been done in order to see the effects of both input and output values. DEA Solver LV was solved for analysis. The results show that 15 out of 25 schools appear as efficient according to the determined input and output variables. The needed changes for the inefficient schools to make them efficient are given at results and suggestions section.

Key Words

Data Enveloping Analysis, Efficiancy

Bildiri No: 50

Dezavantajlı Yüksek Başarılı Öğrencilerin Okul ve Öğrenci Düzeyinde Matematik Başarılarını Etkileyen Değişkenlerin İncelenmesi

Elif Bengi ÜNSAL ÖZBERK, Leyla YILMAZ FINDIK, Eren Halil ÖZBERK

Özet

Çalışmanın amacı, öğrenme fırsatları, okul iklimi ve okul kaynaklarının düşük sosyo-ekonomik statüye sahip olmalarına rağmen yüksek başarı gösteren öğrencilerin matematik başarılarını ne ölçüde açıkladığını araştırmaktır. Bu amaçla araştırma, betimsel tarama modelinde desenlenmiştir. Çalışmada, PISA uygulamasında ekonomik sosyal kültürel statü indeksine göre sıralanan öğrencilerden en alt %33'lük dilimde kalan öğrenciler "sosyo-ekonomik açıdan dezavantajlı öğrenciler", ülke içindeki başarı sıralamasına göre en üst %33'lük dilimde yer alan öğrenciler "yüksek başarı gösteren öğrenciler" olarak tanımlanmaktadır. Bu tanımlamaya göre çalışma grubunu, PISA 2012 uygulamasına göre Türkiye ESKS sıralamasında en alt %33'lük dilimde yer alan ve aynı zamanda Türkiye içindeki Matematik puan sıralamasında %67'nin üstünde başarı gösteren ve "dezavantajlı yüksek başarılı öğrenci" olarak tanımlanan 317 öğrenci oluşturmaktadır. Çalışmanın analizleri, belirlenen bu 317 dezavantajlı yüksek başarılı öğrenciler ile yapılmıştır. Araştırmanın verileri, OECD PISA 2012 veri tabanından elde edilmiştir. Araştırmanın amacı doğrultusunda kullanılan değişkenlere ait veriler, okul ve öğrenci anketinden elde edilmiştir. Okul düzeyinde, okul iklimi ve okul kaynakları boyutundaki değişkenler ile öğrenci düzeyinde, öğrenme fırsatları boyutu altındaki değişkenler alınarak analizler yapılmıştır. Bu çalışmada PISA uygulaması verileri, öğrenci ve okul düzeyinde hiyerarşik yapı gösterdiğinden çözümlemelerde aşamalı doğrusal model (Hierarchical Linear Model-HLM) kullanılmıştır. PISA 2012 uygulamasına ait kayıp veri sorunu çözüldükten sonra aşamalı doğrusal modele ilişkin varsayımlar test edilip; verilerin modeli test etmeye uygun olduğuna karar verilerek, aşamalı doğrusal modele ait denklemler ve modeller kurulmuştur. Çalışmanın bulgularına göre öğrencilerin matematik başarılarının %64'ünün okul içinde farklılaştığı; bir başka deyişle, öğrenciler arasındaki farklardan kaynaklandığı sonucuna varılabilir. Bununla birlikte dezavantajlı yüksek başarılı öğrencilerin matematik başarılarının %43'ünün okul iklimi ve okul kaynakları değişkenleri ile açıklandığı görülmektedir.

Anahtar Kelimeler

Öğrenci Başarısı, Öğrenme Fırsatları, Okul İklimi, Okul Kaynakları, Aşamalı Doğrusal Model

Paper No: 50

An Investigation of Factors Effecting Resillient Students Mathematics Performance According to Student and School Level

Abstract

The study aims to examine the effect of opportunity to learn, school climate and school resources on mathematics performance of resilient students who outperforms their peers in spite of their socio-economically disadvantaged background in Turkey in PISA 2012 assessment. For this reason the study is a descriptive study. The study identifies students as resilient or disadvantaged successful students if they belong to the bottom third of the socio-economic background distribution in the country where they took the assessment and perform in the top third of the achievement distribution in terms of mathematics performance. According to this definition the study defines 317 resilient students in Turkey in PISA 2012 assessment as the ones who were able to overcome their disadvantaged socio-economic background and perform well at school. The data of the study was comprised of the resilient students' mathematics performance and students' questionnaire completed in PISA 2012 assessment. Variables related to opportunity to learn, school climate and school resources are all included in the analysis. Hierarchical Linear Model (HLM) was used to examine the effects of variables on resilient students'mathematics performance. Missing data of PISA 2012 assessment were all excluded and after hypothesis were tested and checked if the model is suitable for this study, equations and models are set related to Hierarchical Linear Model. According to the findings the study concluded that %64 of total variance in resilient students' mathematics performance are differentiated in school that is the mathematics performance of resilient school are explained by difference among students. In addition to this the study also indicated that variables related to school explained approximately %43 of total variance in resilient students' mathematics performance.

Key Words

Students' Performance, Opportunity to Learn, School Climate, School Resources,Hierarchical Linear Models

Bildiri No: 51

Performans Görevleri Hakkında Öğretmen Ve Öğrenci Görüşleri: Meslek Lisesi Örneği

Ayşegül ALTUN, Hülya KELECİOĞLU

Özet

Bu çalışmada meslek lisesi öğretmeni ve öğrencilerinin performans görevleri hakkındaki görüşlerinin belirlenmesi amaçlanmıştır. Çalışma Ankara'daki bir meslek lisesinin 9, 10, 11 ve 12. sınıflarında okuyan 17 öğrenci ve aynı lisede çalışan 13 öğretmen ile gerçekleştirilmiştir. Çalışma nitel araştırma desenlerinden olgu bilim ile planlanmış ve görüşme, gözlem ve dokümanlarını paylaşan bireylerin dokümanları incelenerek gerçekleştirilmiştir. Elde edilen verilerin çözümlenmesinde içerik analizi tekniği kullanılmıştır. Çalışmanın bulgularına göre, öğretmenlerin öğrenci düzeyine ve öğretim programına göre performans konularını belirledikleri, öğrencilerin de kolay ve yapabilecekleri konuları tercih ettikleri belirlenmiştir. Ayrıca öğrenciler, performans görevinden yüksek puan almak ve bazı bilgileri öğrenmek yönünde beklentileri olduğunu ifade etmişlerdir. Öğretmenler ise öğrencilerin görevi yaparken çalışmalarını ve yeni bilgiler öğrenmelerini beklentileri olarak dile getirmişlerdir. Öğrenciler performans görevlerinin ağırlıklı olarak dış görünüşüne göre değerlendirildiğini düşünürken öğretmenler ise içeriğe göre değerlendirdiklerini belirtmişlerdir. Öğrenciler performans görevinin görünüşünü düzgün yapmakta sorun yaşadıklarını belirtirken, öğretmenler öğrencilerin sorumsuz davranması, görevi anlamamaları, başkalarına yaptırmaları gibi daha çok öğrenci kaynaklı sorunlar yaşadıklarını belirtmişlerdir.

Anahtar Kelimeler

Performans görevi, içerik analizi, meslek lisesi

Paper No: 51

Teachers' and Students' Opinions about Performance Task: A Vocational High School Sample

Abstract

The aim of the study is to determine opinions of a Vocational School teachers and students about performance task. This study is carried out with 17 vocational high school students (9th, 10th, 11th and 12th graders) and 13 teachers who work in the same school. This research is planed with the phenomenological qualitative research design and carried out by analyzing the documents of the individuals who shared their own interviews, observations and documents. The data were analyzed through content analysis. The study demonstrates that teachers determine the topic of the performance tasks according to their students' level and curriculum and students prefer topics which are easy and they believe them capable of. In addition, students state that they expect to get high grades and learn the topics. Furthermore, as their expectations teachers state their students study more and learn new information. While students think that their performance task is assessed mostly according to its physical appearance, in contrast teachers stated that they assessed performance task according to its content. While the students stated they had problem with making the physical appearance of the task proper, teachers stated that they had student based problems such as irresponsible behaviors, not understanding the task and also the topic, and making the people around do their own performance task.

Key Words

Performance task, content analysis, vocational high school

Bildiri No: 52

Yapay Test Eşitleme Fonksiyonlarını Oluşturmada KullanılanFarklı Ağırlık Sistemlerinin Karşılaştırılması

Serdar ÇAĞLAK*

Özet

Bu çalışmada farklı ağırlıklandırma sistemleri (AS) ile oluşturulmuş yapay test eşitleme fonksiyonları ürettikleri hata miktarlarına göre denk olmayan gruplar ile ortak soru maddeleri içeren test deseni altında karşılaştırılmıştır. Küçük örneklemler üzerinde yürütülen bu çalışmada, test formlarının zorluk dereceleri ve testi alan grupların kabiliyet düzeyleri de göz önünde bulundurularak, testlerin eşitlenmedinde kullanılacak olan en uygun AS'nin (w = 0.3, 0.5, 0.7) bulunması amaçlanmıştır. Pratikte yaygın olarak kullanılan test eşitleme yöntemlerine ait yapay fonksiyonlar üç farklı AS ile oluşturulmuş ve toplamda 16 farklı durum altında incelenmiştir. Yeni ve eski test formları arasındaki zorluk dereceleri 2 seviyede (δ = -0.05 ve +0.05), yeni ve eski test formlarını alan adayların kabiliyet düzeyleri yine 2 seviyede (θ = -0.03 ve +0.03) ve yeni test formunu alan öğrenci sayıları da 4 farklı seviyede (n = 10, 25, 50, ve 100) sabit tutulmuştur. Bilgisayar ortamında 3-parametreli lojistik model kullanılarak öğrencilere ilişkin test dataları oluşturulmuştur. Yeni ve eski test formları, 36'sı ortak olmak üzere toplamda 120 soruluk çoktan seçmeli soru maddelerinden oluşmaktadır. Öğrenci popülasyonuna ait test puanları Chained Equipercentile (CE) yöntemi ile eşitlenmiş ve Identity (ID), Nominal Weights Mean (NWM) ve Tucker Linear (TL) test eşitleme yöntemleri, ürettikleri sistematik hata (SH) ve rastgele hata (RH) bakımından karşılaştırılmıştır. Çalışmadaki bulgulardan yola çıkarak araştırmacılar için önerilerde bulunulmuştur.

Anahtar Kelimeler

Küçük Örneklemler, Test Eşitleme, Birleşimli Eşitleme Fonksiyonları, Ağırlıklandırma Sistemleri

* Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme ABD. İletişim: scaglak@ogu.edu.tr Tel: +90 222 229 3123 Ext. 1675

Paper No: 52

Comparison of Different Weighting Systems forCreating Synthetic Equating Functions

Abstract

The purpose of this study is to determine the optimum weighting systems used to create synthetic equating functions in order to equate test forms using very small samples under Non-Equivalent Groups Anchor Test Design (NEAT). The synthetic equating functions (SEF) were created using five different weighting systems (w = 0.3, 0.5, 0.7), and compared under variety of simulation conditions. Test forms' mean difficulty levels (δ = -0.05 & +0.05) and examinee groups' mean ability levels (θ = -0.03 & +0.03) were varied in each condition. Using 4 levels of examinee sample sizes (n = 10, 25, 50, 100), a total of 16 simulation conditions were established to compare the SEFs of Identity (ID), Nominal Weights Mean (NWM) and Tucker Linear (TL) equating methods in terms of the amount of the equating error produced. Three parameter logistic IRT model (3-PL IRT) was used to generate item responses of the examinees. Test forms which contain 120 items (36 common and 84 unique) were used in the equating procedures in each simulation iteration. Chained Equipercentile (CE) with 6-moments univariate and 2-moments bivariate loglinear presmoothing was used to establish the criterion equating function between the test forms taken by 50.000 examinees in the population. Small samples were randomly drawn from the population data to equate the test forms using 10 equating methods including the SEFs. The accuracy of equating results obtained through those methods were compared in terms of the magnitudes of the random error (standard error equating), systematic error (equating bias), and the total equating error (root mean squared error). Based on the findings, recommendations were given for practitioners.

Key Words

Small Samples, Test Equating, Synthetic Equating Functions, Weighting Systems

Bildiri No: 53

Test Eşitlemede Çok Boyutluluğun Eş Zamanlı ve Ayrı Kalibrasyona Etkisi

Neşe ÖZTÜRK GÜBEŞ *

Özet

Bu araştırmanın amacı, çok boyutluluğun eş zamanlı ve ayrı kalibrasyon yapılarak elde edilen eşitlenmiş puanlara etkisini incelemektir. Araştırma simülasyon verileri kullanılarak yürütülmüştür. Araştırma kapsamında [4 (çok boyutluluk düzeyi: 1.00, 0.90, 0.75 ve 0.50) x 2 (kalibrasyon yöntemi: eş zamanlı ve ayrı kalibrasyon) x 2 (ölçek dönüştürme yöntemi: Stocking-Lord ve Haebara) x 2 (test eşitleme yöntemi: Madde Tepki Kuramı gerçek puan eşitleme ve gözlenen puan eşitleme)] olmak üzere toplam 32 koşul incelenmiştir. Çok boyutluluk testlerin Θ_1 ve Θ_2 olmak üzere farklı iki yeteneği ölçtüğü varsayılarak oluşturulmuştur. İki yetenek arasındaki korelasyonun 1.00 olduğu durum tek boyutlu koşulu, 0.50 olduğu durum çok boyutluluğun derecesinin en yüksek olduğu koşulu temsil etmektedir. Eşdeğer olmayan gruplar ortak test deseni altında testler birbirine eşitlenmiştir. Elde edilen eşitlenmiş puanlar Bias, RMSE ve SD ölçütleri kullanılarak değerlendirilmiştir. Araştırmanın sonucunda, tek boyutlu koşul altında en yansız eşitleme sonuçlarının Stocking-Lord yöntemi ile en yanlı sonuçların ise eş zamanlı kalibrasyon ile elde edildiği görülmüştür. Çok boyutluluğun derecesinin en yüksek olduğu koşullarda en az eşitleme hatasına Haebara yöntemi sahip olmuştur. Tek boyutlu koşul altında eş zamanlı ve ayrı kalibrasyon yapılarak elde edilen eşitlenmiş puanlar benzer SD değerine bir diğer deyişle random hataya sahip olurken çok boyutluluğun derecesinin ciddi olduğu koşullarda en fazla random hataya Stocking-Lord yöntemi sahip olmuştur.

Anahtar Kelimeler

Test eşitleme, ayrı kalibrasyon, eş zamanlı kalibrasyon, çok boyutluluk

*Mehmet Akif Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: nozturk@mehmetakif.edu.tr
Tel: +90 248 213 4000

Paper No: 53

The Effect of Multidimensionality on Concurrent and Separate Calibration in Test Equating

Abstract

The aim of this study is investigating the effects of multidimensionality on equating results which obtained from separate and concurrent calibration methods. The study was conducted with using simulated data. In the scope of research, totally 32 simulation conditions [4 (degree of multidimensionality: 1.00, 0.90, 0.75, and 0.50) x 2 calibration methods (separate and concurrent) x 2 (scale transformation methods: Stocking-Lord and Haebara) x 2 (test equating methods)] were examined. Multidimensionality was constructed as assuming the two test forms were measuring Θ_1 and Θ_2 abilities. While the simulation condition which has correlation between abilities 1.00 represents the unidimensional case, the correlation between abilities 0.50 represents the severe multidimensional case. Tests were equated under common-item non-equivalent groups design. Equating results were evaluated by using Bias, RMSE, and SD evaluation criteria. The results showed that the Stocking-Lord method had the smallest bias value; the concurrent calibration had the largest bias values when the data were unidimensional.

Haebara method had the smallest RMSE values when the multidimensionality was severe. Under the unidimensional case, the equated results which provided from concurrent and separate calibrations had the similar SD values but as the data got multidimensional Stocking-Lord method had the largest random error.

Key Words

Test equating, separate calibration, concurrent calibration, multidimensionality

Bildiri No: 54

PISA 2012 Türkiye UygulamasındaKararlılık, Problem Çözmeye Açıklık veAkademik Başarı Arasındaki İlişkilerin Belirlenmesi

Ömer KUTLU[*], Seval KULA KARTAL[**], Neslihan Tuğçe ŞİMŞEK[***]

Özet

Bu çalışmada PISA 2012 matematik, fen ve okuma konu alanlarında başarı bakımından alt ve üst %25'lik kesimi oluşturan gruplarda kararlılık ve problem çözmeye açıklık değişkenlerinin öğrenci başarısı üzerindeki yordayıcı rolleri incelenmiştir. PISA 2012 Türkiye örnekleminde yer alan 4848 öğrenciden, kayıp veri içerdiği için indeks değerleri hesaplanmayan öğrenciler çıkarıldıktan sonra çalışma grubunda 3190 öğrenci yer almıştır. Her bir konu alanındaki başarı dağılımının alt ve üst %25'lik kesimini oluşturan gruplarda, kararlılık ve problem çözmeye açıklık değişkenlerinin yordayıcı, başarının yordanan değişken olduğu regresyon modeli, kantil regresyon analizi kullanılarak incelenmiştir. Araştırma sonucunda kararlılık ve problem çözmeye açıklık değişkenleri arasında orta düzeyde pozitif anlamlı bir ilişki bulunmuştur. Kantil regresyon analizi sonuçları tüm konu alanlarında alt %25'lik kesimde kararlılık değişkeninin; üst %25'lik kesimde ise problem çözmeye açıklık değişkeninin başarıyı daha iyi yordadığını göstermektedir. Bu bulguya dayanarak başarı düzeyi düşük öğrencilerin daha üst düzeyde başarı gösterebilmesi için, öğretmenlere öğrencilerin başarıya ulaşmada gösterdikleri çabanın önemini anlamalarını sağlamaya dönük sınıf içi etkinlikler yapmaları, durum belirleme sürecinde ise öğrencilerin farklı stratejileri işe koşmalarını sağlayacak, çaba göstermelerini gerektirecek güçlük düzeyindeki görevlerin kullanılması önerilebilir. Ayrıca, öğretmenlere verdikleri görevlerin geribildiriminde, yalnızca bilişsel becerilerine odaklanarak başarıyı değerlendirmekten kaçınmaları, öğrencilerin görevi tamamlamada gösterdikleri çabaya değer vermeleri, daha fazla çaba ile geliştirilmesi gereken yönler ve başarı ile tamamlanmış noktalara odaklanmaları da önerilebilir.

Anahtar Kelimeler

Öğrenci başarısı, Kararlılık, Problem çözmeye açıklık, PISA, Kantil regresyon analizi

[*] Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı. E-posta: omerkutlu@ankara.edu.tr

[**]Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı. E-Posta: seval.kula@hotmail.com

[***]Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalı. E-posta: simsekneslihantugce@gmail.com

Paper No: 54

Identifying the Relationships BetweenPerseverance, Openness to Problem Solving andAcademic Success in PISA 2012 Turkey

Abstract

In this study, the predictive roles of perseverance and openness to problem solving on success in the lower and upper quartiles of score distribution are analyzed in scientific, mathematical and reading domains of PISA 2012 Turkey sample. Since their index values can't be calculated due to missing values, some of the students are excluded from the Turkish sample of PISA 2012 study included 4848 students and after that the sample of this study is comprised of 3190 students. Regression models in which perseverance and openness to problem solving are predictive variables and success is response variable are formed for the lower and upper quartiles of score distribution of each domain and the models are analyzed by using the quantile regression analysis. At the end of the study, it is founded that there is positive, medium level, significant relationship between the perseverance and openness to problem solving. The results of quantile regression analysis show that perseverance predicts success better in the lower quartile of score distribution while openness to problem solving predicts success better in the upper quartile of score distribution in all domains. Based on these results, planning in class activities enabling students to understand the importance of the effort is recommended for teachers in order to increase performance of students who are in the low-performing group. In the assessment processes, it is suggested to use the tasks that require students to put in effort and use different strategies. While giving feedback to tasks, it is recommended to avoid from evaluating the success by only focusing on students' cognitive skills. Valuing the effort students put in to complete the task and focusing on the points that should be developed by showing more effort are also recommended.

Key Words

Students achievement, Perseverance, Openness to problem solving, PISA, Quantile regression analysis

Bildiri No: 55

Öğretmenlerin Öğretim Becerileri Algısı Ölçeğinin Geliştirilmesi

Ali SABANCI[*], Kemal KAYIKÇI[**], İzzet ÖZDEMİR[***]

Özet

Öğretmenlerin öğretim süreciyle ilgili sahip olması gereken yeterlikler ve bu yeterliklere sahip olma konusunda öğretmenlerin algıları, bazı kaynaklarda öğretmen yeterlikleri, bazı kaynaklarda öğretmen öz yeterlik algısı, bazı kaynaklarda öğretmen öz yeterlik inancı olarak incelenmiştir. Ayrıca öğretmenlerin öğretim uygulamaları konusunda da araştırmaların olduğu görülmektedir. Öğretmen yeterliklerinin sınıf ortamına somut yansıması olarak düşünülebilecek olan öğretim becerileri, öğretmenin öğretimi planlama, uygulama, değerlendirme ve bu süreçte sınıf yönetimiyle ilgili sergilediği davranışlar bütünü olarak tanımlanabilir. Öğretmenlerin etkili bir öğretim ortamı yaratabilmeleri içinöğretmenlik becerisine sahip olduklarına inanmaları gerekir. Bu çalışmanın amacı, öğretmenlerin öğretim becerileri algılarını belirlemeye yönelik geçerli ve güvenilir bir ölçek geliştirmektir. Araştırmanın çalışma grubunu Antalya şehir merkezinde bulunan Aksu, Döşemealtı, Kepez, Konyaaltı ve Muratpaşa ilçelerinde görev yapan yaklaşık 400 öğretmen oluşturmaktadır. Öğretmenlerin öğretim sürecine yönelik yeterlikleri, bu yeterliklere ilişkin öğretmenlerin öz yeterlik inançları ya da algıları, öğretim uygulamaları ve öğretim sürecinin aşamaları konusunda alanyazın incelemesi yapılarak; öğretmenlerin öğretim becerileri konusunda madde havuzu oluşturulmuştur. Ölçeğin kapsam geçerliğini sınamak için Akdeniz Üniversitesi Eğitim Fakültesinde görevli sekiz akademisyenin uzman görüşlerine başvurulmuştur. Uzmanların görüşlerine göre taslak ölçekte gerekli düzenlemeler yapılmıştır. Daha sonra çalışma grubunu oluşturan öğretmenlere uygulanmıştır. Ölçeğin yapı geçerliğini saptamak amacıyla faktör analizi yapılacaktır. Ölçekte yer alan maddeler "Açımlayıcı Faktör Analizi" tekniklerinden "Temel Bileşenler Analizi" ile incelenecektir. Açımlayıcı faktör analizinden sonra elde edilen faktörlerin, kuramsal yapıyı ne ölçüde desteklediği ise "Doğrulayıcı Faktör Analizi" ile test edilecektir. Ölçeğin güvenirlik çalışmaları ise Cronbach Alpha İç Tutarlılık Katsayısı hesaplanacaktır. Yapılan çalışmanın sonucunda elde edilen "Öğretmenlerin Öğretim Becerileri Algısı Ölçeğinin", öğretmenlerin öğretim becerilerine yönelik algılarını ölçebilecek geçerli ve güvenilir bir ölçme aracı olup olmadığı değerlendirilecektir.

Anahtar Kelimeler

Öğretim becerileri, Öğretmen yeterlikleri, Öğretmen öz yeterlik inancı, Ölçek geliştirme

[*] Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: Tel: +90 505 2554300
[**] Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: Tel: +90 505 2381052
[***] Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim: Tel: +90 505 2240415

Paper No: 55

The Development of Teachers' Perception on Teaching Skills Scale

Abstract

Competences related to the teaching process that teachers must haveandtheseperceptions of teachers in having these competences are examined as teacher competences in some sources, teacher self-efficacy perception in other sources or teacher self-efficacy beliefs in some other sources. Also, it is observed that there have been researches on teachers' teaching practices. Teacher competencies in the classroom teaching skills that can be considered as a concrete reflection can be defined as a set for the teacher's behaviour in the process of teacher education planning, implementation, evaluation and classroom management. To create an effective learning environment, teachers need to believe that they have the teaching competencies. The aim of this study is to develop a valid and reliable scale to determine the perceptions of teachers' teaching skills. The study group of this research includes about 400 teachers that work in central districts of Antalya: Aksu, Döşemealtı, Kepez, Konyaaltı and Muratpaşa. An item pool about the teaching competencies of teachers has been formed by conducting a literature researchon competencies of teachers in the teaching process, teachers' self-efficacy beliefs on these competencies or perceptions, teaching practices and phases of teaching process. To test the content validity of the scale, expert opinions of eight academics from Akdeniz University Faculty of Education have been consulted. Necessary arrangements have been made in the draft scale based on the opinions of experts.It has been then applied to the teachers in the study group. Factor analysis will be conducted to determine the construct validity of the scale. The items included in the scale will be examined with "Principal Components Analysis" which is one of the "Exploratory Factor Analysis". After the Exploratory factor analysis, the factors that have been obtained will be tested with the "Confirmatory Factor Analysis" to what extent they support the theoretical structure. The reliability analysis will be calculated with the Cronbach Alpha internal consistency coefficient. As a result of the study, the obtained outcome "the Scale of Teachers' Perception of Teaching Skills" is going to be assessed as to whether it is a valid and reliable instrument to measure the perception of teachers about teaching skills.

Key Words

Teaching skills, Teacher competencies, Teacher self-efficacy belief, Scale development

Bildiri No: 56

Öğretmenlerin Yansıtıcı Düşünme Ölçeğinin Geliştirilmesi

Kemal KAYIKÇI*, İlhan GÜNBAYI**, İzzet ÖZDEMİR***

Özet

Yansıtıcı düşünme; bireylerin geçmiş ve şu anda geçirdiği yaşantılar hakkında derinlemesine düşünerek, kendi öğrenme, öğretme ve düşünme sürecine ilişkin sorgulama yapma, kendini değerlendirme ve bu değerlendirme sonucunda ortaya çıkan sorunları çözmek için neler yapabileceğini düşünmesidir. Yansıtıcı düşünme, uygulamaların iyileştirilmesinin, yeteneklerin geliştirilmesinin ve örgütsel öğrenmenin temelidir. Yansıtıcı düşünme bireye, tecrübeleri ve şimdiki uyguladıklarını birleştirebilmesinde, gelecekle ilgili planlar yapıp içinde bulunduğu durumu daha iyi analiz edip geliştirmesinde yardımcı olabilir. Yansıtıcı düşünme yoluyla öğretmenler, kendilerinin uygulamalarını ve problem çözme becerilerini geliştirirler. Hem alan hem de pedagojik gelişmelerdenhaberdar olur ve yenilikleri eğitim-öğretimdedaha profesyonelce uygularlar. Bu çalışmanın amacı, öğretmenlerin yansıtıcı düşünme düzeylerini belirlemeye yönelik geçerli ve güvenilir bir ölçek geliştirmektir. Araştırmanın çalışma grubunu Antalya şehir merkezinde bulunan Aksu, Döşemealtı, Kepez, Konyaaltı ve Muratpaşa ilçelerinde görev yapan yaklaşık 400 öğretmen oluşturmaktadır. Yansıtıcı uygulama, yansıtıcı düşünme ve yansıtıcı düşünme düzeyleri konusunda alanyazın incelemesi yapılarak; öğretmenlerin yansıtıcı düşünme düzeyleri konusunda madde havuzu oluşturulmuştur. Ölçeğin kapsam geçerliğini sınamak için Akdeniz Üniversitesi Eğitim Fakültesinde görevli sekiz akademisyenin uzman görüşlerine başvurulmuştur. Uzmanların görüşlerine göre taslak ölçekte gerekli düzenlemeler yapılmıştır. Daha sonra çalışma grubunu oluşturan öğretmenlere uygulanmıştır. Ölçeğin yapı geçerliğini saptamak amacıyla faktör analizi yapılacaktır. Ölçekte yer alan maddeler "Açımlayıcı Faktör Analizi" tekniklerinden "Temel Bileşenler Analizi" ile incelenecektir. Açımlayıcı faktör analizinden sonra elde edilen faktörlerin, kuramsal yapıyı ne ölçüde desteklediği ise "Doğrulayıcı Faktör Analizi" ile test edilecektir. Ölçeğin güvenirlik çalışmaları ise Cronbach Alpha İç Tutarlılık Katsayısı hesaplanacaktır. Yapılan çalışmanın sonucunda elde edilen "Öğretmenlerin Yansıtıcı Düşünme Ölçeğinin", öğretmenlerin yansıtıcı düşünme düzeylerine yönelik algılarını ölçebilecek geçerli ve güvenilir bir ölçme aracı olup olmadığı değerlendirilecektir.

Anahtar Kelimeler

Yansıtıcı düşünme, Yansıtıcı düşünme düzeyleri, Ölçek geliştirme

* Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: Tel: +90 505 2381052
** Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: Tel: +90 532 6622951
*** Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim: Tel: +90 505 2240415

Paper No: 56

The Development of Teachers' Reflective Thinking Scale

Abstract

Reflective thinking is that people inquire their own teaching, learning as well as thinking process, evaluate themselves and they think what they can do to solve problems which emerges from as a result of this evaluation by reflecting on past and current experiences. Reflecting thinking is a basis for improving practices, developing skills and organizational learning. Reflective thinking can help people with integrating their experience into their current practices, planning their futures and analyzing actual situation better. Teachers can improve their practices and problem solving skills through reflective thinking. Moreover; they become aware of pedagogical developments as well as field information and they implement educational innovations more professionally. The aim of this study is to develop a reliable and valid scale which is for determining reflective thinking level of teachers. The working group of this study consisted of 400 teachers who are working in Aksu, Döşemealtı, Kepez, Konyaaltı and Muratpaşa districts in Antalya. An item pool was formed on reflecting thinking level of teachers by reviewing literature on reflective thinking and reflective thinking levels. Eight experts who are working at Faculty of Education, Akdeniz University, were consulted to test content validity of the scale. According to experts' opinion, required adjustments were done. After that this scale was implemented to teachers who were in working group. Factor analysis will be done to determine construct validity. The items of the scale will be examined ''Principal Components Analysis'' which is one of ''Exploratory Factor Analysis'' techniques. The factors which will be got after Exploratory Factor Analysis, will be tested what extent they support theoretical background of the topic through ''Confirmatory Factor Analysis''. Cronbach Alpha internal consistency value will be calculated to determine reliability of the scale. As a result of the study, ''The teachers' reflective thinking scale'' will be evaluated whether it is a reliable and valid instrument which can determine the perception of teachers on reflective thinking levels.

Key Words

Reflective thinking, Reflective thinking levels, Scale development

Bildiri No:57

Öğretmenlik Mesleği Sorun Algısı Ölçeği (ÖMSAÖ)'nin Geçerlik Güvenirliğinin Klasik Test Kuramına ve Madde Tepki Kuramına Dayalı Olarak Karşılaştırılması Üzerine Bir Çalışma

Erol KARACA*

Özet

Bu çalışmanın temel amacı Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) altında Öğretmenlik Mesleği Sorun Algısı Ölçeğinin (ÖMSAÖ) kestirilen madde ve test parametrelerinin karşılaştırılarak, araştırmacılara hangi kuram altında kestirilen parametrelerin daha keskin ve güvenilir olduğunu sunmaktır. Bu amaçla çalışmada, öğretmenlerin mesleki sorunlarının saptanmasında kullanılan Likert tipi dereceleme ölçeğinden elde edilen ölçmelerin KTK ve MTK dayalı olarak geçerlik ve güvenirlik düzeyleri saptanmış ve birbiriyle karşılaştırılmıştır. Araştırmanın çalışma grubunu ise, Milli Eğitim Bakanlığı'na bağlı ilköğretim ve ortaöğretim kurumlarında görev yapan 6 ildeki (Kütahya (160), Bursa (38), Burdur (31), Ankara (25), Eskişehir (5) ve Balıkesir (3)) öğretmenlerden basit tesadüfi örneklem yoluyla seçilmiş 274 öğretmen oluşturmaktadır. 12 ölçek ise, yönergeye uygun şekilde doldurulmadığından analiz dışında tutulmuştur. Katılımcıların % 33.20'si (87) sınıf öğretmeni, %17.90'ı (47) matematik öğretmeni, %13.40'ı (35) Türkçe öğretmeni, % 6.90'ı (18) tarih öğretmeni, % 5.70'i (15) kimya öğretmeni, % 4.60'ı (12) İngilizce öğretmeni, % 3.10'u (8) fizik öğretmeni, % 2.70'i (7) biyoloji öğretmeni, % 1.50'si (4) sosyal bilgiler öğretmenidir. Geriye kalan 23 (%9.07) öğretmenin branşlarının ise, din kültürü ve ahlak bilgisi (4), görsel sanatlar (2), Türk dili ve edebiyatı (1), coğrafya (3), felsefe (3), rehberlik (2), beden eğitimi (1), bilgisayar teknolojisi (3), elektrik (1), teknoloji ve tasarım (3) olduğu görülmektedir. Öğretmenlerin mesleki sorunlarının ölçülmesi amacıyla 24 maddeden oluşan 5'li Likert tipi bir ölçek kullanılmıştır. Ölçek, KTK ve MTK yöntemiyle puanlanmış ve elde edilen puanların karşılaştırması yapılmıştır. Sonuç olarak her iki kuram altında kestirilen madde ve test parametrelerinin yüksek düzeyde ilişkili olduğu, model varsayımların karşılandığı ve sınırlılıkların göz önüne alındığı durumlarda her iki kuramında kullanılabileceği sonucuna ulaşılmıştır.

Anahtar Kelimeler

Öğretmenlik mesleği sorun algısı, klasik test kuramı, madde tepki kuramı, aşamalı tepki modeli ve test geliştirme

* Anadolu Üniversitesi Sağlık Bilimleri Fakültesi Sosyal Hizmet Bölümü. İletişim: erolkaraca@anadolu.edu.tr
Tel: +905059111528

Paper No: 57

A Study on Comparison of the Reliability and Validity of the Scale of the Problem Perception of Teaching Profession Based on the Classical Test Theory and Item Response Theory

Abstract

The primary objective of the present research was to investigate the test and item parameters Scale of the Problem Perception of Teaching Profession (SPPTP) under Classical Test Theory (CTT) and Item Response Theory (IRT) and submit the best model which was fitted to real data. Fort this purpose in this study, the levels of reliability and validity of the measurements which were obtained from the Likert type, rating scale used for determining the perceptions of the teachers towards the professional problems based on CTT and IRT and these levels were compared among them. The research was carried out with 262 teachers, working in primary and secondary schools depending on the Ministry of education in 6 cities (Kütahya (160), Bursa (38), Burdur (31), Ankara (25), Eskişehir (5) ve Balıkesir (3)), Turkey, in the 2009. Twelve questionnaires were omitted, because they were not filled out according to the instructions. Participants were 262 teachers working in these cities. Eighty-seven participants (33.20%) were grade teachers, fourty-seven participants (17.90%) were mathematics teachers, thirty-five participants (13.40%) were Turkish teachers, eighteen participants (6.90%) were history teachers, fifteen participants (5.70%) were chemistry teachers, twelve participants (4.60%) were English teachers, eight participants (3.10%) were physics teachers, seven participants (2.70%) were biology teachers, four participants (1.50%) were social studies teachers and the remaining 23 (9.07%) were teachers of religious culture and ethics (4), arts and crafts teachers (2), Turkish Language and Literature teacher (1), geography teachers (3), philosophy teachers (3), guidance teachers (2), physical education and sports teacher (1), computer and instruction Technologies teachers (3),electricity teacher (1) and technology and design teachers (3). By the aim of the measuring, the teachers' professional problems, a 5-point Likert-type scale that was composed of 24 items, was used. Scoring has been made according to CTT and IRT, and then scores have been compared. As a result, it can be concluded that item and test parameters were highly correlated and researchers could use two methods if their assumptions were met and limitations were considered.

Key Words

The problem perception of teaching profession, classical test theory, item response theory, graded response model and test development

Bildiri No: 58

Cronbach Alfa ve Mcdonalds ω Değerlerinin Gerçek Verilerde Karşılaştırılması

Sibel AYDOĞAN[*], Nermin KIBRISLIOĞLU UYSAL[**], Nuri DOĞAN[***]

Özet

Araştırmanın amacı tek boyutlu gerçek verilerde maddelerin betimsel özelliklerinin incelenmesi ve ölçümlerin konjenerik olması durumunda α ve ω katsayılarının karşılaştırılmasıdır. Araştırmada dört farklı veri seti kullanılmıştır. İlk aşamada verilerin yapısı incelenmiş ve konjenerik olup olmadığına bakılmıştır. Sonrasında ise α ve ω katsayıları hesaplanarak karşılaştırma yapılmıştır. α katsayısı varyans (α1) ve faktör yükleri (α2) cinsinden olmak üzere iki farklı yöntemle hesaplanmıştır. Araştırmanın bulguları kullanılan veri setlerinin tamamının konjenerik olduğunu göstermiştir. Güvenirlik katsayıları karşılaştırıldığında tüm veri setleri için en yüksek güvenrilik katsayısının faktör yükleri kullanılarak hesaplanan α2 katsayısı olduğu ve onu sırasıyla ω ve α1 katsayılarının takip ettiği bulunmuştur. Ancak ω ve α1 katsayıları arasındaki farkın pratikte önemli büyüklükte olmadığı görülmüştür.

Anahtar Kelimeler

Cronbach alfa, Mcdonalds ω, Konjenerik ölçme

[*] Hacettepe Üniversitesi Eğitim Bilimleri Bölümü. İletişim: sibeldemirbilek@hacettepe.edu.tr Tel: 05438822404
[**] Hacettepe Üniversitesi Eğitim Bilimleri Bölümü. İletişim: nkibrislioglu@hacettepe.edu.tr Tel: 05543900302
[***] Hacettepe Üniversitesi Eğitim Bilimleri Bölümü. İletişim: nurid@hacettepe.edu.tr

Paper No: 58

Comparison of Cronbach alpha and Mcdonalds ω Reliability Coefficients in Real Data

Abstract

The purpose of this study was to investigate descriptive characteristics of unidimensional real data sets and comparison of α and ω reliability coefficients in case of congeneric data sets. In the study four data sets were used. In the first phase, descriptive characteristics of data sets were examined and determined whether they are congeneric or not. After, α and ω reliability coefficients were calculated and compared. α coefficient were calculated in two ways: variance based (α1) and factor loadings based (α2). The results of the study indicated that all data sets used in the study were congeneric. When reliability coefficients were compared, for all data sets it is indicated that the highest reliability value was factor loading based α2 coefficient following ω and variance based α1 coefficient in order. However, the difference between ω and α1 coefficients was not significant in practice.

Key Words

Cronbach alpha, Mcdonalds ω, Kongeneric measurement

Bildiri No: 59

Türkiye Cumhuriyeti İnkılâp Tarihi ve Atatürkçülük Ders Kitabı Ölçme ve Değerlendirme Sorularının İncelenmesi

Ümmügülsüm CANDEĞER*

Özet

Eğitim, bireyin hayata hazırlandığı süreç olarak görülebilir. Bu sürecin gidişatının ise izlenmesi ve değerlendirilmesi gerekir. Eğitimin izlenmesi ve değerlendirilmesini ölçme olarak nitelendirmek mümkündür. Ölçmeye bağlı olarak edinilmesi gereken bilgi, beceri ve tutumların ne seviyede edinildiği ya da edinilmediği tespit edilebilir. Bu tespit işleminden sonra ise değerlendirme boyutuna geçilir ki burada geriye dönük olarak öğrencinin edinemediği bilgi, beceri ve tutumları edinmesi için öğretmenlerin iyileştirme çalışmaları yapmaya ihtiyacı olur.

Bloom Taksonomisi ölçmenin sağlıklı bir şekilde yapılıp yapılmadığının düzeylere göre değerlendirilmesini sağlayabilir. Bloom Taksonomisinin yenilenmiş formuna göre hatırlama, anlama, uygulama, analiz etme, değerlendirme ve yaratma düzeyleri üzerinden değerlendirildiği zaman öğrencilerin öğrenip, öğrenmedikleri kontrol edilerek yapılacak çalışmalar buna göre yeniden düzenlenebilir.

Okullarda eğitim ve öğretim faaliyetleri bir program dâhilinde yapılmaktadır. Her öğretmenin görevi, derse girdiği sınıfın seviyesine uygun olarak kendisine verilmiş olan programı zamanında uygulamaktır. Programların en can alıcı noktası ise kazanımlardır. Ölçme ve değerlendirme işlemlerinin kazanımlar baz alınarak hazırlanması ise öğrencilerin program dahilinde öğrenmelerinin değerlendirilmesini sağlayabilir.

Bu çalışmada Millî Eğitim Bakanlığı'na bağlı özel ve devlet liseleri 11. sınıf öğrencilerinin ders kitabı olan T.C. İnkılâp Tarihi ve Atatürkçülük ders kitabında bulunan ünite sonu değerlendirme soruları üzerinde durulacaktır. Ünite sonu soruları Bloom Taksonomisi ve kazanımlar açısından incelenecektir. Sorular Bloom Taksonomisi'ne göre düzeylere ayrılacaktır. Ayrıca ünite sonu soruları Talim ve Terbiye Kurulu Başkanlığı tarafından 2012'de yayınlanan T.C. İnkılâp Tarihi ve Atatürkçülük programında bulunan kazanımlarla eşlenecektir. Doküman inceleme yöntemi ile yapılacak olan çalışma T.C. İnkılâp Tarihi ve Atatürkçülük ders kitabında bulunan 7 ünite sonu ölçme ve değerlendirme sorularıyla sınırlıdır. Çalışma sonunda elde edilen veriler soruların düzeyleri ve kazanımlarla uyumuna bakılarak eksik olan düzeylerle ilgili örnek sorular oluşturulacaktır.

Anahtar Kelimeler

Bloom Taksonomisi, T.C. İnkılâp Tarihi ve Atatürkçülük Ders Kitabı Soruları, Ölçme ve Değerlendirme, Kazanım, Program.

* MEB, Çankaya Cumhuriyet Mesleki ve Teknik Anadolu Lisesi, Tarih Öğretmeni, İletişim: gcandeger@gmail.com Tel: +90 532 7991260

Paper No: 59

Analysis of Assesment&Evaluation Questions of The Republic of Turkey and Atatürk Revolution History Textbook

Abstract

Training can be seen as a process in which individuals are prepared for life. The trajectory of this process must be monitored and evaluated. The evaluation and the monitoring of training could be called assessment. By means of assessment, it could be determined what level of knowledge, skills, attitudes are gained or not gained. After the assessment process, the evaluation process begins in which teachers need to make improvement works retrospectively in order to make the student gain knowledge, skills and attitudes that he/she failed to learn before.

Bloom Taxonomy may evaluate according to levels whether the assessment was made properly. According to Bloom Taxonomy of rewised form recall, comprehension, application, analysis, evaluation, and when evaluated on creating levels students learn of work to be done by controlling their learning rearranged accordingly.

Education and educational activities in schools are conducted within a program. The task of each teacher is to implement the program on time given to him/her depending on the level of the classroom. The crucial point of programs are gains. The preparation of assessment and evaluation processes based on gains could ensure the evaluation of the learning processes of students within the program.

In this study, the assessment questions at the end of units of Republic of Turkey and Atatürk Revolution History Textbook of 11th graders of private and public high schools connected to the Ministry of Education is going to be emphasised. Assessment question at the end of units are going to be examined in terms of Bloom Taxonomy and gains. Questions will be divided into levels according to Bloom Taxonomy. Also the questions at the end of units are going to be matched with the gains of Republic of Turkey and Atatürk Revolution History program which was published by the Board of Education in 2012. The study which is going to be done by means of document analysing method is restricted with the assessment and evaluation questions at the end of seven units. Sample questions are going to be prepared on deficits which will be determined by obtained data at the end of the study.

Key Words

Bloom Taxonomy, The Republic of Turkey and Atatürk Revolution History Textbook's Questions, Measurement and Evaluation, Acquisition Program,

Bildiri No: 60

Hong Kong, Finlandiya, Türkiye ve Tayland Ülkelerine Göre PISA 2012 Matematik Alt Testinin Eşitlenmesinde Grup Değişmezliğinin İncelenmesi

Hatice İNAL*, Duygu ANIL**

Özet

Bu çalışmanın amacı; Hong Kong, Finlandiya, Türkiye ve Tayland ülke verilerine ait PISA 2012 Matematik alt testindeki 7. ve 11. kitapçıkların denk olmayan gruplarda Klasik test Kuramına göre doğrusal eşitleme yöntemleri ile Madde Tepki kuramına dayalı eşitleme yöntemlerine göre grup değişmezlik koşulunu sağlayıp sağlamadığı belirleyerek iki kurama ait eşitleme yöntemlerini karşılaştırmaktır. Bu amaç doğrultusunda grup değişmezliğinin belirlenmesinde kullanılan yöntemlerden REMSD ve RESDk grup değişmezliği belirleme yöntemleri kullanılacaktır. Kitapçıklar tüm grup ve her bir ülkeye göre eşitlenecek ve her bir koşula ait grup değişmezliği değerleri hesaplanacaktır. Araştırma sonucunda eşitleme yöntemlerine göre elde edilen REMSD ve REMDk değerleri incelenip, hangi yöntem veya yöntemler için grup değişmezliği değerlerinin anlamlı ya da anlamsız olup olmadığı sonucuna ulaşılacaktır. Ayrıca araştırma sonucunda Her bir ülke için grup değişmezliğinin sağlanıp sağlanmadığına ilişkin yorumlamalar yapılacaktır. Sonuç olarak da araştırmada elde edilen bulgulara dayalı olarak hangi ülkede ve toplamda hangi eşitleme yöntemin en uygun yöntem olduğu belirlenmeye çalışılacaktır.

Anahtar Kelimeler

Eşitleme, Grup Değişmezliği

*Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Bölümü. İletişim:hinal@hacettepe.edu.tr Tel: +0(312) 297-85 71

**Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:aduygu@hacettepe.edu.tr Tel: +0(312) 297-85 71

Paper No: 60

Investigation of Group Invariance of PISA 2012 Mathematics Sub-Test Equating With Regard to Hong Kong, Finland, Turkey and Thailand

Abstract

The aim of this study is to investigate the group invariance condition by equating the 7th and 11th booklets of the PISA 2012 Mathematics subtest according to Tucker and Levine observed score equating methods from among linear equating methods and mean sigma equating methods from item response theory equating methods . Within this scope, REMSD and RESDk group invariance indexes were employed in the study. Booklets were equated according to total group and each country. At the end of the study, when group invariance values, which were obtained according to equating methods, were compared, it was found out that the method which will give lowest group inavariance values and will be the most suitable method.

Key Words

Equating, Group Invariance

Bildiri No: 62

Kayıp Verilerin Varlığında DMF Analizlerinin Yürütülmesi

Duygu Gizem ERTOPRAK*, Didem ÖZDOĞAN**

Özet

DMF belirleme çalışmalarında analiz işlemleri genellikle bilgi kayıpları yaratmaya eğilimli olan odak gruplar üzerinde yürütülür. Örneğin futbol ile ilgili olan bir madde referans gruptaki erkek cevaplayıcılar tarafından yüksek olasılıkla doğru yanıtlanırken, odak gruptaki kız cevaplayıcılar tarafından yanıtsız kalma eğiliminde olurlar. DMF belirleme analizleri ile ölçme aracı içerisindeki DMF'li maddelerin tespiti yapılabilirken kayıp veriler için herhangi bir işlem yapılamaz. Bu nedenle DMF belirleme çalışmalarında kayıp veri probleminin çözümü için uygun olan yöntemlerin kullanılması DMF'li maddelerin doğru tespiti açısından oldukça önemlidir. Bu bağlamda çalışmanın amacı, farklı kayıp veri başa çıkma yöntemlerinin MH yöntemi ile DMF belirleme sürecine olan etkisinin incelenmesidir. Çalışma kapsamında; yetenek dağılımı, DMF etki büyüklüğü, kayıp veri türü, kayıp veri oranı ve kayıp veri başa çıkma yöntemi olmak üzere 5 tane simülasyon faktörü ve bunların koşulları kullanılmıştır. Bu kapsamda, MTK'nın 3 PL modeline uygun olacak şekilde odak ve referans gruplardaki yetenek dağılımları değişen 1000'er bireyin sadece ilk maddesi farklı miktarlarda DMF gösteren toplam 20 maddeye verdikleri ikili puanlanmış yanıtları içeren tam bir veri seti türetilmiştir. Ardından, farklı kayıp veri mekanizmalarına sahip veri setlerini yaratmak için Finch'in (2011a) çalışmasında kullanılan işlem adımları takip edilmiştir. Bu işlem adımları ile sadece DMF'li olan 1. maddede yaratılan farklı türlerdeki kayıplar, farklı kayıp veri başa çıkma yöntemlerinin (SA, DS, LRÇDA, DFÇDA) kullanılması ile tam veri setlerine dönüştürülmüştür. Bu veri setleri üzerinde MH yöntemi ile DMF belirleme analizleri yapılmıştır. Kayıp veri başa çıkma yöntemlerinin performansı RMSE değerleri üzerinden karşılaştırılmıştır. İncelenen tüm koşullar dikkate alındığında, DMF'li maddeyi ve miktarını doğru tespit etmede en başarılı olan yöntemin, grupların yetenek dağılımlarının negatif (odak grup lehine), DMF büyüklüğünün 0.40, kayıp veri türünün ROK ve oranının %30 olduğu durumda DS; en başarısız olan yöntemin ise, grupların yetenek dağılımlarının negatif, DMF büyüklüğünün 0, kayıp veri türünün RK ve oranının %30 olduğu durumda SA yöntemine ait olduğu gözlenmiştir.

Anahtar Kelimeler

Kayıp Veri, Değişen Madde Fonksiyonu

*Amasya Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: duygugizemertoprak@gmail.com Tel: +903582526230 (3332)

**Eskişehir OsmangaziÜniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: dozdogan@ogu.edu.tr Tel: +905467100205

Paper No: 62

Performing DIF Detection Analysis in The Presence of Missing Data

Abstract

DIF detection analysis are generally performed on focal groups that are tend to create missing values. For instance an item about football are most likely answered correctly by male test takers while female test takers omit them. Detecting the items showing DIF can be found with DIF analysis but it can't be used to solve missing data problems in data set. Therefore using suitable missing data handling methods are highly important to detect accurately the items showing DIF in detection analysis. The aim of this study is to examine the effect of different missing data handling methods on DIF detection process with MH method. In this context examinees ability distribution, DIF effect size, missing data mechanism, missing data rate and missing data handling methods are examined as simulation factors. Data set consists of 20 binary items which have just first item showing DIF scored by focal and reference goups which have each of 1000 examinee was generated based on IRT 3 parameter model in R computational software. Then the scheme used in Finch's study (2011a) to create the data sets which have missing data was also used for this study. This data sets which have missing data were turned to complete data sets with using handling methods (Zero Imputation-ZI, Listwise Deletion-LD, Multiple Imputation by Logistic Regression-MILR, Multiple Imputation by Discriminant Function-MIDF). After completing data sets DIF detection analysis with Mantel-Haenszel (MH) method were performed. Root Mean Square Error (RMSE) was used to interpret the performing of missing data handling methods. Consequently, ZI was obtained as the most successful method to detect the item which contains DIF and its amount accurately when DIF effect size was 0.40, ability distribution was negative, missing data type was MNAR and missing data rate was 30%. Besides that ZI was obtained again as the most unsuccessful method to detect the item which contains DIF and its amount when DIF effect size was none, ability distribution was negative, missing data type was MAR and missing data rate was 30%.

