INTERNATIONAL

SYMPOSIUM ON MEASUREMENT, SELECTION AND PLACEMENT

4-6 OCTOBER 2024 ANKARA

Venue: Bilkent Hotel and Conference Center

ULUSLARARASI

ÖLÇME, SEÇME VE YERLEŞTİRME SEMPOZYUMU

4-6 EKİM 2024 ANKARA

Yer: Bilkent Otel ve Kongre Merkezi

ÖZET METİN BİLDİRİ KİTABI

ABSTRACT PROCEEDINGS BOOK

ÖSYM 50 yıl

Bildiri Kitabı Hazırlama Komisyonu
Dr. Yıldız YILDIRIM GÖRGÜLÜ
Dr. Tuba GÜNDÜZ
Dr. Seçil ALACA
Dr. Ayşegül BOZDAĞ KASAP
Dr. İsmail ASLANTAŞ

Dizgi ve Grafik Tasarım
Gökhan ÇAKIR

ISBN
978-975-11-8381-1

Aralık, 2024

Bu yayının tüm hakları T.C. Millî Eğitim Bakanlığına aittir. Millî Eğitim Bakanlığının izni olmaksızın yayının tümünün veya bir kısmının elektronik veya mekanik yollarla basımı, yayını, çoğaltılması veya dağıtımı yapılamaz. Kaynak göstermek suretiyle alıntı yapılabilir.

TAKDİM

Ölçme ve değerlendirme, eğitim sistemlerini şekillendiren, eğitim-öğretim süreçlerinin başarısını ortaya koymayı sağlayan, eğitim süreçlerinin geliştirilmesine ışık tutan ve içinde pek çok farklı araştırma alanını barındıran oldukça önemli bir bilim alanıdır. Bu alanda yenilikler gerçekleştirmek için bilimden, bilimsel gelişmelerden, bilim insanlarından faydalanmamak söz konusu olamaz. Ölçme, seçme ve yerleştirme konularında 50 yıllık tecrübesi olan ve insanların hayatlarını etkileyen kararlar veren bir kurumun bilim insanları ile el ele olması, hem bilimsel gelişmeleri destekleyen, besleyen hem de bilimsel gelişmelerden sürekli faydalanan bir merkez olmalıdır.

Dünya genelinde yapılan ölçme ve değerlendirme uygulamalarına bakıldığında, ÖSYM kadar büyük hacimli ve çeşitlilik içeren sınav uygulaması yapan başka bir kurum bulunmamaktadır. ÖSYM, bu yönüyle bakıldığında sınav uygulamaları açısından küresel ölçekte örnek bir yapılanmadır. Çeşitli ülkeler tarafından uygulamalarımızın yerinde incelendiğine ve benzer uygulamalar geliştirmek üzere çalışmalar yapıldığına şahit olmaktayız. Bugün itibariyle Türkiye, ölçme, seçme ve yerleştirme uygulamalarını ihraç eden ülke konumuna gelmiştir.

50. yılını kutlayan ÖSYM'nin ölçme ve değerlendirme alanındaki tecrübesini günümüzde yaşanan yenilikler ile birleştirerek ÖSYM standardını sağlayan uygulamalarımızı daha iyiye taşımak en önemli hedefimizdir. Bu hedef doğrultusunda yenilikleri takip etmek, alandaki gelişmelerden faydalanmak ve dijitalleşmeyi sağlamak amacıyla geçtiğimiz iki yılda pek çok çalışma gerçekleştirdik. Ölçme ve değerlendirme alanında çalışan çok kıymetli akademisyenlerimiz ile kurumu ileri taşıyacak projeler geliştirdik. Ölçme ve değerlendirme alanında çalışan, alana katkıda bulunan daha fazla bilim insanına ulaşmak ve bilgi paylaşımını çoğaltmak istedik. Bu şiar ile ÖSYM olarak ölçme ve değerlendirme alanındaki araştırmalara alan açmak, bilimsel gelişmeleri takip edebilmek ve kurumda gerçekleştirdiğimiz çalışmaları paylaşabilmek adına Millî Eğitim Bakanlığı (MEB) ve Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı (ÖSYM) iş birliği ile ÖSYM'mizin 50. Yılı kapsamında Uluslararası Ölçme, Seçme ve Yerleştirme Sempozyumu'nun ilkini düzenledik.

Sempozyuma hem ulusal hem de uluslararası düzeyde gösterilen ilgi ÖSYM'nin ölçme ve değerlendirme alanında bilimsel gelişmelere de katkı sunan bir kurum olma hedefini desteklemiştir. Sempozyum boyunca çeşitli bilimsel alanlarda ölçme ve değerlendirme uygulamalarına ilişkin çok kıymetli çalışmaları dinleme fırsatı elde ettik. Sunulan bildiriler

sayesinde bilimsel araştırmalardaki yeni eğilimleri ve yenilikçi ölçme ve değerlendirme uygulamalarını fark ettik. Bunun yanı sıra ÖSYM'de gerçekleştirilen projeleri paylaşma ve araştırmacılardan fikir alma şansına eriştik. Dolayısıyla sempozyumun kurumun gelişmesine, ölçme ve değerlendirme alanındaki bilimsel araştırmaların merkezi haline gelme yolundaki gelişmelere katkısı oldukça büyük olacağına inanıyorum. Sempozyumun gerçekleşmesinde emeği geçen herkese teşekkürlerimi sunuyor, sempozyumun çıktısı olan bildiri kitapçığının alana katkıda bulunmasını temenni ediyorum.

Prof. Dr. Bayram Ali ERSOY

SUNUŞ

Talim terbiye sürecinde okul öncesinden başlayarak lisansüstü düzeye kadar geçen süreçte bir insan hayatında Millî Eğitim Bakanlığı (MEB) ile Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı (ÖSYM) belirgin bir şekilde hem tesirli hem de faaldir. MEB, insanın bu uzun yolculuğunda bütün boyutlarıyla tabi tutulduğu talim terbiye serencamında öne çıkan en belirgin ve yetkili kurumdur. ÖSYM ise bu uzun yolculuğunda öğrenciler için ölçme, seçme ve yerleştirme maksadıyla yapılan sınavları tanzim eden ve uygulayan kuruluştur. Tarihini omuzlarında taşıyabilen ve zamanın taşıyıcı ruhuna bu tecrübe ve müktesebatla mukabelede bulunabilen, zamanın şartları ve imkanları ile teçhiz edilmiş, vasıflı ve ahlâken tekamül eden öğrenciler yetiştirmeyi kendi ufkuna yerleştiren MEB'in bu çalışmalarının sonucunda ortaya çıkan sonucun adil bir şekilde niceliksel olarak ölçülüp değerlendirilerek öğrencilerin/adayların yerleştirilmeleri, adil bir toplumun varlığına ciddi olarak katkı sunması bakımından da oldukça hayatidir. Zira adil bir ölçme ve değerlendirme süreci, yerleştirmenin de adil bir şekilde yapılmasının kapısını açar. Bu durumda ortaya çıkan işbirliği ve dayanışma ülke ve dünya için iyi bir tecrübe olarak sürece mihmandarlık etmeye devam etmektedir. Bununla birlikte her iki alanda da zamanla yaşanan gelişmeler, zaman zaman yeniliklere veya değişikliklere de yol açmaktadır. Gelinen noktada MEB'in Türkiye Yüzyılı Maarif Modeliyle eğitimde beceri temelli bir yaklaşımı odağa almasıyla, gelişen teknolojilerle ölçme ve değerlendirme alanındaki yeniliklerin takibi zorunluluktur. Bu yönüyle ÖSYM'de de önemli derecede yapılan işlerde iyileştirme ve değişiklikler öngörülmektedir. Bu iyileştirme ve değişimlerin paydaşlarla paylaşılması ve paydaşlarımızın ÖSYM'nin çalışmalarına yeni bir bakış açısı kazandırması oldukça önemlidir.

Ölçme ve değerlendirme, eğitim ve öğretim süreçlerinin vazgeçilemez bir temel taşıdır ve bunu ülkemizde büyük oranda geniş kitlelerle icrâ eden kurumlar MEB ve ÖSYM'dir. ÖSYM'nin 50. yılı vesilesiyle MEB ve ÖSYM iş birliği ile tarihinde ilk kez Uluslararası Ölçme, Seçme ve Yerleştirme Sempozyumumuz 4-6 Ekim 2024'te Ankara'da gerçekleştirilmiştir.