Key Words

Missing data, Differential item functioning

Bildiri No: 63

İlkokul 4. Sınıf için Hazırlanan Araştırma Becerileri Öğretim Programının Etkililiğinin İncelenmesi

Betül POLAT DEMİR*, Ömer KUTLU**

Özet

Bu araştırmanın amacı, ilkokul 4. sınıf için bir "Araştırma Becerileri Öğretim Programı" (ABÖP) geliştirmek ve bu programın araştırma yapma sürecini oluşturan becerileri öğrencilere kazandırması açısından etkililiğini incelemektir. Bu amaçla, araştırmada öncelikle ABÖP geliştirilmiştir. ABÖP'nın geliştirilme sürecinde öncelikle Büyük Altı Araştırma Becerileri Modelindeki altı süreç (hedef tanımı, bilgi arama stratejileri, bulma ve erişme, bilgiyi kullanma, bilgiyi organize etme, değerlendirme) dikkate alınarak araştırma becerilerinin alt becerileri belirlenmiştir. Daha sonra ise bu üst düzey zihinsel beceriyi oluşturan altı alt beceri ile ilgili kazanımlar yazılmıştır. Sınıf düzeyi dikkate alınarak öğrenme alanı belirlenmiştir. Öğrenme alanı/içeriği oluşturulurken sınıf düzeyi dikkate alınarak ilgili alanyazın taranmıştır. Öğrenme alanı da dikkate alınarak her bir kazanıma uygun öğretim etkinlikleri tasarlanmıştır. ABÖP kapsamında öğrenme alanı ve öğretim etkinlikleri dikkate alınarak her bir alt beceri için ders planları hazırlanmıştır. Ders planları, ABÖP'nın öğrenci düzeyine uygunluğu ve bilimsel doğruluğu açısından dört uzman tarafından değerlendirilmiştir. Araştırmada yarı deneysel modellerden ön test - son test eşleştirilmiş kontrol gruplu modelden yararlanılmıştır. Deney grubuna süreç boyunca araştırma yapma becerilerini geliştirmek için araştırmacı ABÖP uygulanmıştır. Kontrol grubundaki öğrencilerle hiç bir uygulama yapılmamıştır. Araştırma deney grubunda 34, kontrol grubunda ise 36 öğrenci olmak üzere toplam 70 öğrenciyle yürütülmüştür. Uygulama deney grubunda beş hafta ve toplam 15 ders saati olarak araştırmacı tarafından yürütülmüştür. Araştırmanın verileri Araştırma Becerileri Testi ve her bir alt beceriyle ilgili aşama tamamlandığında her iki gruba da uygulanan izleme testleri ile toplanmıştır. ABT'de öğretim programı kapsamında belirlenen ilk beş alt beceride yer alan 13 kazanımı ölçen 11 açık uçlu maddeye yer verilmiştir. Araştırmadan elde edilen veriler parametrik yöntemlerden Bağımsız Örneklemler t-testi ve İlişkili Ölçümler t-testi ile çözümlenmiştir. Aynı zamanda öğrencilerin izleme testlerinden aldıkları puanların dağılımları incelenmiştir. Araştırmadan edilen bulgular, ABÖP'nın deney grubu öğrencilerinin araştırma becerilerini kullanma düzeylerini kontrol grubunda bulunan öğrencilere göre manidar düzeyde artırdığını ortaya koymaktadır.

Anahtar Kelimeler

Üst düzey zihinsel beceriler, araştırma becerileri, program değerlendirme

*Niğde Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: betul.polat6006@gmail.com Tel: +5067902221

** Ankara Üniversitesi Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü. İletişim:kutluomer57@gmail.com

Paper No: 63

Determining The Effectiveness of Reseach Skills Teaching Program Prepared For Elementary School Grade 4

Abstract

The aim of this research is to develop a Research Skills Teaching Program (RSTP) for elementary school grade 4 and determine the effectiveness of this program in terms of acquiring research skills. To that end, RSTP was primarily developed. In the development process of Research Skills Teaching Program (RSTP), first sub-skills of research skills were identified by taking into consideration the six stages (task definition, information seeking strategies, location and access, use of information, synthesis, evaluation) of Big Six research Skills Model. Then, objectives regarding the six sub-skills that form this high-level cognitive skill were written. While developing the learning area/context, related literature was reviewed by taking the grade level of the students into consideration. Paying regard to the learning area, teaching activities appropriate to each objective were designed. Within the scope of RSTP, lesson plans for each sub-skill were prepared taking the learning area and teaching activities into consideration. The lesson plans were evaluated by four experts in order to evaluate the RSTP for appropriateness to student level and scientific accuracy at different stages. Pretest-posttest paired control group design, one of the the quasi-experimental designs was used in this research. RSTP was administered to the experimental group during the process in order to develop their research skills. No application was carried out with students in control group. The study was conducted with a total of 70 students, 34 students in the experimental group and 36 students in the control group. The research process in the experimental group were carried out for a total of 15 class hours for five weeks by the researcher. The research's datas were collected by Research Skills Test (RST) and Formative Testswere directed in written to the students in the experimental and control group when stages regarding each sub-skill were completed. the RST included 11 open-ended items measuring 13 objectives of the first five sub-skills that were determined within the scope of the learning program. The data obtained from the ressarch were analyzed with parametric methods like Independent Samples T-Test and Paired Samples T-Test. At the same time, the distribution of students' scores from the formative tests were examined. The findings obtained at the end of the research show that RSTP increases the level of use of research skills of the students in the experimental group significantly, compared to the students in the control group.

Key Words

Higher order thinking skills, research skills, program evaluation

Bildiri No: 64

Öğrencilerin Eğitim Bilimleri Enstitüsü Lisansüstü Programlarına Kabul Durumlarının Yordanması

İlhan KOYUNCU, Nuri DOĞAN, Pelin GÖKDEMİR, Müjgen KAHVECİ

Özet

Bu çalışmanın amacı, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü'nde 2011 ile 2016 yılları arasında lisansüstü programlarına başvuran adayların programlara kabul durumlarını hangi değişkenlerin etkilediğinin lojistik regresyon analizi ile yordanmasıdır. Araştırmada, tezli yüksek lisans ve doktora düzeyinde başvuru yapan adayların başvurduğu anabilim dalı, program, yıl ve yarıyıl, cinsiyet, medeni durumu, yabancı dil puanı, diploma notu, ALES puanı, ALES alanı, mezun olduğu okul, yaşı ve programa kabul durumları ile kurulan modellerin performanslarını karşılaştırmalı olarak incelenmiştir. Bu çalışmada, toplam 4990 adaydan (N_{YL}=3818 and N_{DR}=1172) elde edilen veriler kullanılmıştır. Model değerlendirmede doğruluk, hata oranı, özgünlük ve hassaslık ölçütleri; model karşılaştırmada ise iki oran için z testi kullanılmıştır. Araştırma bulgularına göre lojistik regresyon analizi ile yapılan sınıflandırma yüksek lisans düzeyinde %78; doktora düzeyinde ise %73 doğruluk oranına sahiptir. Wald testi sonuçları, her iki akademik derece için de yaş değişkeninin önemli bir faktör olmadığını göstermiştir ($p>0.05$). Ancak, cinsiyet, medeni durum, diploma notu, ALES puanı, ALES alanı ve mezun olduğu okul değişkenlerinin ise önem düzeyi değişmekle birlikte her iki akademik dereceye kabul için önemli faktörler olduğu görülmüştür ($p<0.05$). Özgünlük değerlerinin yüksek lisans için %94 ve doktora için %91 olması, programa kabul etmeme yönünde verilen kararlarının çok büyük oranda yerinde olduğunu göstermektedir. Ancak, hassaslık değerlerinin yüksek lisans için %28 ve doktora için %35 olması ise, programa kabul etme yönünde verilen kararların yeterli düzeyde hassas olmadığını göstermektedir. Modellerin hata oranları karşılaştırıldığında ise iki oran arasında istatistiksel olarak anlamlı fark bulunmuştur ($p>0.05$). Buna göre, aynı değişkenlerle yapılan sınıflandırmada doktora için kullanılan model daha yüksek bir hata oranına sahiptir. Çalışmada, elde edilen bulgular lisansüstü eğitim alımlarında doğru ve uygun ölçütlerin kullanılması ile akademide kalite bağlamında değerlendirilmiş ve bazı önerilere yer verilmiştir.

Anahtar Kelimeler

Lisansüstü eğitim, Lojistik regresyon, Hassaslık, Özgünlük, İki oran için z testi

Paper No: 64

Predicting Acceptance Status of Students for Institute of Educational Sciences Graduate Programs with Logistic Regression

Abstract

The aim of this study was to predict acceptance status of students to the graduate education at Hacettepe University Institute of Educational Sciences between 2011 and 2016 with logistic regression analysis. The performance of models that were formed by using the information of applicants of master's and PhD degrees including applied department, program, years and semester, gender, marital status, score of foreign language exam, diploma grade, ALES score, ALES field, graduate school, age and acceptance status is compared. The data collected from 4990 candidates ($N_{Master's}$=3818 and N_{PhD}=1172) was used to perform analysis of this basic study. Accuracy, error rate, sensitivity and specificity criteria was to evaluate the models and *z* test for two proportions was used to compare the error rate of the models. The findings revealed that the accuracy ratio of classification made for Master's degree applicants was 78% and it was 73% for PhD ones. The results of Wald test showed that ages of applicants are not a significant factor for logistic regression models of both of degrees ($p>0.05$). However, gender, marital status, diploma grade, ALES score, ALES field and graduate school are significant factors for acceptance to the both of graduate degrees ($p<0.05$). Specificity values of 94% for Master's and 91% for PhD degrees means that the decisions given for rejection to graduate programs are proper due to that very high percentages. Nevertheless, sensitivity values of 28% for Master's and 35% for PhD degrees means that the decisions given for acceptation to graduate programs are not proper because of these low percentages. When the error rates of two models are compared, there is not a statistically significant difference between them ($p>0.05$). That is, the model formed for PhD applications with the same variables has a higher classification error rate than the one for Master's. Finally, the findings of the study are evaluated in the context of using accurate and proper criteria for acceptance to the graduate programs and the quality in scientific research. Eventually, some suggestions for further research are provided.

Key Words

Graduate education, Logistic regression, Sensitivity, Specificity, z test for two proportions

Bildiri No: 65

Farklı Yöntemlerle Ve Birden Çok Puanlayıcı Tarafından Puanlanan Açık Uçlu Matematik Sınavının Çok Yüzeyli Rasch Modeli İle İncelenmesi

Hakan ATILGAN*Pelin DAMLAR DEMİRCİ**

ÖZET

Araştırmanın amacı farklı yöntemlerle ve birden çok puanlayıcı tarafından puanlanan açık uçlu matematik sınavının çok yüzeyli Rasch modeli ile incelenmesidir. Araştırmada kullanılan çok yüzeyli Rasch ölçme modelinde üç yüzey bulunmaktadır. Bu yüzeyler; beş puanlayıcı (biri yüksek lisans dördü lisans mezununu olmak üzere beş ilköğretim matematik öğretmeni), 15 açık uçlu matematik maddesi (bu maddeler bir ölçme değerlendirme uzmanı ve iki alan uzmanına gösterildikten sonra uygulanmıştır) ve 260 7. sınıf öğrencisidir. Betimsel bir çalışma olan bu araştırmada çalışma grubunu İzmir ilinin Ödemiş ilçesinde bulunan 50. Yıl Ortaokulunda öğrenim gören 260 7. Sınıf öğrencisi oluşturmaktadır. Araştırmada her bir puanlayıcı (p) her bir maddeyi (m) tüm bireyler (b) için bir birlerinden bağımsız olarak puanlaması nedeniyle Çok Yüzeyli Rasch modelinde X_{bpm} deseni kullanılarak analizler yapılmıştır. Veri girişi ve düzenlenmesi için Facform analiz için ise FACETS 7.71.4 (Linacre, 2015) bilgisayar programı kullanılmıştır. Hem serbest hem de anahtarla puanlamada analizlerde kullanılan 18975 verinin standartlaştırılmış artık değerin +/-2'den büyük olanlarının sayısı % 5'den küçük olduğu için model-veri uyumunun sağlandığı sonucuna varılmıştır. Serbest puanlamada 1ve 5 numaralı puanlayıcıların katı puanlama yaptığını, 2 numaralı puanlayıcının diğer puanlayıcılara göre daha cömert puanlama yaptığı söylenebilir. Anahtarla puanlamada ise puanlayıcıların yaptıkları puanlamalarda istatistiksel olarak katılık/cömertlik bakımından birbirlerinden farklılıklar göstermediği belirlenmiştir. Hem serbest hem de anahtarla puanlamada bireylerce en kolay bulunan maddenin 1 numaralı; en zor bulunan maddenin ise 14 numaralı madde olduğu belirlenmiştir. Maddelerin arasında 13 numaralı maddenin diğer maddelerin puanlarıyla aynı çizgide olmadığı görülmüştür. Ayrıca yapılan analiz sonucunda serbest ve anahtarla puanlamada bireylerin anlamlı iki yetenek tabakasına ayrıldığı gözlenmiştir. Sonuç olarak ortaöğretimde açık uçlu matematik sorularının değerlendirilmesinde Çok yüzeyli Rasch ölçme modelinin etkin olarak kullanılabileceği sonucuna varılmıştır.

Anahtar Kelimeler

Puanlayıcı Tutarlılığı, Çok Yüzeyli Rasch Model, Açık Uçlu Madde Puanlama

* Ege Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı öğretim üyesi. İletişim: atilgan.h@gmail.com, Tel:+90.532.638 84 27

** Ege Üniversitesi, Sosyal Bilimler Enstitüsü Eğitimde Ölçme ve Değerlendirme Bilim Dalı Yüksek Lisans Öğrencisi. İletişim: pelindamlar@hotmail.com, Tel:+90.505.610 06 82

Paper No: 65

Many-Facets Rasch Model Investigation Of An Open Ended Math Exam Scored By Multiple Raters Using Different Methods

Abstract

The aim of the study is to investigate open-ended mathematics exam rated by more than one rater with different methods through many-faced Rasch model. The Rasch many-faced testing model used in the study has three facets. These facets are five raters (five elementary school mathematics teachers; one with a master's degree, four with bachelor's degree), 15 open-ended mathematics items (these items were used after being reviewed by one testing and two mathematics experts), and 260 7th grade students. The study group of this descriptive research consists of 260 7th grade students of "50. Yıl Secondary School" which is a secondary school in Ödemiş county of İzmir province in Turkey. In the research, since each rater (p) rated each item (m) independently for all participants (b), analyses were carried out in many-faced Rasch model using X_{bpm} pattern. Data entrance and organization will be done on Facform, and FACETS 7.71.4 (Linacre, 2015) computer program will be used for the analyses.

The findings showed that since the number of the standardized residual value of 18.975 data used in both free and key scoring analyses which is higher than +/-2 is lower than 5%, it can be inferred that the model-data agreement is reached. In free scoring, raters 1 and 5 were lower-raters while raters 2 was relatively a higher rate. However, in the key scoring, the raters scoring did not show any statistically significant differences in terms of being low and high graders. In both free and key scoring, it was determined that for the participants, the easiest item was item number 1 while the hardest one was item number 14. Among the items, it was observed that item number 13 was not in the same rank with the points of others. Besides, the analysis showed that the participants differed in two significant skill layers. Finally, it was concluded that the many-faced Rasch model could be used effectively in the evaluation of open-ended mathematics items in secondary schools.

Key Words

Rater Agreement, Many-Faced Rasch Model, Open-Ended Test Scoring

Bildiri No: 67

Bilgisayar Ortamında Bireye Uyarlanmış Testlerde Madde Havuzu Teşhir Oranı, Madde Havuzu Büyüklüğü ve Örneklem Büyüklüğü Arasındaki İlişki Üzerine Bir Araştırma

Alper ŞAHİN*

Özet

Bu çalışma, Bilgisayar Ortamında Bireye Uyarlanmış Testlerde, madde havuzu teşhir oranı, madde havuzu büyüklüğü ve yetenek parametrelerinin kestiriminde önemli yer tutan madde parametrelerinin elde edilmesinde kullanılan farklı örneklem büyüklükleri arasındaki ilişkinin ele alındığı bir Monte-Carlo benzetim çalışmasıdır. Bu amaçla 500 maddeye 10.000 kişinin verdiği yanıttan oluşan bir veri seti, madde tepki kuramının üç parametreli lojistik modelinde CATsim bilgisayar programı yardımıyla üretilmiştir. Üretilen bu veri setinden 250, 500, 1000 ve 3000 kişilik örneklemler alınmış, 100, 200 ve 300 maddelik madde havuzları oluşturulmuştur. Böylelikle 16 farklı araştırma durumu araştırmaya dahil edilmiştir. Her bir veri seti için madde parametrelerinin Xcalibre ile yetenek parametrelerinin ise Post-hoc simülasyonlarla yeniden kestirilmesinin ardından 10.000 kişilik veri setinde bulunan bireyler yetenek düzeylerine göre beş ayrı yetenek grubuna ayrılmış ve her madde havuzu için madde havuzu teşhir oranları örneklem büyüklüğü ve yetenek gruplarına göre ayrı ayrı hesaplanmıştır. Elde edilen bulgulara göre herhangi bir madde kullanım sıklığı kontrol yöntemi kullanılmadığında, 100, 200, 300 ve hatta 500 maddeden oluşan madde havuzlarında bazı yetenek gruplarında tüm soruların kullanıldığı belirlenmiştir. Bunun yanı sıra elde edilen bulgular, oluşturulan madde havuzu belirli bir yetenek grubu için yeterli düzeyde bilgiye sahip olmadığında, madde havuzu ne kadar geniş olursa olsun, bu yetenek grubundaki öğrencilerin maddelerin tümüne ulaşma riski olduğunu ve örneklem büyüklüğünün bu riski azaltmaya yardımcı olmadığını göstermiştir. Sonuç olarak madde havuzu teşhir oranları söz konusu olduğunda, madde havuzunun belirli yetenek düzeyleri için içerdiği bilginin madde havuzu büyüklüğünden, madde havuzu büyüklüğünün ise örneklem büyüklüğünden daha kritik öneme sahip olduğu sonucuna varılmıştır.

Anahtar Kelimeler

Bilgisayar Ortamında Bireye Uyarlanmış Testler, Madde havuzu Teşhir Oranı, Madde Havuzu Büyüklüğü, Örneklem Büyüklüğü

*Ortadoğu Teknik Universitesi, Kuzey Kıbrıs Kampusu, Modern Languages Program Tel:+903926612806

Paper No: 67

A study on the Relationship Between the Bank Exposure Rate, Item Pool Size and Sample Size in Computerized Adaptive Testing

Abstract

This is a Monte-Carlo study that was conducted on computerized adaptive tests to investigate the relationship between the bank exposure rate, item bank size and the calibration sample size used to calibrate the item parameters that are criticial for ability parameter estimation. For this purpose, a data set of 500 items and 10.000 simulees was simulated in three-parameter logistic model of Item Response Theory using CATsim. Calibration samples of 250, 500, 1,000, and 3,000 simulees and item banks of 100, 200 and 300 items were drawn from this simulated data. In this way, 16 research conditions were added to the study. After re-estimation of item parameters with Xcalibre and person parameters through post-hoc simulations for each condition, 10,000 simulees were divided into five ability groups and bank exposure rates for each ability group in each item bank was calculated considering the sample size. The findings showed that when no exposure control method was used, all items in 100, 200, 300 and even in 500-item banks were used for some simulees in some ability groups. In addition to this, findings showed that regardless of the item bank size, if there is not enough information in the bank for a specific ability group, there is still a risk that all items to be used and the sample size cannot reduce this risk.As a result, in terms of bank exposure rate, bank information is more important than the bank size and the bank size is more important than the calibration sample size.

Key Words

Computerized Adaptive Tests, Bank exposure rate, Item Pool Size, Sample Size

Bildiri No: 68

Rehberlik ve Araştırma Merkezlerinde Psikolojik Test Uygulama Süreçleri

Bengü BÖRKAN, Şeyda ÇETİNTAŞ, Betül GÜLCAN, Merve ÖZCAN, Gizem ÖZTEMÜR, Ali VAROL, Osman YILMAZ

Özet

Bu araştırma Türkiye'deki Rehberlik Araştırma Merkezlerinde gerçekleştirilen psikometrik test uygulamaları, testlere ait sonuçların raporlanması, öğrenci değerlendirme süreci, farklılıklar konusundaki esneklik, mesleki yardım ve gelişim ve etik uygulamalar konusundaki mevcut durumu incelemek için yapılmıştır. Bu amaçla Rehberlik Araştırma Merkezlerinde çalışan 15 uygulamacıdan Uluslararası Test Kurulu (ITC) Test Kullanımı Kılavuzu temel alınarak hazırlanan yapılandırılmış görüşme formu kullanılarak nitel veri toplanmıştır. Nitel analizler sonucu test uygulama pratikleri arasında farklılıklar görülmüştür. Bazı kurumlarda fiziksel imkansızlıkların test uygulayıcılarının standart uygulamalardan sapmasına neden olduğu, bazı uygulayıcıların ise temel etik prensiplere ve test uygulama standartlarına gerekli özeni göstermedikleri gözlenmiştir. Bu nedenle, test uygulama etik kuralları ülkemiz koşullarına göre iyi uygulama esaslarıyla geliştirilmelidir. Bu etik kurallar ve iyi uygulama esaslarının test uygulayıcıları tarafından benimsenmesi için de gerekli tedbirlerin alınması gerekmektedir.

Anahtar Kelimeler

Test uygulama, değerlendirme, Rehberlik Araştırma Merkezleri, etik ilkeler, iyi uygulamalar

Bildiri No: 69

Çoktan Seçmeli Test ile Açık Uçlu Sınavın Psikometrik Açıdan Karşılaştırılması

Hatice GÜRDİL EGE*, Duygu ANIL**

Özet

Bu araştırmanın amacı; çoktan seçmeli sınav ile aynı kazanımları ölçmek üzere hazırlanan yazılı yoklamaları psikometrik açıdan karşılaştırmaktır. Bu çalışma, 2015-2016 yılında, İstanbul'daki bir ilkokulun 4.sınıfları üzerinden rastgele seçilmiş 100 öğrenci üzerinden yürütülmüştür. Araştırmada test geliştirme basamakları dikkate alınarak geliştirilen bir çoktan seçmeli test ile aynı kazanımları ölçmek üzere hazırlanan açık uçlu sınav kullanılmıştır. Bu iki sınav aynı öğrenci grubuna iki hafta ara ile uygulanmıştır. Bu çalışmada betimsel istatistikler, Pearson Momentler çarpımı korelasyon, ilişkili t testi ve bağımsız örneklemler t testi kullanılmıştır. Sonuç olarak; öğrencilerin çoktan seçmeli sınav sonuçları ortalaması ile açık uçlu sınav ortalamaları arasında manidar fark bulunmuştur. Öğrencilerin çoktan seçmeli testlerdeki ortalaması daha yüksek çıkmıştır. Ayrıca; açık uçlu yazılı yoklamanın güvenirlik ve geçerliği açısından yararını ve katkısını gösteren kanıtlar elde edilebilmiştir.

Anahtar Kelimeler

Çoktan seçmeli test, açık uçlu sınav, geçerlik, güvenirlik

*Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: haticegurdil1985@gmail.com Tel: +90 (542) 428 96 56

** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: aduygu@hacettepe.edu.tr Tel: + 90 (312) 297 68 20

Paper No: 69

Compare Multiple Choice Test And Open-Ended Exam in Terms of Psychometry

Abstract

The purpose of this study is to compare the open-ended test and multiple choice test in terms of psychometry. This study was based on the 100 forth grade (11 years old) students who was choosen random from at a primary school in İstanbul in 2015-2016. The instruments used in the study are the multiple choice math achievement test and open-ended test form of this test. This two test form was implemented to the same students group in two weeks. In this research descriptive statistics, Pearson Correlation Coefficient, paired samples t test and independent sample t test were used. Consequently, significiant difference was found between open-ended test and multiple choice test. Students' multiple choice test mean is higher than their open-ended test mean. Besides, it was founded evidences show the benefit and contribution in terms of reliability and validity of open-ended test.

Key Words

Multiple choice test, open-ended test, reliability, validity

Bildiri No: 70

Derecelendirilmiş Ölçekler İçin Ölçek Değişmezliği Varsayımına Ampirik Bir Bakış: Bireyler-İçi Ve Bireyler-Arası Farklılıklar İçin Soru Sınıflamaları Ve Cevaplayıcı Profilleri

Nilüfer KAHRAMAN, Derya ÇAVDAR, Esra SÖZER

Özet

Bu çalışma, sosyal-duygusal-kültürel özelliklerin ölçülmesinde, genellikle tipik performans testlerinde kullanılan derecelendirme ölçeklerinin kalite kontrolünde kullanılabilecek yeni bir yöntemi önermektedir."Tipik" cevapları puanlayan ölçekler de, diğer ölçekler gibi, elde edilen gözlemlerinin anlamlılığı için, varsaymak zorundadır ki, kullanılan derecelendirme ölçeği, bireyden bireye veya her birey için sorudan soruya aynı şekilde işlemektedir. Bu çalışma,bu ölçek değişmezliği varsayımının test edilmesine olanak sağlayabilecek ve birey-içi ve bireyler-arası benzerlik ve farklılıkların çalışılmasında kullanılabilecek yeni bir yöntem önermektedir.

Önerilen yöntem, veri kalitesinin yükseltilmesi amacı ile geliştirilmiştir ve en az iki aşamalı bir uygulama ve soru ve bireyler için an az iki açılı bir gruplama modelini içermektedir. Uygulamanın 1. aşaması geleneksel derecelendirilmiş ölçek uygulamasıdır. Mevcut araştırmaya konu olan uygulama ölçeği için, bu aşamada, 104 kişilik uygulama grubu, "mutluyum", "yaratıcıyım" gibi ifadelerden oluşan 15 maddelik bir ölçek için, "kesinlikle katılmıyorum" için 1'den, "kesinlikle katılıyorum" için 5'e kadar rakamlar vererek kendi durumlarını derecelendirmişlerdir. Uygulamanın 2. aşaması detaylı derecelendirilmiş ölçek uygulamasıdır. Bu aşamada, aynı uygulama grubu, zamanlarının yüzde kaçını maddede verilen duyguyu çok az, az, nötr, çok, çok fazla hissederek geçirdiklerini yüzdeler vererek derecelendirmişlerdir. Mutluyum ifadesi için, örneğin, bu derecelendirme, toplamda %100 edecek bir şekilde, "çok mutsuz", "mutsuz", "nötr", "mutlu", "çok mutlu" ibarelerine %5, %10, %45, %30, ve %10 gibi, yüzdeliklerin dağıtılmasını gerektirmiştir.

Elde edilen detaylı yüzdelik dağılım puanları, bireyler-arası sorular-içi ve bireyler-içi sorular-arası ölçek değişmezlikleri test etmek amacıyla çalışılmıştır. Sonuçlar göstermektedir ki, geleneksel tek rakam derecelendirme verileri, buz dağının sadece su üstündeki kısmını göstermektedirler. Bu çalışmanın önemli bulgularından biri, öğrencilerin geleneksel 5'li derecelendirme için verdikleri değerin aslında genelde nasıl hissettikleri ile ilgili bir dağılımın tepe noktası olduğudur. Bu göstermektedir ki, en az iki aşamalı ve detaylı ölçek çalışmaları sorular-içi bireyler-arası ve sorular-arası bireyler-içi ölçek farklılıklarının çalışılmasına izin vermeleri açısından oldukça faydalıdırlar. Bulgular, ölçek değişmezliği çalışmalarında aşamalı bir uygulama deseni kullanmasının ve soru sınıflamaları ve öğrenci profili modellerinin çalışılmasının önemini destekler niteliktedir.

Anahtar Kelimeler

21. yüzyıl becerileri, duyuşsal beceriler, ölçekleme, sınıflama analizi, profil analizi

Paper No: 70

An Empirical Study Of Scale Invariance For Ordinal Scales: Classifying Items And Profiling Respondents To Better Understand Within- And Between-Person Scale Differences

Abstract

Whether or not we can properly define and measure 21. century skills is an important topic in recent literature. However, there is a lack of any empirical evaluations as to whether the ordinal scales used in this research field perform equivalently over persons. This study proposes a new methodology to study ordinal measurement invariance when testing with self-report Likert-type scales. The methodology involving a layered testing design and respondent profiling is illustrated using a scale, which is being developed to assess a number of 21. century skills, such as, motivation and resilience. The illustration data consisted of 104 university students who filled out the Scale using two different scales. First, as the conventions dictates, they were asked to mark one single number for each item from 1 to 5, indicating to what degree they agreed with the feeling given in the item, 5 corresponding to the most agreement. Next, they were asked to write how much of their time, in general, was spent feeling the emotion, using percentages, into a 5 point scale, this time, as a distribution that would sum to 100 in total. The resulting layered-ratings were examined for items and persons and were modeled using cluster analyses and profiling to better understand sub-domain, sub-group ties of differences, when observed. The results indicate that the proposed assessment strategy was helpful in providing an ample amount of empirical evidence to assess whether and the degree to which between- and within-item measurement invariance assumption was held across items. Cluster analysis and profiling were useful in assessing whether there were qualitative differences between persons, which, in turn, was helpful in evaluating why scale measurement invariance was held differently for subgroups, when observed. The implications for improving the overall data quality of research that mainly relies on ordinal measurement instruments are discussed.

Key Words

Ordinal scales, Likert, scale invariance, measurement invariance, respondent profiling

Bildiri No: 71

Teog Başarı Puanı, Özgüven, Özyeterlik Ve Motivasyon Düzeyi Arasındaki İlişkinin İncelenmesi

H. Gonca USTA*

Özet

Araştırma kapsamında öğrencilerin TEOG başarı puanları, özgüven, özyeterlik ve motivasyon düzeyleri arasındaki ilişkinin incelenmesi amaçlanmıştır. Bu amaç doğrultusunda Sivas İlinde 2015-2016 öğrenim yılında öğrenim gören 9. Sınıf öğrencilerine ilgili ölçekler uygulanmıştır. Araştırma amacı doğrultusunda kuramsal model kurgulanmış ve bu model analiz edilmiştir. Yola çıkılan kuramsal modelin ilk analizinde dışsal özgüven değişkeni ile özyeterlik değişkeninin alt boyutları olan "Yılmama", "Başlama" ve "Sürdürme" değişkenleri arasındaki ilişki ile "Yılmama" ve "Başlama" değişkenleri ile TEOG puanı arasındaki ilişkinin manidar olmadığı gözlenmiştir. Bu yollar analizden çıkarıldığında model doğrulanmıştır. Modele ait uyum iyiliği değerleri incelendiğinde modelin iyi uyum gösterdiği gözlenmiştir. Aracılık testi için uygulanan Sobel testi sonucuna göre de ilgili değişkenlerin aracılık etkilerinin manidar olduğu gözlenmiştir. Bu doğrultuda t-değerleri, uyum değerleri ve Sobel testi sonuçlarına göre araştırma çerçevesinde kurgulanan model doğrulanmıştır. Araştırma sonuçlarına göre öğrencilerin içsel özgüvenleri ile motivasyon, başlama, yılmama ve sürdürme değişkenleri arasında doğrudan bir ilişkinin olduğu gözlenmiştir. Dışsal özgüven değişkeni ile ise yalnızca motivasyon arasında doğrudan ilişki gözlenmiştir. Motivasyon değişkeni ile başlama, yılmama ve sürdürme değişkenleri arasında doğruda bir ilişki gözlenmiştir. Ayrıca motivasyon, değişkeninin içsel özgüven ile başlama, yılmama ve sürdürme değişkenleri ile dışsal özgüven ile başlama, yılmama ve sürdürme arasında aracılık etkisi manidar bulunmuştur. Başlama değişkeni motivaston ve TEOG puanı ile içsel özgüven ile TEOG puanı arasında aracı bir etkiye sahiptir. Son olarak sürdürme değişkeninin motivaston ve TEOG puanı ile içsel özgüven ile TEOG puanı arasında aracı bir etkiye sahip olduğu görülmüştür.

Anahtar Kelimeler

TEOG, Motivasyon, Özgüven, Özyeterlik, Yapısal Eşitlik Modeli

*Cumhuriyet Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:goncausta@gmail.com
Tel: +903462191010/4731

Paper No: 71

The Examination Relationship Between the Levels of Motivation, TEOG Success Grade, Self-Confidence and Self-Efficacy

Abstract

In the scope of research aimed to examine the relationship between students´s TEOG (transfer exam from basic education to secondary education) success grades, self-confidence, self efficacy and levels of motivation. In accordance with this purpose, in the 2015-2016 school year studying in Sivas province 9th grade students were performed relevant scales. It was observed that there is no meaningful relationship between TEOG (transfer exam from basic education to secondary education)such that initial institutional model´s preliminary external self-confidence and self efficiency sub-dimensions which are encourage, start up, and pursuance. In accordance with the purpose of research, the organizational model was fictionalised and this model was analyzed. When these processes were excluded from the analysis, model was verified. When the model of goodness of fit index was examined, it was observed that the model showed good fit. Also, according to the Sobel test results applied to mediation test, the mediation effects of related versions were found to be meaningful. In t-scores, adaptive value and the sobel test results, the model that constructed in the framework of research were verified. According to results of research, internal self-confidence of students with motivation, starting, elasticising, continuation between were observed a direct relationship. If the external variables of self-confidence between of the only motivation was observed a direct relationship. Also motivation was found meaningful mediation effect starting with internal self-confidence of variables and with elasticising and continuation variables between of starting external self – confidence, elasticising and continuation. Starting variable has a mediator effect with motivation and TEOG grade(transfer exam from basic education to secondary education)between of internal self- confidence and TEOG grade(transfer exam from basic education to secondary education). Finally, it was found a mediator effect between motivation and TEOG grade (transfer exam from basic education to secondary education) of continuation variable, internal self- confidence and TEOG grade (transfer exam from basic education to secondary education).

Key Words

TEOG, Motivation, Self-Confidence, Self-Efficacy, Structural Equation Model

Bildiri No: 72

Eğitimde Ölçme ve Değerlendirme Kongrelerinde (EPOD) Sunulan Bildirilerin Doküman Analizi Yöntemi ile İncelenmesi

Mahmut Sami KOYUNCU,* Mehmet ŞATA**, İsmail KARAKAYA***

Özet

Bu araştırmanın amacı Türkiye'de Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongrelerinde sunulan bildirilerin eğilimlerinin nasıl olduğunun belirlenmesidir. Bu kapsamda 2008, 2010, 2012 ve 2014 yıllarında Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongrelerinde sunulmuş olan bildirilerin konu temaları, anahtar kelimeleri, yazarların kurumları ve unvanları, araştırma türü, araştırma modeli/deseni, örneklem büyüklüğü, kullanılan veri türü ve veri toplama aracı incelenmiştir. Araştırma evren üzerinde gerçekleştirilmiş olup, araştırmanın evrenini 2008, 2010, 2012 ve 2014 yıllarında gerçekleştirilen Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongrelerinde sunulan 248 bildiri oluşturmaktadır. Bildirilere ait bilgilerin incelemelerinin dokuman analizi ile yapıldığı bu araştırmada veriler, araştırmanın amacına bağlı olarak belirlenen kategoriler altında, araştırmacılar tarafından geliştirilen "Bildiri inceleme Formu" ile toplanmıştır. Bildiriler farklı araştırmacılar tarafından puanlandığı için, puanlayıcı güvenirliğini sağlamak amacıyla uzlaşma katsayısı 0.933 olarak hesaplanmış ve puanlayıcılar arası uzlaşma anlamındaki güvenirliğin sağlandığı tespit edilmiştir. Elde edilen veriler SPSS ve Nvivo programında betimsel ve kategorik analizler kullanılarak incelenmiştir. İnceleme sonucunda, sunulan bildirilerde genel olarak unvanların, bildiri türünün, araştırma modelinin/deseninin belirtilmemiş olduğu, aynı eksikliklerin çalışma grubunun, örneklem büyüklüğünü, veri türünde de mevcut olduğu tespit edilmiştir. EPOD kongrelerinde sunulan bildirilerin genel olarak "Ölçek Geliştirme ve Uygulama" , "Eğitim araştırmalarında kullanılan istatistiksel yönelimler" ve "Geniş ölçekli sınavlar, seçme ve yerleştirme sınavları, yüksek riskli sınavlar" konu temalarında yoğunlaştığı belirlenmiştir. "ÖSYM Uygulamaları" ve "Etik" temalarında çok fazla çalışma yapılmadığı belirlenmiştir. "Yanlılık Çalışması" konu temasıyla ilgili bildirilerde EPOD kongrelerinde yıllara göre sürekli bir artış olduğu tespit edilmiştir. Sunulan bildiri türlerinin daha çok nicel araştırma olduğu, buna bağlı olarak tarama ve betimsel araştırma modellerinin kullanıldığı belirlenmiştir. Araştırmalarda genellikle 301-1000 arası örneklem büyüklüğünün tercih edildiği, geliştirilmiş ölçek ve uygulama verisi kullanıldığı sonucuna ulaşılmıştır. Bildirilerde en çok "güvenirlik", "geçerlik" ve "ölçme ve değerlendirme" anahtar kelimelerinin kullanıldığı tespit edilmiştir. Kongrelere en fazla katılımın sırasıyla Ankara, Hacettepe ve Gazi üniversitesinden olduğu belirlenmiştir.

Anahtar Kelimeler

Bildiri, Doküman Analizi, EPOD, Ölçme ve Değerlendirme

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:ms_koyuncu@hotmail.comTel:+905069748289

**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: mehmetwsata@gmail.comTel:+905432617927

***Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü.İletişim: ikarakaya2002@gmail.com Tel:+905053373443

Paper No: 72

The Investigation Of The Papers Presented in Measurement and Evaluation in Education and Psychology (MEEP) Conferences with Document Analysis

Abstract

This study aims to state the trends of the papers presented in Measurement and Evaluation in Education and Psychology (MEEP) Conferences in Turkey. For that purpose, authors title and where they work, the subjects, key words, research methods, sample size, data type and data collection tool of the papers presented in MEEP Conferences in 2008, 2010, 2012 and 2014 were investigated. The population of the study is consisting of 248 papers presented in those years, which are all used in the analysis. In order to investigate the information on these papers, document analysis as a qualitative research method was used. Data were collected with "paper investigation form" developed by the authors under the specific categories determined according to the aim of the study. Because the papers are graded by different raters, interrater reliability coefficient was estimated. The value of it was calculated as 0,933, which showed that interrater reliability was provided for the study. For the data analysis, SPSS and Nvivo software programs were used. The results show that authors' title, type of the paper and research methods are generally unspecified. Likewise, study group, sample size and data type are missing on these papers. The subjects of the papers are generally centred on these topics: "Scale Development and Implementation", "statistics used in educational research", "standardized tests, selection and placement tests, high-stakes tests". However, "student selection and placement centre's implementations" and "ethics" are the subjects that were studied rarely. "Research on bias" is a subject that have been increasingly studied over the years. The papers presented are generally in quantitative research methods, especially survey and descriptive research models were used. Sample sizes are between 301-1000. The most frequent key words are reliability, validity and measurement and evaluation. Authors of the papers are mostly titled research assistant and then assistant professors. People attended the conferences are generally form Ankara University, Hacettepe University and Gazi University, respectively.

Key Words

Papers, Document Analysis, MEEP, Measurement and Evaluation

Bildiri No: 73

Akademik Yazma Beceri Sinavinin Çok Bağlamlı Rasch Modeliyle Değerlendirilmesi

Bengü BORKAN, Aylin ÜNALDI

Özet

Bu çalışmanın amacı, eğitim dilinin İngilizce olduğu bir üniversitenin hazırlık biriminin uyguladığı akademik yazma becerisini ölçme amaçlı kullanılan sınavın ve özellikle puanlama anahtarının ne düzeyde çalıştığını belirlemektir.

Üniversitesi'nin İngilizce Yeterlilik sınavının bir bölümü olan Akademik Yazma Sınavı'nda iki tartışma gerektiren konuyu işleyen metin yazılması istenmektedir. İki metin için toplam 80 dakika süre verilmektedir ve her metinin yaklaşık bir sayfa (yaklaşık 250 sözcük) olması beklenmektedir. Her metin iki ayrı puanlayıcı tarafından puanlanır. İki puanlayıcı arasında değerlendirme anahtarında bir seviyeden fazla fark olduğu durumlarda her iki puanlayıcı da metni yeniden değerlendirir. Uzlaşmaya varamazlarsa üçüncü bir değerlendirici metni puanlar. Birbirine yaklaşık olan puanlar esas kabul edilir. Bu çalışmada 2014-2015 Akademik yılı sonunda yapılan Akademik Yazma Sınavı'ndan elde edilen, puanlayıcı farklılıkları düzeltilmemiş puanlardan oluşan veri seti kullanılmıştır. Bu veri setinde, 29 farklı puanlayıcı 1470 öğrencinin yazdığı iki ayrı görevi puanlamıştır. Her bir puanlayıcı ortalama yaklaşık 200 adet görev puanlamıştır. Yirmidokuz puanlayıcıdan 10'nunun anadili İngilizce'dir. Puanlayıcıların hepsi hazırlık biriminde okutman olarak görev yapmaktadır ve birbirlerinden farklı dil öğretme ve puanlama tecrübesine sahiptirler. Her puanlayıcı hem sene içinde hem de puanlama öncesi standardizasyondan geçmektedir. Bulgular dört ayrı bağlama göre sırasıyla verilmiştir; puanlayıcı, puanlayının anadili, öğrenciler ve yazma görevleri. Bunların yanısıra puanlama ölçeğinin ne derece çalıştığı da raporlanmıştır. Puanlayıcı katılığında farklıklar gözlenmektedir. Minumum ve maximum puanlayıcı logit puanların arasında iki birim fark vardır. Puanlayıcı değerlerleri 0,4 logit civarında dağılım göstermektedir. Şekil 1'de görüldüğü üzere puanlayıcı katılığı puanlayıcının anadilin göre heterojen bir dağılım göstermektedir. Sınavda verilen iki görevin de aynı güçlük değerine sahip olduğu, tutarlı bir şekilde puanlandığı ve puanlamanın puanlayıcının ana diline göre değişmediği bir başka değişle yanlılık olmadığı belirlenmiştir. Puanlayıcılar dokuz kategorili değerlendirme anahtarının her kategorisi kullanmakla birlikte iki numarali category bir alt ve üst kategoriden ayrılamamaktadır. Diğer bütün kategoriler net bir şekilde ayrılmış ve eşik değerleri doğru sıradadır. Tüm uyum indeksleri 1 cıvarındadır. Bu çalışmada İngilizce hazırlık sınıfında okuyan öğrencilerin yazma becerisinin değerlendirilmesi sürecine dahil olan farklı bağlamların (puanlayıcı, puanlayıcının anadili, puanlamada kullanılan ölçek ve görevler) Rasch modeli ile analizi yapılmıştır. Sonuçlar gösteriyorki, genel olarak bu değerlendirme sürecinin psikometrik özellikleri kabul edilebilir düzeydedir. Fakat bu sürecin daha iyi hale gelebilmesi için aşağıdaki öneriler sunulmuştur. Değerlendirme anahtarındaki 2. ve 3.'cü kategorinin tanımları yeniden yapılmalıdır, Öğrenci puanlarındaki ayırt ediciliği artırmak için, iki görevin farklı zorluk derecesinde verilmelidir.Puanlayıcılarda katılık dereceleri farklı olduğu için puanlayıcı eşleşmeleri katılık derecesi göz önüne alınarak eşleştirme yapılmalıdır. Öğrencilere zamanla ilgili yönergeye dikkat etmesi gerektiği uyarısı daha dikkatli yapılmalıdır. Puanlayıcılara bu sonuçlar kullanarak yeniden egitim verilmelidir.

Bildiri No: 74

Hesap Verebilir Eğitim Sisteminde Bir Paydaş Olarak Öğretmenlerin Temel Eğitimden Orta Öğretime Geçiş (Teog) Uygulamasına Yönelik Görüşleri

R. Nükhet DEMİRTAŞLI

Hesap verebilirlik; Okulun erişmeyi istediği hedefleri veya standartları ne ölçüde gerçekleştirdiği konusunda okul içindekilerini ve dışındakileri bilgilendirme sürecidir (Brauen, O'Reilly, &Moore, 1992). Bu bilgilendirme süreci, sürecin asıl paydaşları olan öğrenci, öğretmen, veli ve okul yöneticilerine yönelik olarak gerçekleştirilir. Bu bilgilendirme süreci okul gelişim raporlarıyla sağlanır. Okul gelişim raporları; Okulların, bölgelerin ve/veya ülkenin (İ)eğitimsel girdileri, (ii) eğitimsel süreçlerİ ve (iii)eğitimsel sonuçlarına ilişkin çok kapsamlı bilgi verir. (Ysseldyke, Krentz, Elliott, Thurlow, Erickson ve Moore , 1998). Eğitim-öğretim sürecinin önde gelen paydaşı olan öğretmenlerin özellikleri, eğitim uygulamaları ve politikalarına yönelik tutum ve tavırları, eğitim-öğretim ortamları, eğitim sonuçları üzerinde doğrudan ya da dolaylı doğurgulara sahip Olabilmektedir. Bu araştırmada eğitimsel sonuçların önemli bir göstergesi sayılan TEOG kapsamında yürütülen sınav uygulamasına ilişkin olarak öğretmenlerin görüşleri araştırılmıştır. Eğitim sistemimiz 2013-2014 eğitim öğretim yılından itibaren SBS uygulaması bitmiş Temel Eğitimden Orta Öğretime Geçiş (TEOG) uygulaması başlamıştır. Milli Eğitim Bakanlığı bu uygulamanın amaçlarını şöyle ifade etmiştir (MEB, 2013): "Öğrenci, öğretmen ve okul ilişkisini güçlendirmek; eğitim sürecinde öğretmenlerin ve okulun rolünü daha etkin kılmak; ülke çapında müfredatın eş zamanlı uygulanmasını sağlamak; sınav kaygısını sürece yayarak azaltmak … "Bu amaçların öğretim ve değerlendirme boyutlarında ne denli hayata geçtiğine ilişkin öğretmen görüşleri bu araştırmanın merak konusu olmuştur. Nitel araştırma türündeki bu araştırmada, biri devlet diğeri özel okul statüsünde yer alan iki okuldaki toplam 50 öğretmenin görüşlerine başvurulmuştur. Öğretmenlerin TEOG uygulamalarına yönelik görüşlerini serbest olarak bildirmelerine olanak sağlayan yarı yapılandırılmış, öğretim ve değerlendirme süreçlerine ilişkin görüşlerini belirtecekleri dört açık uçlu soru sorulmuş ve serbest yanıtlarını yazılı olarak vermeleri istenmiştir. Her öğretmene öğretmenler odasında ulaşılmıştır. Öğretmenler soruları yaklaşık 7-8 dakikada yanıtlamıştır. Bu anket formunda, kendilerinden demografik ve mesleki bazı özellikleri dışında herhangi bir kimlik bilgisi alınmamış, araştırma sonuçlarının öğrenecekleri bir iletişim bilgisi de verilmiştir. Elde edilen serbest yanıtlar (veriler) içerik analizine tabi tutulmuştur. Öğretmenlerin açık uçlu dört soruya verdiği yanıtlar, tutarlılık sağlamak üzere, araştırmacı dışında nitel araştırmalar konusunda bir uzman tarafından da her soru ayrı olmak üzere taranmış, temel görüşler/temalar ve görüşlerdeki anahtar sözcükler çıkarılmıştır. İçerik analizi kapsamında, araştırma soruları doğrultusunda birbiriyle örtüşen veriler belirli kavramlar ve temalar çerçevesinde kodlanarak bir araya getirilmiştir Aynı verilerin Nvivo 10 programı kullanılarak da çözümlenmesi sürmektedir. Araştırmanın ön bulguları, genel anlamda öğretmenlerin TEOG uygulamasına yönelik olumsuz görüşleri olduğunu göstermektedir. Öğretmenler, TEOG'un hem öğretim süreçlerini hem de değerlendirme süreçlerini olumsuz yönde etkilediği görüşündedirler. Bulgular ilgili alan TEOG amaçları çerçevesinde yorumlanarak, alanyazın çerçevesinde tartışılacaktır.

Anahtar Kelimeler

TEOG, öğretmen görüşleri, eğitimde hesap verebilirlik

Paper No: 74

As A Stakeholder Teachers' Opınıons About Teog In Educatıonal Accountabılıty System

Educational accountability is kind of informative system which gives feedbacks toward school and out of school environment (Brauen, O'Reilly, &Moore, 1992). This informative process targets main stakeholders of educational system such as students, teachers, parents and school principals.

School progress reports provide evaluative results for educational accountability. School progress reports includes important information and data on (i) educational inputs (ii) educational process (iii) educational outputs (Ysseldyke, Krentz, Elliott, Thurlow, Erickson ve Moore , 1998). Teachers characteristics and their opinions and attitudes towards educational practices and politics have important direct and indirect impacts on educational systems.

the main purpose of this study is to investigate teachers'opinios about TEOG in terms of teaching and assessment practices in our educational system.

TEOG has implemented in national educational system in 2013-2014 academic year. The objectives of TEOG have been reported as follows: to fortify relationships between student, teacher and school; to highlight teachers' role in school; to implement synchronistically teaching programs in nation wide; to decrease test anxiety and so on.

In this respect, the main curious thing is to investigate teacher opinions on these TEOG' objectives. For this purpose, 50 teachers' opinios have been collected by semi-construct survey which has four open-ended questions. These participants were from one public and one private school. Each teacher has responded survey questions approxiamately in 7-8 minutes. In each survey form is questioned teacher's demographic informations except ID information.

Since this study is qualitative research, content analysis is performed on these respons. Each question is scanned by researcher and co-researcher who is experienced in content analysis At the end of this process have been specified main ideas, key words. This coding procesure is checked by two researchers in terms of consistent coding. This coding process will be held by Nvivo.

First results show that teachers have negative opinions towards TEOG. Teachers thinks that TEOG affects negatively their teaching and assessmeny practices. All findings will be summarized with TEOG objectives and compared with related literature.

Key Words

TEOG, teacher opinions, educational accountability

Bildiri No: 76

Çok Kategorili Puanlanan Maddelerde Mantel Test ve Olabilirlik Oran Testi'nin Madde İşlev Farklılığı Belirlenme Gücünün İncelenmesi

Safiye BİLİCAN DEMİR*, R. Nükhet DEMİRTAŞLI**

Özet

Bu araştırmada, çok kategorili tepki gerektiren maddelerde MİF belirleme testlerinden Mantel Test ve MTK-Olabilirlik Oran Testi'nin farklı test koşullarında istatistiksel güç (power) oranları karşılaştırılmıştır. Çalışmada, Monte Carlo simülasyon tekniği yaklaşımıyla araştırma koşullarına uygun yapay veri setleri elde edilmiştir. Bu çalışmada grupların yetenek dağılımı, örneklem büyüklüğü, MİF miktarı ve MİF örüntüsü manipüle edilen değişkenler olarak belirlenmiş; çok kategorili MTK modeli, madde sayısı, MİF içeren madde sayısı ve MİF türü (tek biçimli) bütün koşullar altında sabit tutulmuştur. Son durumda 76 [3 (yetenek dağılımı) x 3 (örneklem büyüklüğü) x2 (MİF miktarı) x2 (MİF örüntüsü) x 2 (MİF belirleme tekniği)] simülasyon koşulu ortaya çıkmıştır. Her bir koşul için 100 tekrar yapılmıştır. Güç oranları MİF içerecek biçimde modellenen üç madde üzerinden hesaplanmıştır. Çalışmada veri üretmek için WinGen1, MTK-OOT karşılaştırmaları için MULTİLOG ve Mantel Test analizleri için DIFAS programı kullanılmıştır. Her iki MİF belirleme testi için de, artan MİF miktarı ve örneklem büyüklüğüne bağlı olarak ilgili testlerin istatistiksel güç oranlarını yükselmiştir. Araştırma bulguları tüm örneklem büyüklüğü ve MİF miktarı koşullarında MTK-OOT'in istatistiksel güç oranlarının, Mantel Test'e göre daha yüksek olduğunu göstermektedir. Yüksek MİF örüntüsü koşullarında ilgili testlerin MİF'i belirlemedeki performansı zayıftır. Mantel Test ve MTK-OOT'nin MİF belirlemedeki gücü odak grubun yetenek ortalamasındaki sapmaya bağlı olarak bir miktar yükselmiştir. Ancak genel olarak farklı sapma koşulları için ilgili testlerin istatistiksel gücü birbirine yakın değerler almıştır.

Anahtar Kelimeler

Çok kategoride puanlanan madde, Madde İşlev Farklılığı, Mantel Test, Olabilirlik Oran Testi, istatistiksel güç

*KocaeliÜniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:safiyebilican@gmail.com
** Ankara Üniversitesi Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü. İletişim:rnukhet@yahoo.com

Paper No: 76

Comparison of Mantel Test and Likelihood Ratio Test for Detection of Differential Item Functioning in Polytomous Item Responses

Abstract

The purpose of this study was to investigate the power and Type I error rate of the likelihood ratio goodness-of-fit (LR) statistic and Mantel Test in detecting differential item functioning (DIF) under Master's (1969, 1972) Partial Credit Model. A multiple replication Monte Carlo study was utilized for simulated data sets. Several variables were manipulated in this study, including the sample size, group mean difference, DIF condition and DIF magnitude. In final study design, there were 76 conditions[3 (sample size) x 3 (group mean difference) x 2 (DIF magnitude) x2 (DIF pattern) x 2 (methods of DIF detection)] for power study. Simulation was replicated for 100 times for each simulation condition. In this study, WinGen3 was used to simulate ability estimates and to generate response data sets. MULTİLOG and DIFAS were used to conduct the Mantel and IRT likelihood-ratio test DIF analyses. Results indicated that for both DIF detection tests, the power to detect DIF increased as the DIF Magnitude and sample size increased. The results also showed that IRT-LR had higher DIF detection rates than Mantel test under all test conditions. For both DIF detection tests, conditions with high-shift DIF pattern had the poorest power. The presence of group mean difference had minimal effect on the power results of both DIF detection tests.