Sempozyumun ana teması "Ölçme ve Değerlendirmede Yenilikçi Yaklaşımlar" olarak belirlenmiştir. Bu ana tema ile yola çıkışımıza özellikle gelişen yapay zekâ teknolojileri, yenilikçi ölçme ve değerlendirme yaklaşımları ve yenilikçi yerleştirme metotları önderlik etmiştir. Sempozyumumuza 247 bildiri başvurusu alınmış, tamamı çift kör hakemlik ile en az iki hakem tarafından hakem değerlendirme rubriği kullanılarak değerlendirilmiş ve bunlar arasından 125 bildiri kabul almıştır. Kabul edilen bildirilerden 40'ı dilde ölç-

me ve değerlendirme, 75'i ölçme ve değerlendirme, 10'u ise yerleştirme alanlarındadır. Beş oturumda ve beş salonda sunulan bu bildirilerin yenilikçi ölçme ve değerlendirmede okul öncesindenyetişkin talim-terbiyesine kadar olan süreci kapsayıcı nitelikte olduğu görülmüştür. Bildiri sunumları sempozyum süresince devam ederken bu ilk günde MEB ve ÖSYM'nin proje özel oturumu yapılmış ve PİKTES, TÜRKDİL ve yapay zekâ tabanlı ölçme ve değerlendirme sunumları gerçekleştirilmiştir. Ayrıca son gün kurumumuzda gerçekleştirilen genel çerçevesiyle bilgisayarlı, bireye uyarlanmış test, otomatik madde üretimi, dört becerili dil sınavları, yenilikçi değerlendirme projelerine ilişkin iki ÖSYM proje özel oturumu daha yapılmıştır. Öne çıkan bu projelerin yanı sıra sempozyumumuzda beş farklı ülkeden 13 davetli konuşmacı ile birlikte beş çalıştay gerçekleştirilmiştir. Sempozyumun üretilen ve paylaşılan içeriklerin öncelikle mevcut durumların iyileştirilmesine ve ardından yeni yaklaşımlar ile daha adil sınavlar yapılabilmesini mümkün kılan yeniliklerin kapılarının aralanmasına katkı sunabilmesini ümit ederiz.

Bu sempozyumun ilk kez gerçekleştirilmesine öncülük eden Sayın Bakanımız Prof. Dr. Yusuf TEKİN'e ve Sayın Başkanımız Prof. Dr. Bayram Ali ERSOY'a şükranlarımı iletmekten büyük bir mutluluk duyuyorum. Düzenleme kurulumuza, danışma kurulumuza ve bilim kurulumuza özverili çalışmalarından dolayı teşekkürlerimi sunarım. Muhtevanın zenginleşmesinde sempozyuma katkı sağlayan davetli konuşmacılarımıza, yaşadığımız heyecan, mutluluk ve gururumuza ortak olan öğretmen ve akademisyenlerimize teşekkürü bir borç bilirim.

Sempozyum Düzenleme Kurulu Başkanı
Prof. Dr. Ahmet KESGİN

SEMPOZYUM BAŞKANLARI

Prof. Dr. Yusuf TEKİN
Milli Eğitim Bakanı - Türkiye

Prof. Dr. Bayram Ali ERSOY
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı Başkanı - Türkiye

SEMPOZYUM DÜZENLEME KURULU BAŞKANI

Prof. Dr. Ahmet KESGİN
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

SEMPOZYUM DÜZENLEME KURULU

Prof. Dr. Mustafa DOĞAN
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Kemal BÜLBÜL
Milli Eğitim Bakanlığı - Türkiye

Prof. Dr. Necip ŞİMŞEK
İstanbul Ticaret Üniversitesi - Türkiye

Dr. Cengiz CEYLAN
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Halil ŞENGÜN
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. Nida KIRÖMEROĞLU
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Ömer OKUMUŞ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. Yıldız YILDIRIM GÖRGÜLÜ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. Tuba GÜNDÜZ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. Seçil ALACA
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. Ayşegül BOZDAĞ KASAP
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı - Türkiye