Key Words

Polytomous items, Differential item functioning, Mantel test, Item Response Theory-Likelihood Ratio test, power

Bildiri No: 77

Bağıl Değerlendirme mi yoksa Mutlak Değerlendirme mi?(Türkiye'den bir Üniversite Örneği)

Erkan Hasan ATALMIŞ*, Alpaslan ALKIŞ**

Özet

Bu çalışmanın amacı öğrencilerin harf notlarının bağıl değerlendirme ve mutlak değerlendirme kullanarak nasıl değiştiğini tespit etmektedir. Bunun için Kahramanmaraş Sütçü İmam Üniversitesi 2015-2016 güz dönemindeki yaklaşık 20000 öğrencinin bağıl değerlendirme ile hesaplanmış harf notları, mutlak değerlendirme kullanılarak tekrar hesaplanmıştır. Yapılan Pearson Korelasyon analizi sonucunda bağıl değerlendirme ve mutlak değerlendirme ile hesaplanan notların arasında pozitif ve anlamlı bir ilişki olduğu görülmüştür. Bağıl değerlendirme ile hesaplanan öğrenci harf notlarının mutlak değerlendirmeyle yeniden hesaplandığında, notların %39'un aynı harf aralığına denk geldiği görülürken, sadece %7'sinde mutlak değerlendirme yönünde artığı, notların %54' ünde ise bağıl değerlendirme yönünde harf notlarının arttığı görülmüştür. Bir sonraki aşamada, bu iki değerlendirmeyle hesaplanan notların istatiksel olarak değişimini görmek için eşleştirilmiş iki grup arasındaki farkların testi uygulanmış, bağıl değerlendirme ile hesaplanan notların mutlak değerlendirmeye göre istatistiksel olarak yüksek olduğu görülmüştür. Genel olarak bağıl değerlendirme ile mutlak değerlendirme kıyaslandığında bağıl değerlendirmenin genel olarak öğrenci faydasına çalıştığını ortaya koymaktadır.

Anahtar Kelimeler

Bağıl değerlendirme, mutlak değerledirme, Türkiye'de yükseköğretim, Kahramanmaraş, Sütçü İmam Üniversitesi

* Kahramanmaraş Sütçü İmam Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: erkanatalmis@gmail.com. Tel: +03442801315

** Kahramanmaraş Sütçü İmam Üniversitesi Öğrenci İşleri Daire Başkanı. İletişim:alkis@ksu.edu.tr. Tel: +0344 2801141

Paper No: 77

Norm-Referenced Assessment or Criterion-Referenced Assesment? (An Evidence from a University in Turkey)

Abstract

The purpose of this study is to examine how student grade letters change when norm-referenced and criterion-referenced grading system were used. Letter grades of over 20000 students calculated using norm-referenced grading system at Kahramanmaras Sutcu Imam University in Spring 2015-2016 semester were recalculated using criterion-referenced grading system. Results of Pearson correlation analysis show that there is a significant and positive correlation between student grade letters calculated using norm-referenced and criterion-referenced grading system. While around 39% of student grades in the norm-referenced grading system did not change when applying criterion-referenced grading system, around %54 of them increased and % 7 of them decreased. The paired sample t-test was used to see whether students' letter grades significantly change across norm-referenced and criterion-referenced grading systems. The findings indicated that norm-referenced grade letters are statistically different from criterion-referenced grade letters. It was concluded norm-referenced grading system is more beneficial for students increasing their scores.

Key Words

Norm-referenced assessment, criterion-refereneced assessment, higher education in Turkey , Kahramanmaras, Sutcu Imam University

Bildiri No: 78

Denk Olmayan Gruplarda Ortak Madde Deseni Kullanılarak Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması*

Bilge GÖK**, Hülya KELECİOĞLU***, Kübra ATALAY KABASAKAL****

Özet

Araştırmada farklı koşullara (örneklem büyüklüğü, yetenek dağılımı, test uzunluğu ve model türü) göre türetilen test formlarını madde tepki kuramına dayalı kestirim yöntemlerini (ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord) kullanarak eşitlemek ve bu yöntemlerden elde edilen sonuçları karşılaştırmak amaçlanmıştır. Bununla birlikte araştırmada hangi koşullarda hangi yöntemin daha iyi sonuç verdiği de araştırılmıştır. Araştırma iki ve üç parametreli lojistik modele uyumlu iki kategorili simülatif veriler kullanılarak yürütülmüştür. Bu araştırmada iki test formunu eşitleyebilmek için "denk olmayan gruplarda ortak madde/test (NEAT) deseni" kullanılmıştır. Verilerin türetilmesinde WINGEN 3 programından yararlanılmış ve araştırmada kullanılan 36 koşulun her biri için 50 tekrar yapılmıştır. Madde ve yetenek parametrelerinin kestirilmesi PARSCALE 4.1 ile ayrı kalibrasyon için test eşitleme ve ölçekleme IRTEQ ile yapılmıştır. Araştırmada madde ve yetenek parametreleri expected a posteriori (EAP) ile kestirilmiştir. Araştırmada elde edilen sonuçlar, eşitleme yanlılığı (BIAS) ve eşitleme hatası (RMSE) ölçütlerine göre değerlendirilmiştir.Hem iki hem de üç parametreli lojistik modelde küçük örneklemlerde (1000 kişilik) ve kısa testlerde (30 maddelik) yöntemlerin daha az hatalı ve yanlı sonuçlar verdiği gözlenmektedir. Küçük örneklemler ve daha kısa testlere gereksinim duyulduğunda moment yöntemlerinin (OO ve OS) SL yönteminden daha iyi sonuçlar verdiği ifade edilebilir. Gerçek puan eşitlemesi durumunda birkaç istisna dışında her koşulda en iyi yöntemin OO yöntemi olduğu tespit edilmiştir.

Anahtar Kelimeler

Test eşitleme, madde tepki kuramı, denk olmayan gruplar, ortak madde deseni.

---

*Bu çalışma ilk yazarın Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı için hazırlanan doktora tezinin bir kısmından özetlenmiştir.

** Hcettepe Üniversitesi Eğitim Fakültesi İlköğretim Bölümü. İletişim:bilge.bekci@gmail.com

*** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: hulyakelecioglu@gmail.com

**** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: kkatalay@gmail.com

Paper No: 78

Comparison of IRT Equating Methods Using the Common Item Nonequivalent Groups Design

Abstract

The purpose of this research was to equate the test forms which were constructed in different conditions (sample size, ability distribution, length of test and model type) through scaling methods based on item response theory (mean-mean, mean-sigma and Stocking Lord) and to compare the results obtained from these methods. In addition, it was also investigated that which method gave better results under different conditions. The research was conducted with using dichotomous simulated data which was consistent with two and three parameter logistic model. In order to equate two test forms "the common-item nonequivalent groups" was used in this research. WINGEN 3 program was utilized for data generation and 50 replication were done for 36 different condition used in this research. PARSCALE 4.1 was utilized for the prediction of item and ability parameters and IRTEQ was utilized for test equating and scaling in separate calibration. Item and ability parameters were predicted through expected a posteriori (EAP) in this research. The results obtained from this simulation study were evaluated based on equating bias (BIAS) and equating error (RMSE) criterions. The results indicated that methods had less biased and less equating errors when small sample sizes together with short tests in two and three logistic models. When the small samples and the short tests were needed, the results indicated that moment methods had better than the SL. Moreover, the results revealed that in true score equating, with a few exceptions in every condition the best method was mean-mean.

Key Words

Test equating, item response theory, equating error, equating bias.

Bildiri No: 79

Uzaktan Eğitim Yoluyla Gerçekleştirilen Ölçme Ve Değerlendirme Dersine İlişkin Öğrenci Görüşleri

Fatma Betül KURNAZ*

Özet

Bu çalışmanın amacı, uzaktan eğitim yoluyla ölçme ve değerlendirme dersini alan öğretmen adaylarının uzaktan eğitime ve uzaktan eğitim yoluyla ölçme ve değerlendirme dersinin etkililiğine ilişkin görüşlerini belirlemektir. Çalışma grubunu 2015-2016 öğretim yılında bir devlet üniversitesinde öğrenim gören 600 öğrenciden 80 öğrenci oluşturmuştur. Veriler öğrencilerin dönem sonunda ölçme ve değerlendirme dersinin etkililiğine ve uzaktan eğitime ilişkin yapmış olduğu değerlendirmelerden elde edilmiştir. Öğrenci değerlendirmeleri bilgisayar destekli nitel veri analiz programı olan Maxqda 11 paket programı kullanılarak içerik analizi yöntemiyle çözümlenmiştir. Sonuç olarak öğrencilerin ölçme ve değerlendirme dersini uzaktan eğitim yoluyla almış olmaktan memnun olmadıkları belirlenmiştir. Öğrenciler sistemin alışkın oldukları eğitimden farklı yürütülmesini, uzaktan eğitimin derse katılım konusunda kendilerini güdülemediğini, öğrenmede güçlük yaşadıklarını gerekçe göstererek bu durumu açıklamışlardır.

Anahtar Kelimeler

Uzaktan eğitim, Ölçme ve değerlendirme, Öğrenci görüşleri

*Karabük Üniversitesi Edebiyat Fakültesi Eğitim Bilimleri Bölümü. İletişim: betulkurnaz@karabuk.edu.tr Tel: +3704338294(Dahili 2602)

Paper No: 79

Students' Opinions On The Measurement And Evaluation Course Given Through Distance Education

Abstract

The aim of this study is to understand and discover the students' opinions on the the measurement and evaluation course given through distance education. The purposive sampling technique was used in this study. The study group has created 80 student who studying at a university in the 2015-2016 academic year. The process of analysis evolved from concrete to more abstract categories. Quantitative data was analyzed using frequency and percentage. Advantages and challenges of the implementation of the the measurement and evaluation course given through distance education were discussed in the light of student's views.

Key Words

Distance education, Measurement and evaluation, Stundents' opinion.

Bildiri No: 80

Öğrencilerin PISA Başarısındaki Gizil Sınıfların ve Bazı Değişkenlerle Gizil Sınıfların İlişkilerinin Belirlenmesi

Seher YALÇIN*

Özet

Bu çalışmanın amacı, Uluslararası Öğrenci Değerlendirme Programı (PISA)'nın 2012 uygulamasına Türkiye'den katılan öğrencilerin, okuma, matematik ve fen başarılarındaki çok düzeyli gizil sınıfları belirlemek ve bu yaklaşımla belirlenen öğrenci ve okul gizil sınıflarının i) öğrencilerin sebatkârlığı, ii) öğrencilerin problem çözmeye açıklığı, iii) öğrencilerin ekonomik, sosyal ve kültürel statüsü ve iv) okul kaynaklarını (fiziksel altyapısı, öğretim çalışanları ve eğitim kaynakları) yordama yeteneğini keşfetmektir. Bu araştırmanın evreni, PISA 2012 uygulamasına Türkiye'den katılan müdürler ve 15 yaşındaki öğrencilerdir. Katılımcılar, iki aşamalı tabakalı örneklem aracılığıyla PISA'ya seçilmektedir. Toplamda bir okul ve 1652 öğrenci, kayıp veri veya uygulanmama nedeniyle dışarıda tutulmuştur. Analizler için 3196 öğrenci ve 169 okul kalmıştır. Veriler, PISA web sitesinden elde edilmiştir. Okuma, matematik ve fen okuryazarlığı testleri ile öğrenci ve okul müdürleri için iki ayrı anketten elde edilen veriler kullanılmıştır. Analizlerin birinci adımında, öğrencilerin okuma, matematik ve fen başarılarında okullar arası çok düzeyli gizil sınıf sayısını incelemek için çok düzeyli gizil sınıf analizi kullanılmıştır. Ardından seçilen değişkenlerin oluşan çok düzeyli gizil sınıfları yordama yeteneğini belirlemek için üç adımlı (three-step) analiz kullanılmıştır. Sonuçlar, 8Kümeli-6Gizil sınıflı modelin en küçük BIC değerine sahip olduğunu göstermektedir. Bu nedenle, bu çalışmada bu model tercih edilmiştir. Bütün seçilen değişkenler, okul düzeyinde gizil sınıf üyeliğini manidar bir şekilde yordamaktadır. Ayrıca, analiz sonuçları, öğrencilerin okuma, matematik ve fen başarılarında en önemli etkinin öğrencilerin ekonomik, sosyal ve kültürel statüsünün olduğunu göstermektedir. Ek olarak, öğrencilerin sebatkârlığı, öğrencilerin problem çözmeye açıklığı, öğrencilerin ekonomik, sosyal ve kültürel statüsü ve okul kaynakları (fiziksel altyapısı, öğretim çalışanları ve eğitim kaynakları) düşükten yükseğe çıktıkça yüksek başarılı okullardaki öğrenci yüzdesi artmaktadır.

Anahtar Kelimeler

Okuma, matematik ve fen başarısı, PISA, çok düzeyli gizil sınıf analizi.

*Ankara Üniversitesi, Eğitim Bilimleri Fakültesi,Ölçme ve Değerlendirme Bölümü. İletişim: yalcins@ankara.edu.tr, Tel: +90312 3633350

Paper No: 80

Determining of the Relationships among Some Variables and Latent Classes and Latent Classes at Students' PISA Achievements

Abstract

The purpose of this study is to determine the multilevel latent classes for reading, mathematics and science achievement of the students who participated in PISA 2012 from Turkey and to find out the predictive ability of i) students' perseverance, ii) students' openness for problem solving, iii)students' economic, social and cultural status, iv) resources of school (physical infrastructure, teaching staff and educational resources) in relationship to according to those determined student and school classes using multilevel approach. The population of this research are principals and all 15-year-old students who attended the Program for International Student Assessment (PISA) in 2012 in Turkey. The participants who attend PISA are selected via two-stage stratified sampling. In total one school and 1652 students were excluded due to missing data and not application, leaving 3196 students from 169 schools for the final analyses. Data was obtained from website of PISA and includes reading, mathematics and science literacy test and two different questionnaires, one for the students, the other for the schools' principals. As a first step, multilevel latent class analysis was used in order to investigate the number of multilevel latent classes in between-schools reading, mathematics and science achievements of the students. Then, a three-step analysis was also used to determine the predictive ability of the chosen variables to the constructed multilevel latent classes. The results indicated that the BIC was smallest for the 8Cluster-6GClass model, thus this model is preferred model in this study. All of the chosen variables significantly predicted the school level latent class membership. Moreover, analyses suggested that the students' economic, social and cultural status is understood to be the most important the effect of a student's reading, mathematics and science achievement. Additionally, as the students' perseverance, students' openness for problem solving, students' economic, social and cultural status and resources of school (physical infrastructure, teaching staff and educational resources) increases from low to high, the percentage of students in the high achievement schools increased.

Key Words

Reading, mathematics and science achievements, PISA, multilevel latent class analysis.

Bildiri No: 81

Öğrencilerin Sınav Kavramına Yönelik Metaforlarının İncelenmesi

Duygu KOÇAK, Çilem DOĞAN-GÜL, Emrah GÜL, Ömay ÇOKLUK

Özet

Bu çalışmada, öğrencilerin sınav kavramına ilişkin algılarının, düşüncelerinin ve deneyimlerine bağlı olarak geliştirdikleri tutumlarının metaforlar aracılığıyla ortaya çıkartılması amaçlanmıştır. Sınavlara bağlı olarak verilen kararlar öğrencinin eğitim yaşantısının şekillenmesinde önemli roller oynadığından, sınavlara yüklenen anlam ve önem de oldukça fazladır. Öğrencilerin yaşamını önemli oranda etkileyen bir olgunun öğrenciler tarafından nasıl algılandığının belirlenmesi, bu süreçleri düzenlemek ve sınava hazırlanan öğrencilere destek sunabilmek adına önem taşımaktadır. Öğrencilerin "sınav" kavramına yönelik metaforlarını belirlemek amacıyla yürütülen bu çalışma, nitel araştırma desenlerinden "olgu bilim (fenomenoloji)" ile uyumludur. Çalışmada 2015-2016 Eğitim Öğretim yılı, güz döneminde Ankara, İstanbul, İzmir illerinde farklı okullarda öğrenim gören ve TEOG sınavına hazırlanan 8. sınıf öğrencileri ile bu sınavı geride bırakıp bir ortaöğretim kurumuna yerleşmiş olan 9. sınıf öğrencilerinden oluşan toplam 603 öğrenciden veri toplanmıştır. Öğrencilerin "sınav" kavramına ilişkin algılarını belirlemek amacıyla, yakın dönemde öğrencilerin karşı karşıya kalacakları ya da yakın zamanda geride bırakmış oldukları ve onlar için büyük önem taşıyan TEOG sınavının,"uyaran" ya da bir başka deyişle "metafor konusu" olarak kullanıldığı form ile veriler toplanmıştır. Çalışmada elde edilen veriler, "içerik analizi" tekniği ile analiz edilmiştir.

Sekizinci ve dokuzuncu sınıf öğrencilerinin "sınav" ile ilgili geliştirdikleri metaforlar incelendiğinde, öğrencilerin sınava ilişkin olarak korku, kaygı, ölüm, dönüm noktası, intihar, kefen gibi olumsuz ve sınavı çok fazla önemseyen algılara sahip oldukları belirlenmiştir. Özellikle hayati önem taşıyan, ölüm kalım meselesi ve diğer sıra dışı benzetmelerin olması, sınavın öğrencilerin üzerinde çok fazla olumsuz etkilere yol açtığını göstermektedir. Geliştirilen metaforlar, sınavın öğrenciler üzerindeki olumsuz etkisini anlama açısından önemli göstergeler olarak kabul edilebilir.

Anahtar Kelimeler

Sınav, Metafor, TEOG

Paper No: 81

Examination of The Metophors for Students Exam Concept

Abstract

It is aimed to uncover the attitudes of students that are related with their perceptions, thoughts and experiences that are linked with the exam concept via metophor.decisions depending on exams play on important role in shoping a student's educational life there fore importance and meaning is pretty much. A phenomen on that effects students lives in a big rate gives importance in terms of how it is perceived by students and setting these processes and supporting the students who are preparing for exams. This study is conducted to determine metophors fort the students to test concept is matching with phenomenology from qualitative research. In this study, the data were collected from totally 603 students who studied in different schools in Ankara, İstanbul and İzmir provinces in 2015-2016 school year. These 603 students were 8th grade students who prepared to take the TEOG exam and 9th grade students who took the excam and got registered to a high school.To determine students perception about exam concept datas were collected through forms. These formsa re used TEOG as "stimılus" or "subject of motophor". In te recent time students will face with TEOG or faced with TEOG. TEOG is of capitol importance for them. When analyzed metophors about exams, developed by eight and ninth grade students, it is identified that they have too much perceptions care about exams and doomy apprehensions such as fear, wory, ending, suicide.

Especially fall within vital importance, a matter of life and death and such unusual analogies show that exam causes too much negative impact on students. Developed metophors acceptable important symptoms to find out negative effects on students.

Key words

Exam, metophor, TEOG

Bildiri No: 82

Bir Makalenin Atıf Sayısını Yordayan Değişkenlerin İncelenmesi

Ercan ÇOBAN*, Okan DEMİR**, Volkan ALKAN***, Ezel TAVŞANCIL****

Özet

Bu araştırmanın amacı, Türkiye'de SSCI kapsamında taranan eğitim dergilerinde yayınlanan makalelerin aldığı atıf sayısını; makalenin sayfa sayısı, yazar sayısı, öz atıf sayısı, ilk yazarın unvanı ve araştırma yaklaşımı gibi değişkenlerden hangilerinin daha fazla yordadığını belirlemektir. Bu amacı gerçekleştirmek amacıyla nicel bir yaklaşımın izlendiği çalışma, tarama modeli ile modellenmiş, ilişkisel tarama modellerinden korelasyonel bir araştırma özelliği taşımaktadır. Çalışma verileri olarak, Türkiye'de eğitim bilimleri alanında yapılan çalışmaları yayınlayan ve SSCI kapsamında taranan; Eğitim ve Bilim Dergisi, Kuram ve Uygulamada Eğitim Bilimleri (KUYEB) Dergisi ve Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (HÜEF)'nde 2010 ile 2012 yılları arasında yayınlanan toplam 283 makale kullanılmıştır. Çalışmanın verileri açık erişime sahip Eğitim ve Bilim Dergisi, Kuram ve Uygulamada Eğitim Bilimleri Dergisi ve Hacettepe Üniversitesi Eğitim Fakültesi Dergisi'nin web sitelerinden elde edilmiştir. Makaleler araştırmacılar tarafından incelenerek; makalenin sayfa sayısı, yazar sayısı, araştırma yaklaşımı ve ilk yazarın unvanı belirlenmiştir. Makalelerin atıf ve öz atıf sayıları, Google Akademik Veri Tabanı kullanılarak elde edilmiştir. Atıf sayısını yordayan değişkenleri belirlemek için, SPSS 20.0 programı kullanılarak çoklu regresyon analizi yapılmıştır. Araştırma sonucunda; öz atıf sayısı, araştırma yaklaşımı, ilk yazarın unvanı, yazar sayısı ve sayfa sayısı değişkenlerinden sadece öz atıf sayısının atıf sayısını yordadığı, öz atıf sayısının da tek başına atıf sayısındaki değişkenliği açıklamakta yetersiz kaldığı görülmüştür. Bu araştırmada ele alınmayan derginin açık erişim durumu gibi değişkenlerin atıf sayısında rol oynadığına dair bulguların varlığı bilinmektedir. Bu yüzden bu değişken gibi diğer değişkenler dikkate alınarak araştırma tekrarlandığında farklı sonuçlara ulaşılabileceği düşünülmektedir. Ayrıca bu araştırmada sadece Social Sciences Citation Index'te taranan dergilere yer verilmiştir. Bu indekste taranmayan dergiler de dahil edilerek araştırma tekrarlanabilir.

Anahtar Kelimeler

Atıf sayısı, Öz atıf sayısı, yazar sayısı

*Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: coban.ercan@gmail.com Tel: 5323781178…..

**İzzet Baysal Teknik ve Endüstri Meslek Lisesi. İletişim: okkandemir@gmail.com Tel: +9053206444811

***Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü.

****Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü.

Paper No: 82

İnvestigation of Factors Predicting Citation Number of An Article

Abstract

The research aims to determine whether article page length, number of authors, journal self-citation, first author's academic title and research design predict the number of citation of articles published in Turkish educational journals which are indexed at Social Sciences Citation Index (SSCI). The research has been conducted as a correlational research. The sample of the study comprised of 283 articles published between 2010 and 2012 in Education and Science Journal, Educational Sciences: Theory & Practice and Hacettepe University Journal of Education. The articles were accessed from the open access web pages of these journals. After investigation of articles, the researchers determined the article page length, number of authors, journal self-citation, first author's academic title and research design for each article. The number of citation and journal self-citation was determined by Google Scholar. To determine the variables predicting the number of citation, the data was analyzed with multiple regression by using SPSS 20.0. The findings showed that only variable predicting significantly the number of citation was journal self-citation. However, the variance explained by journal self-citation was not sufficient. There are some findings that variables such as open access status of journal are related to the number of citation. Since all journals in this research have open access, open access status could not been included as a predictor variable. That is why; new researches including this variable may obtain different results. Moreover, only Turkish education journals indexed at SSCI were included in this study. If researchers increase the number of journal by including journals not indexed at SSCI, different results may be obtained.

KeyWords

Number of citation, Journal self-citation, Number of authors

Bildiri No: 83

Lisans Düzeyi İçin Hazırlanmış Ölçme ve Değerlendirme Kitaplarının İçerik Açısından İncelenmesi

Rana SALMANER*, Sena GENCAN ACAR**, İsmail KARAKAYA***

Özet

Kitaplar kişilerin kendi kendine öğrenmesini, eğitim ve öğretim faaliyetlerinin etkin bir şekilde sürdürülebilmesini ve geliştirilebilmesini sağlayan önemli materyallerdir.Bu araştırmanın amacı Türkiye'de Ölçme ve Değerlendirme alanında uzman olmayanlar tarafından bu alanda yazılan kitaplardaki bilimsel ve teknik hatalar ile eksikliklerin tespit edilmesidir.Çalışmada Ölçme ve Değerlendirme alanında alan dışı uzmanlar tarafından yazılan kitaplar evren olarak belirlenmiştir. Evrende bulunan 9 kitaptan ulaşılabilen 7 kitap incelenmiştir. Kitaplara ait incelemelerin dokuman analizi ile yapıldığı bu araştırmada veriler, araştırmanın amacına bağlı olarak belirlenen kategoriler altında, araştırmacılar tarafından geliştirilen "Doküman inceleme formu" ile toplanmıştır. Kitaplar farklı araştırmacılar tarafından puanlandığı için, puanlayıcı güvenirliğini sağlamak amacıyla Kappa katsayısı hesaplanmış ve puanlayıcılar arası uzlaşma anlamındaki güvenirliğin sağlandığı tespit edilmiştir. Bu çalışmada kitaplar "İçeriğin Niteliği, Kapsamı ve Yeterliliği", "Görsel Ögeler", Örnek Materyaller" ve "Kaynak Kullanımı" olmak üzere 4 açıdan ele alınmıştır. Alan dışı uzmanlar tarafından yazılan kitapların bu açılardan yeterlikleri belirlenmiş ve kitapların yeterlik düzeyleri birbirleriyle karşılaştırılmıştır. Analizler sonucunda kitaplarda genel olarak kavramlardaki eksikliklerin en fazla ölçme ve değerlendirmedeki temel kavramlar ve istatistik konularında olduğu görülmüştür. Kavram yanılgısının en fazla olduğu konuların ise alternatif ölçme araçları ve istatistik olduğu görülmüştür. Kitapların genelinde ölçme ve değerlendirme konusunda en fazla yanlış örnek kullanıldığı belirlenmiştir. Ölçme araçlarında bulunması gereken özellikler ve test geliştirme konularında ise hatalı ifadelerin çoğunlukta olduğu görülmektedir. Örnek materyallere bakıldığında klasik ölçme araçlarında en az hatanın eşleştirmeli testler ve tamamlamalı maddelerde olduğu görünmüştür. Genel olarak performansa dayalı (tamamlayıcı-alternatif) ölçme araçlarına bakıldığında ise en fazla hatanın "dereceli puanlama anahtarı ile ilgili hatalar" ile ilgili olduğu görülmektedir. Elde edilen sonuçlar doğrultusunda alan uzmanı olmayan kişiler tarafından yazılan kitapların geliştirilmesi gerektiği görülmüştür.

Anahtar Kelimeler

Kitap inceleme, Doküman analizi, Ölçme ve değerlendirme.

*Gazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: rasalmaner@hotmail.com
**KastamonuÜniversitesi Kastamonu Meslek Yüksek OkuluTasarım Bölümü. İletişim:senagencan@gmail.com
***Gazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:ikarakaya2002@gmail.com

Paper No: 83

Investigation of Measurement and Evaluation Books Written for Higher Education Level in Term of the Content

Abstract

The books are the most important materials to provide self- learning, and developing educational activities in an effective way. The purpose of the study is to determine scientific and technical mistakes and lacks in the books written by the authors who are not expert in Measurement and Evaluation Field in Turkey. In this study, the books written by non-expert authors in Measurement and Evaluation are identified as population. Seven books from nine books in population are investigated. Examination of the books is made by using document analysis method, and "Document Examination Form" is developed by researchers. Since the books are rated by different raters, Kappa coefficient is used in order to provide rater reliability and it is determined that the reliability meaning as inter rater agreement is provided. In this study, the book are investigated in terms of the categories, such as "feature, scope, andsufficiency of the content", "visual component", "example materials", and "using sources". In these terms, sufficiency of the books is examined and compared with each other. As a result of analysis, it is observed that lacks in concept is seen most in basic concepts and statistics topics. The topics misconception most occurs are alternative measurement tools and statistics. It is seen that the most in correct examples are used in measurement and evaluation topic in the books. In topics, such as the features which must be in measurement tools and test developing, it is observed that wrong using is the most common. Looking at example materials, in classical measurement tools, the most mistakes are observed about matching and completing items. For performance based (alternative) measurement tools, this situation can be seen about rubrics. In the direction of the results obtained, it is seen that the books written by non- expert authors in the field must be developed in terms of some scientific and technical features.

Key Words

Investigation of the books, Document analysis, Measurement and Evaluation.

Bildiri No: 84

Bir Avuç Mutluluk: Ölçme Ve Değerlendirme Dersinde Bir Öğretim Tasarımı

Kenan DEMİR*

Bu çalışmada ölçme ve değerlendirme dersinin öğretme-öğrenme sürecinde tam öğrenme modeli dikkate alınarak bir öğretim tasarımı uygulanmıştır. Ölçme ve değerlendirme dersine özgü bir öğretim tasarımının geliştirildiği, uygulandığı ve değerlendirildiği bu çalışmada Dick-Carey ile ASSURE Tasarım modelleri dikkate alınmıştır. Bu öğretim tasarımı "İhtiyaç analizi, Hazırlık, Uygulama, Değerlendirme, Yeniden düzenleme" olmak üzere 5 aşamada yürütülmüş ve bu süreçte "Kavramlar" ve "İşlemler" olmak üzere iki öğretim tasarımı geliştirilmiştir.

Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi matematik, Türkçe, okulöncesi, fen bilgisi öğretmenliği ile rehberlik ve psikolojik öğretmenliği bölümü öğrencileriyle, 2012-2015 yılları arasında düzenlenen bu çalışmalarda öncelikle dersin adı "Bir Avuç Mutluluk" olarak değiştirilmiştir. Öğretmen adaylarının var olan önyargılarını kırmak ve öğretim sürecini daha da somutlaştırmak, kolaylaştırmak, eğlenceli hale getirmek adına yapılan bu çalışma ile öğrencilerin derse bir avuç kadar hazırlıklı gelmeleri, öğretim sürecine etkin katılmaları ve izleme değerlendirmesi yapmaları sağlanmıştır. Dersin öğretim sürecinde her konuya özgü karma öğretme-öğrenme etkinlikleri eş güdüm içinde uygulanmıştır. Bu öğretme-öğrenme etkinliklerinde temel olarak proje tabanlı ve işbirliğine dayalı öğrenme, yaratıcı drama, çoklu zeka, tartışma gibi öğretme yolları kullanılmıştır. Bu öğretme yolları ile birlikte her bir konuya özel olmak üzere dikkat, tekrar, anlamlandırmayı güçlendirme (açıklama ve soru sorma, eklemleme-ilişkilendirme, benzeştirme, karşılaştırma, gruplama, sıralama, öğrenilen konuyu şekil, grafik ve imgelerle yazıya dökme, kendi kendine soru sorma, bir başkasına anlatma, öğretme, benzetim, not alma, özetleme, şemalaştırma) gibi öğrenme stratejileri öğretmen ve öğrencilerle birlikte kullanılmıştır. Öğretmen adaylarının yürütücü biliş ve duyuşsal stratejileri kullanmalarına imkan tanımak amacıyla öğretme-öğrenme süreci yaratıcı öğrenme merkezi, öğrenci kantini, okul bahçesi gibi çok farklı öğrenme ortamlarda yürütülmüştür.

Öğretme ve öğrenme stratejilerinin birlikte kullanıldığı bu öğretim tasarımının öğretme-öğrenme sürecinde işbirliğine dayalı öğrenmenin ayrılıp-birleşme (jigsaw-II) ve grup araştırması teknikleri; yaratıcı dramanın doğaçlama, rol oynama, oyun, toplantı düzenleme, istasyon, bilgi koridoru, rulman teknikleri kullanılmıştır. Yine bu süreçte ata sözü deyim, karikatür, özlü söz, gerçek yaşamdan örnek durumlar, resim, şekil, hikaye, fıkra, şarkı, ritim, dans, masal vb. olmak üzere konuya uygun, yaşamda yeri olan çok çeşitli öğretme-öğrenme yollarının birlikte kullanılmasına özen gösterilmiştir.

Örneğin ölçeklerin öğretim sürecinde gerçek yaşam durumları, rol oynama, loci, tekrar (açık-kapalı ve bireysel-grupla), şekil, grafik, kafiye, ritim, dans, bedeniyle kodlama, göz bebeklerini kullanarak kodlama gibi birçok öğretme-öğrenme yolu birlikte kullanılmıştır.

Öğrencilerin her ders öncesi hazırlıklı geldikleri, etkin katıldıkları ve konu ile ilgili soru çözdükleri bu süreçte hazırbulunuşluluğun, öğrenmenin, öğretmenin, başarının ve diğer

*Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü, kenandemir@mehmetakif.edu.tr

(sosyal ve psikolojik) özelliklerin değerlendirilmesinde bütüncül bir ölçme ve değerlendirme yaklaşımı izlenmiştir. Öğretmen adaylarının hazırbulunuşluğu 50 maddelik çoktan seçmeli bir test ile belirlenmiştir. Öğretme-öğrenme sürecinde ise yine parametrik özellikleri belirlenmiş ünite testleri ile nitel verilerin toplandığı yazılı, sözlü değerlendirmeler yapılmıştır. Ayrıca vize ve final sınavları "ölçme değerlendirme kavramlarının gerçek yaşamdaki ayak izleri-artık sınavlar da değişmeli" anlayışıyla her öğretmen adayının kendi sınavını kendisinin hazırlamasına imkan tanınmıştır. Bu süreçte öğretmen adayları dersin içeriğinde olduğu gibi ölçme ve değerlendirme kavramlarının gerçek yaşamdaki ayak izlerini, karşılıklarını araştırmıştır. Öğretmen adayları bu kavramların karşılığı olan örnek durum, olay, yazı, hikaye, şiir, tiyatro gösterisi, masal yazma, atasözü, deyim, fotoğraf, karikatür, afiş, slogan, kısa film, kitap vb. hazırlama gibi bildik olmayan yaratıcı yollarla kendi sınavlarını sunmuşlardır. Öğretmen adaylarının kendi yeteneklerine uygun olarak çok farklı yollarla öğrenip-öğrenmediklerini ifade ettiği bu çalışmalar yerel ve uluslararası olmak üzere sergilenmiştir. Yine öğretmen adaylarının ölçme kavramlarına ilişkin yazdıkları sözler bildik şarkılarla eşleştirilmiş ve bu şarkılar yine öğretmen adayları tarafından söylenerek "Ölçme Kavramlarının Müziği" adıyla bir konser düzenlenmiştir.

Öğretmen adaylarının vize ve final sınavı kapsamında yaptıkları bu performans görevleri (ürünler) genel ve dereceli puanlama anahtarı kullanılarak öğretmen adaylarının kendileri, diğer sınıflardaki akranları ve öğretim elemanı tarafından değerlendirmiştir. Dersin değerlendirme yaklaşımı gereği ayrıca 50 maddelik ölçeğin ön ve son test olarak kullanılması, her ders başında bir önceki hafta ile ilgili soru çözümlerinin yapılması, ders sonunda konu testlerinin uygulanması ya da öğrencilerin yazılı görüşlerinin alınması ile de giriş, süreç ve çıkış olmak üzere tüm süreç değerlendirilmiştir.

Bu çalışmalar sonunda öğretmen adaylarının erişi düzeylerinin ortalama olarak 12 puandan başladığı çalışmalar sonunda 92 puana kadar çıktığı belirlenmiştir. Yaratıcı etkinlikler sonunda ortaya çıkan yaklaşık 72 puanlık ilerlemenin ise istatistiksel olarak anlamlı olduğu bağımlı gruplar arası t testi ve ANAVO testi ortaya çıkmıştır.

Öğretmen adayları bu çalışmaların kendilerine ölçme ve değerlendirme dersi sadece bir ders olarak değil aynı zaman da çok farklı öğretme-öğrenme yollarının kullanılmasıyla kendilerine öğretme-öğrenme sürecinin nasıl tasarlanıp, yürütüleceği ve değerlendirileceğinin bütüncül olarak örneklendirildiği bir laboratuvar ortamı sunduğunu ifade etmişlerdir.

Anahtar Kelimeler

Ölçme ve Değerlendirme, Öğretim Tasarımı, Öğretme stratejileri, öğrenme stratejileri, yaratıcı drama, işbirliğine dayalı öğrenme

Paper No: 84

A Handful of Happiness: A Teaching Design in the Measurement and Evaluation Lesson

In this study, an instructional design is applied regarding the whole learning model in the process of teaching and learning of Measurement and Evaluation lesson. Dick-Carey and Assure Desing Models are regarded in this study in which a teaching design, unique for the lesson of Measurement and Evaluation, has been developed, applied, and assessed. This an instructional design has been carried out in 5 processes as ‘‘Needs analysis, Preparation, Application, Assessment, Reorganizing’’ and two instructional design have been developed as ‘‘Concepts’’ and ‘‘Processes’’.

In this studies, organized with the students from the departments of Mathematics, Turkish, Pre-School, Science Teaching, Psychological Counseling and Guidance at Mehmet Akif Ersoy University from 2012 to 2015, the name of the lesson has initially been changed as ‘‘A Handful of Happiness’’. With this study carried out to overcome the prejudice of teacher candidates and make the teaching process more concrete, easier, and funnier, coming ready to the lesson as much as a handful, active participation in the learning process, and making observation assessment of the students have been enabled. During the teaching process of the lesson, mixed teaching and learning activities unique for each subject have been applied with coordination. In this teaching and learning activities, teaching methods such as project-based and collusive learning, creative drama, multiple intelligence, discussion have been used. Apart from these teaching methods learning strategies such as attention, repetition, making stronger of giving the meaning (explaining and questioning, articulating-associating, homogenization, comparing, grouping, grading, performing learned subject with shapes, diagrams, and images in the written form, questioning by oneself, telling someone else, teaching, analogy, taking note, summarization, schematization) as unique for each subject have been used with the teachers and students. With the aim of providing the opportunity of using executive cognition and affective strategies, teaching and learning process has been carried out such different learning environments as creative learning center, school canteen, and schoolyard.

During the process of teaching and learning of this an instructional design, in which teaching and learning strategies are used together, jigsaw-II and group research techniques of collusive learning and improvisation, acting, play, organizing meeting, station, information aisle, roller techniques have been used. Simultaneous use of teaching and learning methods, have a large variety and belong to life, suitable for the subject such as proverb, idiom, caricature, saying, anecdotes, painting, shape, story, joke, song, rhythm, dance, novel has been cared.

For instance, lots of teaching and learning methods such as real-life situations, acting, loci, repetition (open-closed and individual-group), shape, diagram, rhyme, rhythm, dance, coding by body, coding by use of pupils have been used in the teaching process of scales.

In this process, in which the students come ready before each class, participate in classes actively, and answering the questions upon the subject, a totalitarian measurement and evaluation approach has been applied to evaluate readiness, tutor, teaching, success, and other (social and psychological) qualities. Readiness of teacher candidates has been determined by a multiple-choice test which consists of 50 items. During the process of teaching and learning, unit tests, determined its parametric qualities, and written-oral evaluations, in which

qualitative data is acquired, have been carried out. Moreover, it is enabled that every teacher candidate prepares his/her own visa and final exams with the opinion of ‘‘Footprints of assessment and evaluation concepts in real life -exams- must change after now.’’ In this process, teacher candidates have researched footprints of measurement and evaluation concepts in real life -their equivalents- as in the process of the class. Teacher candidates have presented their exams with unknown creative methods, which are the equivalents of these concepts, such as precedent, event, writing, story, poem, theatre, writing novel, proverb, idiom, photograph, caricature, poster, slogan, short film, book. These studies, by which teacher candidates state whether they have learned or not with many different methods suitable for their own capabilities, have been showed as local and international. Lyrics written by teacher candidates upon measurement concepts have been matched with known songs and a concert has been organized named as ‘‘The Music of Measurement Concepts’’ by means of teacher candidates’ singing these songs.

These performance tasks (products), teacher candidates have carried out within visa and final exams, have been assessed by use of general grading key and rubric and by teacher candidates, their peers in the other classes, and the instructor. As a matter of assessment approach of the lesson, the whole process has separately been evaluated as input, process, and output by the use 50 item scale as pre- and final test, answering the questions about former week at the beginning of each class, practice of subject tests at the end of the class, or taking students’ opinions in written form.

At the end of this studies, it has been determined that attainment level of teacher candidates, in average, has started from 12 points and reached to 92 points. It is shown up by t test between dependent groups and ANAVO test that nearly 72 points progression emerged as a result of creative activities is meaningful statistically.

Teacher candidates have stated that these studies has presented the assessment and evaluation lesson not only as a lesson but also as a laboratory environment, in which design, procedure, and assessment of teaching and learning process are sampled as a whole, by use of many teaching and learning methods.

Key Words

Measurement and Evaluation, Instructional Design, Teaching Strategies, Learning Strategies, Creative Drama, Collusive Learning

Bildiri No: 85

DINA Modele Dayalı Madde Parametre Kestiriminde Kayıp Veri Ele Alma Yöntemleri Etkisinin İncelenmesi

Doç. Dr. Tahsin Oğuz BAŞOKÇU[1], Dr. Ömür Kaya KALKAN[2], Doç. Dr. Tuncay ÖĞRETMEN[3]

Özet

Verilerin analizinde karşılaşılan en yaygın problemlerden bir tanesi kayıp veridir. Kayıp veri çoğunlukla uygulama sırasında meydana gelmektedir. Hız testlerinde, büyük ölçekli sınavlarda veya tutum ölçeklerinde kayıp veri sorunu ile karşılaşılabilmektedir. Ancak bir örneklem büyüklüğünde, ne kadarlık bir kayıp verinin kabul edilebileceğine dair bir ölçüt bulunmamaktadır. Türkiye'de yürütülen eğitim araştırmalarında özellikle, kayıp veri sürecinin raporlanmasında önemli eksiklikler bulunduğu, bu konuya kayıtsız kalındığı ve kayıp veri ele alma yeterlik düzeyinin yeterli seviyede olmadığı ifade edilmektedir. Farklı örneklem büyüklüklerinde görülen kayıp veri ve kayıp veri ele alma yöntemlerinin farklı güvenirlik katsayıları, faktörler tarafından açıklanan varyans yüzdesi, t testi ve ANOVA parametreleri, MTK madde parametre kestirimleri üzerinde doğrudan bir etkiye sahip olduğu belirtilmektedir. Bu araştırmada farklı örneklem büyüklüklerinde görülen, farklı miktardaki kayıp verilerin, DINA model ile elde edilen madde parametreleri üzerindeki etkisi ele alınmıştır. Araştırmada, DINA model g ve s parametrelerinin 0,1 de sabitlendiği ve 0,1-0,3 arasında değiştiği, 30 madde ve 4 özellikten oluşan 10000 kişilik 2 simülasyon verisi kullanılmıştır. Bu veriler üzerinden tesadüfi olarak %5, %10 ve %15 düzeyinde 100 farklı örneklem için kayıp veri örneklemi üretilmiştir. Kayıp veri içeren veri setlerine, veri atamada MCMC, EM, ve regresyon tekniğinden yararlanılmıştır. Bu veri setleri ile verilerin silinmemiş halinden elde edilen madde parametreleri ve örtük sınıf kestirimleri karşılaştırılmıştır. Madde parametreleri bias değerleri, örtük sınıf uyumları ise uyum yüzdeleri ile karşılaştırılmıştır. Araştırma sonucunda elde edilen bulgulara göre, EM yönteminin, kayıp madde parametreleri ve örtük sınıfların tutarlılıkları açısından diğer yöntemlere göre daha iyi sonuçlar verdiği görülmüştür.

Anahtar Kelimeler

DINA model, kayıp veri, MCMC, EM

Paper No:85

Examination of Effects of Missing Data Handling Methods in Item Parameter Estimation based on DINA Model

Abstract

Missing data is the most common problem encountered in data analyses. The problem of missing data generally occurs during the application. It can occur in speed tests, large scale examinations or behavior scales. However, there is no established cut-off criterion for an acceptable amount of missing data in a sample size. The prevailing view is, so far as the educational researches carried out in Turkey are concerned, that there are substantial deficiencies especially in reporting the missing data process marked by a general indifference in this field and an insufficient level of missing data handling. It is stated that missing data and missing data handling methods occurring in different sample sizes have a direct effect on different reliability coefficients, variance percentages explained by factors, t test and ANOVA parameters and MTK item parameter estimations. This study explores the effect of different amounts of missing data occurring in different sample sizes on item parameters that have been specified based on DINA model. A simulation data which covers 2000 persons and comprises 30 items and 4 attributes where DINA model g and s parameters varied between 0,1 - 0,3 across the items was utilized in the research. Based on these data, missing data samples were produced for 30 different samples set randomly at the levels of 5 %, 10 % and 15 %. MCMC, EM and regression techniques were used to merge data into data sets containing missing data. Item parameters attained from undeleted records of these data and implied level estimations were compared with these data sets. In respect of the coherency of item parameters and implied levels, the findings attained in the research have shown that while EM method yields better results when there is a low level of missing data compared to other methods, MCMC method is more effective in cases where missing data level turns out to be high. The research has further shown that the regression method, on the other hand, is affected by the missing data rate to a lesser extent compared to other methods providing values close to each other at missing data rates of 5 %, 10 5 and 15 %.

Key Words

DINA model, missing data, MCMC, EM

Bildiri No: 87

PISA 2012 Matematik Alt Testlerinden Seçilen İki Kitapçığın Gözlenen Puan Eşitleme Yöntemleri İle Eşitlenmesi

Sami PEKTAŞ*, Mustafa KILINÇ**

Öğrencilerin başarı düzeylerini artırmak, eğitim politikalarının öğrenci üzerindeki etkisini görmek, eğitim sistemini daha işlevsel hale getirebilmek ve eğitim kalitesini yükseltmek için Türkiye'nin de kurucu üyesi olduğu Ekonomik İş Birliği ve Kalkınma Teşkilatı (OECD)'nın 2012 yılında düzenlediği Uluslararası Öğrenci Değerlendirme Programı (PISA)'nın matematik alt testlerini oluşturan 13 kitapçık bulunmaktadır. Bu kitapçıklardan ortak maddeli iki kitapçık seçilerek klasik test kuramına dayalı eşitleme yöntemlerinden ortalama, lineer, eşit yüzdelikli eşitleme ile eşitlenmiş puanların karşılaştırılması amaçlanmıştır.

Bu amaca yönelik kurulan model, ortak maddeli denk grup deseni olarak nitelendirilebilir. Kullanılan iki testin kapsamının, yapısının, güvenirlik katsayılarının, varyanslarının, ortalama güçlüklerinin ve puan ortalamalarının denk olduğu yani aralarında anlamlı bir farklılık olmadığı saptanmıştır. Bununla beraber her iki teste ait tanımlanan tek faktörlü yapının doğrulanmasıyla eşitleme işlemlerinin gerçekleştirilebileceği söylenebilir.

Eşitleme türleri uygulanarak elde edilen bulgulara göre ortalama eşitleme yapılarak elde edilen eşitlenmiş Form X puanı ile eski form olan Form Y puanı karşılaştırıldığında sadece ortalamalarının eşitlendiği görülmektedir. Lineer (doğrusal) eşitleme yapıldığında ortalama ve standart sapmalarının eşitlendiği görülmektedir. Eşit yüzdelikli eşitleme yapıldığında ise ortalama, standart sapma, çarpıklık ve basıklık değerlerinin eşitlendiği görülmektedir. İki kitapçık üzerinden formların eşitlenmesi işlemleri dikkate alındığında puan dağılımlarının en doğru şekilde eşitlenmesini sağlayan yöntemin eşit yüzdelikli eşitleme olduğu sonucuna ulaşılmıştır. Bu çalışmanın sonuçları doğrultusunda, Türkiye'de seçme amaçlı yapılan birçok sınavın farklı versiyonlarının benzer sonuçlar üretip üretmediğinin kontrol edilmesi gerektiği söylenebilir. PISA gibi uluslararası alanda yapılan sınavların ve sınavların alt testlerine ait kitapçıkların eşitlenmesi için farklı test eşitleme desenleri çalışılarak KTK ve MTK'ya dayalı test eşitleme yöntemleri ile eşitleme işlemleri gerçekleştirilebilir. Araştırmanın genişletilmesi için randum grup deseni kurularak yapılan eşitleme sonuçları ile çalışmanın kapsamına ilişkin kurulan desene ait eşitleme yönteminden elde edilen sonuçların karşılaştırılması yapılabilir.

Anahtar Kelimeler

PISA, Ortak Madde, Ortalama Eşitleme, Doğrusal Eşitleme, Eşit Yüzdelikli Eşitleme

* Gazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı İletişim: pektassami@gmail.com Tel: +905062643401

** Burdur Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı İletişim:mkilinc@mehmetakif.edu.trTel: +905053192868

Paper No: 87

The Equation of Two PISA 2012 Mathematics Subtest Booklets Through Observed Score Equating Method

Abstract

Programme of International Student Assessment (PISA), which was organized by Organization of Economic Cooperation and Development (OECD) in 2012 in order to increase students' academic achievement, determine the effect of educational programmes on students, make educational systems more functional and increase the quality of education, has 13 booklets for Mathematics Subtests. The aim of this study is to compare the scores equated through average, linear, equipercentile method based on classical test theory by choosing two booklets with common items.

The model structured with this aim can be regarded as common items equivalent group design. The content, structure, reliability coefficient, variances, average difficulty and score averages of these two tests were observed as equal to each other, which means that there was no significant difference between these two tests. Besides, it can be said that the equating of these two tests can be realized through the confirmation of the one factor structure defined for these two tests.

According to the results of the study, from the comparison of Form X equated scores obtained through average equating and the old Form Y scores, it can be observed that only their averages were equated. Through linear equating, averages and standard deviations of the tests were equated. By means of equipercentile equating, standard deviations, averages and skewness and kurtosis or the tests were equated. After all the equating processes applied to the two booklets, it can be said that the best method to equate the score distributions in the most correct way was equipercentile equating method. In light of the results of this study, it can be suggested that whether the different versions of many exams in Turkey produce similar results should be checked. To equate the international tests and their subtests booklets like PISA, different test equation models can be studied and the equating can be realized through test equation methods based on classical test theory and item response theory. Further studies can be conducted by comparing the equation results obtained through a structured random group design with the results obtained through the equating model structured in this study.

Key Words

PISA, Common Item, Average Equating, Linear Equating, Equipercentile Equating

Bildiri No: 88

Kopya Belirleme İndekslerinin Madde Ayırt Edicilik ve Güçlük Düzeylerine Göre Performanslarının İncelenmesi

Önder SÜNBÜL*, Seha YORMAZ**

Özet

Kopya, test güvenliği ve buna bağlı olarak test geçerliği için tehdit oluşturmaktadır. Bu nedenden dolayı, kopya belirleme yöntemleri hakkında bir çok araştırma yapılmıştır. Bu çalışmalar bir kısmı yeni yöntemler geliştirmek üzerine odaklanırken bazıları da var olan yöntemlerin çeşitli koşullar altındaki performanslarını incelemeyi amaçlamaktadır. Çoktan seçmeli maddelerden oluşan testlerde kopya belirlemede sıklıkla kullanılan ve en çok itibar edilen indeksler ω (omega) indeksi ve GBT (Generalized Binomial Test) indeksidir. Çalışmalarda çeşitli koşullar altında bu indekslerin I. Tip hata oranı ve kopya belirleme gücü incelenmiştir; ancak bu indekslerin performansını etkileyebilecek olan madde ayırt ediciliğinin ve madde güçlüğün çeşitli düzeylerinde bu indekslerin davranışlarını inceleyen bir çalışmaya rastlanamamıştır. Bu çalışmada ω ve GBT indekslerinin çeşitli α düzeylerinde 40 ve 80 test uzunluğu ve 2000 örneklem büyüklüğü için çeşitli madde ayırt edicilik ve güçlük sınırlandırmaları altındaki I. Tip hata oranı ve kopya belirleme gücü incelenmiştir. Çalışma sonucunda bu koşullar altında her iki indeksin I. Tip hata oranlarının kabul edilebilir sınırlar içinde olduğu görülmüştür. Güç çalışmasında ise madde ayırt edicilik indeksinin, madde güçlük indeksinin ve bunların ortak etkileşiminin kopya indekslerinin güç performansını etkilediği görülmüştür. Beklenenin tersine ayırt edicilikteki artış indekslerin kopya belirleme gücünü düşürücü yönde etki yapmıştır. Güçlük düzeyindeki artış ile kopya indekslerinin gücünün arttığı görülmüştür. Ortak etki açısından incelendiğinde özellikle b parametresinin 0'ın altında olduğu değerlerde ve düşük ayırt edicicilikteki maddelerden oluşan testlerde zayıf performans gözlemlenmiştir. Buna ek olarak b parametresinin 0'ın altında olduğu ve yüksek ayırt edicilikteki maddelerden oluşan testlerde kopya belirleme indekslerinin çok yetersiz performans sergilediği görülmüştür. Madde sayısının 40 olduğu testlerde ω indeksinin GBT indeksine göre daha güçlü olduğu görülmüştür.

Anahtar Kelimeler

Kopya belirleme, ω indeks, GBT indeks, I. Tip hata, güç

*Mersin Üniversitesi Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü. İletişim:ondersunbul@mersin.edu.tr Tel: +903243412815/2018

**MersinÜniversitesi Eğitim Bilimleri Fakültesi Eğitim Bilimleri Bölümü. İletişim:sehayormaz@gmail.com Tel: +903243412815/2018

Paper No: 88

Investigating the Performance of Copy Detection Indices According to Item Discrimination and Difficulty

Abstract

Cheating is a threat for test security and test validity. For this reason, several studies have been conducted for copy detection methods. Some of these studies focused on developing new methods and some aimed to investigate the performance of existing indices under several conditions. ω (Omega) and GBT (Generalized Binomial Test) are the most trustable, frequently used indices as copy detection technique for multiple choice item tests. In several studies, Type I Error rates and copy detection power of these indices were investigated. However, there could be no study related to the performance of these indices for item sets which have different item discrimination and difficulty levels. In this study, Type I Error and the power rates of ω and GBT indices under the restriction of several item discrimination and difficulty levelswere investigated for several nominal alpha levels, 40 and 80 test lengths for 2000 sample size. As a result of study, it has been observed that Type I Error rates of both indices were below the acceptable nominal alpha levels. For the power study, it has been observed that both of item discrimination and difficulty levels were effective for power performance. In contrast to expectations, it has been observed that the increase of discrimination decreased the power. The increase of item difficulty increased power of indices. The results of the interaction effects of discrimination and difficulty showed that for the cases whose b parameters were lower than 0 with weak discrimination, indices had weak power performance levels for both of ω and GBT. Additionally, for the cases whose b parameter levels were below 0 with high discrimination indices, the power performance of both copy detection indices were very weak. Another result related with test length showed that for the cases which have 40 items ω performed better than GBT in terms of power rate.

Key Words:

Copy detection, ω index, GBT index, Type I error, power

Bildiri No: 89

Kayıp Veri Giderme Yöntemlerinin Değişen Madde Fonksiyonu Saptama Teknikleri Üzerindeki İşleyişinin İncelenmesi

Hüseyin SELVİ*, Devrim ALICI**

Özet

Bu çalışmada, farklı kayıp veri giderme yöntemlerinin klasik test kuramına dayalı mantel haenzsel ve standardizasyon ilemadde tepki kuramı'na dayalı olabilirlik oran testi teknikleri üzerindeki işleyişini incelemek amaçlanmıştır.

Bu amaç doğrultusunda, Mersin Üniversitesi (ME.Ü.) tarafından 2016 yılında düzenlenen Yabancı Uyruklu Öğrenci Sınavı (YÖS) Temel Beceriler alt testine katılan toplam 1046 birey ve bu bireylerden elde edilen veri seti üzerinde, farklı kayıp veri giderme yöntemleri kullanılarak, değişen madde fonksiyonu analizleri (mantel haenszel, standardizasyon ve olabilirlik oran testi) yapılmıştır. İlgili alt test ikili (1-0) puanlanan çoktan seçmeli 80 maddeden oluşmakta olup veri setinde 523 kadın ve 523 erkek birey bulunmaktadır.

Çalışma kapsamında elde edilen bulgular, değişen madde fonksiyonu gösteren madde sayılarının kullanılan değişen madde fonksiyonu saptama tekniğiyle birlikte değiştiğini, klasik test kuramına dayalı değişen madde fonksiyonu saptama tekniklerinin kendi içinde madde tepki kuramına dayalı değişen madde fonksiyonu saptama tekniğine göre daha uyumlu oluğunu, kullanılan kayıp veri giderme yaklaşımlarının ise değişen madde fonksiyonu saptanan maddeleri farklılaştırdığını ve bu farkın mantel haenszel tekniği için anlamlı seviyeye ulaştığını göstermiştir.

Anahtar Kelimeler

Değişen madde fonksiyonu, kayıp veri, beklenti maksimizasyonu, regresyona dayalı atama

*Mersin Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı. İletişim: hsyn_selvi@yahoo.com.tr
**Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: devrimo.alici@gmail.com

Paper No: 89

Investigating Differential Item Functioning Methods Under Different Missing Value Reduction Methods Situations

Abstract

The aim of this study is to examine the mechanism of different missing data methods on classical test theory based mantel haenszel and standardization and item response theory based likelihood ratio test.

For this purpose, Foreign Student Examination (YOS) Basic Skills subtest organized by Mersin University at the 2016 was used. There are 523 female and 523 male (Total: 1046) individuals on the data set. Related subtest has binary (1-0) scored 80 multiple choice items. The analysis of differential item functioning (Mantel Haenszel , standardization and likelihood ratio test) was conducted by using different missing value reduction methods on data set.

Findings from the study showed that, the number of items which has differential item functioning has changed with differential item function detection techniques. Classical test theory based methods are more consistent on its own according to method based on item response theory. And the findings showed that the difference has reached to the significant level for mantel haenszel technique between different missing value reduction method.

Key Words

Differential İtem Functioning, Missing Value, Expexted Maximization, Regression İmputation

BildiriNo: 90

Farklı Puanlama Anahtarlarının Güvenirliğinin Genellenebilirlik Kuramı İle İncelenmesi

Mehtap AKTAŞ*, N. Bilge BAŞUSTA**, Devrim ALICI***

Özet

Çalışmanın amacı farklı puanlama anahtarlarının güvenirliğinin genellenebilirlik kuramı ile incelenmesidir. Araştırmanın çalışma grubunu ilkokul 5. Sınıfa devam eden 52 öğrenci ile bu öğrencilerin hikaye yazma becerilerini değerlendirmek üzere Mersin Üniversitesi Sınıf Öğretmenliği ve Türkçe Öğretmenliği bölümü 3. sınıflarında öğrenim gören 20 puanlayıcı oluşturmaktadır. Puanlama anahtarlarının sabit yüzey olarak belirlendiği dengelenmiş ve tümüyle çaprazlanmış hxpxg deseni ile G çalışması yürütülmüştür. Varyans bileşenlerinin kestirilmesinin ardından, görev ve puanlayıcı yüzey koşulları değişimlenerek Karar Çalışması gerçekleştirilmiştir. Sonuç olarak her iki puanlama anahtarında da, ana etkiler içerisinde hikayelere ilişkin değişkenlik kaynağının, ortak etkiler içerisinde ise hp ortak etkisinin yüksek açıklanan varyans yüzdesine sahip olduğu gözlenmiştir. Desen çözümlemesinde açıklanan en büyük varyans hpg ortak etkisine aittir. Bu durum, ortak etkileşimin yanı sıra desende yer almayan çeşitli değişkenlik ve hata kaynaklarının varlığının göstergesi olabilir. Her iki puanlama anahtarı için hesaplanan G ve phi katsayıları incelendiğinde, analitik rubrik ile puanlamanın daha güvenilir ve genellenebilir olduğu belirlenmiştir. Kontrol listesi için G ve phi katsayılarının ikisi de 0,62; analitik rubrik için ise 0,72 olarak bulunmuştur. Karar çalışmalarına göre puanlayıcı sayısının değişimlenmesinin G ve phi katsayısına etkisi, görev sayısının değişimlenmesine göre daha yüksek bulunmuştur.

Anahtar Kelimeler

Hikaye yazma becerileri, performans değerlendirme, kontrol listesi, rubrik, genellenebilirlik kuramı

*Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. E-mail: mhtpaktas@gmail.com +903243412815

**Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. E-mail: n.bilgeuzun@gmail.com +903243412815

***Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. E-mail: devrimo.alici@gmail.com +903243412815

Paper No: 90

The Examination Of Reliability Of Different Scoring Keys According To Generalizability Theory

Abstract

The aim of this study is to determine the reliability of different scoring keys in terms of the generalizability theory. The study group consists of 52 fifth graders from a primary school, who will write stories, and 20 bachelor students in their third year from the departments of Primary School Teacher Education and Turkish Language Teaching in Mersin University, who will evaluate and rate the stories. In this study, G study was conducted through a balanced and fully crossed hxpxg design, where scoring keys are set as the fixed facet. After the estimation of variance components, Decision Study was carried out by varying the surface conditions of the tasks and raters. As a result, it was observed that in both scoring keys, the source of variability related to the stories in the main effects and the common effect of hp in the common effects have a high percentage of variance explained. The common effect of hp has the largest variance explained in the design analysis. This may be indicative of the presence of various sources of error and variability which are not included in the design as well as common interaction. When G and phi coefficients calculated for both scoring keys are examined, it was determined that scoring with analytic rubrics is more reliable and generalizable. G and phi coefficients of the checklist and analytical rubrics were found to be 0.62 and 0.72 respectively. According to the Decision Study, the variance of the number of raters has higher effect on the G and phi coefficients than the variance of the number of tasks.

Key Words

Story writing skills, performance evaluation, checklist, rubric, generalizability theory

Bildiri No: 92

PISA 2012 Türkiye Örnekleminde Okul Özelliklerinin Matematik Okuryazarlığına Etkisi

Melek Gülşah ŞAHİN*, Yıldız YILDIRIM**, Elif SEZER***

Özet

Bu araştırmanın amacı okul iklimi ve kaynaklarının Türkiye'deki 15 yaş grubu öğrencilerin PISA 2012 matematik okuryazarlığına etkisini incelemektir. Bu amaç doğrultusunda PISA 2012 teknik raporunda yer alan okul iklimi indeksleri (öğretmenin odağı ve morali, okul iklimini etkileyen öğrenciyle ve öğretmenle ilişkili faktörler) ve okul kaynakları indekslerinin (okulun eğitimsel kaynaklarının kalitesi, fiziksel altyapının kalitesi) matematik okuryazarlığına etkisi ve bu değişkenler arasındaki ilişkiler yapısal eşitlik modellemesi ile betimlenmiştir. Araştırmada bazı okul değişkenleri ve matematik okuryazarlığı arasındaki ilişkileri belirlemeye yönelik bir model kurulduğu için ilişkisel bir araştırmadır. Araştırmanın çalışma grubunu PISA 2012'ye Türkiye'den katılan 170 okul ve 4848 öğrenci oluşturmaktadır. Araştırma, PISA 2012 Türkiye uygulaması kapsamında 15 yaş grubundaki öğrencilere uygulanan matematik okuryazarlığı testi ve okul yöneticilerine uygulanan okul anketinden elde edilen veriler üzerinden yürütülmüştür. Veri analizinde robust maximum likelihood parametre kestirim yöntemine dayalı yol analizi ve LISREL 8.8 programı kullanılmıştır. Kurulan modelde değişkenler arasındaki yol katsayıları incelendiğinde en yüksek katsayının okul iklimiyle ilişkili öğretmen faktörleri indeksi ve okul iklimiyle ilişkili öğrenci faktörleri indeksi arasında olduğu (0,65); en düşük katsayının ise öğretmenin odağı indeksi ve okul iklimini etkileyen öğretmen faktörleri indeksi arasında (0,24) olduğu gözlenmiştir. Kurulan modelde yer alan tüm yol katsayıları pozitif ve anlamlı bulunmuştur. Ayrıca bu yol katsayılarının orta ve yüksek etki büyüklüğüne sahip olduğu görülmüştür. Modele ilişkin uyum indeksleri (RMSEA= 0,028; X^2/sd=1,1275; GFI=0,96; AGFI=0,92; CFI=1,00; NFI=0,97; NNFI=0,99) ise modelin mükemmel uyum gösterdiğine işaret etmektedir. Bu bağlamda modele dayalı olarak okul iklimi ve okul kaynakları indekslerinin matematik okuryazarlığında etkili olduğu sonucuna varılmıştır.

Anahtar Kelimeler

Yol analizi, PISA 2012, Okul İklimi, Okul Kaynakları, Matematik Okuryazarlığı

* Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü, İletişim:mgulsah@gazi.edu.tr, Tel:+903122021821

** Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü, İletişim:yildizyildirim@gazi.edu.tr, Tel:+903122021821

*** Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü İletişim: eliffszr@gmail.com, Tel:+903122021821

Paper No: 92

The Impact of School Properties to Mathematics Literacy in the PISA 2012 Turkey Sample

Abstract

The aim of this paper was to examine the impact of school climate and resources in the PISA 2012 mathematics literacy for 15 years-old students in Turkey. For this purpose, impact of school climate index (teacher's focus and morale, student and teacher-related factors affecting school climate) and school resources index (quality of the school's educational resources, the quality of physical infrastructure) that situated PISA 2012 technical report to mathematics literacy and the relationship between these variables was described with structural equation model. This research was correlational because of establishing the model for that determine the relationship between some school variables and mathematics literacy. Research group was consisted of 170 schools and 4848 students which constitutes PISA 2012 from Turkey. The research was conducted mathematics literacy test which was administered to 15-years-old students and school questionnaires which was administered to school administrators in context of PISA 2012 Turkey. LISREL 8.8 computer program and path analysis method with robust maximum likelihood parameter estimation method was used in the data analysis. When the path coefficients between variables was examined, it was observed that the highest coefficient was between school climate related teacher factors index and school climate related student factors index (0,65) and the lowest coefficient between teachers focus index and school climate related teacher factor index (0,24) was observed. All path coefficient in the model was founded positively and significantly and this coefficients had been medium and high effect size. Also the fit indices for the model (RMSEA=0,028; X^2/sd=1,1275; GFI= 0,96; AGFI=0,92; CFI=1,00; NFI=0,97; NNFI=0,99) was shown that the model had perfect fit. In this context, it was concluded that school climate and school resources index was effective in mathematics literacy.

Key Words

Path Analysis, PISA 2012, Schools Resources, School Climate, Mathematics Literacy

Bildiri No: 93

Test Geliştirme Sürecinde Madde Seçimi İçin Hiyerarşik Puanlayıcı Model: Sinyal Algılama Puanlayıcı Modelinin Uyarlanması

T. Oğuz BAŞOKÇU*, Elçin MANAP, Duygu GÜNGÖR**

Özet

Hiyerarşik puanlayıcı modelinde, hakem puanları doğrudan öğrenci yeterliliğinin bir göstergesi olarak kabul edilmemektedir. Daha çok öğrenci performansının ait olduğu örtük kategorinin bir göstergesi olarak tanımlanır. Bu durum bir hiyerarşik yapıyı temsil eder. Örtük sınıf, sinyal algılama modeli (Signal Detection Model)'nde ise hakem hassasiyeti ve hakem katılığı-yumuşaklığı gibi çeşitli etkilerin ölçülmesine imkân verir. Bu çalışmada öğrenci performansı yerine geliştirilen maddelere ait hakem puanları; öğrencinin ait olduğu örtük sınıflar yerine ise maddelerin temsil ettiği özellik grupları incelenmiştir. Bu şekilde hiyararşik puanlayıcı modelinin bir türü olan sinyal algılama modelinin, test geliştirme sürecinde madde seçimi aşamasına uyarlanması amaçlanmıştır.

Anahtar Kelimeler

Sinyal Algılama Modeli, Hakem katılığı, Hiyerarşik puanlayıcı modeli

*Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. tahsin.oguz.basokcu@ege.edu.tr
**Dokuz Eylül Üniversitesi Psikoloji Bölümü

Paper No: 93

The Hierarchical Rater Model in Item Selection of Test Development Process: Adopting of Signal Detection Rater Model

Abstract

The raters' scores cannot be considered directly as an indicator of student proficiency in the hierarchical rater model (HRM). It is greatly defined that the indicator of latent category that involves student performance. This situation represents a hierarchical structure. In signal detection model (SDT), the latent class allows measuring of various rater effects as rater precision, severity or leniency. In this study, we examine the developed items' rater scores instead of student performance; attribute groups represented by the items rather than latent classes in which the students belong. In this way, it was aimed to adapt SDM which is a type of HRM to item selection stage in test construction process. According to research results, it is observed that model can be used in item selection which is based on judgments of the rater.

Key Words

Signal detection model, rater severity, the hierarchical rater model

Bildiri No: 94

Çoktan Seçmeli Testlerde Seçenek Ağırlıklandırma Yöntemlerinin Testin Faktör Yapısına Etkisinin İncelenmesi

Başak ERDEM, Levent ERTUNA, Nuri DOĞAN

Sınavların amacı, teste verilen cevaplardan öğrencilerin konuya ilişkin bilgilerini ortaya çıkarmaktır. Testten alınan puanın bireyin 'gerçek' bilgisini mümkün olduğunca doğru şekilde yansıtması gerekmektedir (Ecternacht, 1973; Lesage, Valke ve Sabbe, 2011).

Kurz (1999), çoktan seçmeli maddelerin bilişsel bilginin ölçülmesinde en çok tercih edilen soru formatı olduğunu belirtmiştir. Çoktan seçmeli testlerin en büyük avantajlarından birisi puanlama kolaylığı ve nesnelliği sağlamasıdır. Bu testler genellikle geleneksel doğru sayısı metoduyla puanlanır. Maddeler doğru cevaplar '1' yanlış cevaplar '0' olacak şekilde puanlandırılır. Bu yöntemde, bütün maddeler eşit şekilde ağırlıklandırılır. Geleneksel yöntem kolaylığına rağmen, bazı dezavantajları nedeniyle eleştiri almaktadır. Bu yöntemle bireylerin sahip olduğu bilgi miktarının doğrudan bir kestirimi yapılamaz (Crocker ve Algina, 2008; Kurz, 1999). Kurz (1999), birçok çoktan seçmeli testin sadece sıralama bilgisi sağladığını belirtmiştir. Bu durum özellikle test puanlarının bir programa giriş veya bir sertifika alacak adaylar arasında ayrım yapmak amacıyla kullanıldığı geniş ölçekli sınavlarda sorun yaratmaktadır.

Bu sorunların bazılarına çözüm getirmek amacıyla farklı puanlama yöntemleri önerilmiştir. Bu yöntemler, şans faktörünün etkisini azaltan ve kısmi bilgiyi hesaba katan yöntemleri içermektedir (Kurz, 1999). Frary (1989), bir test maddesinden alınabilecek kısmi bilgiyi, bireyin bazı yanlış seçenekleri eleyebilmesi, böylece tahmini doğru seçeneği de içeren uygun bir seçenek kümesine indirmesi olarak tanımlamıştır. Bireyin doğru cevabı bilememesi durumunda şansla cevaplaması veya boş bırakması yerine, çeldiriciler arasından eleme yoluyla seçim yapması varsayımına dayanır. Şans faktörü ise, bireyin bilgi sahibi olmadan doğrudan şans yoluyla seçim yaptığı durumda ortaya çıkar.

Seçenek ağırlıklandırma yöntemlerinde kısmi bilginin çeldiricilere farklı ağırlıklar verilmesi yoluyla ölçülebileceği varsayılmıştır. Her madde için birey, işaretlediği seçeneğe bağlı olarak bir puan alır (Lord ve Novick, 1968'den aktaran Çıtak, 2010). Seçenek ağırlıklandırma, madde cevap seçeneklerinin doğruluk derecelerinin farklılaştığı ve 'daha doğru' cevap seçeneğini seçen bireylerin 'daha az doğru' seçeneği tercih edenlere göre daha fazla bilgi sahibi olduğu varsayımına dayanır (Crocker ve Algina, 2008). Frary (1989), seçenek ağırlıklandırmasını (1) önsel ağırlıklandırma ve (2) deneysel ağırlıklandırma olarak 2 şekilde ele almıştır. Önsel ağırlıklandırmada, alan uzmanlarından her bir madde için seçenekleri doğru cevaba yakınlığına göre sıralamaları ve puanlamaları istenir. Sonrasında uzmanların seçeneklere verdikleri puanların ortalaması alınır ve alınan ortalamalarla orantılı olarak her bir seçeneğe ilişkin atanacak ağırlıklar belirlenir (Echternacht, 1973; Cross, Ross ve Geller, 1980). Deneysel ağırlıklandırmada ise, bir içsel ya da dışsal kriter aracılığıyla ağırlıklandırma yapılır. Seçenek ağırlıkları, bireylerin verdikleri cevaplara dayalı olarak belirlenir. Her bir madde için, cevaplayıcıların seçtiği seçeneklere bağlı olarak seçenek ağırlıkları belirlenir. Bu yöntemin altında yatan mantık, testten yüksek puan alan bireylerin seçtiği seçeneklere, düşük puanlı bireylerin seçtiği seçeneklerden daha fazla ağırlık verilmesidir (Echternacht, 1973; Crocker ve Algina, 2008).

Seçenek ağırlıklandırma yöntemlerinden önsel ağırlıklandırmanın, testin geliştirilmesi aşamasında emek ve maliyet gerektirmesi, uzmanların madde seçeneklerini ağırlıklandırmada her zaman tutarlı olmaması gibi dezavantajları bulunmaktadır. Bu nedenle, genellikle deneysel ağırlıklandırma yönteminin kullanılması tercih edilmektedir (Echternacht, 1973; Frary, 1989; Özdemir, 2004). Bu çalışma kapsamında, seçenek ağırlıklandırma yöntemlerinden deneysel ağırlıklandırma kullanılacaktır.

Deneysel ağırlıklandırma farklı yöntemlerle yapılabilmektedir. Bu yöntemlerden en çok tercih edilenleri doğrusal yaklaşımlar içerisinde yer alan seçenek-toplam korelasyonu ve karşılıklı (reciprocal) ortalamalar yöntemleridir (Haladyna, 1990). Seçenek-toplam korelasyonu yönteminde, her bir madde için her bir cevap seçeneği ile toplam puan arasında nokta çift serili korelasyon hesaplanır. Hesaplanan değere göre, seçeneklere ağırlık atanır. Bu şekilde cevaplayıcı doğru cevabı işaretlemese bile, işaretlediği cevabın yüksek puan alan bireyler arasında popüler olması ölçüsünde kısmi puan alır (Diedenhofen ve Musch, 2015). Karşılıklı ortalamalar yönteminde ise, her bir madde için her bir seçeneği işaretleyen bireylerin standart toplam puan ortalamaları alınır ve bu ortalamalarla doğru orantılı olacak şekilde seçeneklere ağırlıklar atanır. Ardından, iç tutarlık katsayısı maksimum değere ulaşıncaya kadar bu süreç iteratif olarak devam ettirilir (Haladyna, 1990). İç tutarlık katsayısının maksimum olduğu durumdaki seçenek ağırlıkları kullanılır.

Deneysel ağırlıklandırmanın, testin iç tutarlığını olumlu yönde etkilediği birçok çalışmada görülmüştür (Hendrickson, 1971; Echternacht, 1973; Reilly ve Jackson, 1973 ve Haladyna, 1990). Bu yöntemlerin testin psikometrik özellikleri üzerindeki etkileri birçok çalışmada incelenmişken, faktör yapısının ne şekilde etkilendiğinin incelendiği yalnızca bir çalışma bulunmaktadır (Hendrickson ve Green, 1972). İlgili çalışmada yalnızca, karşılıklı ortalamalar ve 'doğru-yanlış' puanlama yöntemleri ele alınmıştır. Bu çalışmada ise, bu yöntemlere ek olarak seçenek-toplam korelasyonu yöntemi de ele alınacaktır. Bu yönüyle çalışmanın literatüre katkı sağlayacağı düşünülmektedir.

Sonuç olarak bu çalışmanın amacı, 1-0, seçenek-toplam korelasyonu ve karşılıklı (reciprocal) ortalamalar yöntemleriyle yapılan puanlamaların testin faktör yapısını ne şekilde etkilediğini incelemektir. Araştırma sonucunda, farklı puanlama yöntemlerinin faktör yapısına etkisi konusunda elde edilecek bulguların literatüre katkı sağlayacağı düşünüldüğünden, bu araştırma temel araştırma olarak nitelendirilebilir.

Araştırma verileri, 2014-2015 güz döneminde 8. Sınıf öğrencilerine uygulanan TEOG Türkçe testinden elde edilmiştir. Türkçe testini alan 2129 öğrenciye ilişkin doğru cevapların 1, diğer cevapların 0 ile puanlandığı '1-0' matrisi hesaplanmıştır. Öğrencilerin TEOG Türkçe testinden aldıkları puanlar, toplam doğru cevap sayısı ile elde edilmiştir.

Verilerin analizi aşamasında; '1-0', seçenek-toplam korelasyonu ve karşılıklı ortalamalar yöntemleri kullanılarak seçenek ağırlıklandırması yapılacak ve her bir puanlama durumu için faktör yapısı incelenecektir. Bu amaç doğrultusunda, açımlayıcı ve doğrulayıcı faktör analizi yapılacaktır. Faktör yapılarının karşılaştırılması amacıyla, açımlayıcı faktör analizinde faktör yüklerindeki ve açıklanan varyans yüzdelerindeki değişim incelenecektir. Doğrulayıcı faktör analizinde ise, uyum istatistikleri karşılaştırılacaktır.

Bildiri No: 95

Thurstone ve Bradley-Terry Ölçekleme Yöntemlerinin Karşılaştırılması

Osman TAT*, Duygu ANIL**

Özet

Günümüzde sekiz tane önemli, küresel yükseköğretim kurumları sıralama listesi ve özellikle gelişmekte olan ülkelerde uzmanlık alanlarını, bölgesel farlılıkları dikkate alan 60'tan fazla yerel sıralama listesi vardır. Yükseköğretim kurumları listelerinin hazırlanmasında listeleri hazırlayan kuruluşların tercih ettikleri kriterler ve yöntemler bir birinden oldukça farklı olabilmektedir. Thurstone'nun karşılaştırmalı yargı kanununun farklı varsayımlar altında kullanılan beş hali olduğunu bilinmektedir. Thurstone'un ikili karşılaştırmaya dayalı en bilinen ölçekleme yöntemleri V. ve III. hal denklemleridir. İkili karşılaştırma yargılarına dayalı ölçekleme tekniklerinden bir diğeri ise psikolojik deneyler, spor müsabakaları ve genetik gibi geniş bir kullanım alanı olan Bradley-Terry modelidir. Bu çalışmada iki farklı ölçekleme yöntemi ile üniversite sıralama kuruluşlarının kullandıkları kriterler önem düzeyine göre sıralanmış ve elde edilen kriter sıralamasının tutarlılığı incelenmiştir. Bu çalışmada Thurstone'un III. hal denklemi ile Bradley-Terry modelinin sonuçlarının karşılaştırılması amaçlanmıştır. Çalışma, nicel araştırma tekniklerinin kullanıldığı betimsel araştırma modelinde desenlenmiştir. Araştırmanın örneklemini Türkiye'deki kamu ve özel/vakıf yükseköğrenim kurumlarında akademisyen olarak çalışmakta olan 247 kişi oluşturmaktadır. III. Hal Denklemi ve Bradley-Terry modeli ile yapılan ölçeklemede akademisyenlerin SCI, SSCI ve AHCII İndekslerinde Taranan Makale Sayısı'nı en önemli kriter olarak gördüğü buna karşın Üniversitenin Nobel Sahibi Üye veya Mezun Sayısı'nı ise en önemsiz kriter olarak gördükleri belirlenmiştir. Her iki ölçekleme yönteminde de üniversite sıralama kriterlerinin sıralarının değişmediği tespit edilmiştir.

Anahtar Kelimeler

Ölçekleme, Thurstone, Bradley-Terry, İkili Karşılaştırma

*Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:osmntt@gmail.com
** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:duygu.anil73@gmail.com

Paper No: 95

A Comparative Analysis of Thurstone and Bradley-Terry Scaling Techniques

Abstract

Higher education institution rankings are monitored seriously by universities, sponsors of private and state institutions, potential students, parents and also by public. Nowadays there are eight important global ranking lists and more than 60 local lists considering regional differences. The methods and criteria preferred by the institutions publishing these league tables are specific to the institutions. There are five cases of the Thurstone's law of comparative judgement that can be benefited from under different conditions. The best known cases of Thurstone's law of comparative judgement are case III. and case V. Bradley and Terry's model which has a broad use from sports to genetics is another paired comparative scaling method. In this study university ranking criteria were scaled using two different methods and consistence of emerged order of criteria was investigated. In this study it is aimed to compare the results of Thurstone's III. Case and Bradley-Terry model. The research is a descriptive research in which quantitative research techniques used. The sample of the research consists of 247 academician working at state and private higher education institutions in Turkey. According to the both scaling methods the most important ranking criterion is Number of Papers Indexed by SCI, SSCI and AHCII. Moreover, academicians think the least important criterion is Number of Graduate of Member Who Has Nobel Prize. In both scaling methods the order of university ranking criteria does not change.

Key Words

Scaling, Thurstone, Bradley-Terry, Paired Comparison

Bildiri No: 96

CART, CHAID ve Lojistik Regresyon Analizi Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması

Bayram ÇETİN*, Murat KAYRİ**, Görkem CEYHAN***

Özet

Yapılan bu çalışmayla, sınıflama ve regresyon modelleri arasında en yaygın olarak kullanılan yöntemler arasında yer alan CART (Sınıflandırma ve Regresyon Ağaçları), CHAID (Ki-Kare Otomatik Etkileşim Belirleme) ve LRA (Lojistik Regresyon Analizi) yöntemlerinin karşılaştırılması amaçlanmıştır. Araştırma kapsamında kullanılan veri seti, 2012-2013 eğitim-öğretim yılı bahar döneminde Muş Alparslan Üniversitesi Eğitim Fakültesi'nde öğrenim gören 954 öğrenciden "Yansıtıcı Düşünme Düzeyini Belirleme Ölçeği" (YDDBÖ), "ve "Kişisel Bilgi Formu" ile elde edilmiştir. Ölçek kullanılarak elde edilen bağımlı değişkene ait puanlar, analizler esnasında kategorik veriler kullanılacağı için iki aşamalı kümeleme analizi ile düşük ve yüksek olacak biçimde iki kategoriye ayrılmıştır. Karşılaştırılması düşünülen veri seti üzerinde üç yönteme ilişkin analizler yapılmış ve elde edilen sonuçlar çerçevesinde bu üç yöntemin sınıflandırma performansları karşılaştırılmıştır. Analizler sonucunda elde edilen bulgulara göre, veri setinin tamamında her üç yönteminde orta düzeyin üzerinde bir sınıflandırma doğruluk oranı verdiği tespit edilmiştir. Araştırmada elde edilen diğer bir bulguda, sınıflandırması performansı bakımından en iyi doğruluk oranı sonucunu CART analizi, en kötü doğruluk oranı sonucunu ise LRA gösterdiği belirlenmiştir. Ayrıca elde edilen analizler sonucunda sınıflandırma işleminin tamamındaki doğruluk oranlarına bakıldığında CART ve CHAID analizlerinin birbirine çok yakın değerlere sahip olduğu, ancak az bir farklılıkla da olsa CART analizinin daha iyi doğruluk oranı sonucu verdiği tespit edilmiştir. Her üç yönteme ait modeller incelendiğinde bağımsız değişkenlerin modele dâhil edilmesi sonucunda sahip oldukları önem düzeyi bakımından CART analizinin 6 değişkeni, CHAID analizinin 4 değişkeni ve LRA'nın ise 3 değişkeni modele dâhil ettiği en önemli değişkeni belirleme noktasında ise her üç yönteminde aynı değişkeni en önemli değişken olarak modele dâhil ettiği tespit edilmiştir.

Anahtar Kelimeler
Sınıflandırma, CART, CHAID, Lojistik Regresyon, Veri Madenciliği

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:bcetin@gmail.com
** Batman Üniversitesi Mühendislik ve Mimarlık Fakültesi İletişim:muratkayri@gmail.com
*** Gazi Üniversitesi Gazi EğitimFakültesi Eğitim Bilimleri Bölümü. İletişim:gorkemceyhan23@gmail.com
Tel:05324383023

Paper No: 96

The comparison of CART, CHAID and Logistic Regression Analyses in terms of Classification Performances

Abstract

The purpose of this study is to compare CART (Classification and Regression Trees), CHAID (Chi-square Automatic Interaction Detector) and LRA (Logistic Regression Analysis), which are the most common methods used in classification and regression models. The data used in the research were obtained from undergraduates studied at Faculty of Education of Muş Alparslan University in 2012-2013 educational year via "The scale of Determining Reflective Thinking" and "Personal Information Form". The scores belonging to dependent variable were classified into two categories as low and high by using two-step cluster analysis because categorical data were used in the analysis. The data were investigated in terms of these three methods and the classification performances were compared in terms of the results. According to findings obtained from the analyses, it was ascertained that each methods had classification accuracy rate more than moderate. In another finding, it was determined that the best and worst accuracy rates in terms of classification performance belonged to CART analysis and LRA respectively. When accuracy rate in whole classification were investigated, it was observed that CART analysis had slightly more accuracy rate than CHAID although CART and CHAID analyses had almost same values. When models belonging to each method were examined, CART, CHAID and LRA included 6,4 and 3 variables in terms of significance level respectively. Besides, these three methods included the same variable to the model as the most significant variable.

Key Words:
CART, CHAID, Logistic Regression, Data Mining

Bildiri No: 97

Amerika'da Görme Engelli Öğrenciler için Uygulanan Bireyselleştirilmiş Bilgisayar Testlerinin Yapısal Açısından İncelenmesine İlişkin Bir Derleme Çalışması

Derya ÇOBANOĞLU AKTAN[*], Mehmet Taha ESER[**],Gökhan AKSU[***]

Özet

Bu çalışmada, görme engelli oldukları belirlenen öğrenciler için uygulanan bireyselleştirilmiş bilgisayar testlerinin yapısal anlamda incelenmesi amaçlanmıştır. Bu amaç kapsamında Amerika'nın Teksas, Oregon ve Virjinya eyaletlerinde kullanılmakta olan görme engelli öğrenciler için oluşturulmuş bireyselleştirilmiş bilgisayar testleri (GÖBBT) madde havuzu, grafik içeren soruların varlığı ve sınav için ayrılan süre bakımından incelenmiştir. Çalışma sonucunda, ilgili üç eyalette de uygulanan GÖBBT'lerinin yapısal olarak engelli olmayan öğrencilere uygulanan BBT'lerden çok farklı olmadığı gözlemlenmiştir. Ülkemizde görme engelli öğrenciler için yerel ya da ulusal anlamda herhangi bir BBT uygulanmamaktadır. Dolayısıyla bu çalışmada incelenen örnekler aynı zamanda ülkemiz için birer iyi uygulama örneği şeklinde düşünülebilir. Elde edilen bulgulardan yola çıkılarak, ülkemizde GÖBBT oluşturmak için ihtiyaç duyulan araştırmalara yön verilebilir ve benzer uygulamaların ülkemizde de yavaş yavaş gerçekleştirilmesine katkı sağlanabilir.

Anahtar Kelimeler

Engelli öğrenci, Geniş ölçekli sınav, Eşitlik, Uyumsama çalışmaları

[*] Hacettepe Üniversitesi, Eğitim Fakültesi, ANKARA, dcaktan@hacettepe.edu.tr
[**] Türkiye Belediyeler Birliği, Eğitim Şube Müdürlüğü, tahaeser@gmail.com
[***] Adnan Menderes Üniversitesi, Aydın Meslek Yüksekokulu, gokhanaksu1983@hotmail.com

Paper No: 97

A Review On Structural Examination of Computer Adaptive Tests for Visually Impaired Students in US

Abstract

This study aims to examine the Computer Adaptive Tests for visually impaired students in US. In this context , item pool of the Computer Adaptive Tests created for visually impaired students used in Texas, Oregon and Virginia states are examined in terms of existence of questions containing graphics and time given for the examinations. As a result of the studies it is observed that the computer adaptive tests for visually impaired students implemented in three states are not much different than the tests implemented for structurally non-disabled students. There are not any computer adaptive tests implemented for visually impaired students in Turkey. Therefore, examples viewed in this study can be considered as good implementation examples for Turkey. Based on the findings obtained, the researchs needed to form a computer adaptive tests for visually impaired students in our country can be directed and gradual realisation of similar implementations in Turkey can be supported.

Key Words

Visually impaired student, Computer adaptive test, Validity

Bildiri No: 98

Farkı Sayıda Değişen Yargıcı Kararlarının Farklı Madde Türlerine Göre Tutarlılığının İncelenmesi

Fazilet TAŞDEMİR*, D. Bahar ŞAHİN**

Özet

Güvenirlik, ölçümlerin tekrarlanabilirliği ya da tekrarlı ölçümlerin tutarlılığı olarak tanımlanır. Eğitim alanında en yaygın kullanılan güvenirlik çalışmaları, gözlemci-içi ve gözlemciler arası uyumdur. İkili ve ikiden fazla yargıcı kararlarının tutarlığı farklı şekilde puanlanan madde türlerinde araştırılması gereken bir sorundur. Açık uçlu sorularda objektif puanlama sağlanması için kullanılan rubrik ile elde edilen puanlarla, çoktan seçmeli maddeler için yapılan yargıcılar arası puanlama güvenirliğinin incelenmesi bu araştırmanın amacını oluşturmaktadır. Bu amaç doğrultusunda sınıf öğretmenliği yapan yargıcıların farklı madde türlerindekisorulara verdikleri yanıtların değerlendirmesinde yargıcılar arası verilen kararların güvenirliği incelenmiştir. Araştırma grubu seçkisiz olmayan örnekleme yöntemlerinden Amaçlı Örnekleme başlığı altında yer alan uygun örnekleme yöntemi ile belirlenmiştir. Araştırma grubunu 2015-2016 eğitim-öğretim yılında sınıf öğretmenliği yapan yargıcı olarak belirlenen 46 öğretmen oluşturmaktadır. Yargıcılardan açık uçlu sorularda ikili uyumuna bakıldığında Cohen Kappa güvenirlik katsayısının (0.52-0.74) arasında değiştiği bulunmuştur (p=0.00). Bu değer Kappa istatistik sonuçlarından yargıcılar arasında "ortanın altında-iyi düzeyde" uyumaların olduğunu göstermiştir. Açık uçlu ifadelerde yargıcıların verdikleri kararların farklılığı, bu madde türlerinin tek yargıcı tarafından değerlendirilmesinde ne denli doğru kararlar verilebileceği sorusunu da beraberinde getirmiştir. Çoktan seçmeli sorularda ikili uyum katsayısının (0.78) daha yüksek olduğu bulunmuştur. Teorik olarak Kappa sürekli veriden elde edilen sıralı ölçekteki kategorilere uygulanabilir. Ancak bu durumda Kappa değerinin düşük bir anlamlılığa sahip olmasına neden olmaktadır. Bununla birlikte bu işlem orijinal ölçme aracında bilgi ve istatistiksel güç kaybına neden olabilir. Çoktan seçmeli soruların puanlanmasında yargıcı sayısı 46 kişi olduğunda çoklu değerlendirici arası sınıf-içi korelasyon katsayısı (r=0.85) ile mükemmel düzeyde uyum gösterdiği sonucuna ulaşılmıştır. Bu sonuç doğrultusunda sınıf-içi korelasyon katsayısı kullanılarak orijinal sürekli ölçekte verinin uyumluluğunu analiz etmenin çok daha iyi sonuçlar ürettiği söylenebilir.

Anahtar Kelimeler

Güvenirlik katsayısı, Çoklu değerlendirici, Sınıf-içi korelasyon katsayısı

*Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: fazilettasdemir@yahoo.com; fazilet.tasdemir@erdogan.edu.tr Tel: +4645328454

**Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: bahar.sahin@erdogan.edu.tr Tel: +4645328454

Paper No: 98

Examination of Coherency of Different Number of Changing Judgers' Decisions according to Different Article Types

Abstract

Reliability is defined as the repeatability of measurements or the consistency of repeated measurements.The most common type of reliability in the field ofeducation research is the intra/interrater reliability. Coherency of dual and more than two decisions of judgers is an issue required to be researched in item types graded differently. The examination of the scores obtained by rubric used for ensuring an objective scoring in open-endedquestions and the reliability of scoring between judgers done for multiple choice items constitutes the purpose of this study. In line with this purpose, the reliability of the decisions given among judgers in the evaluation of the answers given by judgers who are classroom teachers to the questions in different item types was examined. The research group was determined by convenience sampling method included under the title of Purposeful Sampling out of sampling methods that are not random. 46 teachers who are classroom teachers in 2015-2016 school year designated as judgers constitute the research group. When dual concordance of judgerswas examined in open-ended questions, Cohen's Kappa reliability co-efficient was found to vary between 0.52-0.74 (p=0.00). This value showed that "below average – above average" concordances out of Kappa statistical results among judgers. The difference of the decisions given by judgers in open-ended statements brought along the question of how correct decisions may be given by one judger in the evaluation of such item types, as well. Dual concordance co-efficient was found to be higher (0.78) in multiple choice questions. Theoretically, Kappa may be applied to the categories in ordinary scale obtained from continuous data. However, in this case, it causes Kappa value to have a low significance. On the other hand, this process may cause loss of information and statistical power in original measurement tool. When the number of judgers was 46 in scoring multiple choice questions, it was concluded that it accorded at a perfect level with an intra-class correlation coefficient (r=0.85) between multiple evaluators. In line with this result, it may be said that analyzing the coherence of data in original continuous scale by using intra-class correlation coefficient gives much better results.

Key Words

Reliability coefficient, Multiple evaluators, Intra-class correlation coefficient

Bildiri No: 99

Eğitim Araştırmalarında Güç Analizi Ve Bir Uygulama

Hikmet ŞEVGİN*, Bayram ÇETİN**

Özet

Bir araştırma için evreni temsil edecek örneklem sayısı daha planlama aşamasında karar verilmesi gereken bir durumdur. Örneklem sayısının ilk baştan belirlenecek olmasının birçok sebebi bulunmaktadır. Çünkü yapılacak araştırma sonucunda ilgili topluluk ya da grup hakkında bir fikir edinme, karar verme, ya da bir çıkarımda bulunma gibi risk taşıyan çok önemli sonuçlara ulaşılmaktadır. Bu sonuçların geçerli ve güvenilir olması bilimsel açıdan büyük bir önem taşımaktadır. Çalışmamızda istatistiksel güç analizi konusunda farkındalık yaratmak, alan yazına katkı sağlamak ve eğitim araştırmalarında istatistiksel güç analizinin kullanıldığı uygulamalara yer vermek amaçlanmıştır. Araştırma kapsamında eğitim bilimleri alanında yayın yapmakta olan dergilerden üç tanesi seçkisiz olmayan örnekleme yöntemlerinden ölçüt örnekleme yöntemi ile seçilmiş olup bu dergilerde yer alan toplamda 50 makale incelemeye alınmıştır. Araştırmaya sadece çözümleyici istatistik kullanılan 25 makale dâhil edilmiştir. Betimsel istatistikler ve güç analizi için bazı istatistiksel programlardan yararlanılmıştır. Araştırma kapsamında ilk aşamada eğitim bilimleri alanında yayınlanan makaleler incelenerek eğitim üzerine yapılan araştırmalarda istatistiksel güç analizi konusu irdelenmiştir. Sonrasında gücün hesaplanmasında ihtiyaç duyulan anlamlılık (alfa) seviyesi, etki büyüklüğü ve örneklem büyüklüğü ile ilgili incelemeler yapılmış olup analizlere ilişkin güç hesaplamaları yapılmıştır. Bu İnceleme kapsamında sadece bir makalede etki büyüklüğünün hesaplanıp raporlandığı ve analizlere ait istatistiksel gücün ise hesaplanmadığı tespit edilmiştir. Benzer şekilde örneklem büyüklüğü belirlemede de güç analizi yöntemine rastlanmamıştır.

Anahtar Kelimeler

Etki büyüklüğü, Örneklem büyüklüğü, Anlamlılık (alfa) seviyesi, Güç analizi

* Gazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: hikmetsevgin@gmail.com Tel: (+90) 0312 2023697

** Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:bcetin27@gmail.com

Paper No: 99

Power Analysis In Educational Research And An Application

Abstract

The number of samples representing the population of a research should be decided in the planning stage. There are many reasons to determine the number of samples in advance. Because many important results with risks such as getting an idea, making decision or making a conclusion about a community or group can be achieved. Having reliable and valid results is very important in scientific terms. In the present study, it was aimed to create awareness about statistical power analysis, contribute to the literature and give place to educational studies using statistical power analysis. In the study, three educational science magazines were selected by criteria sampling method, one of random sampling methods, and a total of 50 articles were studied. 25 articles using analytical statistics were included in the study. For descriptive statistics and power analysis, some statistical soft wares were used. In the first stage of the present study, articles published in the area of educational science were investigated and statistical power subject covered in these articles were studied. Then, significant (alpha) level used in the calculation of the power, effect size and sample size were examined, and power calculations about the analyses were made. As a result, it was determined that in only one essay, effect size was calculated and reported, while statistical power was not calculated. Similarly, power analysis method was not encountered in determining sample size.

Key Words

Effect Size, Sample Size, Significance (Alpha) Level, Power Analysis

Bildiri No: 100

Örtük Markov Modellerinde Örtük Statü Sayısının Belirlenmesi

Duygu GÜNGÖR*

Özet

Örtük değişkenlerin zamana bağlı değişimi çeşitli modellerle incelenebilmektedir. Örtük sınıf analizinin boylamsal versiyonu olarak da adlandırılabilecek örtük Markov model kategorik değişkenlerin analizi için güçlü bir yaklaşımdır. Örtük sınıf olasılıkları, madde tepki olasılıkları ve örtük geçiş olasılıkları olarak üç temel parametrenin tahminlendiği modelde, örtük statü sayısının doğru şekilde belirlenmesi modelin yorumlanması için önem taşımaktadır. Bu çalışmanın amacı farklı örneklem büyüklükleri (200, 300, 600), madde tepki olasıkları (.90, .70) ve ölçüm alınan zaman sayısına (2, 3, 4) bağlı olarak çeşitli bilgi kriterlerinin doğru modeli seçme gücünü simülasyon çalışması ile belirlemektedir. Çalışma kapsamında öncelikle 18 farklı koşul için veri setleri Latent GOLD 5.0 (Vermunt ve Magidson, 2015) istatistik paket programının Monte Carlo simülasyon bölümü kullanılarak üretilmiştir. Elde edilen veriler bir, iki, üç, dört ve beş örtük statü olacak şekilde analiz edilmiştir. Analizler sonucunda her koşul için elde edilen L^2 değerleri kullanılarak bilgi kriterlerine ait değerler hesaplanmıştır. Bu bilgi kriterlerinin işaret ettiği modeller seçilerek raporlanmıştır. Örtük statü sayısı gerçekte iki olarak belirlendiğinden bu değerden sapma oranı yüzde olarak hesaplanmıştır. Elde edilen sonuçlara göre ölçüm alınan zaman sayısı, örneklem büyüklüğü ve madde tepki olasılıklarının büyüklüğünden bağımsız olarak BIC, ve CAIC bilgi kriterlerinin doğru modeli işaret etme olasılığı 1 olarak bulunmuştur. Benzer şekilde AIC3 kullanıldığında sadece madde tepki olasılıklarının gücü düştüğünde bilgi kriterinin gücü de .99 da düşmüştür. AIC bilgi kriterinin ise bu gücü .78 ile .93 arasında değişmektedir.

Anahtar Kelimeler

Örtük Markov model, model seçimi

*Dokuz Eylül Üniversitesi Edebiyat Fakültesi Psikoloji Bölümü. İletişim:duygu.gungor@deu.edu.tr

Paper No: 100

Identifying Number of Latent States in Latent Markov Models

Abstract

Change of latent variables can be investigated with different latent variable models. Latent Markov model which can be identified as the longitudinal version of latent class models is a comprehensive approach in analyzing categorical data. Three parameters are estimated in latent Markov models as latent state probabilities, item response probabilities and latent transition probabilities. true identification number of latent states is important to interpret the model. The aim of this study is to investigate the power of different information criteria in detecting number of different latent states for several conditions. For this purpose we varied the sample size (200, 300, 600), item response probabilities (.90, .70) and number of time occasions (2, 3, 4). Total 18 conditions was tested after simulating data with the syntax module of Latent GOLD 5.0 (Vermunt & Magidson, 2015). We tested five models: one to five states. We used L^2 statistics and calculated each information criteria based on this statistic to report which model they pointed. Results indicate that regardless the number of time occasions, sample size and magnitude of item response probabilities power of BIC and CAIC was 1. Similarly, AIC3 had high power except in the case of low item response probabilities. The power of AIC ranged between .78 and .93, however.

Key Words

Latent Markov model, model selection

Bildiri No: 101

Ortalama Kovaryans Yapılarının Değişmezliği ve Örtük Sınıf Analizi Yöntemleri ile Ölçme Değişmezliğinin İncelenmesi

Alperen YANDI*, İ. Alper KÖSE**, Ömür UYSAL***

Özet

Gruplar arası farkların belirlenmesinin amaçlandığı çalışmalarda kullanılan ölçme araçları için ölçme değişmezliği test edilmesi gereken en önemli özelliklerdendir. Bu tür çalışmalarda "gerçek" farklılıklara dayalı sonuçların rapor edilebilmesi için ölçme değişmezliğinin sağlanması gerekmektedir. Ölçme değişmezliği, bir çalışma sonucunda ortaya çıkan sonuçların bireylerin bir gruba olan üyeliklerinden bağımsız olması anlamına gelmektedir. Ayrıca bu durum ölçme aracına ilişkin en önemli geçerlilik kanıtlarından biridir. Ölçme değişmezliğinin test edilme süreciyle ilgili olarak farklı görüşlerin varlığı ve yöntemlerle ilgili karışıklıklar yapılan araştırmalarda ölçme değişmezliğinin göz ardı edilmesine yol açmaktadır. Alanyazındaki çalışmalarda yapısal eşitlik modellemesi (YEM) ve madde tepki kuramı (MTK) altındaki yöntemler ölçme değişmezliğinin test edilmesinde sıklıkla kullanılmaktadır. Belli özellikler bakımından diğerlerinden farklılaşan bir diğer yöntemde örtük sınıf analizidir. Ölçme değişmezliği yöntemlerinin arasındaki benzerlik ve farklılıkların ortaya konmasıyla, diğer araştırmacılar tarafından ölçme değişmezliğinin bu yöntemlerle uygun bir şekilde test edilebilmesi sağlanabilir.

Bu çalışmanın amacı PISA 2012'de uygulanan ve ST94 kodlu Problem Çözmeye Açıklık Ölçeği'nin ortalama kovaryans yapılarının değişmezliği (OKYD) ile çoklu grup örtük sınıf analizi (ÇGÖSA) yöntemlerini kullanarak ölçme değişmezliği sonuçlarının incelenmesi ve yöntemlerin karşılaştırılmasıdır. Çalışmada Türkiye ve Şangay – Çin örneklemine ait veriler kullanılmıştır. Problem Çözmeye Açıklık Ölçeği 5'li likert tipi puanlanan ve 5 maddeden oluşan tek boyutlu bir ölçektir. Türkiye ve Finlandiya örneklemlerinde sırasıyla 3159 ve 5676 öğrenci yer almaktadır. OKYD için Lisrel 8.72 programı, ÇGÖSA için ise Latent Gold 5.1 programı tercih edilmiştir. Araştırma sürecinde analizler devam etmekte ve bulgular raporlaştırma aşamasındadır. Araştırma bulguları alanyazında bulunan çalışmalarla karşılaştırılarak tartışılacaktır. Ayrıca bulgular doğrultusunda sonraki çalışmalar için önerilerde bulunulacaktır.

Anahtar Kelimeler

Ortalama kovaryans yapılarının değişmezliği, çoklu grup örtük sınıf analizi, ölçme değişmezliği, PISA 2012.

*Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: alperenyandi@gmail.com Tel: +905056758867

**Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: i.alper.kose@gmail.com

*** Ankara Üniversitesi Eğitim Bilimleri Fakültesi Bilgisayar Öğretmenliği Bölümü.

Paper No: 101

Examination of Measurement Invariance with Mean and Covariance Structures and Multi group Latent Class Analyses Methods

Abstract

Measurement invariance is one of the most important characteristics of the measurement instruments used in the studies aiming to specify the differences between groups. In these studies, measurement invariance must be established to report the "real" differences. Measurement invariance means that the results obtained from a study are independent from the group memberships of the individuals. Also measurement invariance is one of the most important evidences related to the validity of measurement instruments. The existence of different views about the testing process of measurement invariance and the methodological intricacies lead to the negligence of measurement invariance in the studies. The methods under the title of structural equation models and item response theory are frequently used for testing measurement invariance in the studies in literature. Another method differing from other methods with certain characteristics is the latent class analysis. Revealing the similarities and differences between the methods of testing measurement invariance helps the researchers test measurement invariance appropriately.