Dr. İsmail ASLANTAŞ
Milli Eğitim Bakanlığı - Türkiye

SEMPOZYUM DANIŞMA KURULU

Prof. Dr. Bayram ÇETİN
Gaziantep Üniversitesi - Türkiye

Prof. Dr. Çetin SEMERCİ
Bartın Üniversitesi - Türkiye

Prof. Dr. Dilara BAKAN KALAYCIOĞLU
Gazi Üniversitesi - Türkiye

Prof. Dr. Ercan ÖZTEMEL
Marmara Üniversitesi - Türkiye

Prof. Dr. Vatan KARAKAYA
Yıldız Teknik Üniversitesi - Türkiye

Doç. Dr. Ahmet Salih ŞİMŞEK
Kırşehir Ahi Evran Üniversitesi - Türkiye

Doç. Dr. Erkan Hasan ATALMIŞ
Manisa Celal Bayar Üniversitesi - Türkiye

Doç. Dr. Murat Doğan ŞAHİN
Anadolu Üniversitesi - Türkiye

SEMPOZYUM BİLİM KURULU

Prof. Dr. Ahmet KOÇAK
İstanbul Medeniyet Üniversitesi - Türkiye

Prof. Dr. Akihito KAMATA
Southern Methodist University - ABD

Prof. Dr. Alex TEYTELBOYM
University of Oxford - Birleşik Krallık

Prof. Dr. Allan S. COHEN
University of Georgia - ABD

Prof. Dr. Alparslan OKUR
Sakarya Üniversitesi - Türkiye

Prof. Dr. Andreas FREY
Goethe University Frankfurt am Main - Almanya

Prof. Dr. Anthony GREEN
University of Bedfordshire - Birleşik Krallık

Prof. Dr. Ayşegül DALOĞLU
Orta Doğu Teknik Üniversitesi - Türkiye

Prof. Dr. Bayram BIÇAK
Akdeniz Üniversitesi - Türkiye

Prof. Dr. Chun WANG
University of Washington - ABD

Prof. Dr. Cihad DEMİRLİ
Milli Eğitim Bakanlığı - Türkiye

Prof. Dr. Çiğdem SAĞIN ŞİMŞEK
Orta Doğu Teknik Üniversitesi - Türkiye

Prof. Dr. David J. WEISS
University of Minnesota - ABD

Prof. Dr. Deniz MELANLIOĞLU
İstanbul Üniversitesi - Türkiye

Prof. Dr. Dilek PEÇENEK
Ankara Üniversitesi - Türkiye

Prof. Dr. Edward L. BOONE
Virginia Commonwealth University - ABD

Prof. Dr. Fatma BÖLÜKBAŞ KAYA
İstanbul Üniversitesi / Cerrahpaşa - Türkiye

Prof. Dr. Hakan ATILGAN
Ege Üniversitesi - Türkiye

Prof. Dr. Hakan TUTAR
Gazi Üniversitesi - Türkiye

Prof. Dr. Herve MOULIN
University of Glasgow - Birleşik Krallık

Prof. Dr. Hülya KELECİOĞLU
Hacettepe Üniversitesi - Türkiye

Prof. Dr. İbrahim DEMİR
Türkiye İstatistik Kurumu - Türkiye

Prof. Dr. İbrahim Alper KÖSE
Abant İzzet Baysal Üniversitesi - Türkiye

Prof. Dr. İrfan YURDABAKAN
Dokuz Eylül Üniversitesi - Türkiye

Prof. Dr. İsmail KARAKAYA
Gazi Üniversitesi - Türkiye

Prof. Dr. James Enos PURPURA
Columbia University - ABD

Prof. Dr. Jimmy de la TORRE
The University of Hong Kong - Hong Kong

Prof. Dr. Julien COMBE
Paris School of Economics - Fransa

Prof. Dr. Kaan BÜYÜKİKİZ
Gaziantep Üniversitesi - Türkiye

Prof. Dr. Kaan Zülfikar DENİZ
Ankara Üniversitesi - Türkiye

Prof. Dr. Kadriye ERÇİKAN
ETS Canada Inc. - Kanada

Prof. Dr. Kasım KILIÇARSLAN
Sağlık Bilimleri Üniversitesi - Türkiye

Prof. Dr. Kutlay YAĞMUR
Tilburg University - Hollanda

Prof. Dr. M. David MILLER
University of Florida - ABD