The purpose of this study was twofold: (a) to examine the results about measurement invariance of the Openness to Problem Solving scale coded as ST94 and applied in PISA 2012 by using mean and covariance structures invariance and multi group latent class analysis methods and (b) to compare the aforementioned methods. In the study, the data belonging to Turkey and Shanghai – China samples were analyzed. Openness to Problem Solving Scale was a-five point likert type scale and consists of five items. It was a unidimensional scale. There were 3159 and 5676 students in the samples of Turkey and Finland respectively. Lisrel 8.72 program for MACS and Latent Gold 5.1 for MG – LCA were preferred. The data are currently in the process of analysis. The findings will be reported and argued in line with the others studies in the related literature. Also based on the findings, suggestions will be made for further studies.

Key Words

Mean and covariance structures invariance, multi group latent class analysis, measurement invariance, PISA 2012

Bildiri No: 102

Bifaktör Yapının Tek Boyutlu, Çok Boyutlu ve İkinci Dereceden Yapılarla Karşılaştırılması

Murat Doğan ŞAHİN*, Murat AKYILDIZ**

Özet

Bu çalışmada Anadolu Üniversitesi Yabancı Diller Yüksek Okulunda İngilizce hazırlık eğitimi gören öğrencilere öğretim yılı bitiminde uygulanan dinleme, okuma, kelime bilgisi ve gramer olmak üzere dört alt beceriyi ölçen dört farklı İngilizce yeterlik sınavından elde edilen veriler kullanılmıştır. Yapılan açımlayıcı faktör analizi sonucu tüm veri setlerinde birinci faktörden ikinci faktöre yüksek özdeğer düşüşü, maddelerin çoğunlukla tek faktöre yüksek yük vermeleri gibi baskın tek boyuta işaret eden sonuçlar alınmıştır. Bunun yanında sınavların dört alt yetenek için kestirim yapması durumu göz önüne alınarak, bu sınav verilerinin bifaktör yapıya uygun olabileceği düşünülmüştür. Daha sonra tüm veri setlerine sırasıyla tek boyutlu, çok boyutlu (boyutlar arası korelasyon olan/hiyerarşik olmayan), ikinci dereceden ve bifaktör doğrulayıcı faktör analizi (DFA) uygulanmıştır. Tüm veri setlerinde, WLSMV kestirim yöntemine dayanarak uygulanan tüm DFA analizlerinin model veri uyumlarının oldukça yüksek olduğu sonucuna ulaşılmıştır. Her bir sınav kendi içinde değerlendirildiğinde ise diğer DFA modellerine göre en iyi uyum değerlerine bifaktör modelin sahip olduğu görülmüştür. Modeller arasındaki farkın incelenmesi içinse MLR kestirim yöntemiyle tüm veriler tekrar analiz edilmiştir. Yapılan analizler sonucu her bir sınav için diğer üç DFA modelinden elde edilen BIC uyum istatistikleri bifaktör modelin BIC istatistikleri ile karşılaştırılmıştır. Elde edilen sonuçlara göre dört sınav verisi için de bifaktör modelin en uygun model olduğu sonucuna ulaşılmıştır. Bifaktör modeller, çok boyutlu yapılara dayanan faktör analitik modellerde gruplar arasındaki yüksek korelasyonların baskın bir genel faktörden kaynaklandığını iddia eder. Dolayısıyla bifaktör modellerde alt boyutlar düzeyinde daha dakik (pure) yetenek kestirimleri elde edilmiş olur. Tüm bu sonuçlar göz önüne alındığında, her ne kadar dinleme, okuma, gramer ve kelime bilgisi becerilerini ölçen ayrı soru setlerinden meydana gelse de, bireyin İngilizce yeteneğini ölçen bu sınavlara bifaktör yapının uygun olduğu sonucuna varılabilir. Ayrıca araştırma sonuçları doğrultusunda, farklı alt boyutlardan oluşmasına rağmen faktör yapısı tek boyutlu modellere uyum gösteren psikolojik yapılardaki boyutluluk sınamalarına bifaktör modelin de dahil edilmesinin gerekliliği tartışılmıştır.

Anahtar Kelimeler

Bifaktör model, doğrulayıcı faktör analizi, çok boyutluluk

* Anadolu Üniversitesi Açıköğretim Fakültesi Test Araştırma Birimi. İletişim: mdsahin@anadolu.edu.tr
** Anadolu Üniversitesi Açıköğretim Fakültesi. İletişim: muratakyildiz@anadolu.edu.tr

Paper No: 102

The Comparison of Bifactor Construct with Unidimensioanl, Multidimensional and Second-Order Constructs

The study uses data obtained from four proficiency tests administered at Anadolu University School of Foreign Languages. Each of these tests measures four different language skills; listening, reading, vocabulary and grammar. The results of the exploratory factor analysis showed that the eigenvalue considerable decreased from the first factor to the second, and the items often loaded into a single factor, which clearly indicates a dominant single dimension. In addition, this data was thought to be suitable for bifactor model since the tests can predict four skills mentioned above. Later, all data sets were applied single dimensional, multidimensional, second-order and bifactor confirmatory factor analyses respectively. In all data sets, it was found that model data fit of DFA analyses which were done based on WLSMV prediction method was quite high; the best fit being for bi-factor model. In order to examine the differences among the models, all data sets were reanalyzed by using MLR prediction method, and BIC fit statistics obtained were compared, and as a result, bifactor model was concluded to be the best method for the data obtained from these four tests.

Bifactor models claim that high correlations among groups in factor analytic models based on multidimensional structures are due to a dominant general factor. Thus, it is possible to make pure ability predictions for subskills in bifactor models. When all these results are concerned, it can be concluded that bifactor model is convenient for the tests measuring English proficiency. Although the tests consist of separate items measuring listening, reading, grammar and vocabulary skills, all these items as a whole measure English proficiency level of individuals. Finally, the inclusion of bifactor model to test dimensionality used for psychological constructs, whose factor structure fits to unidimensional models although they consist of different sub-dimensions, will be discussed in the study.

Key Words

Bifactor model, Confirmatory factor analysis, Multidimensionality

Bildiri No: 103

DinasoruS Değerlendirme Sistemi Yardımıyla Bilgisayar Ortamında Dinamik Olarak Özgün Soru Maddeleri Üretilmesi

Erol KARAKIRIK*, Mustafa DOĞAN**

Özet

Aynı soru kalıbı veya modelinden sınırları iyi belirlenmiş belli değişkenler yardımıyla anlık olarak farklı içerikli sorular üretilebilmesini sağlayan birçok farklı dinamik soru üretim sistemi bulunmaktadır. Ancak mevcut sistemlerdeki değişken tanımı genellikle soru modellerindeki belli bir örüntü veya fonksiyonla tanımlanabilen sayısal değerleri veya bir şablon içinde tanımlanmış kısıtlı metin öğelerini kapsamaktadır. Bu dar kapsamlı değişken tanımı birbirinden çok az farklılıklar içeren soru maddelerinin soru bankalarında farklı soru modelleri ile ifade edilmesini gerektirmekte ve aynı kazanıma odaklı yapısal olarak birbirine benzeyen birçok sorunun ayırt edilebilmesini güçleştirmektedir. Ayrıca soru metinlerinde dinamik kullanılabilecek alan bilgisi sistemlerde daha önceden tanımlanmış değişkenlerle sınırlı kalmaktadır. Bunun yanda değişken kavramının sadece soru metninin bazı kısımlarının dinamik olarak değiştirilmesi için kullanılması bilgisayar destekli ölçme ve değerlendirme sistemlerinin sağlayabileceği imkânların yeterince anlaşılamamasına yol açmaktadır. Bu maksatla DinaSoruS adı verilen ve mevcut değişken tanımının ötesinde kendine has bir soru hazırlama diline sahip bir dinamik soru hazırlama ve değerlendirme sistemi geliştirilmiştir. DinaSoruS sisteminin diğer sistemlerden en büyük farkı soru dili yardımıyla soru modellerindeki tün unsurların (kategori, fonksiyon, grafik, resim, şema vesaire) değişken olarak tanımlanabilmesi ve değişkenler yardımıyla yeni alan bilgisinin anlık olarak tanımlanabilmesidir. Bu çalışmada soru modelindeki metnin belli bir kısmının değiştirilmesinin ötesinde DinaSoruS sisteminin değişkenler ve parametreler yardımıyla nasıl özgün soru maddeleri üretebildiği gösterilecektir. Sistemde değişkenler ve parametreler yardımıyla birbirinden bağımsız olarak tanımlanmış önermeler kullanılarak farklı soru metinlerinin nasıl üretilebileceği gösterilecektir.

Anahtar Kelimeler

DinaSoruS, otomatik soru üretme, dinamik özgün soru modelleri

*Abant İzzet Baysal Ü.Eğitim Fakültesi İlköğretim Bölümü. İletişim:karakirik@gmail.com Tel: (374) 254 10 00
**Yıldız Teknik Ü.Eğitim Fak.İlköğretim Mat. Eğitimi. İletişim:mdogan69@hotmail.com Tel: (212) 383 48 03

Paper No: 103

Creating Dynamic Novel Item Models Through Computer Based Dinasorus Evaluation System

Abstract

There are a number of dynamic question generation systems that enable the creation of question items with different content or wordings through well-defined item models or certain variables. However, in the most of the systems, the concept of variable generally reduced to numeric values that can be identified with a certain pattern or function or certain text or string elements defined in a template in an item model. This narrow-scoped definition of the question variable leads to expressing question items containing slightly different content with different item models in item pools and also makes it difficult to distinguish structurally identical question items related to same objective. Furthermore, subject knowledge that could be changed dynamically in item models is limited to the predefined variables in the system. Besides, using variables just to replace certain textual parts of an item model leads to the underestimation of the capabilities and possibilities of computer-aided measurement and evaluation systems. Hence, DinaSoruS question generation and evaluation system, having a unique question preparation language that goes beyond the existing variable definitions, has been developed for this purpose. The main difference of DinaSoruS system from other counterparts is the ability to define and change all elements in an item model (including categories, functions, graphics, pictures etc.) and new subject knowledge instantly through variables. In this study, it will be shown how the system could produce novel question formats through parameters and variables beyond the replacement of a certain portion of the text in item model. It will be specifically demonstrated how to create new question formats through parameters and variables by combining predefined item models in the system defined as independent propositions.

Key Words

DinaSoruS, automatic item generation, novel dynamic item models

Bildiri No: 104

Anadolu Üniversitesi İngilizce Yeterlilik Sınavının Bireyselleştirilmiş Bilgisayarlı Test (CAT) Olarak Uygulanması

Murat AKYILDIZ*, Murat Doğan ŞAHİN**

Özet

Psikolojik yapıların ölçülmesi amacıyla geliştirilen ölçekler çoğunlukla Klasik Test Kuramı (KTK) temel alınarak hazırlanmıştır. KTK'da, sabit sayıdaki madde setinin belirli bir gruba uygulanmasıyla elde edilen sonuçlar madde analizi için kullanılarak madde parametreleri elde edilir. Bu nedenle madde parametreleri KTK'da gruba bağımlıdır. Madde Tepki Kuramı (MTK), test tarafından ölçülmek istenilen yetenek ile bireyin bu maddeyi doğru yanıtlama olasılığı arasındaki ilişkiyi matematiksel bir formülle ifade etmektedir. Bu kurama göre, geliştirilen testlerden elde edilen yetenek parametreleri yetenek ranjının sadece bir kısmına sahip bir gruptan, ranjın tamamına dair bilgi verebilecek şekilde gruptan bağımsız olarak elde edilebilir. MTK ve KTK'ya dayalı yetenek kestirimleri arasındaki korelasyon katsayılarının yüksek olduğu çok defa bildirilmiştir. Pek çok araştırmacı, karmaşık istatistiksel işlemlere dayanan MTK'nın, KTK ile benzer sonuçları vermesine rağmen tercih edilmesini, farklı maddeleri alan bireylerin yeteneklerini birbirleriyle karşılaştırma imkanı getirmesine bağlamaktadır. Günümüzde bilgisayarda bireye uyarlanmış testler (BBT) MTK sayesinde yaygınlaşmaktadır. Bu çalışmada da Anadolu Üniversitesi Yabancı Diller Yüksekokulunda uygulanan İngilizce yeterlilik sınavı bilgisayarda bireye uyarlanmış teste dönüştürülecektir. Bilgisayarda bireye uyarlanmış test oluşturmak için yazılımların sayısı sınırlıdır. mirtCAT yazılımı, R dili kullanılarak yazılmış bu sınırlı yazılımlardan birisidir. mirtCAT yazılımı, R dili için bir sunucu ara yüzü olan Shiny web sunucu kullanılarak İngilizce sınavını alacak öğrencilere sınav ortamı olarak sunulacaktır. mirtCAT maddelerin, a, b ve c parametre değerlerinin ve faktör yüklerinin elle girilmesine imkan vermektedir. Ayrıca MTK'ya dayalı, bireye uyarlanmış yetenek kestirimi yapabilmektedir. Bu çalışmanın ana amacı Madde seçimi, test sonlandırma kuralı gibi MTK süreçlerini elle kontrol etme imkanı veren mirtCAT paketinden kestirilen yetenek ölçüleriyle madde havuzunu oluşturan maddelerden oluşan kağıt kalem testinden elde edilen yetenek ölçüleri arasındaki ilişkiyi ortaya koymaktır. Bu sayede web üzerinden uygulanabilir ve kalıcı bir bireye uyarlanmış test uygulamasına ulaşılması planlanmaktadır.

Anahtar Kelimeler

Bireyselleştirilmiş bilgisayarlı test uygulaması, madde tepki kuramı

* Anadolu Üniversitesi Açıköğretim Fakültesi. İletişim: muratakyildiz @anadolu.edu.tr
** Anadolu Üniversitesi Açıköğretim Fakültesi Test Araştırma Birimi. İletişim: mdsahin@anadolu.edu.tr

Paper No: 104

The Administration of Anadolu University English Proficiency Test as a Computerized Adaptive Test (CAT)

Abstract

The scales developed to measure psychological constructs are mostly based on Classical Test Theory (CTT), in which the results obtained by administering a fixed set of items to a certain group are used for item analysis in order to obtain items parameters. Therefore, item parameters are dependent on the "group" in CTT. Item Response Theory (IRT) uses mathematical formulas to show the relationship between the ability aimed to be measured by a test item and the possibility of individuals to give the right answer to this item. According to this theory, the ability parameters can independently be obtained from a group having only some ranges of this ability to make predictions about the whole range. It has often been reported that correlation coefficients between ability predictions based on CTT and IRT are quite high. According to many researchers, IRT, which requires complex statistical calculations, is preferred more than CTT although they both give similar results because IRT enables researchers to compare the abilities of individuals who replied different items. Today, tests adapted to individual use on computers are getting more and more popular thanks to IRT. In this study, proficiency test administered at Anadolu University School of Foreign Languages will be transformed into CAT (Computerized Adaptive Test).

The number of software available to prepare Computer Adaptive Tests is quite limited. Mirt CAT software, which was developed by using R language, is one of the software available in the market. The students who will take the test will be provided a test environment by using mirtCat software and Shiny Web Server, a server interface for R language. It is possible to enter items, a, b, and c parameters of items and factor loads manually by using mirtCAT software. In addition, individualized ability predictions can be made based on IRT. The main aim of the study is to explore the relationship between ability criteria predicted through mirtCAT package, which makes it possible to manually control IRT processes such as item selection and test termination rules, and ability criteria obtained from paper-based test, which also involves the items from the same item pool. By doing so, it is planned to develop a permanent test application adapted to individual and applicable via the internet.

Key Words

Computerized adaptive testing, item response theory

Bildiri No: 105

PISA 2006 Öğrenci Anketinde Yer Alan Tutum Maddelerine Uygulanan Faktör Analizi, Çok Boyutlu Ölçekleme ve Kümeleme Analizi Sonuçlarının Karşılaştırılması

Seda KOZAN *, N. Bilge BAŞUSTA **

Özet

Bu çalışmanın amacı PISA 2006 öğrenci anketi tutum maddelerinin sınıflandırılmasında açımlayıcı faktör analizi, çok boyutlu ölçekleme ve kümeleme analizi sonuçlarını oluşan boyut sayısı ve maddelerin boyutlara dağılımı bakımından karşılaştırmaktır. Faktör analizinin sayıltılarının her zaman sağlanamaması araştırmacıyı benzer sınıflama, boyut indirgeme işlevini yerine getirebilecek analizler kullanmaya yöneltmiştir. Bu nedenle değişkenleri aralarındaki uzaklıklara dayalı olarak sınıflayan çok boyutlu ölçekleme analizi ve kümeleme analizleri karşılaştırmaya dahil edilmiştir. Çalışmanın araştırma grubunu PISA 2006 Türkiye öğrenci anketi maddelerini yanıtlayan, 7 coğrafi bölgemizden bölgelere ve okul türlerine göre tabakalandırılarak OECD tarafından rastgele seçilen 160 okuldan 15 yaş grubu toplam 4942 öğrenci oluşturmaktadır. PISA 2006 öğrenci anketi tutum maddelerine kayıp veri, örneklem büyüklüğü, aykırı değerler, değişkenlerin normalliği, çoklu bağlantı tekillik faktör analizi sayıltıları sınandıktan sonra faktör analizi uygulanmıştır. Faktör sayısına karar vermede özdeğer, açıklanan toplam varyans ve yamaç eğim grafiğinden yararlanılmıştır. Çok boyutlu ölçekleme ve kümeleme analizi herhangi bir sayıltı sınanmadan uygulanmıştır. Çok boyutlu ölçekleme analizinde boyut sayısına karşılık gelen stress ve RSQ değerleri incelenerek çözüm için uygun boyut sayısı belirlenmiş, çizdirilen Öklid mesafe modelinden kümeler belirlenmiştir. Kümeleme analizinde kümelere arası uzaklık ve ağaç grafiğinden yararlanılarak uygun küme sayısına karar verilmiştir. Faktör analizi, çok boyutlu ölçekleme analizi ve kümeleme analizi sonucunda her birinde 7 faktör/küme elde edilmiştir. Maddelerin kümelere dağılımları incelendiğinde her üç analizinde benzer madde gruplandırmalar yaptığı görülmüştür. Bu madde gruplandırmaları PISA 2006 Ulusal nihai raporu gruplandırmaları ile tutarlıdır. Bu araştırmanın sonucunda faktör analizine ek olarak çok boyutlu ölçekleme ve kümeleme analizlerinin oluşturulan yapıyı desteklemede, analiz sonuçlarının geçerliğini artırmada kullanılabilecek birer analiz olabilecekleri sonucuna ulaşılmıştır.

Anahtar Kelimeler

PISA, PISA 2006, Faktör Analizi, Kümeleme Analizi, Çok Boyutlu Ölçekleme

*Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: n.bilgeuzun@gmail.com
** Mersin Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: kozanseda@gmail.com

Paper No: 105

Comparison of Factor Analysis, Multidimensional Scaling and Cluster Analysis For PISA 2016 Student Questionnaire Attitude İtems

Abstract

The aim of this study is to compare factor analysis, multidimensional scaling and cluster analysis in PISA 2006 student questionnaire attiude items in terms of number of dimension and distrubition of items in this dimensions. Because of difficulty for providing factor analysisassumptions, researchers use similar classification and size reduction analysis. Therefore cluster analysis, multidimensional scaling analysis which classify variables based on distance between them are included in the comparison.Research group constituted of 15 year old 4942 student who answered PISA 2006 Turkey student questionnaire.PISA 2006 Student questionnaire items tested missing value, sample size, outliers, the normality of variables, multicollinearty singularity, factorability assumptions. After testing assumptions factor analysis applied.To decide number of factor have been benefited from eigenvalues, explained total variance, scree plot.Multidimensional scaling and clustering analysis is performed without any assumption tested.Stress and RSQ values are examined for multidimensional scaling appropriate size for the number of solutions. Dendogram and agglomerative schedule examined to set appropriate number of cluster for cluster analysis. Factor analysis, multidimensional scaling analysis and cluster analysis were obtained in each of the 7 factors / clusters. Distribution of items in the clusters are similar for all three analyzes. This grouping of items is consistent with the PISA 2006 national final report groupings.The result of this research multidimensional scaling and cluster analysis can be used for supporting structure of factor analysis results and used to increase the validity of analysis results in addition to factor analysis.

Key Words

PISA, PISA 2006, Factor Analysis, Cluster Analysis, Multidimensional Scaling

Bildiri No: 106

Yordayıcı Değişkenlerin Belirlenmesinde Chaıd Analizi İle Yol Analizi Karşılaştırılması

Çiğdem REYHANLIOĞLU KEÇEOĞLU*, Emine Gülen ULUSOY**, Didem ÖZDOĞAN***

Özet

Bu çalışmada bağımlı değişkeni en iyi açıklayan bağımsız değişkenleri ortaya koyabilecek regresyona dayalı iki yöntem karşılaştırılmıştır. Her iki yöntem de regresyon analizine dayalı bir açıklama yaptığı için benzer sonuçların elde edileceği öngörülmüştür. Yöntemlerden biri kategorik veri analizinde çok değişkenli istatistiksel yöntemler arasında önemli bir yere sahip olan CHAID analizidir. Araştırmada kullanılan bir dizi bağımsız değişkenle kategorik bir bağımlı değişken arasında istatistiksel olarak anlamlı bir etkinin olup olmadığını belirleyen CHAID analizi bu etkiyi regresyona dayalı sınıflamayla gerçekleştirir. Yöntemlerden diğeri ise bağımsız değişkenlerin bağımlı değişken üzerindeki doğrudan ve dolaylı etkilerini regresyon analizine dayalı yapan yapısal eşitlik modellerinden yol analizidir. Bu çalışmada 186 ortaöğretim düzeyindeki bireyin matematik başarılarını açıklamak üzere 7 değişken belirlenmiş ve bu değişkenlerin matematik başarısını yordama düzeyleri ortaya konmuştur. Etkisi incelenen bağımsız değişkenler; matematik başarı puanı hesaplanan bireyin anne eğitim düzeyi, baba eğitim düzeyi, ailesinin sosyo-ekonomik düzeyi, evde birlikte yaşadığı kişi sayısı, bir eğitim döneminde okuduğu kitap sayısı, yaşadığı şehrin nüfusu ve okul dışında bir haftada matematiğe ayırdığı zaman olarak belirlenmiştir. Bu amaçla kullanılan yöntem CHAID analizidir. CHAID analizinde yordayıcı değişken olarak belirlenen bağımsız değişkenler ile yol analizi modeli oluşturulmuş ve model veri uyumu incelenmiştir. CHAID analizi sonucunda bireylerin matematik başarısını olumlu etkileyen en önemli değişken okunan kitap sayısı, okunan kitap sayısını en çok etkileyen değişken ise anne eğitim düzeyidir. CHAID analizinde okunan kitap sayısını etkileyen en önemli ikinci değişken sosyo-ekonomik düzey ve yine CHAID analizine göre okunan kitap sayısını en az etkileyen değişken baba eğitim düzeyidir. Yol analizinde kurulan modele göre, nüfus değişkeninin okunan kitap sayısı üzerindeki dolaylı etkisi baba eğitim düzeyi ile sağlanmaktadır. Nüfus ve baba eğitim düzeyinin birlikte matematik başarısına olan etkileri ise okunan kitap sayısı üzerinden sağlanmaktadır. Ayrıca CHAID analizi ile elde edilen sonuçlara dayalı olarak kurulan model yol analizinde zayıf uyum göstermiştir.

Anahtar Kelimeler

Matematik Başarısı, CHAID Analizi, Yol Analizi, Kategorik Değişken

* Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Bölümü Doktora Öğrencisi

**Marmara Üniversitesi, Atatürk Eğitim Fakültesi, İlköğretim Matematik Öğretmenliği Bölümü

***Osmangazi Üniversitesi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Bölümü

Paper No: 106

Comparing Chaid Analysis And Path Analysis While Defining Predictor Variable

In this study, two methods which are the most suitable to define dependent variable via independent variables based on regression models are compared. It is predicted that because both of the methods based on regression analysis, their results may be similar. One of them is CHAID analysis which is used frequently in multivariate statistics in categorical data analysis. CHAID analysis is studied the importance of relations between dependent and independent variables by regression based on classification analysis. The other method is path analysis, a kind of structural equation model, also based on regression analysis is aimed to disclose direct and indirect effects of independent variables on dependent variable. In this research, 7 independent variables are specified to predict mathematics achievement of 186 high school students and predictive level of these variables to mathematics achievement are mentioned. Independent variables that effects are researched are determined as education level of the mother, education level of the father, social economic level of the family, number of people living together with student in the same house, number of books that being read during one semester, population of the city where the student live, and length of study to practice mathematics except school hours. CHAID analysis is used to analyse those relations. Model of path analysis is formed with independent variables which are defined as predictive variables in CHAID analysis, and model data fit indices are examined. As a result of CHAID analysis, the most significance variable that effects mathematics achievement positively is defined as number of books that being read during one semester, in addition, number of books that being read during one semester is effected mostly by education level of mother. Second variable that effects number of books that being read during one semester is found out as social economic level of family, moreover, education level of father is determined as the least effective variable on number of books that being read during one semester. Based on the model of path analysis, indirect effect of population of the city on number of books that being read during one semester is provided by education level of father. Effects of population of the city and education level of father together on mathematics achievement are conducted by number of books that being read during one semester. Furthermore, the model that formed by results obtained from CHAID analysis is showed poor fit in path analysis.

Key Words

Mathematics Achievemet, CHAID Analysis, Path Analysis, Categorical Variable

Bildiri No: 107

Bireye Uyarlanmış Bilgisayarlı Test Ortamında Maksimum Fisher Madde Seçme Yönteminin İncelenmesi

Mevra KAÇAR*, Sema SULAK**

Özet

Bireyselleştirilmiş bilgisayarlı test uygulamalarında Maksimum Fisher Bilgisi madde seçme yönteminin etkililiğinin incelenmesi amaçlanmıştır. Bu amaçla farklı madde havuzları simule edilerek MFB madde seçme yönteminin madde sayısına göre nasıl farklılaştığı araştırılmıştır. MFB madde seçme yönteminin, test başlangıcında yüksek a parametresine sahip maddeleri seçme eğiliminde olduğu çeşitli araştırmalarda ele alınmıştır. Bu araştırmada da madde havuzu genişliğinin MFB madde seçme yönteminin bu durumu üzerindeki etkisi incelenmiştir. Eldeki araştırma simülasyon çalışması olarak gerçekleştirilmiştir. Araştırmanın verisi SIMULCAT bilgisayar programı aracılığıyla üretilmiştir. Bilgisayar ortamında bireyselleştirilmiş test uygulamalarında Maksimum Fisher Bilgisi madde seçme yöntemine göre kestirilen yetenek düzeyinin standart hatası, RMSE, BIAS ve yanlılığın, madde havuzu genişliğine göre nasıl bir değişim gösterdiği incelenmiştir. Tahminin standart hatasının madde havuzu genişliğine göre incelenmesi sonucunda; en yüksek değerin madde havuzu 500 iken, en düşük değerin ise madde havuzu 1500 iken elde edildiği görülmüştür. Bilgisayar ortamında bireyselleştirilmiş test uygulamalarında Maximum Fisher Bilgisi madde seçme yöntemine göre elde edilen madde havuzu kullanım sıklığının, madde havuzu genişliğine göre nasıl bir değişim gösterdiği araştırılmıştır. Bu problemin analizi için bütün simülasyon koşullarında her bir cevaplayıcının cevap örüntüsü kullanılmıştır. Cevap örüntülerinde madde havuzundan çekilen maddelerin frekansları ve a parametre değerleri incelenmiştir. Madde havuzu genişliğinin bu durumu nasıl etkilediğini belirlemek için her üç simülasyon koşulunda ilk on maddenin a parametre değerleri bulunmuştur. Madde havuzu genişliği 500, 1000 ve 1500 iken elde edilen a parametre değerlerinin anlamlı farklılık gösterdiği bulunmuştur. Madde havuzu genişliği arttıkça yüksek a parametresine sahip maddelerin kullanıldığı görülmüştür. MFB madde seçme yönteminin farklı yetenek kestirim yöntemleri ve durdurma kurallarına göre incelenmesi önerilmektedir.

Anahtar Kelimeler

Bireyselleştirilmiş Bilgisayarlı Test Uygulamaları, madde seçme yöntemleri, Madde Havuzu, Maximum Fisher Bilgisi

* Yüksek Lisans Öğrencisi, Mevra KAÇAR, Doğa Okulları, mewra_akturk@hotmail.com
** Yrd. Doç. Dr. Sema SULAK, Bartın Üniversitesi, semasulak@gmail.com

Paper No: 107

Investigation of Maximum Fisher Item Selection Method on Computerized Adaptive Testing

The effectiveness of Maximum Fisher Information item selection method on computerized adaptive testing was investigated. For this aim, different item pools were simulated in order to search as to whether MFB item selection method differ based on item number. Several research concluded that MFB item selection method has tendency to choose items that are higher a parameters at the beginning of the test. In this research, the impacts of MFB item selection method and item pool extension on this aforementioned situation was investigated. The data of the study was created utilising SIMULCAT computer program. It was investigated how the extension of item pool effect standard error of estimation, RMSE, BIAS, and inaccuracy of estimated ability levels, that are based on MFB item selection method, during computerized adaptive testing. As a result of how estimated standard error differ based on item selection method, it was found that the maximum was when item pool was 500 and the minimum was when the item pool was 1500. Moreover, how item pool extension differs based on item pool selection frequency, that are based on MFB item selection method during computerized adaptive testing was investigated. In order to analyse this question, each responders' answer patterns in every simulation situation was used. on these answer patterns item frequencies and a parameters were searched. In order to determine how item pool extension impacts this situation, a parameters of first ten questions from three simulation situations were found.It was found that when item pool extension was 500, 1000, and 1500; a parameters were significant. Additionally, higher a parameters items were used when item pool extension increases. It was further recommended to investigate MFB item selection method according to different ability estimation method and stopping rules.

Key Words

Adaptive testing, item selection, article pool, Maksimum Fisher Information

Bildiri No: 110

Okul Öncesi Eğitimde Ölçek Uyarlamaya Dayalı Çalışmaların İncelenmesi

Esra SÖZER*, Seyhan SARITAŞ**, İsmail KARAKAYA***

Özet

Okul öncesi dönemde çocukların dinamik bir yapıda olması ve henüz okuma yazma bilmiyor olmaları ölçme sürecinde çevresel faktörlerin etkisini arttırabilmektedir. Bu nedenle de okul öncesi eğitimi alanında kullanılan ölçeklerin uyarlanması aşaması daha dikkatli bir çalışmayı gerektirmektedir. Uyarlama sürecinin hatalı olması, sonuçların hatalı yorumlanmasına neden olacaktır. Bu çalışmada okul öncesi eğitimi alanında yapılmış olan ölçek uyarlama çalışmalarının belirli ölçütler kapsamında geçerlik ve güvenirlik süreçlerinin incelenmesi amaçlanmıştır.Çalışmanın örneklemini 2005-2015 yılları arasında Türkiye'de okul öncesi eğitimi alanında yayınlanan, SSCI ve ULAKBİM ulusal veri tabanında indekslenen dergilerde yayınlanmış makaleler ve Ulusal Tez Merkezi veri tabanında yayınlanmış tezler oluşturmaktadır. Çalışmada 18 makale ve 11 tez doküman incelemesi yoluyla incelenmiştir. Betimsel içerik analizi kullanılmıştır. Veri toplama aracı olarak "Ölçek Uyarlama Süreci Kontrol Formu" kullanılmıştır. Araştırmacılar tarafından yapılan kodlamaların güvenirliğini belirlemek için randum seçilen makaleler analiz edilmiş ve kodlamalar arası tutarlılık 0.896 olarak hesaplanmıştır. İncelenen tez ve makalelerin tümünde ölçme aracının amacının ve ölçülen yapının kuramsal temelleri belirtilmiştir. İncelenen çalışmalarda ağırlıklı olarak kaynak formun güvenirlik katsayılarına yer verilmiştir. Çalışmaların dilsel eşdeğerliği için sıklıkla geri çeviri yöntemi kullanılmıştır. Çalışmaların genelinde çeviri işleminden sonra uzman kişilerin görüşüne başvurulduğu belirlenmiştir. Hedef formun hedefe gruba uygulandığı veri toplama deseni seçildiği görülmüştür. Geçerlik için yapı geçerliğine yönelik çoğunlukla AFA ve AFA ile DFA'nın beraber kullanıldığı, ölçüt geçerliğinin ise çoğunlukla incelenmediği belirlenmiştir. Ölçüt geçerliğinin incelendiği çalışmalarda ise genel olarak uygunluk (convergent) geçerliği incelenmiştir. Çalışmaların bir kısmında ise uyarlanan ölçeklerin özelliğinden dolayı içerik analizi kullanılmıştır. AFA ve DFA' nın birlikte yapıldığı çalışmalarda, uygulanan analizlerin büyük kısmında aynı örneklem üzerinden analizlerin yapıldığı sonucuna ulaşılmıştır. Çalışmalarda genel olarak madde analizi yapılmamıştır. Yapılan güvenirlik çalışmalarında sıklıkla iç tutarlılık ile test tekrar testin beraber kullanıldığı görülmüştür. AFA yapılan çalışmaların genelinde uyarlanan ölçeğin orijinal ölçekle aynı boyut sayısına sahip olduğu görülmüştür.

Anahtar Kelimeler

Ölçek Uyarlama, Geçerlik, Güvenirlik, Okul Öncesi Eğitimi

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: e.sozer89@gmail.com
Tel: +90 0312 202 36 97
**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: seyhangenom@gmail.com
Tel: +90 0312 202 19 25
***Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: ikarakaya2002@gmail.com
Tel: +90 0312 202 17 49

Paper No: 110

A Review of Studies Based on Scale Adaptation in the Field of Preschool Education

Abstract

Adaptation of the scales used in preschool education field is more than the impact of the environment and children compared to other groups so scale adaptation requires more careful study. A wrong adaptation process will result in the incorrect interpretation of the results. This study aims at reviewing the realibility and validity process of studies written on scale adaptation in the field of preschool education based on certain criteria. A total of 18 articles published in journals in Turkey indexed in SSCI and/or ULAKBIM and 11 thesis published in National Thesis Center during 2005-2015 were analyzed for this study. The studies were examined by document analyis. In this study, descriptive content analysis is used. The "Scale Adaptation Process Check Form" was used in the analysis of the articles/thesis. Random selected articles were analyzed to determine the realibilty of the coding done by researchers and consistency between coding is calculated as 0.896. The results demonstrated that the purpose of the tools of measurement was stated and that the theoretical bases of the structures to be measured were described in the studies all of scale adaptation. In the form of the source of the realibilty coefficient is mainly given in examined studies. In studies of linguistic equivalence, the back translation method was often used. After the translation process, it was determined that experts consulted. It is determined that the translated target forms were usually administered only to the target groups. It was concluded that the EFA and EFA and CFA together were used in combination in most of the studies for determining the construct validity, whereas the criterion validity was not studied generally. Convergent validity was examined in the study of the examined criterion validity. Because of the examined scale characterictics, some of the studies were used content analysis. The method of internal consistency and retest realibilty were usually preferred in calculating the reliability of the scale adapted. In the general studies of EFA was shown that adjustment scale have the same number of dimensions to the original scale.

Key Words

Scale Adaptation, Validity, Realibility**,** Preschool Education

Bildiri No: 111

Çok Değişkenli Regresyon, Yapay Sinir Ağları ve Bulanık Mantık Tahminleme Modellerinin Karşılaştırılması

Cem Oktay GÜZELLER, Rasim ÖNDER, Caner DURSUN

Özet

Bu çalışmanın amacı Yapay Sinir Ağları (YSA), Bulanık Mantık (ANFIS) ve Çoklu Regresyon tahminleme modellerini karşılaştırmaktır. Çalışma ilişkisel tarama modelinde gerçekleştirilmiştir. Verileri PISA 2012 datasında yer alan öğretmen yeterlikleri ile öğrenci matematik başarısı oluşturmaktadır. Örneklemi ise PISA 2012'ye Türkiye'den katılan 4848 öğrenci arasından eksik ve kayıp değer sorunu yaşamayan 2954 kişiden oluşturmaktadır. Analizler sonucunda YSA (R^2=%81,68) ve ANFIS (R^2=%79,40) doğru tahminleme olasılıklarının Çoklu Regresyon (R^2=%71,8) modeline göre daha yüksek olduğu görülmüştür. Bu doğrultuda tahminleme için kurulacak modellerde Çoklu Regresyon yerine YSA ve ANFIS tercih edilebilir.

Anahtar Kelimeler

Yapay Sinir Ağları, Bulanık Mantık, Çoklu Regresyon, Tahminleme

Bildiri No: 112

Faktör Analizi Kullanılan Yayınlamış Bilimsel Makalelerde Örneklem Özellikleri

Hakan ATILGAN*, Burak AYDIN**, Sungur GÜREL***, Mehmet KAPLAN****

Özet

Ölçek uyarlama ve geliştirme çalışmalarında kullanılan istatistiksel çözümleme tekniklerinden bir tanesi faktör analizidir. Faktör analizleri ile tanımlanan ya da doğrulanan faktörlerin genellebilirliği örneklemin evreni temsil edebilirliği ile doğru orantılıdır. Bu çalışmanın amacı Turkiye'den yayınlanan uluslararası indekslerce taranan dört dergide yayınlanmış faktör analizi kullanan makalelerin örneklemlerini incelemektir. Bu dergiler; Eurasian Journal of Educational Research (EJER), Kuram ve Uygulamada Eğitim Bilimleri (KUYEB), Hacettepe Üniversitesi Eğitim Fakültesi ile Eğitim ve Bilim dergileridir. Çalışma kapsamında yer alan dergilerde faktör analizi kullanılan makalelerinde taranan bilgiler şunlardır; (a) çalışmanın türü (ölçek uyarlama, ölçek geliştirme), (b) faktör analizine taabi tutulan madde sayısı, (c) faktör analizi çerçevesi (doğrulayıcı ya da açımlayıcı), (d) örnekleme yöntemi, (e) örneklem büyüklüğü, (f) veri türü (kategorik, dereceleme),(g) kullanılan kestirici, (h) kullanılan yazılım, (ı) kayıp veri yüzdesi, (j) popülasyon tanımın mevcut olup olmaması, (k) alt faktör sayısı, (l) karşılaştırılan model sayısı.

Çalışma kapsamında edinilen öncül sonuçlar, incelenen makalelerin çok büyük bir çoğunluğunun uygun örnekleme yöntemini kullandığını göstermiştir. Uygun örneklemin kendinden büyük herhangi bir evreni temsil edebileceği dahi şüpheliyken çalışmalarda bu kadar sık kullanılması kaygı vericidir. Bir diğer öncül sonuç çalışmalarda kayıp veriye neredeyse hiç rastlanmamış olmasıdır. Çalışmalarda kayıp veriye rastlanmaması ya da kayıp verinin nasıl işlendiğinin açıkça belirtilmemesi de kaygı vericidir. Çalışma sonunda örneklem kalitesinin önemi hakkında farkındalık oluşturulması beklenilmektedir.

Anahtar Kelimeler

Faktör analizi, Örneklem, Geçerlik

*Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü: atilgan.h.@gmail.com
**Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü : burak.aydin@erdogan.edu.tr
***Siirt Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü: sungur5@hotmail.com
**** Milli Eğitim Bakanlığı. İletişim: mehmet.kaplan2@gmail.com

Paper No: 112

Sampling Characteristics of Factor Analyses Articles Published in Turkey

Abstract

Factor analyses are commonly used statistical procedures when developing or adapting a scale. Generalizability of extracted factors is directly related to sample's quality and representativeness. This study aims to examine the factor analyses sample characteristics within the articles published by four respected Turkish journals; EJER, KUYEB, Hacettepe, and Education & Science.

We examined (a) study type (scale development or adaptation), (b) number of items, (c) framework (confirmatory or explanatory), (d) sampling procedure, (e) sample size, (f) item types, (g) software, (h) estimator, (i) missing data occurrence, (j) population information, (k) number of factors and (l) number of competing factor models.

Our preliminary findings showed that the vast majority of inspected articles used convenience sampling. This finding raises concerns about the external validity of factor analytic procedures conducted in Turkey. We are also concerned about the non-existence of missing data in our preliminary inspection. The ultimate purpose of this study is to raise awareness for the importance of good sampling.

Key Words

Factor Analyses, Sampling, Validity

Bildiri No: 113

PISA Sınavlarında Sorulan Soruların Math Taksonomi Çerçevesinde Analizi

H. Sevgi MORALI*, Hakan KARADUMAN**

Özet

Matematik, eğitim-öğretimin her aşamasında tartışılmaz bir öneme sahiptir. Her alanda olduğu gibi matematikte de doğru dönütler alabilmek için detaylı bir ölçme değerlendirme aşamasından geçmek gerekmektedir. Ölçme ve değerlendirmenin alana özgü yöntemlerle yapılması daha sağlıklı sonuçlar alınmasına katkıda bulunabilir. Matematik bilgisini ölçen uluslararası bir sınav olması bakımından PISA sorularının MATH Taksonomisi(**M**athematical **A**ssessment **T**ask **H**ierarchy) kapsamındaki yerlerinin ne olduğu merak uyandırıcıdır. MATH taksonomisi, Bloom taksonomisinin matematik alanına özgü şekilde geliştirilmesi ile oluşturulmuştur. Doğrudan matematik alanında ölçme yapmak için Bloom Taksonomisinin matematiğe uyarlanmış halidir. Araştırmanın amacı; PISA sınavlarında sorulan soruların Smith ve arkadaşları tarafından geliştirilen MATH Taksonomideki yerini ortaya koymaktır. Bu amaçla, yayınlanmış 3 adet PISA sınavının soruları MATH Taksonominin gruplarına göre incelenerek değerlendirilmiştir. Sınavların seçiminde güncellik ön planda tutulmuştur. Araştırma nitel bir özel durum çalışmasıdır. Araştırmanın güvenirliğinin sağlanması için, birinci aşamada iki araştırmacı soruları bireysel olarak MATH Taksonomi kategorilerine göre incelemişlerdir. İkinci aşamada bir araya gelerek bulgularını tartışmışlardır. Bulguların karşılaştırılmasında %85 oranında tutarlılık görülmüştür. Bireysel değerlendirmelerde farklılık gösteren sonuçlar tartışılıp kategorizasyona son hali verilmiştir. Üç sınavda da soruların büyük oranda rutin işlemler kategorisine ait olduğu temel sonucuna ulaşılmıştır.

Anahtar Kelimeler

MATH Taksonomi, PISA, ölçme-değerlendirme, matematik dersi bilgileri.

* Dokuz Eylül Üniversitesi, sevgi.morali@deu.edu.tr

** Dokuz Eylül Üniversitesi, hakankaraduman.deu@gmail.com

Paper No: 113

Analysıs Of PISA Exam Questıons Wıthın The Framework Of The Math Taxonomy

Abstract

Mathematics is an impartant topic in every stage of education.As in all fields,attaining correct feedback requires a detailed Assesment and evaluation process. Using methods sepecific to the field may lead to more accurate results. MATH TAxonomy (Mathematical Assessment Task Hierarchy) has been developed based on Bloom's taxonomy specifically for mathematics. Being on international exam, designed to assess mathematical knowladge, the place of PISA exam questions within MATH Taxonomy is a metter of interest.The research objective is to evaluate PISa exam questions within the framework of MATH Taxonomy, devoloped by Smith and co. To this end, thequestions of three puplished PISA exams have been examined according to their places in MATH Taxonomy. The exam used were chosen for being most recent. The research is a qualitative case study. For reliability, two researchers indivicually within categorized the questions MATH Taxonomy. They then come together to compare and discuss their seperate results when compared, it was seen that the two categorizations were consistent to degree of 85%. The categorization was given its final form after the items that were categorized differently had been discussed. The main conclusion reached is that all three exams consist mostly of questions belonging to the routine operations category.

Key words

MATH Taxonomy, PISA, Evaluation and assesment, mathematics field knowledge.

Bildiri No: 115

Ölçme Araçlarının Psikometrik Özellikler Açısından İncelenmesi

Esra Sözer*, Derya Çavdar**, Nilüfer Kahraman***

Özet

Bilişsel olmayan beceriler olarak tanımlanan 21. yüzyıl yaşam becerileri kapsamında öğrencilerin yaratıcılık, iletişim, işbirliği gibi becerilerinin ölçülmesi amaçlanmaktadır. Bilişsel olmayan becerilerin, duyuşsal özellikler gibi, doğru-cevap için geliştirilmiş ölçeklerle ölçülmesi her durumda mümkün olmamaktadır. Ölçülmek istenilen özelliklerin tipik cevaplar aracılığıyla çalışıldığı durumlarda farklı yaklaşımlar izlenmelidir. Bu becerilerin ölçülmesi için genellikle bireylerin kendilerini değerlendirmesine dayalı ölçümlerden yararlanılmaktadır. Bu nedenle, kullanılan ölçme araçlarında yer alan maddelerin istenilen becerileri ölçmede ne derece katkı sağladığının psikometrik açıdan daha detaylı olarak incelenmesi önemlidir. Bu çalışmada, 21.yüzyıl becerilerini ölçmek için geliştirilmiş bir ölçeğin geçerliliğinin çalışılması amaçlanmıştır.

Özetlenen uygulama, bireylerin 21.yüzyıl becerilerini ölçmek için derecelendirilmiş cevap ölçeği kullanan maddelerden oluşan çok-aşamalı bir test uygulamasıdır. Maddelerden, bu makaledeki geçerlik çalışmasına konu olan iki madde, kültürel yapının etkili olup olmadığının çalışıldığı, geleneksel ve muhafazakar maddelerdir. Test desenine ek olarak uygulanmış olan, Büyük 5 ölçeğine verilen cevaplar dikkate alındığında, bulunmuştur ki bireylerin diğer bireylerin düşüncelerine saygı gösterme özelliğinin belirlenmesinde geleneksel aile yapısı önemlidir. Yine Büyük 5 kişilik ölçeğinde yer alan diğer maddede ayrımcılık ve önyargıdan uzağım ifadelerinin hem geleneksel hem de muhafazakar aile yapısından geldiğini söyleyenlerde anlamlı olduğu gözlenmiştir. Aile yapısının, bireylerin geleneksel/yaratıcı olmayan özelliğinin daha güvenilir ölçülmesinde katkı sağladığı bulunmuştur. Ancak aile yapısı özelliğinin belirlenmesi, bireylerin mutluluk, iyimserlik, özgüven gibi özelliklerinin belirlenmesinde herhangi bir katkı sağlamadığı görülmüştür. Bu da göstermektedir ki ölçek içerisinde yer alan maddelerin bir kısmı ilgili özelliğin ölçülmesinde daha güvenilir sonuçlar oluştururken bazı maddeler yeterli olmamaktadır.

Sonuçlar incelendiğinde, bireylerin kendilerine uygun cevapları seçmelerine dayalı kullanılan derecelendirilmiş ölçeklerde ilgili özelliğin belirlenmesine yönelik daha fazla bilgi sağlaması için ölçek geliştirme aşamasında daha titiz çalışılması gerektiğini göstermektedir. Bireylerin derecelendirilmiş ölçeğe verdikleri cevapların yanında ek ölçme araçlarının da kullanılması ilgili özelliği belirlemede daha güvenilir ve geçerli sonuçların elde edilebileceğini göstermektedir.

Anahtar Kelimeler

21.Yüzyıl Becerileri, Geçerlik, Güvenirlik

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: e.sozer89@gmail.com
Tel: +90 0312 202 36 97
**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: drycvdr1902@gmail.com
Tel: +90 0312 202 36 97
***Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: nkahraman6@gmail.com

Paper No: 115

Investigation of Psychometric Properties of Measurement Tools

Abstract

Defined as non-cognitive skills that 21st century life skills aims to measure students' creavity, communication, collaboration skills. Such as affective characteristics, non-cognitive skills is not measure possible in every case with improved scale for measuring the correct answer. In cases where different approaches attempt to measure the desired characteristics through the typical responses are monitored. To measure this ability, generally utilized for measuring based on the assessment of the individuals themselves. Therefore, it is important that the measurement tools used in the items must be examined in terms of psychometric measurements more details that contribute to what extent the required skills. In this study an improved scale to measure the 21st century skills aimed to study the validity.

Outlined in the application , consisting of agents using graded response scale to measure the individuals' the 21st century skills is a multi- stage test. Two of items that are subject to validation study in this article that studied whether effective cultural structures are traditional and conservative materials. In addition to the test pattern has been applied, taking into account the answers to the big 5 scale, it is important that traditional family structure to determination of characteristic respecting the opinions of others people. In other an item located in big 5 scale, people said to come from both traditional and conservative family structure, discrimination and away from prejudice expression is meaningful. Family structure of individuals' has been found to contribute to the measurement traditional / non- creative feature more reliable. But the determination of the structure of the family property has been shown to provide any additional knowledge for such as features determination of individual happiness, optimism, self-confidence. Indicated that, some of the items used to measure a portion of the related feature is creating more reliable results while the other scale items is not enough to measure this feature. When the results were analyzed, to provide more information that graded scale used in individuals based on their answers to select the appropriate degree of the scale of development for the determination features indicates that it should be studied more carefully. The use of additional measurement tools to determine of the feature shows that obtain more reliable and valid results.

Key Words

21st Century Skills, Realibility, Validity

Bildiri No: 116

Madde Parametre Kaymasının Test Eşitleme Üzerindeki Etkisi

Kadriye Belgin DEMİRUS*, İbrahim UYSAL**

Özet

Madde Tepki Kuramı'nın (MTK) varyans değişmezliği özelliğine göre madde parametre değerleri evrendeki tüm örneklemler için aynı olmalıdır. Ancak uygulamalarda varyans değişmezliği varsayımı her zaman karşılanmayabilir. Bir testte madde parametre değerleri aynı yetenek düzeyinde bulunan farklı alt gruptaki bireyler için farklılaştığında bu durum değişen madde fonksiyonu (DMF) olarak adlandırılmaktadır. Madde parametreleri zaman boyunca veya farklı test uygulamalarında değiştiğinde ise bu durum madde parametre kayması (MPK) olarak tanımlanmaktadır. Madde parametre kaymasının program değişmesi, kapsam etkileri, örneklem istatistikleri ve madde içeriği gibi olası nedenleri bulunmaktadır. MPK testlerin geçerliğini negatif yönde etkilemektedir. Bununla birlikte, MPK'lı maddeler eşitleme katsayılarını da etkilemektedir ve genellikle bağlama yapılan maddelerden çıkarılmaktadır. Bu araştırmanın amacı MPK'nın test eşitlemeye etkisini farklı eşitleme yöntemleri (Ortalama-ortalama, ortalama-sigma,, Haebara ve Stocking-Lord) kullanarak ortaya koymaktır. Bu amaç doğrultusunda MPK'nın bulunduğu farklı koşullar ile MPK'nın bulunmadığı koşulda elde edilen RMSD eşitleme hataları karşılaştırılmıştır. Aynı zamanda MPK'nın etkisini en aza indiren eşitleme yöntemi belirlenmiştir. Araştırma bir Monte Carlo simülasyon çalışmasıdır. Araştırmada MPK yönü (tek yönlü MPK: sadece artan, sadece azalan ve çok yönlü MPK: artan ve azalan birlikte) simülasyon koşulu olarak ele alınmıştır. Araştırmada sadece b güçlük parametresinde yüksek düzeyde kayma ele alınmıştır. MPK büyüklüğü $\delta=0.5$ olarak seçilmiştir. Çalışmanın madde tepki verisi sadece çoktan seçmeli maddeler için düşünülmüş ve bir yıl arayla iki farklı test durumu üç-parametreli lojistik model kullanılarak simüle edilmiştir. Araştırma sonuçlarına göre, ortak maddelerin b parametresinde tek yönlü azalan MPK olduğunda en büyük RMSD eşitleme hatası Stocking-Lord yönteminde, en küçük RMSD hatası ise ortalama-ortalama yönteminde elde edilmiştir. Tek yönlü artan MPK olduğunda ise en büyük RMSD eşitleme hatası ortalama-sigma yönteminde elde edilirken, en küçük hata Stocking-Lord yönteminde bulunmuştur. Ortak maddelerde çok yönlü MPK olduğunda en büyük RMSD eşitleme hatası ortalama-sigma yönteminde, en küçük hata ise ortalama-ortalama yönteminde bulunmuştur.

Anahtar Kelimeler

Madde Tepki Kuramı, madde parametre kayması, varyans değişmezliği, test eşitleme

*Milli Eğitim Bakanlığı, Fen Bilgisi Öğretmeni. İletişim: belgindemirus@gmail.com Tel: 0505 742 36 12

**Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: ibrahimuysal06@gmail.com Tel: 0312 780 7259

Paper No: 116

The Impact of Item Parameter Drift on Test Equating

Abstract

According to the invariance property of Item Response Theory (IRT), item parameter values should be the same for all samples from a population. However, in practice, this assumption of invariance does not always hold. When item parameter values in the same test behave differently for different subgroups with the same degree of abilty, it is called differential item functioning (DIF). When item parameter values change on different testing occasions or testing time it is called item parameter drift (IPD). There are some possible sources of drift, such as change in curriculum, content affects, sample statistics, and content of items. IPD affects the validity of tests negatively. However, items flagged as IPD affect equating coefficients as well and are often removed from the linking items. The purpose of the present study is to investigate the effect of item parameter drift on test equating using different equating methods (Mean-mean, mean-sigma,, Haebara ve Stocking-Lord). For this purpose, root-mean-square difference (RMSD) equating errors were compared under the different simulation conditions that include IPD and non-IPD. Also equating method that most decrease the affects of IPD was assessed. This research is a Monte Carlo simulation study. Direction of IPD (unidirectional IPD: always increasing, always decreasing and multidirecional IPD: both increasing and decreasing) considered as simulation conditon. In the research large amount of drift on difficulty parameter b was simulated. Magnitude of IPD was chosen as δ=0.5. Item response data were considered for only multiple-choice items and two testing occasions one year apart were simulated for the three-parameter logistic model. According to the results of the research, when there was unidirectional decreasing IPD on b parameter , Stocking Lord method produced the biggest equating error while mean-mean method produced the smallest. When there was unidirectional increasing IPD on b parameter , mean-sigma method produced the biggest equating error while Stocking Lord method produced the smallest. When there was multidirectional IPD on b parameter of anchor items, mean-sigma method produced the biggest equating error while mean-mean method produced the smallest.

Key Words

Item Response Theory, item parameter drift, invariance, test equating

Bildiri No: 117

Bir Bilişsel Tanı Modeli Uygulaması: Niteliklerin Belirlenmesi ve Geçerlik Analizi

Mehmet KAPLAN[*], Hakan ATILGAN[**], Sungur GÜREL[***], Burak AYDIN[****]

Özet

Bilişsel tanı modellemesi (BTM) teknikleri öğrenme eksikliklerinin belirlenmesinde kullanılan biçimlendirici değerlendirme tekniklerinden biri olup, son zamanlarda eğitimde ölçme literatüründe önemi giderek artmıştır. Bu çalışmada Temel Kabiliyetler Testi (TKT) 7-11 kelime testi için, açımlayıcı faktör analizi ve konu alanı uzman görüşleri yardımıyla testte bulunan her bir madde için nitelikler belirlenmiştir. Daha sonra BTM'de bulunan genel bir model kullanılarak, testi alan bireyler uzman oldukları ya da yetersiz kaldıkları niteliklere göre sınıflandırılmıştır. Ayrıca, her bir madde için belirlenen niteliklerin geçerlik analizi yapılmıştır. Araştırmanın bir sonraki aşamasında geçerlik sonucu elde edilen maddelerin nitelik vektörleri tekrar uzman görüşleri yardımıyla değerlendirilecektir. Ayrıca, yorumlanması genel modellere göre daha basit olan kısıtlayıcı modeller kullanılarak, TKT 7-11 verisinin model uyumu her bir madde için incelenecektir.

Anahtar Kelimeler

Bilişsel Tanı Modellemesi, Biçimlendirici değerlendirme, G-DINA

[*]Milli Eğitim Bakanlığı. İletişim: mehmet.kaplan2@gmail.com
[**]Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: atilgan.h.@gmail.com
[***]Siirt Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: sungur5@hotmail.com
[****]Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: burak.aydin@erdogan.edu.tr

Paper No: 117

An Application of Cognitive Diagnosis Modeling: Attribute Identification and Validation

Abstract

Cognitive diagnosis modeling (CDM) is one of the formative assessment techniques that can determine strengths and weaknesses of individuals in a particular content area, and its importance has been increased in the recent educational measurement literature. In this study, attributes have been identified for each item in Primary Mental Abilities 7-11 test using an exploratory factor analysis and content expert opinions. Then individuals have been classified whether they have mastered the defined attributes using a general model in the context of CDM. In addition, the attribute validation has been done for each item in the test. In the next step of this research, the attribute validation for each item will be performed with the help of expert opinions again. Also, the model fit using reduced models will be investigated because the interpretation of reduced models is much easier compared to general models.

Key Words

Cognitive diagnosis modeling, Formative Assessment, G-DINA

Bildiri No: 118

Farklı Kategori Sayısı ve Yetenek Dağılımı Durumlarında Kategoriler Arası Psikolojik Uzaklıkların İncelenmesi

Gizem UYUMAZ*, Gözde SIRGANCI**, Ömay ÇOKLUK***

Özet

Psikolojik özelliklerin ölçülmesinde kullanılan ölçme araçlarının birçoğu sıralama ölçeği düzeyinde olmakla birlikte, üst düzey matematiksel işlemlerin yapılabilmesi için eşit aralıklı ölçek düzeyinde olduğu varsayılır. Ancak aralıkların eşit olmaması durumu, bireyler arasındaki farkların doğru belirlenememesine, elde edilen sonuçların kullanılabilirliği ve bu ölçeğe dayalı olarak verilen kararların sorgulanır hale gelmesine neden olur. Aynı zamanda güvenirlik de sonuçların doğruluğunu etkileyecek bir başka psikometrik niteliktir. Bu doğrultuda bu çalışmada, dereceleme ölçeklerinde göreli olarak birimlerin, yani kategoriler arası psikolojik uzaklıkların eşitliği varsayımının uygulamada sağlanıp sağlanmadığının ve bireylerin yetenek dağılımlarına ilişkin özellikler ile kategori sayısında yapılan değişimlerin bu varsayımı ve güvenirliği nasıl etkilediğinin incelenmesi amaçlanmıştır. Kategoriler arası psikolojik uzaklıkların belirlenmesi amacıyla çoklu puanlanan MTK modellerinden Genelleştirilmiş Kısmi Puanlama Modeli (GPCM) kullanılmıştır. Çalışma, farklı koşullar altında tepki seçenekleri arasındaki psikolojik uzaklıkların nasıl değiştiğini inceleyen bir simülasyon çalışmasıdır. Çalışmada, Wakita, Ueshima ve Noguchi (2012)'nin geliştirdikleri algoritma kullanılarak, dereceleme ölçeklerinde kategori sayısının (dörtlü, beşli, altılı, yedili ve sekizli) ve yetenek dağılımı özelliklerinin (standart normal dağılım, sağda-solda normal dağılım, basık-sivri normal dağılım, normal beta dağılımı, sağa-sola çarpık beta dağılımı, basık-sivri beta dağılımı) farklılaşmasının ölçekteki tepki seçenekleri arasındaki uzaklığı nasıl değiştirdiği incelenmiştir.Söz konusu özelliklere sahip veriler WinGen3 yazılımı kullanılarak üretilmiştir. Çalışma kapsamında, bireylerin yetenek dağılımına ilişkin tüm durumlar için güvenirlik katsayılarının kategori sayısından bağımsız olduğu belirlenmiştir. Hem yetenek dağılımının, hem de kategori sayısının farklılaştırıldığı tüm durumlarda kategoriler arası psikolojik uzaklıkların eşitliği varsayımının tam olarak sağlanamadığı ve her bir kategoride kategori değerinden sapmalar olduğu belirlenmiştir. Kategori değerinden sapma miktarı dağılımın özelliklerine göre farklılaşmaktadır. Ancak, kategori sayısının az olduğu durumların, kategoriler arası psikolojik uzaklıkların eşitliği varsayımının sağlanmasına daha yakın durumlar olduğu saptanmıştır.

Anahtar Kelimeler

Dereceleme Ölçekleri, Eşit Aralıklılık, Kategori Sayısı, Psikolojik Uzaklık

*Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri BölümüÖlçme ve Değerlendirme Anabilim Dalı. İletişim: gizemuyumaz@akdeniz.edu.trTel: 02422274400 / 4645

**Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: sırgancı@ankara.edu.tr Tel: 0312 3633350 / 3105

*** Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümü. İletişim: cokluk@education.ankara.edu.tr Tel: 0312 3633350 / 3105

Paper No: 118

Examination of The Psychological Distance Between Categories In Differences Of Category Number And Ability Dispersion

Abstract

The measurement scales are assuemd as equal interval scale in order to make high order mathematical calculations while they are mostly in grading scale. But, the unequality of the psychological distance of categories could be resulted in not only fail to determine the differences of examinees but also the doubt in the usage of findings and the reliability of the decision related to scale. Indeed, reliability is another psychometris speciality that will effect the accuracy of the results. Accordingly, in this study, it is analysed whether the assumption, that the equilavance of psychological distance between the categories is provided in the application, and than, how the changes in ability dispersion characteristics and the number of categories effect this assumption The Generalized Partial Credit model (GPCM) based on IRT is used in order to determine the psychological distance between categories. The study is a work of smulation that examines how to change psychological distance between the options response under various conditions. In the study it was examined that how differantiation of when number of categories of the scale (Quartet, Quintet, Sextet, Sevens and eights) and ability distrubition characteristics according to normal distrubitions (the standard normal, right-left normal, low-pointed) and beta distrubitions (normal, right-left twisted, flattened-pointed)can change psychological distances of categories.It was calculated based on an algorithm as developed by Wakita, Ueshima, and Noguchi (2012). These properties have been produced using the WinGen3 software data. When the results were analyzed, it was found that reliability coefficients were independent from the number of categories. In all cases that both ability distrubitions and number of categories are different, the assumption of equality in the psychological distance between categories could not be provided exactly and it was also found that the deviation value of each category in the category. The amount of the deviation from the value of category varies according to the distribution of the category. But if there is less number of categories,it shows that it is closer to secure the assumptions.

Key words

Rating Scales, Equivalent İnterval, Number of Categories, Physhological Distance

Bildiri No: 119

Küçük Örneklemlerde Ortak Maddeli Eşdeğer Olmayan Gruplar Deseni Altında Farklı Eşitleme Yaklaşımlarının Karşılaştırılması

Vildan BAĞCI*, Ayşenur ERDEMİR**, Derya ÇAVDAR***

Özet

Çalışmanın amacı, küçük örneklemlerdeki eşitleme yöntemlerinin doğruluğunun ortak maddeli eşdeğer olmayan gruplar deseni altında incelenmesidir. Bu amaçla farklı sayılardaki küçük örneklemler (50, 100, 150, 200) için eşitleme çalışmaları yapılmıştır. Araştırmada TIMMS 2011 sekizinci sınıf matematik alt testine ait 19 ortak maddeye sahip olan sekizinci ve dokuzuncu kitapçıklar kullanılmıştır. Sekizinci kitapçığı alan 494, dokuzuncu kitapçığı alan 502 kişi bulunmaktadır. Sekizinci kitapçık X Formu ve dokuzuncu kitapçık Y Formu olarak ele alınmıştır. X formu puanlarının Y formu puanlarına eşitlendiği çalışmada, her bir örneklem büyüklüğü için ortalama eşitleme, lineer eşitleme, circle-arc eşitleme yaklaşımları altında farklı yöntemlerin (Tucker-lineer, Tucker-ortalama, Tucker-circle, Levine-lineer, Levine-ortalama, Levine-circle, Braun-Holland lineer, Braun-Holland ortalama, Braun-Holland-circle, nominal weights-ortalama, nominal weights-circle, eşit yüzdelikli zincirleme, lineer-zincirleme, ortalama-zincirleme, zincirleme-circle, frekans kestirim ve Kernel eşitleme) uygunluğu karşılaştırmalı olarak incelenmiştir. Çalışmadaki eşitleme sonuçlarının karşılaştırılmasında; eşitlenmiş puanların ortalama standart hata kestirimleri kriter alınmıştır. Araştırma sonuçlarına göre örneklem büyüklüklerinin tamamı için, tüm eşitleme yöntemlerinde, circle-arc yaklaşımı en doğru sonucu vermiştir. Özellikle Tucker-circle ve Levine-circle eşitlemelerinin tüm örneklem büyüklüklerinde, en küçük standart hataya sahip oldukları ve birbirlerine oldukça yakın eşitleme sonuçları ürettikleri görülmüştür. Eşit yüzdelikli zincirleme, Braun-lineer, Levine-lineer ve frekans kestirimine dayalı eşit yüzdelikli eşitleme yöntemleri ile yapılan eşitlemeler sonucu ise daha fazla hata değerleri elde edilmiştir. Bu bulgular doğrultusunda, ortak maddeli eşdeğer olmayan gruplar deseni için, eşit yüzdelikli ve lineer eşitlemeye dayalı yaklaşımların daha hatalı kestirimler yaptığı; daha az hataya sahip olan ortalama ve circle-arc eşitleme yaklaşımlarının ise küçük örneklemler için daha uygun olabileceği sonucuna varılmıştır.

Anahtar Kelimeler

Ortak Maddeli Eşdeğer Olmayan Gruplar Deseni, küçük örneklemlerde eşitleme, circle-arc

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: vildanbagci@gmail.com Tel: +905078101399

**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:erdemiraysenur@gmail.com Tel: +905057753840

***Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:cavdarderya@yandex.com Tel: +905543572810

Paper No: 119

Comparison Of Equating Methods For Small Samples In A Common Item Non-Equivalent Groups Design

Abstract

This study aims to investigate the accuracy of equating methods in small samples in a common item non-equivalent design. For this purpose, test equating was conducted for different small sample sizes (50, 100, 150 and 200). TIMSS 2011 mathematic achievement test data (eighth and ninth booklets) of Turkey were used. Each test has 34 items and 19 of them are common items which mean they are the same for both booklets. 494 students took the 8th form named Form X and 502 students took the 9th booklet named Form Y. Form X scores were equated to Form Y scores for all sample size separately. For each equating methods (Tucker, Levine, Braun-Holland, Nominal weights, Chained, Frequency estimation ad Kernel) with each related equating types (equipercentile, linear, mean and circle-arc) were analysed. These are Tucker linear, Tucker mean, Tucker circle-arc, Levine linear, Levine mean, Levine circle-arc, Braun-Holland linear, Braun-Holland mean, Braun-Holland circle-arc, nominal weights mean, nominal weights circle, chained equipercentile, chained linear, chained mean, chained circle-arc, frequency estimation and Kernel equating. Accuracy was described in terms of standard error of equating. According to the results, for all sample size and across all methods, circle-arc equating type created more accurate results than others. Especially it was found that Tucker circle-arc and Levine circle-arc methods had the least standard errors of all. However, equating with chained equipercentile, Braun-Holland linear, Levine linear and frequency estimation methods produced more standard errors than others. Accordingly, in a common item non-equivalent design, linear and equipercentile equating types had higher standard errors that others; mean and circle-arc equating types estimated equated scores more accurately. Therefore, it is suggested that mean equating and especially circle-arc equating types would be favourable for small sample test equating in common item non-equivalent design.

Key words

Test equating with small samples, common item non-equivalent design, circle-arc

Bildiri No: 122

Performans Takip Sınavında (PTS) Kullanılan Alt Testlerin Psikometrik Özelliklerinin İncelenmesi

Dilek GÜR*, Özge GÖÇER**, Evrim YALÇIN***

Özet

Bu çalışmanın amacı 2015-2016 eğitim- öğretim yılı bahar dönemi içerisinde Antalya ili kapsamında 8.sınıfa devam eden öğrenciler için yapılan Performans Takip Sınavında (PTS) kullanılan çoktan seçmeli maddelerin psikometrik özelliklerini saptamak ve aynı zamanda maddelerin seçildiği konu alanlarının zorluk derecelerini belirlemektir. Bu çalışma PTS'de kullanılan alt testlerin psikometrik özellikleri incelendiği için bir tarama araştırmasıdır. Çalışmanın evrenini PTS kapsamında toplanmış olan 30800 8.sınıf öğrencisine ait veriler oluşturmaktadır. Örneklem seçerken basit tesadüfi yöntem kullanılmış ve seçilen 380 öğrenciden elde edilen veriler üzerinden madde ve test istatistikleri yapılmıştır. Yapılan analizler sonucunda matematik alt testindeki üslü sayılar konusunun diğer konulara oranla öğrencilere daha zor geldiği görülmüştür. Türkçe alt testinde noktalama işaretlerinin, fen ve teknoloji alt testindeki asit-baz, kalıtım, kimyasal tepkimeler ve periyodik tablo, din kültürü alt testinde ise zekat, kader ve sadaka, inkılap tarihi testinde milli bilincin uyandırılması, liderlik yolunda adımlar, cumhuriyetin ilanı, menemen olayı ve dayanışma günleri ve İngilizce testinde yer alan communication konu alanlarının diğer konulara oranla öğrencilere daha zor geldiği belirlenmiştir. Ayrıca testlerde kullanılan maddelerin ayırt etme gücünü belirlemek amacıyla çift serili korelasyon katsayıları (Rbis) hesaplanmış ve genellikle .40 üzerinde değer almaları nedeniyle testlerin ayırt etme gücü yüksek iyi maddelerden oluştuğunu söylemek mümkündür. Test istatistikleri sonucunda elde edilen standart sapma değerleri incelendiğinde tüm testlerde öğrencilerin maddelere verdikleri cevaplar arasında değişimin fazla olduğu belirlenmiştir. Testlerin ortalama güçlük değerleri incelendiğinde öğrencilere en zor gelen testin matematik testi olduğu ve din kültürü testinin daha kolay geldiği belirlenmiştir. İç tutarlılık katsayısı olarak değerlendirilen güvenirlik katsayısı KR20, alt testler için hesaplandığında değerlerin .70 üzerinde olduğu görülmüştür. Güvenilir bir başarı testi için elde edilen değerlerin yeterli olduğu söylenebilir.

Anahtar Kelimeler

Performans Takip Sınavı (PTS), Çift serili korelasyon katsayısı (Rbis), KR20

*Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü.İletişim: dilekgur.dilek@gmail.com
**Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: ozgegocer07@hotmail.com
*** Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: yalcin_evrim@hotmail.com

Paper No: 122

Examining psychometric features of subtests used in Performance Follow-Up Exams (PFE)

Abstract

The aim of this study is to determine the psychometric features of multiple-choice items and level of difficulties of the fields from which items were chosen in Performance Follow-Up Exams (Performans Takip Sınavı - PTS) that was implemented for 8th grade students in Antalya. In this study, survey research method was preferred because of examination of psychometric features of sub-tests, which were used in PTS. Target population of the study is consisting of datas, which are belonging to 30.800 8th grade students who entered PTS. While choosing target population, simple random technique was implemented, and items and test statistics were produced from datas gathered from 380 students. In respect to the results of the analysis, the subject of powered numbers in sub-tests of Mathematics is more difficult for students than other subjects. Some topics are seen more difficult for students such as in Turkish sub-test punctuation, in science sub-test acid-base, heredity, chemical reactions and periodic table, in moral arts sub-test zakat, faith and alms-giving, in revolution history sub-test raising the national awareness, steps in leadership, declaration of republic, Menemen occasions and unity days and in English sub-test, communication. Also two series correlation coefficient (the Rbis) were calculated to select the items those were used in the test and usually they have degree over .40; according to these datas it can be said that these are good items with their ability to distinguish. When the standard deviation figures were analyzed, a significant variation among the answers of the students was seen. When the average difficulty level was reviewed, it was seen that students are having difficulty in mathematics exam most and find moral art easier to solve. When the reliability coefficient which is considered as internal consistency coefficient (KR20) was calculated for the sub-testes, the values were seen to be above .70. These values that are obtained may be considered sufficient for a reliable test.

Key Words

Performance Follow-Up Exams (PFE), Two Series Correlation Coefficient (Rbis), KR20

Paper No: 123

Many small or one large study? - Detection of effects and a prevention to the file drawer problem

Can GÜRER

Abstract

We take a look at the file drawer problem which arises when only certain studies are reported and others, mostly non-significant ones, are not. The file drawer is fueled by researchers doing many small studies instead of a single large study, believing that the chance for getting significant results is higher with many studies. This is true under the assumption that no effect is present, which seems rarely appropriate. We show that in the presence of an effect, the power of a single combined large study can by far outweigh the probability for at least one significant effect with several small studies. These advantages are shown to be most common for situations with a small effect and strict alpha-levels, e.g. .001. We conclude that researchers considering this can improve their chances of significant findings easily, and that editors can help prevent the file drawer by demanding strict levels of significance.

Key Words

Power, file drawer problem, crisis of replicability, null hypothesis significance testing (NHST), sample size

Bildiri No: 125

Yönlendirmeye ve Biçimlendirmeye Dayalı Çok Aşamalı Bilgisayarlı Testlerde Yetenek Kestirimlerinin Karşılaştırılması

Murat Doğan ŞAHİN*, Murat AKYILDIZ**

Özet

Klasik Test Kuramına (KTK) dayalı geleneksel testlerde, tüm yanıtlayıcılar testin tamamını cevaplamak zorundadırlar. Bu durum testi alanların kendi yetenek düzeylerine uymayan soruları da yanıtlaması mecburiyetini doğurmaktadır. MTK'nın kağıt-kalem sınavlarından ziyade bireyselleştirilmiş bilgisayarlı testlerde (BBT) kullanılması çok daha az maddeyle, daha dakik (pure) yetenek kestirimleri elde edilmesini sağlar. BBT uygulamalarında, her bir maddenin yanıtlaması sonrası yeni bir yetenek kestirimi ve bu yetenek düzeyine en uygun maddenin havuzdan çekilerek bireye sunulması söz konusudur. Bu nedenle BBT'ler, geniş yetenek ranjına hitap eden çok sayıda maddeden oluşan bir havuza ihtiyaç duyarlar. BBT'lerdeki bu sınırlılıklar nedeniyle son dönemde Çok Aşamalı Bilgisayarlı Testlerin (ÇABT - Multistage Testing by Routing- MST-R) kullanımı yaygınlaşmaktadır. Araştırmacılar, ÇABT'nin geleneksel testler ve BBT'lerin birçok avantajını bir araya getirirken bir çok dezavantajı ise minimize ettiğini ortaya koymuşlardır (Jan, Lewis &von Davier, 2014). Bir ÇABT uygulamasında önceden bir araya getirilmiş madde grupları vardır. Bu madde gruplarına "modül" adı verilir. Bireyler, öncelikle bir sonraki aşamada yanıtlayacakları modülü belirleyecek ortak bir modülü yanıtlarlar. Geleneksel ÇABT sürecinde her modülün test bilgi fonksiyonu kendine özgüdür ve ilgili yetenek grubu için maksimize edilmiştir. ÇABT uygulaması sırasında nihai modül kararından önce, gerçekte bulunduğu yetenek düzeyinin bir üst ya da altındaki modüle yönlendirilen birey için test bilgi fonksiyonu, yetenek düzeyi uyumsuzluğu nedeniyle yetenek kestiriminin standart hatasının büyümesine yol açabilir. Bunu engellemek amacıyla Çok Aşamalı Bilgisayarlı Test-Biçimlendirmeyle (Multistage Testing by Shaping, MST-S) ismi verilen yeni bir yöntem önerilmiştir (Han & Guo, 2013). Bu çalışmada Anadolu Üniversitesi YDYO yeterlilik sınavında uygulanmış olan 100 kelime bilgisi sorusu kullanılacaktır. MST-R yöntemi 1-3-3 yönlendirme formatında olacak şekilde düzenlenecektir. MST-S yöntemi için 5,10 ve 50 iterasyon için sonuçlar elde edilecektir. İki yöntemden elde edilen yetenek kestirimlerinin standart hataları ve koşullu ortalama mutlak hata değerleri karşılaştırtılacaktır.

Anahtar Kelimeler

Çok Aşamalı Bilgisayarlı Test, Yetenek Kestirimi, Test Bilgi Fonksiyonu

* Anadolu Üniversitesi Açıköğretim Fakültesi Test Araştırma Birimi. İletişim: mdsahin@anadolu.edu.tr
** Anadolu Üniversitesi Açıköğretim Fakültesi. İletişim: muratakyildiz @anadolu.edu.tr

Paper No: 125

Comparing the Ability Estimates of Multistage Testing by routing (MST-S) and Multistage Testing by Shaping (MST-R)

Abstract

In traditional tests based on Classical Test Theory (CTT), test takers have to reply all the items in the test. In this situation, test takers have to reply even the items that are out of their ability range. The use of IRT in Computer Adaptive Tests (CAT) rather than in paper-pencil tests enables educators to make pure ability predictions by using fewer number of items. In CAT applications, when a test taker replies an item, a new ability prediction is made and the most appropriate item for this new prediction is provided for the test taker in the next phase. Thus, CATs require a large size item pool prepared for a wide range of abilities. Due to such limitations of CATs, the use of Multistage Testing by Routing (MST-R) has been more common nowadays. Researchers report that MST-R combine the advantages of traditional tests and CATs and minimizes certain disadvantages (Jan, Lewis &von Davier, 2014). Prior to an MST-R administration, a number of item groups are formed, which are called “module”. Test takers first reply a common module that will determine the module they will reply in the next phase. In traditional MST-R process, test knowledge function of each module is unique and is maximized for the related ability group. During MST-R practice, test knowledge function for the test taker who is directed to one level higher or lower level than his real level before the final level decision can lead to high standard deviation for ability prediction. In order to avoid this problem, a new method has been suggested, which is called Multistage Testing by Shaping, MST-S. In this study, a total number of 100 vocabulary items previously asked in Anadolu University School of Foreign Languages proficiency tests will be used. MST-R method will be organized as 1-3-3 model. 5,10 and 50 iteration results will be obtained for MST-S. The ability predictions obtained from these two methods will be compared in terms of conditional standard errors and conditional mean absolute errors for the final theta estimate.

Key Words

Computerized multistage testing, ability estimate, test information function

Bildiri No: 126

Madde Tepki Kuramı'na Dayalı Madde Uyum İndekslerinin I. Tip Hata ve Güç Oranlarının Çeşitli Değişkenler Açısından İncelenmesi

Seçil ÖMÜR SÜNBÜL, Semih AŞİRET

Özet

Bu çalışmada Madde Tepki Kuramı'na göre ikili puanlanan ve 1, 2 ve 3 parametreli lojistik modellere uygun olarak üretilen maddelerde çeşitli madde-uyum indekslerinin çeşitli koşullardaki (örneklem büyüklüğü, test uzunluğu ve uyumsuzluk (misfit) yüzdesi) I. tip hata ve güç oranlarının incelenmesi amaçlanmıştır. Bu çalışmada, indekslerin I. tip hata ve güç oranlarının belirlenmesi simülasyon çalışmasıyla yapılmıştır Çalışmada, madde uyumu için geleneksel indekslerden χ^2, $Q1$ ve G^2 indeksleri ile alternatif indekslerden $S-\chi^2$ indeksi kullanılmıştır. Çalışmada yer alan dört farklı madde uyum indeksinin I. tip hata ve güç oranları, örneklem büyüklüğü (1000, 2000, 4000), test uzunluğu (20, 30, 40) ve uyumsuzluk yüzdesi (%0, %10, %30 ve %50) değişimlenerek incelenmiştir. Veriler R 3.3.1 yazılımı kullanılarak üretilmiştir ve "mirt" paketi kullanılarak analiz edilmiştir. Çalışmada iki tür model kullanılmıştır. Üretilen model (GM) için, veriler 2PL ve 3PL modellere göre üretilmiştir. Analiz edilen modellerde (CM) ise, veriler 2PL model için 2PL (GM=CM) ve 1PL modele (GM>CM) ve 3PL model için 3PL (GM=CM), 2PL-1PL (GM>CM) modellere göre üretilmiştir. Üretilen modele uygun madde tepkileri ile analiz edilen modele uygun madde tepkileri için R 3.3.1 programı kullanılarak, madde uyum indekslerinin p değerleri ve serbestlik dereceleri hesaplanmıştır. Eğer uyum indeksinin p değeri 0.05'den küçük ise madde modele uyumsuz olarak yorumlanmıştır. Uyum indeksleri I. Tip hata ve güç oranlarına göre değerlendirilmiştir. Uyum indekslerinin I. tip hatalarını değerlendirmek için 0.05 anlamlılık düzeyi kullanılmıştır. Her uyum indeksinin tüm koşullardaki I. tip hata ve güç oranları hesaplanarak bu indeksler karşılaştırılmıştır. Çalışma sonucunda, tüm faktörlerde $S-\chi^2$indeksinin diğer indekslere göre daha düşük hataya sahip olduğu görülmüştür. 2000 ve üzeri örneklem büyüklüğünde ve 20 ve daha uzun maddeden oluşan testlerde $S-\chi^2$indeksi diğer indekslerden daha düşük I. tip hata oranına ve daha yüksek güce sahip olduğu görülmüştür.

Anahtar Kelimeler

Madde Uyum, $S-\chi^2$, Madde Tepki Kuramı, Ki-kare

Paper No: 126

Investigating Type 1 Error and Power Rates of Item Fit Indices Based on Item Response Theory Through Various Factors

Abstract

In this study, it was aimed to investigate type I error and power rates of the item fit indices through various conditions (sample sizes, different test lengths and different magnitudes of misfit) for dichotomously generated items based on one-, two-, and three-parameter logistic models in Item Response Theory. In this study, the type I error and power rates of these item fit indices were assessed in a simulation study.χ^2, $Q1$ and G^2 indices as traditional item fit indices and S-X2 index as an alternative indices were assessed. The performance of four different item fit indices in study were compared my manupulating three different logistic models(1PL, 2Pl, 3PL), three different sample size (1000, 2000, 4000), three different test lengths (20, 30, 40) and four different misfit magnitude (%0, %10, %30 and %50). Item responses were generated using the R 3.3.1 software program and analyzed by using "mirt" package in R software . In the study, item responses were generated based on 2PL and 3PL models for generated model (GM). Analysis model (CM) included 2PL (GM=CM) and 1PL model (GM>CM) for 2PL generated model and 3PL (GM=CM), 2PL-1PL (GM>CM) models for 3PL generated model. The p value of item fit indices and their degrees of freedom were calculated by using R 3.3.1 software program for both item responses fit for generated model and the item resonses fit for analysis model. When the p value of the index was smaller than 0.05, it was interpreted as the item was misfit to the model. Item fit indices were evaluated according to their type I error and their power rates. Significance levels of .05 was used to examine the type I error rates of the indices. All item fit indices in this study were compared by calculating the type I error and power rates of each item fit indices under all conditions. The findings indicated that $S-\chi^2$index yields lower type I error to detect misfit than the other indices. It can be concluded that in the case where the sample size was 2000 or more and the number of items in test are 20 and more, $S-\chi^2$ index has lower type I error rates than traditional indices and has adequate power to detect misfit items.

Key Words

Item Fit, $S-\chi^2$, Item Response Theory, Chi-Square

Bildiri No: 127

Antalya İli Performans Takip Sınav Puanları İle Yüksek Öğretime Geçiş Sınav Puanları Arasındaki İlişkinin İncelenmesi

Şerife ZEYBEKOĞLU*, Şevki Yetkin ODABAŞI**

Özet

Bu araştırmanın amacı 2015-2016 eğitim öğretim yılı içerisinde Antalya ilinde yapılan Performans Takip Sınavları (PTS) ile Yüksek Öğretime Geçiş Sınavları (YGS) arasındaki ilişkinin incelenmesidir. Bu çerçevede öğrencilerin PTS ve YGS alt testleri olan Türkçe, Sosyal Bilgiler, Matematik ve Fen Bilgisi testlerine ait netler ile YGS puan türleri incelenmiştir. Öncelikle, her alt test için ayrı ayrı PTS ve YGS netleri arasında anlamlı bir ilişkinin olup olmadığı incelenmiştir. Daha sonra altı faklı puan türünde PTS ve YGS puanları arasında anlamlı bir ilişki olup olmadığı incelenmiştir. Ayrıca yine altı farklı puan türünde PTS puanlarının YGS puanlarının anlamlı bir yordayıcısı olup olmadığı test edilmiştir. Araştırmailişkisel taramatüründedir. Araştırmanın evrenini Antalya ilinde öğrenim gören PTS ve YGS sınavlarına katılan öğrenciler oluşturmaktadır. Araştırmanın örneklemini ise tabakalı örnekleme yöntemi ile seçilen 6 lisede öğrenim gören 350 öğrenci oluşturmaktadır. Elde edilen sonuçlara göre PTS ve YGS alt testlerine ait netler arasında istatistiksel olarak pozitif ve yüksek bir ilişki bulunmuştur. En yüksek ilişkinin matematik dersinde en düşük ilişkinin ise fen bilgisi dersinde olduğu belirlenmiştir. PTS ve YGS puanları arasındaki ilişki incelendiğinde altı farklı puan türünde de pozitif ve yüksek bir ilişki bulunmuştur. En yüksek ilişkinin YGS-1 puan türünde, en düşük ilişkinin ise YGS-4 puan türünde olduğu belirlenmiştir. Ayrıca PTS puanlarının YGS puanlarının anlamlı bir yordayıcısı olduğu belirlenmiştir.

Anahtar Kelimeler

PTS, YGS, Korelasyon, Regresyon

* Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü.Taşlıca Ahmet Hamdi Akseki Ortaokulu. İletişim: serifezeybekoglu79@gmail.com

**Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. Kepez İsmet İnönü Mesleki ve Teknik Anadolu Lisesi. İletişim: syetkin.odabasi@gmail.com

Paper No: 127

An Investigation of the Relationship Between the Transition to Higher Education Examination (YGS) Scores and Performance Monitoring Examination (PTS) Scores in Antalya

Abstract

This study examines the relationship between the Transition to Higher Education Examination (Yüksek Öğretime Geçiş Sınavı – YGS) and Performance Monitoring Examination (Performans Takip Sınavı – PTS) those were applied in 2015-2016 school year in Antalya. To do so the raw scores of the subtests that are Turkish, Social Studies, Mathematics and Science, and the different six score types of that are YGS-1 to YGS-6 were investigated. First, the correlations between raw scores of YGS and PTS were checked for each subtest. Then the correlations between PTS and YGS score types were examined. Moreover whether the scores of PTS are significant estimator of the scores of YGS or not is tested for each single score types. As a research design, correlational research was used in this study. The interest of population in this study is the students who entered the both PTS and YGS in Antalya. The sample of this study is 350 students who are studying six different high schools in Antalya that were selected by stratification method. The results showed that there is a strong positive relationship between the raw scores of the subtests of PTS and YGS. The strongest relationship is emerged in Mathematics subject and the weakest relationship is determined in Science subject. In addition to this, when the correlation between PTS and YGS score types, it seemed that the whole six score types have a strong positive relationship. The strongest relationship is emerged in YGS-1 score type, and the weakest relationship is determined in YGS-4 score type. Furthermore, it is revealed that PTS scores are significant estimator of YGS scores.

Key Words

PTS, YGS, Correlation, Regression

Bildiri No: 128

Çapraz Yüklerin Sıfırdan Farklı Olduğu Durumlarda Maksimum Olabilirlik ve Bayesçi Doğrulayıcı Faktör Analizi Sonuçlarının Karşılaştırılması

Derya ÇAVDAR, Ergün Cihat ÇORBACI, Vildan BAĞCI, Ayşenur ERDEMİR

Özet

Ölçme araçlarında yer alan maddeler bağlı olduğu faktörlerin dışında diğer faktörlerden de bir kısım bilgiler içerebilmektedir. Bu durum, analizler sırasında maddelerin ana yüklerinin yanında çapraz yüklerinin de gözlenmesine sebep olmaktadır. Ancak, sıklıkla kullanılan Maksimum Olabilirlik doğrulayıcı faktör analizleri çapraz yüklerin 0'a eşit olması varsayımını içermektedir. Pratikte bu durumun karşılanması zor olduğundan çapraz yükler için ön bilgilerin tanımlandığı Bayesçi doğrulayıcı faktör analizi iyi bir alternatiftir. Bu doğrultuda araştırmanın amacı, çapraz yüklerin 0'a eşit olduğu ve olmadığı durumlarda ML doğrulayıcı faktör analizi ile Bayesçi doğrulayıcı faktör analizi sonuçlarını karşılaştırmaktır. Bunun için, araştırmada her faktör için üçer maddenin bulunduğu 2 faktörlü ve 6 maddeli bir desen oluşturulmuş ve buna göre Monte Carlo simülasyonu kullanılarak veri üretilmiştir. Örneklem büyüklüğü olarak n = 200 alınıştır. Maddelere ait ana yükler 0.8 ve çapraz yükler 0.1, 0.2 ve 0.3 olacak şekilde tasarlanmıştır. Faktör varyansları 1 ve faktörler arası korelasyon 0.3 olacak şekilde veri üretilmiştir. Analizler 4 aşamada gerçekleştirilmiştir. İlk aşamada (Analiz 0), yalnızca ana yüklerin yer aldığı çapraz yüklerin bulunmadığı durum için ML ve Bayes doğrulayıcı faktör analizleri yapılmıştır. Ardından, analiz 1, 2 ve 3 için çapraz faktör yüklerinin 0.1, 0.2 ve 0.3 olduğu durumlar ve çapraz yüklere sahip madde sayısının arttığı durumlar için doğrulayıcı faktör analizleri yapılmıştır. Böylece, hem çapraz yük sayısının artışı hem de çapraz yük değerlerinin artışının analiz sonuçlarına etkisi incelenmiştir. Sonuç olarak, çapraz yüklerin sıfıra eşit olduğu durumlarda ML ve Bayesçi doğrulayıcı faktör analizlerinin benzer sonuçlar ürettiği ancak modele çapraz yükler eklendikçe ve çapraz yük değerleri arttıkça ML analizleri sonucunda model reddedilme oranlarının arttığı görülmüştür. Ancak, Bayesçi doğrulayıcı faktör analizleri sonucunda modelin reddedilme oranı çapraz yük sayısının ve değerlerin artışından etkilenmemiştir ve tüm analiz sonuçlarında oldukça düşük olduğu görülmüştür.

Anahtar Kelimeler

Maksimum Olabilirlik, Bayesçi, doğrulayıcı faktör analizi

Paper No: 128

Comparison between Results of Maximum Likelihood and Bayesian Confirmatory Factor Analyses when Cross Loadings are Different from Zero

Abstract

Items of an instrument might have major loadings as well as cross loadings because items could include some information about other factors. However, confirmatory factor analyses using Maximum Likelihood, which are widely used, are very restrictive since they assume that cross loadings are fixed at zero. Since this restriction does not always hold true in practice, Bayesian analysis is a good alternative as it uses the information about distribution of parameters. Bayesian analysis can combine the information that cross loadings are not exactly zero but approximately zero so that model fit is not influenced by violation of assumptions. Therefore, the purpose of this study is to compare the results of Maximum Likelihood and Bayesian confirmatory factor analyses when cross loadings are exactly 0 and different from zero. In order to achieve this aim, a Monte Carlo simulation study was conducted. In the simulation, 2 factors and 6 items, 3 for each factor, were used. Sample size was taken as 200. Moreover, main loadings were fixed at 0.8 and cross loadings were taken as 0.1, 0.2 and 0.3. Factor variance was 1 whereas correlation between factors was 0.3. Analyses were conducted in 4 steps. In the first step (Analysis 0), the model with no cross loadings was tested. Within following steps (Analysis 1, 2 and 3), one, two and three cross loadings were added to the model by increasing magnitude of cross loadings. To do so, impact of both number of cross loadings and magnitude of them were examined. As a result of Analysis 0, outcomes of ML and Bayesian confirmatory factor analyses were very similar. In other analyses (Analysis 1, 2, 3), the results indicated that model rejection rate for ML confirmatory factor analyses increased and it was low and stable for Bayesian confirmatory factor analyses when number and magnitude of cross loadings increased.

Key Words

Maximum Likelihood, Bayesian, confirmatory factor analysis

Bildiri No: 130

Likert Tipli Envanterler Nasıl Puanlanmalı?

Sungur GÜREL[*], Hakan ATILGAN[**], Burak AYDIN[***], Mehmet KAPLAN[****]

Özet

Eğitimsel araştırmacılar tarafndan sıkıkla başvurlan Likert tipli soru içeren envanterler ile toplanan verilerin istatistiksel analizi konusunda görüş ayrılkları mevcuttur. Ancak sıralndirma seviyesinde ölçüm yapabilen Likert tipli maddelerin eşit aralıklı kabul edilip analiz edilmesinin yanı sıra verilerin sıralı doğasını da göz önüne alan seçenkler mevcuttur. Bu araştırmanın amacı, bir gerçek veri seti üzerinde çeşitli puanlamaları karşılaştırarak, eğitimsel araştırmacıları alt özellik puanı hesaplama yönteminin olası etkileri konusunda bilgilendirmektir. Toplama yöntemi ve doğulayıcı faktör analizi ile elde edilen alt boyu puanları arasıda yüksek korelasyon mevcut olup, doğrulayıcı faktör analizi ile elde edilen alt boyu puanları ile anlamlı ilişki tespit etmek daha kolaydır ve bu nedenle tercih edilebilir. Son olarak, doğrulayıcı faktör analizi yerine toplama yöntemi ile doğrulayıcı faktör alizine çok yakın sonuçlara ulaşılabilmiştir.

Anahtar Kelimeler

Likert, Dogrulayici Faktor Analizi, Siralma, Esit Arlik, Envanter

[*] Siirt Üniversitesi Egitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: sungur5@hotmail.com
[**] Ege Üniversitesi Egitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: atilgan.h@gmail.com
[***]Recep Tayyip Erdogan Üniversitesi Egitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: burak.aydin@erdogan.edu.tr Tel: +90 (464) 532 8454 / 2329
[****]Milli Egitim Bakanligi. İletişim: mehmet.kapan2@gmail.com

Paper No: 130

How Likert Type Inventories are Supposed to Be Scored?

Abstract

There are differences among statistical analysis of data that are gathered using Likert Type inventories that are frequently employed by educational statisticians. In addition to the analysis procedures that assume Likert type data is interval data rather than ordered data, there are statistical procedures that address the ordered nature of the Likert type data properly. The purpose of this study is comparing various scoring procedures for Liker type data and informing educational statisticians about apparent consequences of using each procedure. Correlations among factor scores that are obtained using summation method and confirmatory factor analysis are high but concluding significant relationship is relatively easier using factor scores that are obtained via confirmatory factor analysis, and therefore, confirmatory factor analysis is preferable over summation method. Finally, results that are obtained using summation method are really similar to the results that are obtained using confirmatory factor analysis.

Key Words

Likert, Confirmatory Factor Analysis, Ordered, Interval, Inventory

Bildiri No: 132

PISA 2012 Türkiye Örneklemi Yüksek ve Düşük Performans Gösteren Okulların KNN Analizi ve Kümeleme Analizi ile İncelenmesi

Halime YILIDIRIM*, Sinem ŞENFERAH**, Menekşe UYSAL***, Şener BÜYÜKÖZTÜRK****

Özet

Bu çalışmada, Uluslararası Öğrenci Değerlendirme Programı (PISA, Programme for International Student Assessment) 2012 Türkiye verilerine dayalı olarak düşük ve yüksek matematik performansı gösteren okulların, sosyoekonomik statü indeksi (ESCS) ve öğretmen özellikleri ile ilişkili formative değerlendirme (TCHBEHFA_1), öğretmen odaklı öğretim (Teacher-directed Instruction, TCHBEHTD_1) ve öğrenci yönelimli (Student Orientation, TCHBEHSO_1) değişkenlerine göre nasıl modelleneceklerinin ve kümeleneceklerinin KNN analizi ve kümeleme analizi ile belirlenmesi ve elde edilen sonuçların karşılaştırılması amaçlanmıştır. Bu çalışma, Türkiye örneklemi PISA 2012 uygulamasında düşük ve yüksek matematik performansı gösteren okulların belirlenen değişkenler bakımından nasıl gruplanacağının belirlenmesini amaçladığından betimsel düzeyde bir tarama çalışmasıdır. Çalışma, PISA 2012 Türkiye verileri kullanılarak 8. sınıf düzeyinde matematik başarı puanlarına göre düşük performans gösteren alt %10'luk dilimde kalan 17 okul ile yüksek performans gösteren üst %10'luk dilimde kalan 17 okul kapsamında gerçekleştirilmiştir. Alt ve üst yüzde onluk dilimlerle belirlenen 34 okul için düşük performans gösteren okulda 241 öğrenci ve yüksek performans gösteren okullarda ise 585 öğrenci olmak üzere toplam 826 öğrenci bulunmaktadır. Çalışmada KNN algoritması kullanılarak düşük ve yüksek performans gösteren okulların belirlenen değişkenlere göre modellenmesi ve okul performanslarının tahmin edilmesi amaçlanmış elde edilen bulgulara göre analize dâhil edilen ESCS, TCHBEHFA, TCHBEHTD ve THCHBEHSO değişkenlerinin yüksek ve düşük performans gösteren okulları 2 kümeye ayırdığı görülmüştür. Kümeleme analizi kapsamında ise hiyerarşik kümeleme yöntemlerinden "Tam bağlantı en uzak komşuluk" yöntemi ve uzaklık ölçütü olarak "Öklid uzaklığı" kullanılarak okulların belirlenen değişkenlere göre 2 kümeye ayrıldığı belirlenmiştir. Çalışmada, KNN analizi ve kümeleme analizinden elde edilen sonuçların benzer nitelikte olduğu görülmüş olup bu durum ilgili değişkenlerin yüksek ve düşük performans gösteren okulları ayrıştırmada etkili değişkenler olduğunu göstermektedir.

Anahtar Kelimeler

PISA, Matematik okuryazarlığı, KNN analizi, Kümeleme analizi

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:halime.yldrm@gmail.com Tel: +905078907991

**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: ssnfrh@gmail.com Tel: +905059359388

***Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:menekseysl@gmail.com Tel: +905057026029

**** Hasan Kalyoncu Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: senerbuyukozturk@gmail.com Tel: +905375543450

Paper No: 132

Investigation of High- and Low-Performing Schools in PISA 2012 (Turkey) through KNN and Cluster Analyses

Abstract

The present study aims to determine how high- and low-performing schools in mathematics according to PISA (*Programme for International Student Assessment*) 2012 Turkey data will be modeled and organized into clusters in line with the Index of Economic, Social and Cultural Status (ESCS), Teacher Behavior: Formative Assessment (TCHBEHFA), Teacher-Directed Instruction and Teacher Behavior: Student Orientation (TCHBEHSO) variables through KNN and cluster analyses and compare the obtained results. The study is a descriptive survey as it aims to determine how high- and low-performing schools in mathematics in PISA 2012 - Turkey Data will be grouped in terms of the specified variables. It was carried out across 17 low-performing schools at the bottom 10% and other 17 high-performing schools at the top 10% according to mathematics achievement scores at the 8th grade using PISA 2012 - Turkey Data. There were 826 students in total across 34 schools chosen from the top and bottom 10% - 241 of them were from low-performing schools while 585 students were from high-performing schools. In the study, it was aimed to model high- and low-performing schools according to the specified variables and to predict school performances using KNN algorithm. Besides, it was seen that the ESCS, TCHBEHFA, TCHBEHTD and THCHBEHSO variables included in the analysis divided the high- and low-performing schools into 2 clusters. Within the scope of the cluster analysis, it was found out that the schools were divided into 2 clusters according to the specified variables using the "Complete linkage (furthest neighbor)" method, which is a hierarchical clustering method, and "Euclidean Distance" as a distance measure. Lastly, it was seen that the results obtained from KNN and cluster analyses were similar. This situation demonstrated that the relevant variables were effective in differentiating between high- and low-performing schools.

Key Words

PISA, mathematics literacy, KNN analysis, cluster analysis

Bildiri No: 133

Farklı Örneklem Büyüklüğü ve Dağılımı Koşullarında WLS ve Robust WLS Yöntemlerinin Karşılaştırılması

Halime YILIDIRIM*, Menekşe UYSAL**, Sinem ŞENFERAH***, Şener BÜYÜKÖZTÜRK****

Özet

Yapısal eşitlik modellemesinde kullanılan parametre kestirim yöntemleri verinin sürekli, sıralı olup olmamasına ve dağılımın normalliğine göre farklılık göstermektedir. Sıralı verilerle çalışıldığında en sık kullanılan parametre kestirim yöntemi WLS (Ağırlıklandırılmış en küçük kareler) olup, dağılıma ilişkin herhangi bir varsayım gerektirmemesi avantajı iken büyük örneklemler gerektirmesi dezavantajı olarak karşımıza çıkmaktadır. Son yıllarda Robust WLS kestirim yöntemleri WLSM ve WLSMV sıklıkla kullanılmakla beraber, küçük örneklemlerde ve farklı dağılım koşullarında WLS kestirim yöntemine alternatif olup olamayacağı önemli görülmektedir. Bu çalışmada PISA 2012 öğrenci anketinin A ve B formunda yer alan matematiğe yönelik tutum maddelerinden; matematiğe yönelik ilgi, araçsal motivasyon, matematik özyeterliği, algılanan kontrol ve matematik çalışma etiği olmak üzere 5 faktör ve 28 maddeden oluşan model temel alınmıştır. Söz konusu maddelere ilişkin madde parametreleri MULTILOG 7.03 programı kullanılarak elde edilmiştir. Kestirilen madde parametreleri kullanılarak WinGen 3.1 simülasyon programıyla 3 tür dağılım için 25 replikasyonla 4 farklı büyüklükte (N=100, 200, 500 ve 1000) veri seti oluşturulmuştur. Bu veri setlerindeki çarpıklık katsayıları yaklaşık olarak 0,00; 1,00 ve 1,50 olan dağılımlardır. Bu çalışmada farklı örneklem büyüklüğü ve dağılımları koşullarında WLS, WLSM ve WLSMV yöntemlerinin performansları uyum indeksleri (RMSEA, CFI ve TLI) bakımından karşılaştırılmıştır. Farklı örneklem büyüklüğü koşulu altında WLSMV yönteminin WLSM ve WLS yöntemlerinden daha iyi uyum indeksleri ürettiği belirlenirken, özellikle küçük örneklem koşullarında WLS yöntemine bir alternatif olabileceği belirlenmiştir. Dağılımın çarpıklığına göre WLS, WLSM ve WLSMV kestirim yöntemleri incelendiğinde, dağılımın çarpıklığına karşı en az duyarlı kestirim yönteminin WLSMV olduğu, dağılımın çarpıklığı arttıkça WLS ve WLSM kestirim yöntemlerinin ürettiği uyum indekslerinin düştüğü fakat en az düşüşün WLSMV yönteminde olduğu görülmüştür.

Anahtar Kelimeler

WLS, WLSM, WLSMV, Parametre Kestirim Yöntemleri, Dağılım ve Örneklem Büyüklüğü, Simülasyon

*Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:halime.yldrm@gmail.com Tel: +905078907991

**Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:menekseysl@gmail.com Tel: +905057026029

*** Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: ssnfrh@gmail.com Tel: +905059359388

****Hasan Kalyoncu Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: senerbuyukozturk@gmail.com Tel: +905375543450

Paper No: 133

The Comparison of WLS and Robust WLS Methods in Different Sample Size and Distribution Conditions

Abstract

In structural equation modeling parameter estimation methods vary according to if data is continuous or ordinal and the normality of distribution. When working with ordinal data, the most commonly used parameter estimation method is WLS (weighted least square), the advantage of which is not requiring any assumption, as the disadvantage of it is not requiring large samples. Recently, while Robust estimation methods, WLSM and WLSMV, are commonly used, it is important to see whether they are alternative to WLS in different distribution and sample size conditions. In this study, it is based on the model with five factor including mathematics interest, instrumental motivation for mathematics, mathematics self-efficacy, perceived control, mathematics work ethic and 28 items about students' attitudes towards mathematics in A and B form of PISA student survey. Item parameters were estimated by using MULTILOG 7.03. Data sets with four different sample sizes (N=100, 200, 500 ve 1000) and 3 different distribution types were produced by using estimated item parameters in WinGen 3.1 simulation program with 25 replications. The skewnesses of distributions are about 0,00; 1,00, and 1,50. The performance of parameter estimation methods including WLS, WLSM, and WLSMV in different sample size with different distributions were compared according to fit indices (RMSEA, CFI and TLI). As a resultit was seen that WLSMV method has better fit indices than WLSM and WLS methods in different sample size conditions,, especially in small sample size condition it is an alternative to WLS method. When it was examined WLS, WLSM, and WLSMV estimation methods according to skewness of distribution, it was seen that the least sensitive method to skewness of distribution is WLSMV and when the skewness of distribution increases, WLS and WLSM methods' fit indices decreases, but the least decrease occurs in WLSMV method.

Key Words

WLS, WLSM, WLSMV, Parameter Estimation Methods, Distribution and Sample Size Conditions, Simulation

Bildiri No: 134

Madde Seti Büyüklüğü ve Eğim Parametresinin CAT Test Uzunluğuna Etkisi

Ahmet Salih ŞİMŞEK*, Ezel TAVŞANCIL**

Özet

Bu araştırmanın temel amacı çoklu puanlanabilen maddeler için kullanılan CAT uygulamalarında madde seti büyüklüğünün ve eğim parametresinin test uzunluğuna etkisinin incelenmesidir. Psikolojik özelliklerin ölçülmesi için yaygın olarak kullanılarak likert tipi maddeler çoklu puanlanabilen maddeler için tipik birer örnektir. Kağıt-kalem uygulamalı likert tipi ölçeklerde en az dört maddeden oluşmak koşuluyla ilgili boyutun ölçülebileceği belirtilmektedir. Ancak CAT uygulamaları için bu alt sınırın ne olduğu hakkında literatürde bir bilgi bulunmamaktadır. Bu araştırma kapsamında madde seti büyüklüğü için n_1=10, n_2=20, n_3=30 olacak şekilde üç koşul, madde setindeki maddelerin eğim parametresi için a_1=[0.65-1.00], a_2=(1.00-1.35), a_3=[1.35-1.70] olacak şekilde üç koşul belirlenmiştir. CAT uygulaması için FI madde seçme yöntemi, EAP kestirim yöntemi ve SH<.385 durdurma kuralı kullanılmıştır. Simülasyon çalışması R kütüphaneleri kullanılarak yapılmıştır. Verilerin üretilmesi için ltm (1.0) R kütüphanesi, CAT simülasyonları için catIrt (0.5) R kütüphanesi kullanılmıştır.

Simülasyon çalışmaları sonucunda CAT uygulaması test uzunlukları betimsel istatistikleri elde edilmiştir. Eğim parametresi a_1=[0.65-1.00] olduğu durumda n_1=10, n_2=20, n_3=30 için CAT test uzunlukları μ_{11}=10.00 (%100), μ_{21}=19.97 (%100), μ_{31}=27.68 (%92) şeklinde olmuştur. Eğim parametresi a_2=(1.00-1.35) olduğu durumda ise CAT test uzunluğu μ_{12}=9.99 (%100), μ_{22}=14.53 (%73), μ_{32}=13.60 (%45) şeklinde olmuştur. Son olarak eğim parametresi a_3=[1.35-1.70] olan maddeler için incelendiğinde CAT test uzunluğu μ_{13}=9.52 (%95), μ_{23}=9.10 (%45), μ_{33}=8.56 (%29) değerlerine düşmüştür. Araştırma kapsamında CAT yetenek kestirimleri ile tüm madde seti yetenek kestirimleri arasındaki korelasyon katsayıları .95 ile 1.00 arasında değişmiştir. Tüm maddelerin kullanıldığı CAT uygulamalarında korelasyon katsayıları 1 olurken, madde setinin %29'unun kullanıldığı durumda .95 olarak hesaplanmıştır.

Araştırmanın temel sonuçlarından biri eğim parametresi arttıkça CAT test uzunluğunun azalması olmuştur. Araştırma sonucunda, düşük eğim parametresine sahip maddelerden oluşan madde setlerinin kullanılmasının CAT uygulamasında ölçme işleminin kullanışlılığına katkı getirmediği görülmüştür. Bununla birlikte yüksek eğim parametresine sahip maddelerden oluşan madde setlerinin kullanılması CAT test uzunluğunu önemli ölçüde azaltmıştır. Elde edilen sonuçlar literatürle uyumludur (Betz ve Turner, 2011; Gnambs ve Batinic, 2011; Hol, Vorst, ve Mellenbergh, 2007; Reise ve Henson, 2000).

Anahtar Kelimeler

CAT, MTK, Simülasyon, Madde Seti, Eğim Parametresi, Test Uzunluğu

* Cumhuriyet Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü. E-posta: asalihsimsek@gmail.com

** Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü (Emekli Öğretim Üyesi). E-posta: etavsancil@gmail.com

Paper No: 134

Effect of Item Pool Size and Slope Parameter onto Test Length in CAT

Abstract

The main purpose of this research is to examine the effect of the item pool size and the slope parameter onto the test length in CAT used for polytomous items. For the item pool size, three condition (n1=10, n2=20, n3=30) and also for the slope parameter in the item pool three condition (a_1=[0.65-1.00], a_2=(1.00-1.35), a_3=[1.35-1.70) have been determined. FI as a method of item choice, EAP as a estimation method and SH<.385 as stopping rule have been used for CAT. Simulation study was carried out by using R libraries. In order to generate data ltm (1.0) R library was used and for CAT simulations catIrt (0.5) R library was used.

As a result of simulation studies, descriptive statistics of test length in CAT were obtained. In cases where the slope parameter was a1= [0.65-1.00] CAT test length was realized as μ_{11}=10.00 (%100), μ_{21}=19.97 (%100), μ_{31}=27.68 (%100) for n_1=10, n_2=20, n_3=30. However, in cases where the slope parameter was a_2= (1.00-1.35) CAT test length was realized as μ_{12}=9. 99 (%100), μ_{22}=14. 53 (%73), μ_{32}=13. 60 (%45). Once for all in cases where the slope parameter was a_3=[1.35-1.70] CAT test length has turned out as μ_{13}=9.52 (%95), μ_{23}=9.10 (%45), μ_{33}=8.56 (%29) value. Within this research the correlation coefficients between CAT theta and full item bank theta have been varied between .95 and 1.00. While correlation coefficients were 1 in CAT where all items were used, it has been calculated as .95 in the cases where 29 percent of item pool was used.

One of the main results of this research is that as the slope parameter increased, the CAT test length evenly decreased. As a result of the research, it is understood that the use of item pools composing of items that have low slope parameter didn't have any contribution to the suitability of measuring in CAT. Further, the use of the use of item pools composing of items that have high slope parameter considerably reduced test length in CAT. The results obtained were in harmony with literature (Betz and Turner, 2011; Gnambs and Batinic, 2011; Hol, Vorst, and Mellenbergh, 2007; Reise and Henson, 2000).

Key Words

CAT, IRT, Simulation, Item pool, Slope parameter, Test Length

Bildiri No: 135

Hiyerarşik Nitelik Yapı Seçiminde Model Seçme Metotlarının Kullanılabilirliği

Lokman AKBAY

Özet

Nitelikler hiyerarşik bir yapıya sahip olduğunda, bilişsel tanı modelleri (BTM) hiyerarşik yapıyı gözardı etmemelidir; aksi halde, BTM'ler uygun ve kullanışlı olmayabilirler (Templin & Bradshaw, 2014). Uzman görüşü ve sözel data analizi yöntemleri birden fazla hierarşik yapının önerilmesine yol açabilir. Böyle bir durumda, BTM'ler belirlenen hierarşiye göre yeniden yapılandırılacağı için en doğru/uygun yapıyı seçmek oldukca önemlidir. Bu çalışma, Likelihood'a dayalı model seçme yöntemlerinin niteliklerin hierarşik yapısının seçilmesinde kullanılabilirliğini değerlendirmektedir. Çalışma sonuçları model seçme yöntemlerinin, özellikle LRT ve AIC kriterlerinin birçok durumda doğru hierarşik yapının seçilmesinde kullanılabileceğini ortaya koymuştur. Sonuclar, yanlış Q-matrix spesifikasyonlarının LRT, AIC, ve BIC kriterlerine bağlı olarak hierarşi seçimine negatif etki gösterdiğini ve bu etkinin boyutunu ortaya koymuştur.

Anahtar Kelimeler

Hiyerarşik Nitelikler, BTM, Bilişsel tanı modelleri

Paper No: 135

Viability of Model Selection Methods for Attribute Hierarchy Selection

Abstract

When attributes are hierarchically structured, cognitive diagnosis models (CDMs) need to take the hierarchical structure into account; otherwise, they may not be appropriate and useful (Templin & Bradshaw, 2014). Based on expert opinions and verbal data analyses may result in more than one possible hierarchy. In this case, selection of the most appropriate hierarchy is essential as the CDM is structured in accordance with the proposed hierarchy. This study examined the viability of likelihood-based model selection methods in hierarchical structure selection. Results suggested that model selection methods, especially LRT and AIC, can successfully selects the generating hierarchy in many conditions. Furthermore, although all three types of misspecifications in the Q-matrix degrade selection of correct hierarch in some degree, negative impact of underspecified Q-matrix was observed to be the largest.

Key Words

Attribute hierarchy, CDM, Cognitive diagnosis

Bildiri No: 136

Performans Takip Sınavının (PTS) Psikometrik Özelliklerinin Rasch Analizi ile İncelenmesi

Hakan KOĞAR*

Özet

Bu araştırmanın amacı, 2015-2016 eğitim öğretim yılı içerisinde Antalya ilinde uygulanan Performans Takip Sınavları'na (PTS) ait Matematik, Fen ve Teknoloji, Türkçe ve Sosyal Bilgiler alt testlerinin psikometrik özelliklerini Rasch analizi ile incelemektir. Bu araştırmada veri toplama aracı olarak PTS kullanılmıştır. Bu araştırmanın evrenini oluşturan 30800 8.sınıf öğrencisinden rastgele seçilerek belirlenen 2000 öğrenci bu araştırmanın örneklemini oluşturmaktadır. Testin psikometrik özelliklerinin belirlenmesinde bir Madde Tepki Kuramı (MTK) uygulaması olan Rasch analizi kullanılmıştır. Fen ve Teknoloji ile Sosyal Bilgiler alt testlerindeki bazı maddelerin uyum değerlerinin beklenen değer aralıklarının dışında olduğu belirlenmiştir. Fen ve Teknoloji dersindeki 2 ve 14 numaralı maddeler ile Sosyal Bilgilerde ise 6, 7, 11, 12, 15, 16 numaralı maddeler uyum değerlerinin 0,8 - 1,2 aralığında olmamasından dolayı, testten çıkartılması gerekmektedir. Diğer maddelerin örtük (gizil) değişken ile uyum gösterdiğine karar verilmiştir. Güçlük parametreleri incelendiğinde, güçlük düzeyi en düşük alt testin Türkçe testi olduğu belirlenmiştir. Fen ve Teknoloji ile Sosyal Bilgiler alt testleri ise diğer testlere göre güçlük düzeyinin yüksek olduğu belirlenmiştir. Testteki en zor soru, Sosyal Bilgiler, en kolay soru ise Türkçe dersinde yer almaktadır. Sosyal Bilgiler alt testindeki madde güçlüğü tahminleri, diğer alt testlere göre daha yüksek hata ile tahmin edilmekle birlikte, tüm hata değerleri kabul edilebilir sınırlar içerisindedir. Testin EAP/PV güvenirlik katsayısı 0,574 ve ayırma güvenirliği 0,998 olarak hesaplanmıştır. Teste ait soruların büyük çoğunluğunun gizil değişkenle yüksek uyum gösterdiği, madde güçlük parametrelerine göre Türkçe alt testinin kolay, diğer alt testlerin görece zor olduğu ve testten ölçülen özellik açısından yüksek bilgi elde edilmesi dolayısıyla testin geçerliğinin yüksek olduğuna karar verilmiştir.

Anahtar Kelimeler

Performans takip sınavı, rasch analizi

*Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Öğretim Üyesi, İletişim: hkogar@gmail.com

Paper No: 136

An Investigation of Psychometrical Properties of Performance Monitoring Examination (PTS) with Rasch Analysis

The purpose of this study, investigation the psychometric properties of Performance Monitoring Examination's (PTS) Mathematics, Science and Technology, Turkish and Social Studies subtests, which is applied at Antalya, 2015-2016 school year. The data collection tool fort his study is PTS. The population of this study is 30800 8. Grade students. Sample is 2000 students who is randomly chosen. For determining the psychometric properties of tests, Rasch analysis, an Item Response Theory application were used. Some items from Science and Technology and Social Studies subtest's fit indices are out of defined value ranges. 2. and 14. items from Science and Technology and 6., 7., 11., 12., 15. and 16. items from Social Studies are out of defined (0,8 – 1,2) value ranges, because of this, these items must be remove from test. Other items fits good with latent variables. When difficulty parameter's investigated, Turkish subtest have the lowest difficulty indices of test. Science and Technology and Social Studies subtests have the largest difficulty parameters. Social Studies subtest have the hardest item of all test. Social Studies subtest have the higher standard errors, when comparing the other subtests; but these standard errors are in acceptable ranges. The EAP/PV reliability coefficient is 0,574; discrimination reliability coefficients is 0,998. A great majority of items have higher fit with latent variables, Turkish subtest are easier from other subtests, other subtests are relatively difficult and high test information are obtained about measured trait. Because of all this findings, PTS is a valid assessment tool.

Key Words

Performance monitoring examination, rasch analysis

Bildiri No: 137

Madde Takımları İçeren Testlerde Tek ve Çok Boyutlu Madde Tepki Kuramı Modellerinden Elde Edilen Madde ve Yetenek Parametrelerinin Karşılaştırılması

Esin YILMAZ KOĞAR*, Hülya KELECİOĞLU**

Özet

Bu araştırmanın amacı, tek boyutlu madde tepki kuramı (TBMTK), iki faktör (2FM) ve madde takımı tepki kuramı (MTTK) modellerinden elde edilen madde ve yetenek parametreleri ile bu parametrelere ait standart hata değerlerini örneklem büyüklüğünü göz önüne alarak belirlemek ve elde edilen sonuçları karşılaştırmaktır. Bu amaç doğrultusunda PISA 2012 uygulamasındaki matematik alanına ait 10. kitapçık veri toplama aracı olarak seçilmiş ve bu kitapçıkta yer alan 36 maddeden elde edilen veriler kullanılmıştır. Verilerin analizinde IRTPRO 2.1 programı kullanılmıştır. Araştırmanın bulguları incelendiğinde bütün koşullarda yerel bağımlılığı göz önüne alan modellerin TBMTK'den daha iyi model-veri uyumu sağladığı ve 2FM ile MTTK arasında anlamlı farklılık olmadığı; her iki modelin de kullanılabileceği belirlenmiştir. Farklı örneklem büyüklüklerinde, farklı modellerden elde edilen madde ve yetenek parametreleri için genel olarak en düşük ortalama hata değerlerinin TBMTK'den elde edildiği; MTTK'nin 2FM'den daha düşük ortalama hata kestirimi yaptığı belirlenmiştir. Örneklem büyüklüğü arttıkça farklı modellerden elde edilen madde ve yetenek parametreleri daha benzer sonuçlar vermiştir. Çalışmanın sonuçlarına dayanarak yerel bağımlı maddeler içeren veri setleri için standart MTK modelleri yerine, maddeler arasındaki yerel bağımlılık durumunu göz önüne alan modellerin kullanılması önerilmektedir.

Anahtar Kelimeler

Madde tepki kuramı, Yerel bağımlılık, İki faktör modeli, Madde takımı tepki kuramı modeli

* Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: esinyilmaz@hacettepe.edu.tr
** Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: hulyaebb@hacettepe.edu.tr

Paper No: 137

Comparison of Item and Ability Parameters Obtained from Unidimensional and Multidimensional Item Response Theory Models on Tests Composed of Testlets

Abstract

The purpose of this research is to estimate the item and ability parameters and the standard error values related to those parameters, obtained from unidimensional item response theory (UIRT), bifactor (BIF) and testlet response theory models (TRT), when sample size change, and to compare the obtained results. In accordance with this purpose, the 10th booklet applied in the field of mathematics in PISA 2012, has been selected as a data collection tool, and data obtained from 36 items in that booklet have been used. IRTPRO 2.1 program has been used in analyses of the data. It has been determined that, under all conditions, models which take into consideration the local dependency have provided better model - data compatibility than UIRT, generally there is no meaningful difference between BIF and TRT; and both models can be used for those data sets. When the findings of the research have been investigated, it was determined that, for the item and ability parameters estimated under different samples sizes from different models, generally the lowest mean error values were obtained from UIRT; and TRT has made a mean of error estimation lower that BIF. It has been concluded that, the sample size increases, the item and ability parameters are more similar. When the data set containing local dependency items, the models of considering the local dependency between items are recommended instead of the standard IRT models based on the results of the study.

Key Words

Item response theory, local dependency**,** , bi-factor model, testlet response theory

Bildiri No: 138

Çok Boyutlu İş Motivasyonu Ölçeğinin Türk Kültürüne Uyarlanması

Aydın ÇİVİLİDAĞ*, Güçlü ŞEKERCİOĞLU**

Özet

Bu araştırmanın amacı endüstri ve örgüt psikolojisi alanında sıkça kullanılan ve yedi farklı dilde (İngilizce, Fransızca, Almanca, Çince, Endonezya dili, Norveççe ve Felemenkçe) uyarlaması yapılan çok boyutlu iş motivasyonu ölçeğini (ÇBİMÖ) Türk kültürüne uyarlamak ve araçtan elde edilen puanların psikometrik niteliklerine ilişkin kanıtları ortaya koymaktır. İlişkisel araştırma modeli bağlamında yürütülen araştırmada çalışma grubu, farklı hizmet iş kollarında çalışan katılımcılardan (akademisyen, öğretmen, bankacı, otomotiv pazarlama, turizm ve üniversitede çalışan memurlar) oluşmaktadır. Likert tipi 19 maddeden oluşan ölçme aracı, orijinal kültürde altı faktörden oluşmaktadır. Bu alt ölçekler motive olmama, dış düzenleme-sosyal, dış düzenleme-maddesel, içe yansıtılan düzenleme, belirlenen düzenleme ve içsel motivasyondur. Uyarlama sürecinde çeviri, geri çeviri yapıldıktan sonra uzman görüşleri alınmış ve öneriler doğrultusunda maddelerde gerekli düzenlemeler yapılmıştır. Bu çerçevede hazırlanan deneme formu için dilsel eş değerlilik çalışması yapılmış ve eşdeğerliliğin sağlandığı belirlenmiştir. Bu aşamada uygulamalar pilot ve asıl uygulama biçiminde yapılmıştır. Pilot uygulamada 166 katılımcıdan elde edilen veri seti üzerinden doğrulayıcı faktör analizi ve açımlayıcı faktör (DFA ve AFA) analizi yapılmış; madde test korelasyonları ve Cronbach alfa iç tutarlılık katsayıları hesaplanmıştır. Yapılan analizler sonucunda çalışmayan maddeler uzman görüşleri de alınarak gözden geçirilmiştir. Bu analizler ışığında yeniden düzenlenen ÇBİMÖ'nün asıl formu 506 katılımcıya uygulanmış ve aracın psikometrik nitelikleri test edilmiştir. DFA sonucunda orijinal kültürdeki araç ile aynı faktör deseni elde edilmiş, uyum indeksleri S-Bχ^2(119)=330.07, S-Bχ^2/sd=2.77, GFI=.93, RMSEA=.059, SRMR=.059, CFI=.96 ve NNFI=.095 olarak belirlenmiştir. Ayrıca alt ölçekler düzeyinde, iç tutarlılık bağlamında güvenilirlik katsayılarının .72 ile .80 arasında değiştiği belirlenmiştir. Bu çerçevede ÇBİMÖ'den elde edilen puanlar doğrultusunda geçerlilik ve güvenilirliğin belirli ölçüde kabul edilebilir bir düzeyde olduğu söylenebilir. Dolayısıyla bu çalışmada uyarlaması yapılan ölçme aracının özellikle endüstri ve örgüt psikolojisi alanındaki işgörenlerin motivasyonlarının ölçülmesine yönelik çalışmalarda ve hizmet sektöründeki işletmelerde kullanılması beklenmektedir.

Anahtar Kelimeler

Çok Boyutlu İş Motivasyonu, Psikoloji, Endüstri ve Örgüt Psikolojisi, Psikometri, Uyarlama

*Yrd.Doç.Dr., Akdeniz Üniversitesi, Edebiyat Fakültesi, Psikoloji Bölümü. İletişim: aydinc@akdeniz.edu.tr, Tlf: +90242 227 44 00 (3257)

** Yrd.Doç.Dr., Akdeniz Üniversitesi, Eğitim Fakültesi, Ölçme ve Değerlendirme Anabilim Dalı. İletişim: guclus@akdeniz.edu.tr, Tlf: +90242 227 44 00 (4641)

Paper No: 138

Studying of Adaptation to Turkish Culture The Multidimensional Work Motivation Scale

Abstract

The purpose of this research is to adapt the multidimensional work motivation scale (MWMS) into Turkish culture which is frequently used in industrial and organizational psychology and has already been adapted into seven different languages (English, French, German, Dutch, Indonesian, Chinese, Norwegian) and to reveal evidence in respect to psychometric characteristics of the scale those were obtained from the collected data. The study was carried out in the context of relational research model and the research group of the study included the participants from different service sectors (academic, teacher, bank employer, employee who works in marketing, automotive, tourism sector and officer staff in the university). The scale is consisted of 19 Likert type items and six factors in the original culture. These sub-dimensions are amotivation, extrinsic regulation-social, extrinsic regulation material, introjected regulation, identified regulation and intrinsic motivation. In the adaptation process of the scale, translation from English to Turkish and back translation from Turkish to English was done by English lecturers and necessary regulations were done in accordance with the suggestions of the experts. Therefore, linguistic equivalence study of the pilot form of the scale was done and the linguistic equivalency was provided. The applications were made as pilot and master practices by researchers in this stage. Confirmatory and exploratory factor analysis (CFA and EFA) were conducted on the data set those were obtained from 166 participants and also item test correlations and Cronbach alpha coefficients were calculated in the pilot study. In consequence of the analysis dysfunctional items were revised based on the expert's opinions.Revised master form of MWMS was performed on 506 participants and the psychometric qualifications of the scale was tested in the light of these analysis. As a result of CFA, the same factor pattern as it is in the original culture was obtained. It was determinated that compliance indices meet the acceptable levels as S-Bχ^2(119)=330.07,S-Bχ^2/sd=2.77, GFI=.93, RMSEA=.059, SRMR=.059, CFI=.96 and NNFI=.095. Furthermore, it was determined differentiate that the reliability coefficients were between .72-.80 in the acceptable levels at subscales grade in the context of internal consistency. In this frame, it can be reported that with the points obtained from MWMS, acceptable levels of validity and reliability were provided. Therefore, the scale which was adapted in this study is expected to be used in the field of industrial and organizational psychology and service sectors in order to measure the motivations of the servers.

Keywords

Multidimensional Work Motivation, Psychology, Industrial and Organizational Psychology, Psychometry, Adaptation

Bildiri No: 139

Kağıt-Kalem Testi ve Web Temelli Test Olarak Uygulanan Ölçme Araçlarının Ölçme Değişmezliği ve Evren Ayrışıklığı Bakımından Karşılaştırılması

Güçlü ŞEKERCİOĞLU*

Özet

Son yıllarda bilgisayar yazılım teknolojilerinin gelişmesi ve internetin yaygınlaşması ile araştırmacıların ölçek uygulamalarını web temelli (internet ortamında) yapma tercihinde bir artış gözlenmektedir. Ölçme araçlarının kağıt-kalem testi ya da web temelli test olarak uygulanması ile araçların ölçmeye yöneldiği psikolojik özellikte bir farklılık oluşup oluşmadığı, diğer bir ifadeyle aynı yapıyı ölçüp ölçmediği; ölçme aracından elde edilen puanlar çerçevesinde varyansların, kovaryansların ve örtük ortalamaların eşit olup olmadığı psikometrik bir problem olarak ele alınabilir. Bu araştırmanın genel amacı çok faktörlü (yurttaşlık aidiyeti ölçeği) ve tek faktörlü (kadercilikölçeği) iki ölçme aracının, web temelli ve kağıt-kalem testi olarak uygulanması çerçevesinde, ölçme değişmezliği ve evren ayrışıklığı bakımından karşılaştırılmasıdır. Bu çalışmanın araştırma grubu Akdeniz Üniversitesi, Eğitim Fakültesindeki Rehberlik ve Psikolojik Danışmanlık, Fen Bilgisi Öğretmenliği, Sınıf Öğretmenliği ve Okul Öncesi Öğretmenliği bölümlerinde öğrenimine devam eden toplam 438 öğrenciden oluşmaktadır. Araştırma grubunun oluşturulmasında öncelikle araştırmaya katılmak için gönüllü olan öğrenciler belirlenmiştir. Bu belirlemenin ardından sınıf listeleri yeniden düzenlenmiş ve gruplar rassal sayılar tablosuna göre ikiye ayrılmıştır. Buna göre 222 katılımcı kağıt-kalem testi uygulamasına ve 216 katılımcı ise web temelli test uygulamasına dahil edilmiştir. Çalışmanın amacı doğrultusunda araştırma sorularına yanıt bulabilmek amacıyla doğrulayıcı faktör analizi (DFA), Cronbach alfa analizi, kovaryans matrisleri eşitliği testi ve çoklu-grup DFA uygulanmıştır. Elde edilen iki veri seti üzerinden ölçme değişmezliğine ilişkin yapılan çoklu grup DFA sonuçlarına göre, hem çok faktörlü desende hem de tek faktörlü desende dört değişmezlik modeli içinde en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğu, dolayısıyla ölçme değişmezliğinin sağlanamadığı sonucuna ulaşılmıştır. Diğer taraftan evren ayrışıklığına ilişkin yapılan analizlere göre, çok faktörlü desende kovaryansların eşitliği ve tek faktörlü desende ise örtük ortalamaların eşitliği hipotezi ile kurulan modellerin daha iyi çalıştığı, bir başka ifadeyle grupların tek bir örtük özellik altında tanımlanamadığı sonucuna ulaşılmıştır. Bu çerçevede sonuç olarak, araştırmacıların ölçme araçlarının kağıt-kalem testi ya da web temelli test yöntemleriyle uygulanması sonucunda, araçların aynı psikolojik özelliği ölçtüğüne ilişkin kabule şüpheyle yaklaşması gerektiği düşünülebilir.

Anahtar Kelimeler

Kağıt-kalem Testleri, Web Temelli Test, Ölçme Değişmezliği, Evren Ayrışıklığı, Psikometrik Nitelikler

* Akdeniz Üniversitesi, Eğitim Fakültesi, Ölçme ve Değerlendirme Anabilim Dalı

Paper No: 139

A comparison of Paper-and-Pencil Test and Web-based test administered as measurement instruments in terms of Measurement Invariance and Population Heterogeneity

Abstract

In recent years, researchers mostly prefer to perform their studies in web-based areas thanks to developments of computer software and internet. Measurement instruments are administered as paper-and-pen tests or web-based tests, so it can be seen whether there is a change psychological trait that instruments aim to measure or not. In other words, within the framework of scores obtained from measurement too it can be discussed whether variances, covariance and implicit means are equal or not as a psychometric problem. The aim of this research is to compare multi-factor instrument (citizenship belonging scale) and single-factor instrument (fatalism scale) applied as paper-and-pen and web-based tests in terms of measurement invariance and population heterogeneity. The research group of this research consists of 438 students who study in Faculty of Education of Akdeniz University in the departments of Psychological Counseling and Guidance, Science Teaching, Primary School Teaching and Pre-school Teaching. Volunteer students were primarily determined to research group. After this determination, classroom lists were re-designed and the groups split in half according to random numbers table. According to this, 222 participants were included paper-and-pen test and 216 participants were included web-based test application. Confirmatory factor analysis (CFA), Cronbach alpha analysis, the equality of covariance matrices test and multi-group CFA were applied in order to find answers of research' questions. According to the results of multi-group CFA which was performed for the two sets of data obtained through the measurement invariance, models strong factorial invariance model is the best working model in both multi-factor design and single-factor design, thereby measurement invariance could not be provided. On the other hand, according to population heterogeneity analysis the models which were set up with the hypothesis of covariance equality in multi-factor design and implicit means equality in single-factor design work better and the groups could not be defined under one latent trait. Consequently, it can be claimed that researchers should be doubtful about administrating paper-and-pencil tests or web based tests while measuring the same psychological trait.

Keywords

Paper-and-pen Tests, Web-based Tests, Measurement Invariance, Population Heterogeneity, Psychometric Properties

Bildiri No: 141

Tutum Ölçeği Geliştirmede Genel Tutum İfadelerinin Kullanımı Üzerine Bir Araştırma

Nuri DOĞAN*, Yusuf KARA**

Özet

Bu çalışmada, geliştirilen tutum ölçeklerinde sıklıkla kullanılan ortak tutum ifadelerinin bir araya getirilmesi ile bu ifadelerin yeni bir tutum ölçeğinin maddeleri olarak kullanılabilme durumları araştırılmıştır. Bu amaç doğrultusunda ülkemizdeki çeşitli dergilerde yayınlanmış olan ve içeriğinde geliştirilen, uyarlanan veya hazır olarak kullanılan bir tutum ölçeğine yer verilen toplam 176 çalışma detaylı olarak incelenmiştir. Bu çalışmalarda yer alan tutum maddelerindeki "…'yı severim, …. ilgi çekicidir " gibi ifadelerin frekansları hesaplanarak toplam 2128 ifadeden en çok frekansa sahip 22'si belirlenmiştir. Belirlenen bu tutum ifadeleri formasyon dersleri olan Eğitim Programları ve Öğretim, Rehberlik, Öğretim Yöntem ve Teknikleri, Eğitimde Ölçme ve Değerlendirme, Gelişim Psikolojisi ile Öğrenme Psikolojisi için tutum ifadeleri haline getirilerek toplam 6 ders için ön uygulama formları oluşturulmuştur. Her bir ders için düzenlenen denemelik tutum ifadelerine açımlayıcı faktör analizleri uygulanmıştır. Gerçekleştirilen analizler sonucunda, her bir formasyon dersi için kullanılan ortak ifadeler ile oluşturulan ölçeklerin belirgin biçimde tek boyutlu bir yapıya sahip oldukları gözlenmiştir. Maddelerin bu tek faktöre verdikleri yük miktarı 0.35 ve üzerinde olmuştur. Her bir ölçeğe ait iç tutarlılık bağlamındaki güvenirlik katsayıları oldukça yüksek olup 0.87 ile 0.95 arasında değişmiştir. Sonuç olarak herhangi bir amaç için bir tutum ölçeğinin geliştirilmesine ihtiyaç duyulması halinde bu ortak ifadelerin ölçeğin amacına uygun biçimde düzenlenerek kullanılabileceğine dair güçlü kanıtlar elde edilmiştir. Böylece faktör yapısı tek boyutlu olan ve iç tutarlılığı oldukça yüksek olan bir tutum ölçeğinin hızlı ve pratik bir biçimde elde edilebileceği söylenebilir.

Anahtar Kelimeler

Tutum ölçeği, ortak tutum ifadeleri, ölçek geliştirme

* Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü. İletişim: nurid@hacette.edu.tr
** Anadolu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü. İletişim: yusufkara@anadolu.edu.tr

Paper No: 141

A Study on Using Common Attitude Expressions for Attitude Scale Development

Abstract

This study aims to reveal the usability of the common attitude expressions that are widely used in attitude scales developed previously. 176 studies that were published in various Turkish journals and contained an attitude scale which is developed, adapted by the author(s) or borrowed from other researchers were examined in detail. Frequencies of the total 2128 expressions such as "I like …" or "… is attractive" was calculated and 22 expressions which had the most frequencies were selected among them. Selected expressions were used in order to develop new attitude scales for teacher certification courses which are Curriculum and Instruction, Counseling, Instruction Methods and Techniques, Educational Measurement and Evaluation, Development Psychology and Learning Psychology. Explanatory and confirmatory factor analyses were conducted for each of the 6 certification courses attitude scales. According to explanatory factor analyses, all of the attitude scales showed a dominant single factor structure. Factor loadings for each item were higher than 0.35, which is also a good sign of the unidimensionality. The reliabilities in terms of the internal consistency were quite high for every attitude scale and varied between 0.87 - 0.95. As a result of the analyses, it was concluded that these common attitude expressions can be used for development of an attitude scale after a revision based on the aim of the scale. By this way, it is possible to create unidimensional and highly reliable attitude scales very conveniently.

Key Words

Attitude Scale, common attitude expressions, scale development

Bildiri No: 142

Açımlayıcı Faktör Analizi ve Hiyerarşik Kümeleme Analizi Yöntemlerinin Farklı Örneklem Büyüklüklerine Göre Karşılaştırılması

Murat KAYRİ*, Fuat ELKONCA**, Görkem CEYHAN***

Özet

Bu araştırmanın genel amacı ölçeklerin faktör yapılarını ortaya çıkarmada kullanılan Açımlayıcı Faktör Analizi (AFA) ve Hiyerarşik Kümeleme Analizi (HKA) yöntemlerini farklı örneklem büyüklüklerine göre karşılaştırmaktır. Araştırma kapsamında 2012 yılında Muş Alparslan Üniversitesinde öğrenim gören 2300 üniversite öğrencisine uygulanan ve Kember vd. (2000) tarafından geliştirilen ve Başol & Evin Gencel (2013) tarafından Türkçeye uyarlananan "Yansıtıcı Düşünme Düzeyini Belirleme Ölçeği" den elde edilen veri seti kullanılmıştır. Verilerin analizinde SPSS ve Lisrel paket programı kullanılmıştır. Analizler sonucunda, 50 ve 200 kişilik örneklemlerde her iki yöntemden de aynı sonuçlar elde edilmiş olup elde edilen uyum değerleri iyi düzeydedir. Örneklem sayısı 500 ve üstünde olduğunda (1000, 1500, 2000) AFA dan elde edilen sonuçlar değişmemektedir. 500'den büyük örneklemlerde AFA ve Ward yöntemlerinden elde edilen faktör yapıları ve bu yapılara ait uyum değerleri genel olarak çok yakındır. Ancak, kısmen de olsa Ward yöntemi kullanılarak elde edilen faktör yapılarına ait uyum değerlerinin daha iyi düzeyde olduğu söylenebilir. Bu sonuçlara ek olarak 50 kişilik küçük örneklemlerde bile hem AFA hem de Ward yönteminden elde edilen faktör yapılarının ve bu faktör yapılarına ait uyum değerlerinin iyi düzeyde olduğu söylenebilir. Yapılan bu araştırma da gerçek veri kullanıldığı için maddelerin bir biriyle olan yüksek ve düşük ilişkileri, dağılım normal olup olmaması, faktörlerdeki maddelerinin tutarlılığı ve faktörlerin anlamlılığı gibi araştırma sonuçlarını etkileyebilecek etkenler göz ardı edilmiştir. Bu konu ile ilgili yapılacak diğer araştırmalarda veri setini çeşitli özelliklere göre kontrol altına almak için similasyon verisi kullanılabilir.

Anahtar Kelimeler

Açımlayıcı Faktör Analizi, Hiyerarşik Kümeleme Analizi, Örneklem büyüklüğü

* Batman Üniversitesi Mühendislik ve Mimarlık Fakültesi İletişim: muratkayri@gmail.com
** Gazi Üniversitesi Gazi EğitimFakültesi Eğitim Bilimleri Bölümü. İletişim: fuat.elkonca@gmail.com
*** Gazi Üniversitesi Gazi EğitimFakültesi Eğitim Bilimleri Bölümü. İletişim: gorkemceyhan23@gmail.com

Paper No: 142

The Comparison of Exploratory Factor Analysis and Hierarchical Cluster Analysis Methods According to Different Sample Size

Abstract

The general aim of this study is to compare Exploratory Factor Analysis and Hierarchical Clustering Analysis methods, which are used to find out the factor structures of the scales, according to the different sample sizes. In the study, the data set obtained from "Reflective Thinking Scale", which was administered to 2300 students in Muş Alparslan University in 2012, developed by Kember et al. (2000), adapted into Turkish by Başol & Evin Gencel (2013), was used. In data analysis, SPSS and LISREL package programs were used. Analysis results show that the same results were obtained from both two methods in the 50-person and 200-person samples and the goodness of fit values obtained are at good level. When the sample size is 500 or above (1000, 1500, 2000), the results obtained from EFA do not change. Although the factor structures and goodness of fit values belonging to these factors, which were obtained from EFA and Ward's methods, are generally too close in the samples larger than 500, it can be said that the goodness of fit values, which belong to the factor structures obtained by using Ward's method, are at relatively better level. In addition to these results, it can be said that the factor structures obtained from both EFA and Ward's methods and the goodness of fit values belonging to these factor structures are at good level in even 50-person small sample. As the real data were used in this study, the elements, which might affect the results, such as the high and low relationships among the items, whether the distribution was normal or not, and the consistency and significance of the items in the factors, were not taken into consideration. In other related research studies on this topic, the simulation data can be used to get the data set under the control according to various traits.

Key Words: Exploratory Factor Analysis, Hierarchical Cluster Analysis, Sample size

Bildiri No: 143

Bileşke Puan Güvenirlik Hesaplama Yöntemlerinin İki Ve Üç Boyutlu Yapılarda Karşılaştırılması

Nuri DOĞAN*, Yusuf KARA**, Seda DEMİR***, Buket EREN JANSSEN****

Özet

Bir test alt boyutlara sahip olmakla birlikte, bireylere bileşke puanlar verildiği durumlarda tüm test için bileşke puanların güvenirliğinin bilinmesine de ihtiyaç duyulmaktadır. Pratik uygulamalarda başvurulduğuna az rastlansa da çok boyutlu testlerden elde edilen ölçme sonuçları için geliştirilmiş güvenirlik hesaplama yaklaşımları bulunmaktadır ve genel olarak bunlara "bileşke puan güvenirliği" (composite reliability) denmektedir. Çok boyutlu testlerde bileşke puanların güvenirliğini belirlemek için geliştirilen bu teknikler az kullanıldığı gibi bunların hangi koşullarda tercih edilmesi gerektiğine ilişkin pratik karşılaştırmalarda yetersizdir. Dolayısıyla çok boyutlu testlerde bileşke puanlar için güvenirlik belirleme tekniklerinin aynı koşullarda nasıl işlediklerine ilişkin karşılaştırmalı araştırmalara ihtiyaç duyulduğu düşünülmektedir. Gözlenen bu ihtiyaca dayanarak iki ve üç boyutlu yapılarda farklı (Felt-Brenan, Wang-Stanley, Reykov ve McDonald'ın) bileşke güvenirlik tekniklerivle elde edilen güvenirlik değerlerinin hata ve yanlılık değerleri araştırılmıştır. Bu tekniklerin eşit koşullarda karşılaştırılmasını sağlayabilmek için simülatif veriler üzerinde çalışılmıştır. Araştırmada testlerin boyutu, boyutlar arası korelasyonlar, boyutlardaki soru sayılarının farklı olması, puanlamanın ikili ve Likert tipi olması, dağılımın normal veya çarpık olması araştırmanın koşulları olarak ele alınmıştır. Araştırma 2(boyut)*3(boyutlar arası korelasyon)*2(puanlama türü)*3(dağılım biçimi)*30(iterasyon) = 1080 veri seti üzerinde yürütülmüştür. Veriler R ve Excel programları kullanılarak analiz edilmiştir. Bu analizlerden her bir güvenirlik tekniği için elde edilen 30 tekrarlı sonuçlar için hata karelerinin ortalamasının karekökü (Root Mean Squre of Errors) ve yanlılık (bias) değerleri hesaplanmıştır. Elde edilen küçük değerler tekniklerin yeterlik düzeyi olarak yorumlanmıştır. Bulgular çoğu veri setinde faktör analizi sonuçlarına dayanan bileşke güvenirlik yaklaşımlarının hata ve yanlılık değerlerini düşük olduğunu; dereceli puanlamanın hata ve yanlılık değerlerinde artışa neden olduğunu; boyut sayısındaki artışın hata ve yanlılık değerlerinde artışa neden olduğunu; boyutlar arası korelasyonlardaki azalışın hata ve yanlılık değerlerinde artışa neden olduğunu göstermektedir.

Anahtar Kelimeler

Bileşke güvenirlik, simülatif araştırma, çok boyutlu yapılarda güvenirlik, conjenerik yapılarda güvenirlik

* Hacettepe Üniversitesi, Eğitim. Fakültesi, Eğt. Bil. Böl. İletişim: nurid@hacettepe.edu.tr Tel: +905055882459

** Anadolu Üniversitesi, Eğitim. Fakültesi, Eğt. Bil. Böl. İletişim: ykara88@gmail.com Tel: +905424296698

***Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Doktora Öğrencisi. İletişim: seda.demir@hacettepe.edu.tr Tel: +90 312 297 8550

****Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Doktora Öğrencisi. İletişim: buketin2003@yahoo.com Tel: +905369384261

Paper No: 143

Comparison of Composite Score Reliability Methods in Two and Three-Dimensional Structures

Usually test are multi-dimensional. When a multi-dimensional test produces composite score for examinees, we need investigate composite reliability. Although it is applied rarely, there are several reliability calculation method for multi-dimensional test's composite scores, it is called composite reliability. Also, there are not enough studies about which method must be used on which situation. So, methods must be compared to learn how they are perform in multidimensional tests. With this respect, methods (Felt-Brennan, Wang-Stanley, Reykov and McDonald) were performed on two and three dimensional structures and errors and biases of composite reliabilities were searched. Research was performed on simulated, 2 dimensions* 3 correlation between dimensions * 2 scoring methods * 3 distributions * 30 replications, totally 1080 datasets. Data were analyzed by using R and Excel. For every data root mean of square errors and bias were calculated. Results showed, in most of data analysis, factor analytic method had lower RMSE and bias. Likert type scoring, higher correlation between dimensions and increasing number of dimensions caused higher bias and RMSE.

Key Words

Composite reliability, simulation study, reliability in multidimensional structures, reliability in congeneric structures

Bildiri No: 144

Çok Düzeyli Modellerle PISA 2012 Matematik Testinde Değişen Madde Fonksiyonu ve Potansiyel Kaynak Olarak Öğretmen Davranışı Özelliklerinin İncelenmesi

Özen YILDIRIM*, Ömer KUTLU**, Hatice KUMANDAŞ***

Özet

Bu araştırmanın amacı çok düzeyli modelleme yaklaşımına dayalı olarak PISA 2012 matematik testinde yer alan maddelerin cinsiyete göre değişen madde fonksiyonlarını ve varsa öğrenci düzeyinde SED ve sınıf içi öğretmen davranışlarına dayalı DMF kaynaklarını incelemektir. Araştırmanın örneklemini PISA 2012 uygulamasına katılmış 15 yaş grubundan 1458 öğrenci oluşturmaktadır. Veriler PISA 2012'nin matematik bilişsel alan testinden ve öğrenci anketinden elde edilmiştir. İki düzeyli GHLM ve geleneksel lojistik regresyon yaklaşımına dayalı olarak cinsiyete göre DMF incelenmiş ve sonuçlar karşılaştırılmıştır. GHLM'de üç madde, lojistik regresyonda ise altı madde DMF içermektedir. GHLM sonuçları lojistik regresyon yaklaşımına göre daha tutucudur. DMF'li üç madde her iki analizde de ortaktır. Bu maddelerin biri kızlar diğer ikisi erkekler lehine DMF'lidir. Madde özellikleri incelendiğinde erkekler lehine olan maddenin bilişsel düzeyleri uygulama ve formüle etme, kızların ki ise yorumlamadır. Bununla birlikte erkekler lehine olan maddelerin günlük yaşama dayalı problem çözme niteliği taşıdığı belirlenmiştir. Potansiyel kaynaklar olarak öğrenci düzeyi değişkenleri modele eklendiğinde, DMF büyüklüğünün azalmakla birlikte ortadan kalkmadığı belirlenmiştir. Öğrencinin SED düzeyinin ve öğretmen davranışları değişkenlerinin öğrencinin maddeleri yanıtlama olasılığını farklılaştırdığı da elde edilen bulgular arasındadır.

Anahtar Kelimeler

PISA 2012 matematik performansı, Değişen madde fonksiyonu, Genelleştirilmiş hiyerarşik doğrusal modeller, Lojistik regresyon, Öğretmen davranışları

*Niğde Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü Ölçme ve Değerlendirme ABD, ozen19@gmail.com

** Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, omerkutlu@ankara.edu.tr

*** Artvin Çoruh Üniversitesi, Eğitim Fakültesi Eğitim Bilimleri Bölümü Ölçme ve Değerlendirme ABD, haticekumandas@yahoo.com

Paper No: 144

Multi-Level Models to Investigate Differential Item Functioning and Teacher Behaviors as Potential Sources in PISA 2012 Math Test

Abstract

A multilevel approach was proposed for the assessment of differential item functioning and for determining student's characteristics as the sources of DIF. Data from PISA 2012 for 1458 15 years old students were analyzed. Existing data were used from PISA 2012 and they were obtained from math cognitive domains tests and student questionnaire. The multilevel approach used two-level hierarchical generalized linear models and logistic regression was used as traditional approach. 13 dichotomously coded items from math assessment were analyzed from Booklet1, 3, 4 and 6. According to results, it was not surprising to see that HGLM was more conservative in identifying DIF in this study. HGLM showed that three items were including gender DIF and logistic regression found six items with DIF. Three of them were common. It was also found that one item was favoring for girls, the rest of them were for males. Gender DIF favoring male students was more noticeable in the cognitive levels of employing and formulating. It is interpretation for girls. Female students also performed less well on items requiring daily-life experience and problem solving. When student-level variables as potential sources of DIF were added, it was seen that the magnitude of DIF narrowed but DIF not disappeared. Two level model results suggested that student's ESC and some teacher variables at the student level may be the cause of DIF variations.

KeyWords

PISA 2012 math performance, Differential item functioning, Hierarchical generalized linear models (HGLM), Logistic regression, Teacher behaviours

Bildiri No: 145

PISA 2012 Matematik Okuryazarlığı Testinde Farklı Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması

Şeyma UYAR, Burcu AKSEKİOĞLU, Neşe ÖZTÜRK GÜBEŞ[3]

Özet

Test eşitleme, test formlarından alınan puanları eşitlemede kullanılan istatistiksel bir süreçtir. Büyük ölçekli sınavların yılda birden fazla yapılması ya da aynı uygulamada güvenlik sebebiyle farklı formların kullanılması ile test eşitleme önem kazanmıştır. Aynı anda birden fazla formun kullanıldığı test uygulamalarında kitapçıkların eşdeğer olup olmadığının sınanması gerekmektedir. Bu çalışmanın amacı PISA 2012 matematik okuryazarlığı testinde kullanılan iki kitapçığın istatistiksel olarak eşdeğer olup olmadığını farklı yöntemlerle incelemek ve en az hata veren yöntemi tespit etmektir. Bu amaçla iki kitapçıktan elde edilen puanlar madde tepki kuramına dayalı ölçek dönüştürme yöntemleri (ortalama-ortalama, ortalama-standart sapma, Stocking-Lord, Haebara) ile eşitlenmiş ve kullanılan yöntemler hata miktarlarına göre karşılaştırılmıştır.

Bu kapsamda öncelikle kitapçıklar incelenmiş ve en fazla maddenin yer aldığı 4 ve 11 numaralı iki kitapçık analize dahil edilmiştir. Kitapçıklarda 12'si ortak olmak üzere toplam 35'er madde yer almaktadır. Her iki kitapçıkta da çok kategoride puanlanan 3 madde bulunmaktadır. Buna göre araştırmanın çalışma grubunu Türkiye örnekleminde 4 numaralı kitapçığı cevaplayan 348 ve 11 numaralı kitapçığı cevaplayan 368 olmak üzere toplam 716 öğrenci oluşturmaktadır. Çalışmada test eşitleme için denk olmayan gruplarda ortak madde deseni kullanılmıştır. Verilerin analizinde öncelikle madde tepki kuramının varsayımları test edilmiştir. Ardından PARSCALE 4.1 programı ile madde ve yetenek parametreleri kestirilmiştir. Parametre kestiriminde 2 parametreli model ve genelleştirilmiş kısmi kredi modeli kullanılmıştır. Daha sonra STUIRT programı ile dört farklı yönteme göre ölçek dönüştürme işlemi gerçekleştirilmiştir. Son aşamada ise her iki formdan elde edilen test puanları POLYEQUATE programı ile MTK gerçek puan ve MTK gözlenen puan eşitleme yöntemleri kullanılarak eşitlenmiştir. Farklı yöntemlerden elde edilen hata miktarları ise WMSE katsayısı ile hesaplanmıştır.

Çalışma sonucunda en az hata miktarına sahip yöntemin gerçek puan eşitlemede Stocking-Lord yöntemi, gözlenen puan eşitlemede ise Haebara yöntemi olduğu bulunmuştur. En yüksek eşitleme hatasını ise ortalama-standart sapma yönteminin verdiği tespit edilmiştir. Ayrıca genel olarak gerçek puan eşitleme yönteminin gözlenen puan eşitlemeden daha düşük hata oranı verdiği görülmüştür.

Anahtar Kelimeler

Test Eşitleme, Madde Tepki Kuramı, Karma Test, Ölçek Dönüştürme Yöntemleri

[33]Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: syuksel@mehmetakif.edu.tr. Tel: 02482134135

[2]Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: baksekioglu@mehmetakif.edu.tr. Tel: 02482134131

[3]Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: nozturk@mehmetakif.edu.tr. Tel: 02482134039

Paper No: 145

Comparison of Different Scale Linking Methods in PISA 2012 Mathematics Literacy Test

Abstract

Test equating is a statistical process to equate scores obtained from different test forms. Because of the application of large-scale tests several times a year or using different forms in a same application due to security reasons, test equation process has gained great importance. In test applications that is used different forms at the same time are made requires equivalence of different booklets to be assessed. The purpose of this study is to examine to equivalence of two booklets which is used in PISA 2012 mathematics litaracy test via different methods statistically and determine the method which have the lowest error. For this purpose, the scores obtained from different booklets were equated with different scale linking methods (mean-mean, mean-sigma, Stocking-Lord, Haebara) based on item response theory and methods were compared in terms of the amount of errors.

In this context, booklets were examined firstly and 4th and 11th booklets, which have maximum items, were included to analyze. In booklets there were 35 items and 12 items were common in these booklets. There are 3 polytomous items in these booklets. The sample consists of 716 students in Turkey; 348 of these participants are the takers of booklet 4, 368 of them are the takers of booklet 11. In order to equate test forms "the common-item nonequivalent groups" design was used in this research. In data analysis process, firstly, the assumptions of item response teory were tested. Then PARSCALE 4.1 was used to estimate item and ability parameters. Generalized partial credit and 2 PLM model were used to estimate parameters. Afterwards STUIRT program was used for scale transformation for four different methods. In the last step IRT true score and IRT observed score equating were conducted and test scores obtained from different forms were equated by using POLYEQUATE program. Equating error obtained from different methods calculated with WMSE index.

Results showed that Stocking-Lord method had the smallest equating error in true score equating and Haebara method had the smallest equating error in observed score equating . The amount of maximum error has been established that of the mean-sigma method. In addition,generally IRT true score equating method have lower amount of error than IRT observed score equating.

Key Words

Test Equating, Item response Theory, Mixed Test, Scale Linking Methods

Bildiri No: 146

Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Örneklem Büyüklüğünün Ölçme Kesinliğine Etkisinin İncelenmesi

Hülya KELECİOĞLU, Begüm ÖZTEMUR, Melike ÖZER TAYMUR, Şule ÖTKEN

Özet

Bireyselleştirilmiş testlerde, geleneksel testlerden farklı olarak test algoritması söz konusudur. Test algoritması; teste başlama, devam etme ve testi sonlandırma olmak üzere üç bölümden oluşmaktadır. Bu çalışmanın amacı, bireyselleştirilmiş bilgisayarlı test uygulamasında farklı örneklem büyüklükleri, farklı madde seçme yöntemleri, farklı yetenek kestirim yöntemleri ve farklı durdurma kurallarıyla oluşturulan koşullar altında ölçme kesinliğinin ve yetenek kestirimi ile madde seçme yöntemlerine ait istatistiklerin nasıl değiştiğini incelemektir. Araştırma simülasyon çalışması olarak yürütülmüştür. Araştırma kapsamında 1000 ve 5000 kişilik iki örneklem, dört madde seçme yöntemi, iki yetenek kestirim yöntemi, sabit uzunluk, SH<0,2 ve SH<0,4 olmak üzere üç farklı sonlandırma kuralı kullanılmıştır. Her bir örneklemde farklı koşullar söz konusu olup toplam 48 koşul birbiriyle karşılaştırılmıştır. Ayrıca örneklemlerin karşılaştırılmasında BBT'de test algoritmasında önemli yere sahip olan yetenek kestirim yöntemleri (Maksimum Olabilirlik Kestirimi ve Beklenen Sonsal Dağılım) seçilmiştir. Her bir BBT uygulamasında ölçme kesinliği için RMSE, yanlılık ve uyum değerleri hesaplanmış ve farklı koşullardan elde edilen bu değerler birbirleriyle karşılaştırılmıştır. Araştırmanın sonucunda genel olarak, 5000 kişilik örneklemde EAP yetenek kestirimi yönteminin ve sonlandırma kuralı olarak sabit uzunluk 40 madde kullanıldığı koşullarda RMSE, yanlılık değerlerinin düşük elde edildiği ancak uyum katsayılarının önemli oranda etkilenmediği belirlenmiştir.

Anahtar Kelimeler

Bireyselleştirilmiş Bilgisayarlı Test Uygulamaları, Ölçme Kesinliği

Paper No: 146

Investigating The Effect Of Sample Size On Measurement Precision In Computerized Adaptive Testing

Abstract

Different from classical tests, there is a test algorithm in adaptive testing. The test algorithm consists of three parts which are starting, resuming and termination of the test. The aim of this study is to investigate the changes in measurement precision and statistical changes in ability estimation and item selection methods under conditions specified with different sample size, different item selection methods, different ability estimation methods and different termination rules in computerized adaptive testing. The research has been implemented as a simulation study. Two different sample sizes (a sample of 1000 participants and another of 5000 participants), four different item selection methods, two ability estimation methods and three different termination rules which are fixed length, SE<0.2 and SE> 0.4 have been used for the study. Different conditions are in place in each sample size and a total of 48 conditions are compared. Additionally, for comparing the two sample sizes ability estimation methods (Maximum Likelihood Estimation and Expected Posterior Distribution) that are critical in the algorithms of Computer Adaptive Testing (CAT) have been used. For the measurement precision of each CAT applications RMSE, bias and fidelity values are calculated and these values which are obtained in different conditions are compared with each other. As a result, for the sample of 5000 participants and in the condition specified with EAP ability estimation method and 40 item fixed length termination rule, it is found out that RMSE and bias values are small but fidelity factors are not significantly affected.

Key Words

Computerized Adaptive Testing (CAT), Measurement Precision

Bildiri No: 149

Altıncı Sınıf Fen Bilimleri Dersi Elektriği İletelim Ünitesinin Değerlendirilmesinde Kullanılan Ölçme Araçlarının Güvenirliğinin Genellenebilirlik Kuramına Göre Analizi[4]

Banu BAŞAT[5] **Doç. Dr. Gülşah BAŞOL[6]**

Özet

Araştırmanın temel amacı; Genellenebilirlik Kuramına göre öğrencilerin performanslarının ölçülmesinde kullanılan çoktan seçmeli test, yapılandırılmış grid ve performans görevi gibi ölçme araçlarının Genellenebilirlik (G) ve Karar (K) çalışmaları sonuçlarını karşılaştırmaktır. Araştırmanın evrenini 2014-2015 Eğitim-Öğretim yılında Akdeniz bölgesindeki bir ilde öğrenim görmekte olan ortaokul altıncı sınıf öğrencileri oluşturmaktadır. Araştırmanın örneklemi iki okulda altıncı sınıf düzeyinde öğrenim gören rastgele seçilmiş iki sınıftan 23 kız 17 erkek öğrenci oluşturmaktadır. Araştırmada öğrencilere araştırmacılar tarafından geliştirilen çoktan seçmeli test, yapılandırılmış grid ve performans görevi uygulanmıştır. Yapılandırılmış grid ve performans görevi değerlendirilmesinde üç puanlayıcı görev almıştır. Elde edilen sonuçlara göre uygulanan çoktan seçmeli testin KR-20 güvenirlik katsayısı .68 bulunurken G katsayısı .68, Phi (Φ) katsayısı .64 bulunmuştur. Yapılandırılmış gridin Cronbach Alfa değeri .96, G katsayısı .91 ve Phi(Φ) katsayısı .83 bulunmuştur. Performans görevi sonuçlarına bakıldığında Cronbach Alfa değeri .89, G katsayısı .89 ve Phi (Φ) katsayısı .86 bulunmuştur. Ölçme araçları ile yapılan karar çalışmaları sonucunda her üç ölçme aracı için bulunan G ve Phi katsayılarının değişiminin paralellik gösterdiği görülmüştür. Yapılandırılmış grid ve performans görevini üç puanlayıcı değerlendirmiş olup puanlayıcılar arasındaki korelasyona bakılmıştır. Puanlayıcıların öğrencileri puanlamada aralarında farklılıklar olmadığı bulunmuştur.

Anahtar Kelimeler

Ölçme Araçları, Genellenebilirlik Kuramı, Güvenirlik.

[4] Bu çalışma, ilk yazarın aynı isimle Doç. Dr. Gülşah BAŞOL danışmalığında hazırlamış olduğu yüksek lisans tez çalışmasının bir kısmından oluşmaktadır.

[2] Halil Akyüz Ortaokulunda Kepez /ANTALYA Elektronik posta: banusevran@hotmail.com

[6] Gaziosmanpaşa Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Taşlıçiftlik Yerleşkesi, 60100 Tokat. Elektronik posta: gulsah.basol@gop.edu.tr & gulsahbasol@gmail.com.

Paper No: 149

A Generalizability Analysis Of The Reliability Of Measurements, Applied In The Sixth Grade Science Course "Let's Circuit Electric" Unit.

Abstract

The purpose of the study was to compare the results and reliability of G and K studies of measurement instruments which were utilized in the current study; a multiple-choice test, a structured grid, and a performance task applied for measuring students' performance based on the generalizability theory. The target population of the study was the 6th grade students in a province in Mediterranean region of Turkey during the 2014-2015 school years. Forty students (23 girls and 17 boys) from two 6th grade classes from two middle schools were randomly selected for the study. In the study, a multiple-choice test, a structured grid, and a performance task, developed by the researchers, were utilized to measure students' achievement in each. Three teachers graded the structured grids and performance tasks. Later, G and K studies were implemented on the results. According to the results, the KR 20 reliability coefficient of the multiple-choice test was found as .68; meanwhile G and Phi(Φ) coefficients were .68 and .64, respectively. The Cronbach's Alpha coefficient of structured grid was .96 and G coefficient was .91, whereas Phi(Φ) coefficient found as .83. The results of performance utility indicated a Cronbach's Alpha coefficient of .89; G coefficient was .89, and Phi(Φ) coefficient found as .86. According to results of decision studies on measurement instruments, variation of G and Phi coefficient for all three measurements were in line.

The assessments by three graders for the structured grid and performance task were evaluated and the correlations among them were analyzed. It was found that there were no differences among graders while grading students.

KeyWords

Measurement Instruments, Generalizability, Reliability.

Bildiri No: 150

Pısa 2012'deki Matematiğe Yönelik Duyuşsal Özelliklerin Bölge, Okul Türü Ve Cinsiyete Göre Sınıflama Doğruluğunun İncelenmesi

Emrah BÜYÜKATAK*, Duygu ANIL**

Özet

Bu araştırmanın amacı, öğrencilerin bölge, okul türü ve cinsiyet değişkenlerinin PISA 2012 Türkiye Öğrenci Anketi verilerinden elde edilen Matematiğe Yönelik Duyuşsal Özellikler açısından ne düzeyde doğru sınıflama yaptığının belirlenmesidir. Çalışma değişkenlerin duyuşsal özellikler açısından ne düzeyde doğru sınıflama yaptığının belirlenmesine yönelik olarak var olan durumun tanımlanması yönüyle betimsel araştırmadır.

Araştırmada Uluslararası Öğrenci Değerlendirme Programı 2012 çalışmasında yer alan 4848 öğrenci anketi kullanılmıştır. Birinci aşamada alt problemler dikkate alınarak araştırmanın amacı kapsamında diskriminant analizinin varsayımları test edilmiştir. İkinci aşamada Matematiğe Yönelik Duyuşsal Özelliklere ait değişkenler çerçevesinde yapısal model oluşturularak incelenmiştir. Üçüncü aşamada ise alt problemlerin sınıflama doğruluğu diskriminant analizi ile incelenmiştir. Araştırmadan elde edilen verilerin istatistiksel çözümlemeleri SPSS paket programı ile yapılmıştır.

Araştırma sonunda; Matematiğe Yönelik Duyuşsal Özelliklerin öğrenci performansı üzerindeki açıklayıcılıklarının düşük olması sebebiyle öğrencilerin bölge, okul türü ve cinsiyete göre sınıflandırmasında çok etkili olmadığı görülmektedir. Bu kapsamda Matematiğe Yönelik Duyuşsal Özelliklerin öğrencileri yaşadıkları bölgelere göre ayırmada kullanılamayacağı, bu özelliklerin okul türleri açısından anlamlı bir fark yaratmadığı fakat duyuşsal özelliklerin cinsiyete göre çok anlamlı olmasa dahi şansın ötesinde doğru sınıflandırma yapabilmesi sebebiyle öğrencileri ayırmada kullanılabileceği sonucuna varılabilir.

Anahtar Kelimeler

PISA, duyuşsal özellikler, sınıflama doğruluğu, diskriminant analizi, lojisitk regresyon.

* Hacettepe Üniversitesi yüksek lisans öğrencisi, Cep: 0 555 271 32 86 Mail: ebuyukatak@hotmail.com
** Hacettepe Üniversitesi. Cep: 0 505 8134873 Mail : duygu.anil73@gmail.com

Paper No: 150

Investıgatıon Of Classıfıcatıon Accuracy Of Mathematıc Affectıve Characterıstıcs As To Regıon, School Type And Gender In Pısa 2012 Data

Abstract

The purpose of this research was to determine at what level variables of regions, the school types and gender of students classify the correctly in terms of Mathematic Affective Characteristics obtained from PISA 2012 Student Survey. The research is a descriptive study in that level of correct classification of variables in terms of affective characteristics was identified expecting of existing situation.

In the research, 4848 student surveys located in Program for International Student Assessment 2012 were used. In the first phase, assumptions of discriminant analysis were tested considering sub-problems in the scope of the purpose of the research. In the second phase structural model were created in the framework of features of Mathematic Affective Characteristics and missing data and extreme values were examined. In the third stage, the accuracy of the sub-classification problem was investigated by discriminant analysis. Statistical analysis of the data obtained from the study was conducted with SPSS.

At the end of the research, it was seen that due to the low explanatory of Mathematic Affective Characteristics on student performance, Mathematic Affective Characteristics were not very effective in the classification according to regions, type of school and gender of students In this context, it can be concluded that Mathematic Affective Characteristics cannot be used to differentiate students according to the area they live and these features did not make a significant difference in the types of schools. But since they classify correctly beyond the chance, Affective Characteristics can be used to differentiate students according to their gender even though it is not very significant.

Key Words

PISA, affective characteristics, classification accuracy, discriminant analysis logistic regression.

Bildiri No: 151

8. Sınıf Öğrencilerinin Temel Eğitimden Ortaöğretime Geçiş Sınavı (TEOG) Kavramı Algıları Metaforları

Şemsettin ERCAN*, Devran BAYADUR**, Mürşet ÇAKMAK***

Özet

Bu çalışmanın amacı 8. sınıf öğrencilerinin Temel Eğitimden Ortaöğretime Geçiş Sınavı (TEOG) kavramı ile ilgili algılarını metaforlar ile incelemektir. Çalışma, Mardin ilinin Kızıltepe ve Artuklu ilçelerindeki 2015-2016 eğitim-öğretim yılında dört ortaokulda okuyan öğrencilerle (n=160) gerçekleştirilmiştir. Nitel olarak yapılan bu çalışmadaki veriler, "TEOG Sınavı ……..gibidir, çünkü ……." şeklindeki bir formu ile toplanmıştır. Bu formda yer alan ilk boşlukta "TEOG Sınavı" ile ilgili bir benzetme yapmaları, ikinci boşlukta ise bu benzetmenin nedenini açıklamaları istenmiştir. Veriler içerik analizi ile analiz edilmiştir. Öğrencilerin yapmış oldukları bütün metaforlar ve açıklamaları incelenerek ortak özellikleri belirlenmiş ve bunların temsil edebileceği düşünülen temalar oluşturulmuştur. 160 öğrencinin TEOG sınavı kavramı ile ilgili 118 yaptıkları benzetmeler ve temalar çıkartılarak yorumlanmış ve önerilerde bulunulmuştur.

Anahtar Kelimeler

TEOG sınavı, metafor, öğrenci, algı.

** Mardin Artuklu Üniversitesi Sosyal Bilimler Enstitüsü Eğitim Bilimleri Bölümü. Email:dbayadur@gmail.com
***Mardin Artuklu Üniversitesi Edebiyat Fakültesi Eğitim Bilimleri Bölümü.Email:mursetcakmak@artuklu.edu.tr

Bildiri No: 152

Gelişimin Ölçülmesinde Örtük Gelişim Modelleme ve Aşamalı Doğrusal Modellemenin Model-Veri Uyumlarının Örneklem Büyüklüğüne Göre Karşılaştırılması

Gürkan CÜVİTOĞLU*, Hakan ATILGAN**

Özet

Birçok bilim dalında ölçülmek istenen yapıların nitelikleri gereği ya da ilgili yapılardaki gelişimin izlenmek istenmesinden toplanan veriler hiyerarşik ve/veya boylamsal özelliktedir. Bu tür verilerin analizinde ilişkili örneklem için t testi, RMANOVA, MANOVA gibi geleneksel yöntemlerin yanında son 30 yılda hızlı bir gelişim gösteren Gelişme Modelleri (GM) kullanılmaktadır. Geleneksel yöntemlerde gözlenen değişkenlerdeki hata varyansı göz ardı edilmekte ve örtük yapılardaki değişişim yerine gözlenen değişkenlerdeki değişim modellenmektedir. GM ise hata varyansı açıklanması gereken bir etki olarak ele alınmakta ve modele eklenen birtakım açıklayıcı değişkenler aracılığıyla yordanabilmektedir. Zamana bağlı değişimin incelenmesi amacıyla sıklıkla kullanılan yöntemler Aşamalı Doğrusal Modelleme (ADM) ve Örtük Gelişme Modelidir (ÖGM). Her iki yöntem de kesişim ve eğim katsayılarını kullanarak bireyiçi ve bireylerarası değişimi analiz eder. ÖGM tüm bireyler için alınan tekrarlı ölçümler alındıkları ölçüm zaman noktasına göre ölçüm modeline dâhil edilirken, ADM yaklaşımda ise öğrencilerden alınan tekrarlı ölçümler öğrenci düzeyinin içine yuvalanır. Bu yöntem farklı kestirim ve modelleme yaklaşımlarına sahip olmasına rağmen benzer sonuçlar üretmekle birlikte modellerin birbirlerine göre üstünlükleri ve sınırlıkları bulunmaktadır. Bu çalışmada örneklem büyüklüğü ile birlikte ele alındığında, bu üstünlük ve sınırlıkların model veri uyumunu nasıl etkilediği araştırılacaktır.

Çalışmada Joop Hox'un kitabında kullandığı GPA2 veri seti kullanılacaktır. GPA2 veri seti 200 öğrencinin 6 dönem boyunca not ortalamalarının takip edildiği boylamsal ve simülatif bir veri setidir. Veri seti; 6 farklı zaman noktasında elde edilmiş not ortalamaları dışında cinsiyet, çalışma saati ve lise ortalaması gibi açıklayıcı değişkenler içermektedir. Analizleri R Studio ve Mplus programları aracılığıyla yapılacaktır.

Çalışmada farklı örneklem büyüklükleri altında tekrarlı 4, 5 ve 6 ölçümlü modeller, açıklayıcı değişkenlerin bulunduğu ve bulunmadığı ile kayıp verinin olduğu ve olmadığı durumlarda 2 ve 3 düzeyli modeller model veri uyumu kay kare, CFI, TLI, RMSEA, SRMR, BIC ve AIC gibi indeksleri ile test edilecektir. Gelişim eğrilerinin kesişim ve eğim katsayıları yorumlanarak bireyiçi, bireylerarası ve gruplar arası gelişim yorumlanacaktır.

* Öğretmen, Ege Üniversitesi, Sosyal Bilimler Enstitüsü Eğitimde Ölçme ve Değerlendirme Bilim Dalı Yüksek Lisans Öğrencisi. Eposta: gurkancuvitoglu@gmail.com

** Doç. Dr. Ege Üniversitesi Eğitim Fakültesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı öğretim üyesi. Eposta: atilgan.h@gmail.com

Paper No: 152

Comparison of the Model Data Fit of Latent Growth Modeling and Hierarchical Linear Modeling in terms of Sample Size for the Measurement of Growth

Abstract

In most scientific studies, data obtained are hierarchical and/or longitudinal due to the properties of the structures to be measured or because it is aimed to observe the development in the concerning structures. In the analysis of such data, in addition to the traditional models such as Paired-Samples t test, RMANOVA and MANOVA; Growth Models (GM) have been used, which have had rapid improvement in the last three decades. In the traditional methods, error variance in the observed variables is ignored and the variation in the observed variables is modeled instead of that in the latent structures. In GM, error variance is dealt with as an effect that is to be explained and can be predicted via several explanatory variables added to the model. GM that are frequently used to examine the time-bound variation are Hierarchical Linear Modeling (HLM) and Latent Growth Modeling (LGM). Both methods analyze intra-individual and inter-individual variations using intercept and slope coefficients. In LGM all repeated measures taken for all individuals are included in the model according to their measurement time points while repeated measures taken from students are nested into student levels in HLM. Although this method has different predicting and modeling approaches, it produces similar results while models have advantages and limitations over each other. The present study will examine how these advantages and limitations affect model fit considered together with the sample size.

GPA2 data set used in Joop Hox's book will be employed in the study. GPA2 data set is a longitudinal and simulative one on which grade averages of 200 students are followed during 6 semesters. In addition to the grade averages obtained at 6 different time points, the data set includes such explanatory variables as gender, study time and high school averages. Analyses will be carried out on R Studio and Mplus programs.

GMs that have 4, 5 and 6 time points of measurement, with missing and complete data, with and without explanatory variables, 2 and 3 levels will be analyzed (fitted) under different sample size conditions. For both GM approaches model fit will be tested with Chi-Square test, RMSEA, RMSR, AIC, BIC, TLI, CFI indexes in this study. Interpreting the intercept and slope coefficients of the development curves, intra-individual, inter-individual and inter-groups development will be evaluated.

Bildiri No: 153

Hakem Katılığının Yapılandırılmış-Yanıtlı Maddelerde Yetenek ve Parametre Kestirimine Etkisi

T. Oğuz BAŞOKÇU*, Sevcan GÖKÇE

Özet

Performans değerlendirme, cevaplama sınırlarının daha geniş olması ve puanlama çeşitliliği bakımından tercih edilen bir ölçme yöntemidir. Ancak, Messick'in (1994) de belirttiği gibi, diğer değerlendirme yöntemlerine göre, yapı ve konu dışı değişkenliğinin bozucu etkisi daha fazla görülmektedir. Bu nedenle performans değerlendirme literatüründe, hata kaynakları önemli bir çalışma alanıdır (Cantor & Hoover, 1986; Engelhard, 1992, 1994; Engelhard, Gordon & Gabrielson, 1991; Gabrielson, Gordon & Engelhard, 1995; McNamara, 1996; ve Ruth & Murphy, 1988). Bu hata kaynakları arasında en çok öne çıkan puanlayıcılar ile ilgili olan değişkenliktir. Ne kadar dikkatle yapılandırılsa da, bir puanlama ölçeğinin güvenirliği, o ölçeği kullanan puanlayıcılara önemli düzeyde bağımlıdır (Overall & Magee, 1992). Dunbar, Koretz ve Hoover (1991) hataya düşebilen puanlayıcıların, puanların objektifliğini etkileyebileceğini ve objektif şekilde puanlanabilen testlerde ortaya çıkmayan bir terimi güvenirlik denklemine ekleyebileceğini belirtmişlerdir.

Bu araştırma hakem puanlamalarında Muraki'nin puanlayıcı etkisi modeli ve Yao'nun puanlayıcı modelini kullanarak, hakem katılığının puanlama üzerindeki etkisini belirlemeyi amaçlamaktadır.

Anahtar Kelimeler

Hakem katılığı, Performans değerlendirme**,** Muraki'nin puanlayıcı etkisi modeli, Yao'nun puanlayıcı modeli

*Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. tahsin.oguz.basokcu@ege.edu.tr

Paper No: 153

The Effects of Raters Severity on Ability and Parameter Estimations for Constructed-Response Items

Abstract

Performance assessment is a preferred method for that it has extended response boundries and variety of scoring. However, as Messcik (1994) points out, according to other forms, in performance assesment distorting effect of construct-irrelevant variance is mostly seen. Therefore, sources of error is important field of study in the literature on performance assesment (Cantor & Hoover, 1986; Engelhard, 1992, 1994; Engelhard, Gordon & Gabrielson, 1991; Gabrielson, Gordon & Engelhard, 1995; McNamara, 1996; and Ruth & Murphy, 1988). Prominent among these sources is the variance associated with raters. This is a reflection of the concern that, no matter how carefully constructed, the reliability of a rating scale is critically dependent on the raters who operate it (Overall & Magee, 1992). As Dunbar, Koretz ve Hoover (1991) put that fallible raters can wreak havoc on the trustworthiness of scores and add a term to the reliability equation that does not exist in the tests that can be scored objectively.

This study aims to determine effect of rater severity on scoring by using Muraki rater effect model and Yao rater model.

Key Words

Rater severity, Performance assessment, Muraki rater effect model, Yao rater model

Bildiri No: 154

BTM' de Ankor Test Deseni İle Gözlenen Puanlara Dayalı Lokal Eşitleme Çalışması

T. Oğuz BAŞOKÇU*, Arzu CANPOLAT

Özet

Zincir-Eşitleme ve tabaklama gibi ankor test desenli geleneksel gözlenen puan eşitleme yöntemlerinde, ölçüm değişmezliğini ve eşitlik kriterini sağlama konusundaki zorluklar nedeniyle, yapılan eşitleme çalışmaları yanlılık içerme riski taşımaktadır. Lokal eşitleme, bu yanlılığa çözüm bulmak için geliştirilmiş bir yöntemdir. Lokal eşitleme (van der Linden, 2000, 2006a, 2010), farklı test cevaplamaları sonucunda ortaya çıkan dağılımların eşitlenmesi yerine dağılımlar ile ilişkili önsel bilgileri kullanarak bu dağılımları kestirmeyi amaçlamaktadır. Bu tahminler, X ve Y gibi iki farklı testi cevaplayan bireylere ait tek bir dönüşüm kestirmek yerine, iki popülasyon için gerçek puan dönüşümleri üzerinden dağılım fonksiyonu kestirimleri belirlemek için kullanılmaktadır. Çalışmada, lokal eşitleme yöntemi, Bilişsel Tanı Modellerine (BTM) göre geliştirilmiş ve öğrenci yeteneklerinin bir örtük değişken olarak tanımlandığı durumlar için simülasyon çalışması ve gerçek uygulama verileri kullanılarak incelenmiştir.

Anahtar Kelimeler

Lokal eşitleme, Bilişsel tanı modeli, Ankor test deseni

*Ege Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. tahsin.oguz.basokcu@ege.edu.tr

Paper No: 154

Study in Local Equating of Cognitively Diagnostic Modeled Observed Scores with Anchor-test Design

Abstract

For traditional methods of observed-score equating with anchor-test designs, such as chain and poststratification equating, since it is difficult to satisfy the criteria of equity and population invariance, their equating's have risk of bias. Local equating (van der Linden, 2000, 2006a, 2010) aim to approximate individual distributions using whatever information is available about the test takers rather than to compromise between them. The set of transformations for the individual test takers calculated from these approximations is expected to be closer to their true equating transformations than any single transformation based on the two populations that took X and Y. In this study local equating method was developed based on cognitive diagnosis models. For the situations like students ability defined as an implicit variable was analyzed with the help of using real and simulation data.

Key Words

Local equating, Cognitive diagnosis model, Anchor test design

Bildiri No: 155

Öğretmen Adaylarının Akranları ve Öğretim Elemanları Tarafından Değerlendirilmelerinden Elde Edilen Puanların Güvenirliğinin Karşılaştırılması

Gizem UYUMAZ*, Nimet AKBEN**

Özet

Öğrenci merkezli eğitim anlayışı öğrencilerin öğrenme etkinliklerine dahil edilerek, öğrenmelerinin farkında ve öğrenmelerinden sorumlu olmalarının sağlanmasını gerektirir. Bunu sağlayabilmenin temeli öğrencilere hangi ölçütlerde değerlendirilecekleri ve bu değerlendirmelerin nasıl yapılacağı bilgisinin verilmesidir. Öğrencilerin değerlendirme sürecine aktif katılmalarının sağlanmaya çalışıldığı bu çalışma, fizik öğretmen adaylarının öğretmenlik uygulaması dersi çerçevesinde gerçekleştirdikleri sunumlardaki performanslarının, aynı dersi alan akranları ve dersi izleyen öğretim elemanları tarafından belirli ölçütlere göre değerlendirilmelerinden elde edilen puanların güvenirliğinin genellenebilirlik kuramı ile karşılaştırılmasını amaçlayan betimsel bir araştırmadır. Değerlendirmeler Yüksek Öğrenim Kurumu (YÖK) tarafından oluşturulan Ders Gözlem Uygulaması Değerlendirme Formu'nda yer alan bazı ölçütlere göre yapılmıştır. Sunumlarını hazırlama sürecinde adaylar bu ölçütlere göre değerlendirilecekleri konusunda bilgilendirilmiştir. Her bir öğretmen adayının performansı 2 öğretim elemanı ve 10 akranı tarafından değerlendirilmiştir. On altı öğretmen adayının yaptıkları sunumlardaki yeterlilikleri, puanlayıcılar tarafından 3 farklı boyutta yer alan on dört madde için puanlanmıştır. Bu boyutlar konu alanı bilgisi (4 madde), alan eğitimi bilgisi (4 madde) ve dersin planlanması (6 madde) boyutlarıdır. Çalışmada tümüyle çaprazlanmış desen (öğretmen adayı × puanlayıcı × boyut) kullanılarak genellenebilirlik ve karar çalışmaları yapılmıştır. Öğretim elemanlarının puanlayıcı olduğu durumda en yüksek varyans açıklayan varyans kaynağının öğretmen adaylarından kaynaklanan varyans olduğu, akran öğretmen adaylarının puanlayıcı olduğu durumda ise en yüksek varyans açıklayan varyans kaynağının, üç kaynağının birlikte etkileşimine yani öğretmen adaylarının farklı boyutlarda farklı puanlayıcılardan aldıkların puanlara ve belirlenemeyen (tesadüfi) hatalara ilişkin varyans olduğu bulunmuştur. Öğretim elemanlarının yaptığı değerlendirmeler, akran değerlendirmelerinden daha yüksek güvenirlik değerine sahiptir. Öğretmen adayları akranlarını puanlarken yansız davranamamıştır. Yapılan karar (K) çalışmalarında, değerlendirilen boyut sayısı ve değerlendirme yapan puanlayıcı sayısı arttıkça güvenirliğin arttığı bulunmuştur.

Anahtar Kelimeler

Genellenebilirlik, güvenirlik, akran değerlendirme, performans değerlendirme

*Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Ölçme ve Değerlendirme Anabilim Dalı. İletişim: gizemuyumaz@akdeniz.edu.tr Tel:02422274400 / 4645

**Ankara Üniversitesi Eğitim Bilimleri Fakültesi İlköğretim Bölümü. İletişim: nakben@ankara.edu.tr Tel:+903633350/5206

Paper No: 155

Comparison of Reliability of the Points Obtained through Evaluations Made by the Peers of Pre-Service Teachers and Lecturers

Student-centered education requires to include students in learning activities for awaring of and being responsible for their learning. In order to provide this, it should be given information of which criterions are used in students' assessment and how these assessments are made. This study, which is provided to enable students in assessment process actively, is a descriptive research which aims to compare pre-service physics teachers' presentation performances in teaching practice course, reliability of points assessed with some criterions by peers taking same course and instructors following course with the generazibility theory. Assessments were made by some criterions in Course Observation Application Assessment Form which are formed by Board of Higher Education (YÖK). In period of preparing presentations, pre-service physics teachers were informed that they were going to be assessed according to these criterions. The performance of each preservice teacher were assessed by 2 lecturers and 10 peers. 16 preservice teachers' sufficiency of presentations were given points by rater for 14 items which are located 3 different aspects. These aspects are subject area knowledge (4 items), field training knowledge(4 items) and lesson planning(6 items). In study, using completely crossed design (preservice teacher × rater × aspect), generalizability and decision studies were made.It was found that when the lecturers were raters, the variance source which declared the highest variance, was the one arising from the lecturers. In cases where the peers of the pre-service teachers were raters, the highest variance was the source of explanatory variance which was formed through co-interaction of three sources; in other words, it was obtained from pre-service teachers in different dimensions and raters and unidentified (random) errors. The evaluations of the lecturers were higher reliability value than peers review. The pre-service teachers were unsuccessful in scoring objectively. The decision (D) studies showed that the reliability increased as long as the number of dimension and rater of evaluation increased.

Keywords

Generalizability, reliability, peer assessment, performance assessment

Bildiri No: 158

Genel Öğretmen Yeterliliklerinin İkili Karşılaştırmalar Yöntemi ile Ölçeklenmesi

Haydar KARAMAN*, Atilla ÖZDEMİR**, Rasim ÖNDER***, Duygu ANIL****

Özet

Bu araştırmanın amacı Milli Eğitim Bakanlığı tarafından belirlenen genel öğretmen özelliklerinden hangisinin daha çok önemli olduğunun öğrenciler ve öğretmen adaylarına göre belirlenmesidir. Çalışmada MEB (2008) tarafından belirlenen 6 genel öğretmen yeterliliği, ikili olarak karşılaştırılmıştır. Bu araştırma 2015-2016 bahar döneminde Ankara ilinde 269 lise 1,2 3 ve 4. sınıflarında öğrenim gören öğrenci ve üç devlet üniversitesinin eğitim fakültesinin 2, 3 ve 4. sınıflarında 297 üniversite öğrencisinin katılımıyla yapılmıştır. Veri toplama aracı olarak araştırmacılar tarafından hazırlanan 6 yeterliliğin tüm ikili olarak karşılaştırıldığı yazılı ve elektronik araçlar kullanılmıştır. Araştırmanın yönteminde Thurstone tarafından geliştirilen ikili yargılar kanuna göre V. Hal ölçekleme tekniği kullanılmıştır ve elde edilen ölçeğin uyum indekslerine ve iç tutarlılığına bakılmıştır. V. Hal varsayımları sağlanmayan durumlar için III. hal denklem çözümlerine göre ölçekleme yapılmıştır. Elde edilen bulgulara göre tüm çalışma gruplarına göre en önemli genel öğretmen özelliği öğrenmeyi, gelişimi izleme ve değerlendirme olarak elde edilmiştir. Bu özelliğe sahip bir öğretmen değişik ölçme araçları kullanarak süreci ve öğrencileri değerlendirebilen, verileri analiz edebilen ve yorumlayabilen ve elde ettiği sonuçlara göre süreci yorumlayabilen öğretmendir. Program ve içerik bilgisi öğretmen adayları için en az öneme sahip olan özellik olarak elde edilmiştir. Öğrenciler için ise en az öneme sahip olan genel öğretmen özelliği kişisel ve mesleki gelişim olarak elde edilmiştir. Buradan öğretmenlerin seçim kriterleri gözden geçirilmeli, öğretmen adaylarına kavramsal eğitimden daha çok uygulamalı bir eğitim verilmesi önerilmektedir.

Anahtar Kelimeler
Ölçekleme, Öğretmen Yeterlikleri

*Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim:hkaraman@hacettepe.edu.trTel: +90543 915 93 05

** MEB. İletişim: atimaths06@gmail.comTel: +905063812683

***Süleyman Demirel Üniversitesi Eğitim Fakültesi İlköğretim Bölümü. İletişim: rasimonder@sdu.edu.trTel: +90532 060 10 08

****Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü. İletişim: aduygu@hacettepe.edu.trTel: +90505 813 48 73

Paper No: 158

Scaling of General Teacher Competencies by Using Pair-Wise Comparison Method

Abstract

The purpose of this research is to determine which of general teacher competence determined by MEB is much important with respect to students in high school and teacher candidates. In this study, Pair-Wise comparison method was used to analyze 6 general teacher competencies. This research was made in Ankara during 2015-2016 fall semester and the study group of this research consisted of 269 high college students between 9th and 12th class and 297 university students from three public universities in Ankara. Data collection tool prepared by researchers consisted of all possible binary comparison of 6 teacher competencies. It was prepared as written and electronic way, also. Data was analyzed by using the scaling methods of Thurstone's V according to law of comparative judgement and internal consistency and fit index was checked. If assumption of scaling methods of Thurstone V. law is not held, data was analyzed by scaling methods of Thurstone III. Law. Although internal consistency was high for students, it was low teacher candidates. According to results, the most important general teacher competence for two study groups was observing and evaluating the education process and development. This kind of teachers is able to evaluate development and education process by using different measurement tools, to analyze data and evaluate them and to evaluate the whole learning process with respect to analysis of data. Program and content knowledge was found mostly the least important competency of teachers. At this point, the selection of teachers should be revised and students should be given mostly practical education rather than conceptual education.

Key Words

Scaling, Teacher Competencies

Bildiri No: 159

Doğrulayıcı Faktör Analizinde Farklı Parametre Kestirim Yöntemi, Örneklem Büyüklüğü ve Modelin Hatalı Kurulmasının Uyum İndekslerine Etkisi

Ayfer SAYIN

Özet

Ölçme araçlarının uyarlama ve test aracı geliştirme sürecinde çapraz geçerlik çalışmalarında doğrulayıcı faktör analizi hesaplanmaktadır. Çok değişkenli bir istatistik olan doğrulayıcı faktör analizinin hesaplanmasında kullanılan birçok parametre kestirim yöntemi bulunmaktadır ve model-veri uyumunun değerlendirilmesinde hesaplanan farklı uyum indeksleri mevcuttur. Bu çalışmada, doğrulayıcı faktör analizinde farklı parametre kestirim yöntemleri ve farklı örneklem büyüklüklerinin uyum indekslerine etkisinin araştırılması; aynı zamanda model-veri uyum indekslerinin modelin hatalı kurulmasından ne düzeyde etkilendiğinin belirlenmesi amaçlanmıştır. PISA 2011 çalışmasında "matematiğin önemine yönelik görüşler" maddelerine 4848 öğrenci cevap vermiştir. Eksik veriler, veri setinden çıkarılmış ve varsayımların incelemesi gerçekleştirilmiştir. Kalan 3033 öğrenci cevapları içerisinden rastgele seçilen 60, 100, 250, 500 ve 1000 kişilik veri setleri seçilmiştir. Seçilen her bir veri setinin doğrulayıcı faktör analizinin varsayımları olan normallik, uç değer, eksik veri, eşvaryanslılık ve doğrusallık varsayımlarını karşılayıp karşılamadığı incelenmiştir. Ardından analiz aşamasında ilk olarak açımlayıcı ve doğrulayıcı faktör analizleri hesaplanarak maddelerin iki boyutta toplandığı belirlenmiştir. Ardından en çok olabilirlik, ağırlıklandırılmamış en küçük kareler ve genelleştirilmiş en küçük karelerparametre kestirim yöntemleri ile hesaplamalar gerçekleştirilmiştir.Ardından boyutlar içerisinde yer alan ve madde faktör yük değeri en küçük olan maddeler yer değiştirilmiş ve hatalı bir model oluşturulmuştur. Üç farklı parametre kestirim yöntemi ile hatalı model üzerinde kestirimler yinelenmiştir. Uyum indekslerinin değişimin incelenmesinde ki kare, RMSEA, GFI, CFI ve NFI uyum indeksleri incelenmiştir. Araştırmanın sonucunda varsayımların karşılanması durumunda en iyi kestirimlerin en çok olabilirlik yöntemi ile gerçekleştirildiği tespit edilmiştir. Uyum indekslerinin örneklem büyüklüğünden etkilendiği belirlenmiştir. Bazı uyum indekslerinin modelin hatalı kurulmasından etkilendiği, bazı uyum indekslerinin ise daha az etkilendiği saptanmıştır. Bu doğrultuda analizler için önerilerde bulunulmuştur.

Anahtar Kelimeler

Doğrulayıcı faktör analizi, örnek büyüklüğü, uyum indeksi, ölçek geliştirme ve uyarlama

Paper No: 159

TheEffect of MisspecifiedModels, ParameterEstimationMethods, andSampleSizes on Fit Indexes in ConfirmatoryFactor Analysis

Abstract

Inadaptational works of measurement practices and in terms of improving crossvalidity, confirmatory factor analysis is utilized. Confirmatory factor analysis, a multivariate statistic, is estimated via various parameter estimation methods and fit index for interpreting the model. In this work, the effect of different parameter estimation methods and various sample sizes on the fit indexutilized in confirmatory factor analysis are explored, and how faulty settings of models affect the model-data fit index is determined. There search group consists of students who attended the PISA 2011 in Turkey. In PISA, 4848 students answered the questions in the "views on the importance of mathematics" section. The missing data are removed from the data set and analysis of the assumptions are done. From the remaining 3033 student answers, clusters of 60, 100, 250, 500 and 1000 answers were pulled and processed within there search.Estimations were based on methods of maximum likelihood (ML), unweighted least squares (ULS) and generalized least squares (GLS). In there search, analysis of assumptions were carried out be for eanalysis of the data. Then the factors with the least factor loadings were determined and replaced to reach a misspecified, which the estimations were based upon. At the end of the research, maximum likelihood (ML) method is found to make the most accurate estimations. Fit indexes are found to be affected by the sample size, however some fit indexes were affected less than others and viceversa. Inorder to analyse these, some suggestions were made.

KeyWords

Comfirmatoryfactoranalysis, sample size, fit index, scaledevelopmentandadjustment

Bildiri No: 160

Performans Takip Sınavında (PTS) Kullanılan Alt Testlerin Psikometrik Özelliklerinin Resmi ve Özel Okullara Göre İncelenmesi

Bayram BIÇAK, Dilek GÜR, Evrim YALÇIN, Özge GÖÇER

Özet

Bu çalışmanın amacı Antalya ili kapsamında 8.sınıfa devam eden öğrenciler için yapılan Performans Takip Sınavı'nda (PTS) kullanılan çoktan seçmeli maddelerin psikometrik özelliklerini saptamak ve aynı zamanda sınavda kullanılan alt testlerin özel okul ve devlet okulu öğrencilerinden elde edilen veriler üzerinden psikometrik farklılıklarını belirlemektir. Bu çalışma PTS'de kullanılan alt testlerin psikometrik özellikleri incelendiği için bir betimsel tarama araştırmasıdır. Çalışmanın evrenini PTS kapsamında toplanmış olan 30800 8.sınıf öğrencisine ait veriler oluşturmaktadır. Verilerin %91,75'i devlet okullarında okuyan öğrencilerden, %8,25'i özel okullarda okuyan öğrencilerden oluşmaktadır. Örneklem, basit tesadüfi yöntem kullanılarak devlet okulları içinden seçilen 372 öğrenciden ve özel okullardan seçilen 307 öğrenciden oluşmaktadır. İki ayrı örneklemden elde edilen veriler üzerinden madde ve test istatistikleri yapılmıştır. Yapılan analizler sonucunda Matematik alt testinin devlet okulunda okuyan öğrencilere zor geldiği, özel okulda okuyan öğrencilere ise kolay geldiği belirlenmiştir. Türkçe alt testi ile din kültürü ve ahlak bilgisi alt testinin devlet okulunda okuyan öğrencilere kolay geldiği, özel okulda okuyan öğrencilere ise çok kolay geldiği belirlenmiştir. Fen ve Teknoloji, İnkılap Tarihi ve İngilizce alt testleri incelendiğinde ise devlet okulunda okuyan öğrencilere orta güçlükte geldiği, özel okulda okuyan öğrencilere ise çok kolay geldiği belirlenmiştir. Ayrıca testlerde kullanılan maddelerin ayırt etme gücünü belirlemek amacıyla çift serili korelasyon katsayıları (Rbis) hesaplanmış ve çoğunlukla .40 üzerinde değer almaları nedeniyle testlerin ayırt etme gücü yüksek çok iyi maddelerden oluştuğunu söylemek mümkündür. Her iki örneklem için hesaplanan madde varyans değerleri incelendiğinde özel okulda okuyan öğrencilerin Türkçe, İngilizce ve Din Kültürü ve Ahlak Bilgisi alt testlerindeki maddelere verdiği cevapların çok fazla değişmediği saptanmıştır. Yapılan ölçme sürecinde alt testlerin KR-20 iç tutarlılık güvenirlik katsayılarının .78ile .89 aralığında oldukları görülmüştür.

Anahtar Kelimeler

Performans Takip Sınavı, PTS, Psikometrik Özellik

Paper No: 160

The Psychometric Attributions of the Subtests of Performance Monitoring Examination Used to Compare State and Private Schools

Abstract

The aim of this study is to determine the psychometric attributions of multiple-choice items asked in Performance Monitoring Examination(PME) implemented for 8th grade students in Antalya and to figure out the psychometric differences of subtests through datas gathered from both private and state schools students. This study is a scanning study as psychometric features of sub-tests which are used in PFE are examined. Target populations of the study are consist of datas which are belong to 30.800 8th grade PME students. %91,75 of datas are collected from state school students, %8,25 of datas are collected from private school students. Target population consists of 372 state school students and 307 private school students that are chosen through simple random technique. Items and test statistics is produced from datas gathered from two different target population. In respect to the results of the analysis, it is found out that the sub-test of Maths is much more difficult for state school students rather than private school students. While the sub-test of Turkish and Religion and Ethics is easy for state school students, they are very easy for private school students. When the subtests of Science and Technology, Revolution History and English are examined, it determined that those sub-tests are reasonably difficult for state school students, however the same subtests are very easy for private school students. Also in order to figure out the ability to distinguish of items used in tests, two series correlation coefficient (the Rbis) are calculated and because of the fact that they usually have degree over .40, it can be said that these are good items with their ability to distinguish. It is determined that answers given by private schools students to the items in the subtests of Turkish, English, Religion and Ethics do not change too much. In the process of assessment, it is observed that coefficients of KR-20 internal consistency reliability of subtests are between the range of .78 and .89

Keywords

Performance Monitoring Examination, PME, Psychometric

Bildiri No: 161

Tahmin Parametresinin Yetenek ile Düzenlenmesi

Fatih ORÇAN*

Özet

Çoktan seçmeli testlerde 3 parametre lojistik (3PL) modelinin kullanılması kaçınılmazdır. 3PL modeline göre konu hakkında hiç bir bilgisi olamayan bir birey maddeye şans parametresi (g) değerinde bir ihtimalle doğru cevabı verebilir. Yani, şans parametresi en az yetenekli bir bireyin dahi doğru cevabı bulmadaki minimum ihtimalini gösterir. Tanımsal olarak g-parametresi maddenin bir özelliğidir. Fakat, bireyin bir özelliği olarak yetenek temelli şans parametresi son zamanlarda gündeme gelmiştir. Bu çalışmanın amacı 3PL'deki şans parametresinin bireylerin yetenek parametresi ile ilişkilendirerek yeni bir prosedür tanıtmak (İki Basamaklı Madde Tepki Kuramı - TSIRT), ve bir simulasyon çalışması aracılığıyla 2PL, 3PL, C3PL ve FG3PL modellerinin madde yanlığı (bias), ortalama standart hataların karekökü (RMSE) ve test bilgi fonksiyonu bakımından karşılaştırmaktır. Sonuçlara göre FG3PL modeli birçok durumda diğer moddellere göre daha kötü sonuçlar vermiştir (yüksek bias ve RMSE değerleri). C3PL modeli verinin 2PL'de üretilmesi durumunda iyi sonuçlar vermişken 3PL'de üretilen veriler için sonuçlar daha yüksek bias ve RMSE'ye sahiptir. Ayrıca, TSIRT prosedürü diğer modelle kıyasla daha bilgi vericidir ve tahmin edilen yetenek parametrelerinin dağılımı 3PL, C3PL ve FG3PL'ye göre daha normal görünmektedir.

Anahtar Kelimeler

Madde Tepki Kuramı, 3PL şans parametresi, Yetenek temelli şans, İki Basamaklı MTK, Simülasyon

*Karadeniz Teknik Üniversitesi, Fatih Eğitim Fakültesi Eğitim Bilimleri Bölümü. E-mail: fatihorcan@ktu.edu.tr Tel: +90 (462) 377 6760.

Paper No: 161

Adjusting the Guessing Parameters by Ability

Abstract

Using three parameter logistic (3PL) model is inevitable for multiple choice exams. Based on 3PL model, an examinee with no knowledge about the topic can correctly answer an item with the probability of value of g parameter. That is, the guessing parameter shows the probability of having a correct answer for the least able examinee. Recently, the idea of ability based guessing in which guessing parameter was linked to ability of examinees was asserted. One of the purposes of this study was to introduce a new procedure (called Two Step Item Response Theory - TSIRT) while linking the guessing parameter of a 3PL model with the ability parameters. Also, comparing item response theory models (e.g., 2PL, 3PL, C3PL, and FG3PL) in which the guessing parameters were defined differently via a simulation study. The models were compared based on item parameter estimation bias and Root Mean Squared Errors (RMSE). Based on the results, FG3PL model gave the worst results (i.e., larger bias and RMSE) compared to other models. C3PL model was fine when the artificial data were generated by 2PL model, but not for 3PLdata. Besides, TSIRT procedures were more informative with respect to other models, and showed more normal theta estimates.

Key Words

Item response theory (IRT), 3PL guessing parameter, Ability based guessing, Two step IRT, Simulation

ÖLÇME VE DEĞERLENDİRME KONGRESİ'NİN ARDINDAN

İlki 2008 yılında Ankara Üniversitesinde, İkincisi 2010 yılında Mersin Üniversitesinde, Üçüncüsü 2012 yılında Bolu İzzet Baysal Üniversitesinde, Dördüncüsü 2014 yılında Hacettepe Üniversitesinde yapılan "Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongreleri"nin, Beşincisi,1 – 3 Eylül 2016 tarihlerinde Antalya'da Akdeniz Üniversitesi Eğitim Fakültesi'nin ev sahipliğinde gerçekleştirilmiştir. .

Kongreye, Üniversiteler, Milli Eğitim Bakanlığı Merkez ve taşra teşkilatı başta olmak üzere, Türkiye'nin her tarafından, çeşitli resmi ve özel kurumlarda ölçme ve değerlendirme alanında çalışan uzmanlar, akademisyenler, Lisansüstü program (Yüksek Lisans ve Doktora) öğrencilerinden oluşan yaklaşık 400 kişilik geniş bir grup katılmıştır.

Dört "Çağrılı Konuşmacı – Bildiri" ile 139 "Sözlü Bildiri"nin sunulduğu ve tartışıldığı kongre; "Türkiye'de Merkezi Olarak Yürütülen Sınav Uygulamalarının Özellikleri, Eğitim Uygulamalarındaki ve Toplumdaki Etkileri" konulu, beş panel üyesinin katıldığı bir Panel çalışması ile tamamlanmıştır. .

Prof. Dr. Nizamettin KOÇ' un yöneticiliğinde; Prof. Dr. Nükhet DEMİRTAŞLI(Ankara Üniversitesi), Doç. Dr. Hakan ATILGAN(Ege Üniversitesi), Prof. Dr. Mehtap ÇAKAN(Gazi Üniversitesi), Prof. Dr. Selahattin GELBAL'ın(Hacettepe Üniversitesi) katılımıyla gerçekleştirilen ve kongrenin genel bir değerlendirmesinin de yapıldığı panelde; .

YGS, LYS, KPSS başta olmak son yıllarda, Merkezi sınavlarda yaşanan sorunların (soruların önceden bazı adayların eline geçmesi, şifre skandalı vb) öğrencilerde, velilerde, öğretmenlerde, toplumun tüm kesimlerine büyük bir güvensizliğe ve adalet duygularının zedelenmesine yol açtığı, .

Sınavlarda tekniğine uygun hazırlanmamış, ağırlıklı olarak hatırlama ve kavrama düzeyindeki sorular sorulmasının, üst düzey becerilere yönelik ölçme yapmada sınırlı kalınmasının önemli sorunlara yol açtığı, .

Çoktan Seçmeli soruların yanı sıra ÖSYM tarafından kullanılmaya başlanan "açık uçlu" (serbest cevaplı) soruların tekniğine uygun hazırlanmadığı, .

Merkezi sınavlarda çeşitli engel gruplarına mensup bireylere (görme engelliler, işitme engelliler, ortopedik engelliler vb.), teknolojinin de bize sunduğu imkânlardan yararlanarak, onların bireysel özelliklerine uygun ve gerçek performanslarını ortaya koyabilecekleri sınav ortamları yeterince sağlanamamasının önemli sorunlara yol açtığı, .

"Soru İptali" sorunlarına sıklıkla rastlandığı, bu durumlarda ilgililerce uygulanan çözümlerin (İptal edilen soruların tüm adaylarca doğru cevaplandırıldığının kabul edilmesi! gibi), ölçme ve değerlendirmenin temel ilkelerine ters düştüğü, .

Merkezi sınavları gerçekleştiren kurumların "ölçme ve değerlendirme uzmanlık alanı"nın destek ve hizmetlerinden yeterince yararlanmamalarının basına yansıyan, yargıya taşınan ve toplumun geniş kesimlerini ciddi düzeylerde rahatsız eden, güven duygularını zedeleyen sorunların ana nedeni olduğu, .

Karşılaşılan eğitim sorunlarının çözümünde ilgililerin ilk düşündüğünün yeni bir "Sınav" ya da "Sınavlar" olmasının ve "sınav odaklı öğretim" in öğrencilerin sanatsal, sportif ve kültürel etkinliklere katılmalarını sınırladığı, bunun da onların çok yönlü ve kişilik gelişimlerini olumsuz olarak etkilediği, .

Merkezi ölçme sonuçlarından eğitimi geliştirmede etkili bir biçimde yararlanılamadığı,

Merkezi ölçme sonuçlarının "okul ve öğretmenlerin performansını değerlendirme"de yanlış bir biçimde kullanılmalarının önemli sorunlara yol açtığı, .

Eğitimde "Sınıf-içi ölçmelerin" ihmal edildiği, öğretmenlerin öğrencilere başarıları için verdiği "not"ların, karar süreçlerinde yeterince kullanılmamasının eğitimi-öğretimi, öğretmenleri ve öğrencileri olumsuz biçimde etkilediği, .

Türkiye'nin de katıldığı; PISA, ve TIMSS gibi uluslar arası düzeyde uygulaman sınav sonuçlarından, eğitimi geliştirme çalışmalarında yeterince ve etkili bir biçimde yararlanılmadığı tespitlerinde bulunulmuştur.

Kongre sununda şu önerilerde bulunulmuştur: .

Merkezi sınavlar başta olmak üzere çeşitli amaçlarla uygulanan "Test Programları" nın; "ilgili tarafların işbirliğine dayalı olması", "işin uzmanlarınca geliştirilmesi ve uygulanması" vb başta olmak üzere "iyi bir test programı"nın ayırıcı özelliklerini karşılaması sağlanmalıdır.

TEOG'da yaşanan ciddi sorunlar; öğrenciler, veliler, öğretmenler başta olmak üzere toplumun tüm kesimleri üzerindeki olumsuz etkileri dikkate alınarak bu merkezi sınavın yeniden değerlendirilmesi, gerekirse TEOG uygulamalarına tamamen son verilmesi uygun olur.

Merkezi sınavlarda kullanılan testleri de içermek üzere, ülkemizde değişik amaçlarla kullanılabilecek testleri, ölçme araçlarını geliştiren – uyarlayan, bunları bünyesinde toplayan, bu testler üzerinde araştırma, geliştirme çalışmaları yapan; ülke düzeyinde kullanılan tüm testler üzerinde bir denetim sistemi geliştirerek uluslararası kuruluş ve uygulamalarla da bütünleşen, bu kapsamda, çağdaş ülkelerdeki uygulamalarda olduğu gibi, psikologların, ölçme ve değerlendirme uzmanlarının testler üzerindeki denetimine olanak sağlayan güçlü bir kuruluşa ihtiyaç vardır. .

"Türkiye Test ve Araştırma Kurumu" veya benzeri bir adla kurulabilecek bu kuruluşun (ÖSYM'nin böyle bir kuruluşa dönüştürülmesi düşünülebilir); a) Öğretimi Geliştirme Amacıyla Kullanılan Testler Birimi, b) Rehberlik Amacıyla Kullanılan Testler Birimi, c) Endüstriyel Amaçlarla Kullanılan Testler Birimi, d) Klinik Amaçlarla Kullanılan Testler Birimi, olmak üzere dört ana birimi içerir biçimde örgütlenmesi yararlı görülebilir.

Yine böyle bir kuruluş, topluma sunacağı hizmetler (özellikle insan gücü kaynağının etkili kullanımı, geliştirilmesi, yönetimi ve değerlendirilmesi çerçevesinde toplumun gelişmesine katkıları) dikkate alınarak, Devlet Bütçesi'nden önemli ölçülerde parasal destek alırken; sunduğu hizmetler uyarınca geliştirebileceği bir siteme dayalı olarak, büyük ölçüde kendi kendini finanse edebilen "ÖZERK" bir kuruluş olarak örgütlenmesi sağlanabilir.

Prof. Dr. Nizamettin KOÇ