Prof. Dr. Marek PYCIA
University of Zurich - İsviçre

Prof. Dr. Mark J. GIERL
University of Alberta - Kanada

Prof. Dr. Mark RECKASE
Michigan State University - ABD

Prof. Dr. Mehmet KALAYCI
Ankara Üniversitesi - Türkiye

Prof. Dr. Mehtap ÇAKAN
Gazi Üniversitesi - Türkiye

Prof. Dr. M. Ali AKÇAYOL
Gazi Üniversitesi - Türkiye

Prof. Dr. M. Eyüp SALLABAŞ
Yıldız Teknik Üniversitesi - Türkiye

Prof. Dr. M.Fatih DEMİRCİ
Ankara Yıldırım Beyazıt Üniversitesi - Türkiye

Prof. Dr. Mustafa DURMUŞ
Hacettepe Üniversitesi - Türkiye

Prof. Dr. Neal KINGSTON
University of Kansas - ABD

Prof. Dr. Nermin YAZICI
Hacettepe Üniversitesi - Türkiye

Prof. Dr. Neşe GÜLER
İzmir Demokrasi Üniversitesi - Türkiye

Prof. Dr. Nezir TEMUR
Gazi Üniversitesi - Türkiye

Prof. Dr. Nilüfer KAHRAMAN
Gazi Üniversitesi - Türkiye

Prof. Dr. Nuri DOĞAN
Hacettepe Üniversitesi - Türkiye

Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU
Ankara Üniversitesi - Türkiye

Prof. Dr. Rodrigo VELEZ
University of Texas at San Antonio - ABD

Prof. Dr. Sadegül AKBABA ALTUN
Başkent Üniversitesi - Türkiye

Prof. Dr. Selahattin GELBAL
Hacettepe Üniversitesi - Türkiye

Prof. Dr. Şener BÜYÜKÖZTÜRK
Hasan Kalyoncu Üniversitesi - Türkiye

Prof. Dr. Stephen G. SIRECI
University of Massachusetts Amherst- ABD

Prof. Dr. T. Oğuz BAŞOKÇU
Ege Üniversitesi - Türkiye

Prof. Dr. Thayer MORRILL
North Carolina State University - ABD

Prof. Dr. Theo EGGEN
University of Twente - Hollanda

Prof. Dr. Tommy ANDERSSON
Lund University - İsveç

Prof. Dr. Tuncay ÖĞRETMEN
Ege Üniversitesi - Türkiye

Prof. Dr. Umut Mert DUR
North Carolina State University - ABD

Prof. Dr. Vikram MANJUNANTH
University of Ottawa - Kanada

Prof. Dr. Yasin TOKLU
Ankara Yıldırım Beyazıt Üniversitesi - Türkiye

Prof. Dr. Walter L. LEITE
University of Florida - ABD

Doç Dr. Alper ŞAHİN
Atılım Üniversitesi - Türkiye

Doç. Dr. Asiye ŞENGÜL AVŞAR
Recep Tayyip Erdoğan Üniversitesi - Türkiye

Doç. Dr. Ayfer SAYIN
Gazi Üniversitesi - Türkiye

Doç. Dr. Aysel SARICAOĞLU AYGAN
Ankara Sosyal Bilimler Üniversitesi - Türkiye

Doç. Dr. Azar ABİZADE
ADA University - Azerbaycan

Doç. Dr. Burak AYDIN
Leuphana University - Almanya

Doç. Dr. C. Deha DOĞAN
Ankara Üniversitesi – Türkiye

Doç. Dr. Dilek FİDAN
Kocaeli Üniversitesi - Türkiye

Doç. Dr. Duygu KOÇAK
Alanya Alaaddin Keykubat Üniversitesi - Türkiye

Doç. Dr. Emrah BOYLU
Bartın Üniversitesi - Türkiye

Doç. Dr. Eren Can AYBEK
Pamukkale Üniversitesi - Türkiye

Doç. Dr. Eren Halil ÖZBERK
National Foundation for Educational Research - Birleşik Krallık

Doç. Dr. Erkan ER
Orta Doğu Teknik Üniversitesi – Türkiye

Doç. Dr. Fuat ELKONCA
Muş Alparslan Üniversitesi- Türkiye

Doç. Dr. Gökhan AKÇAPINAR
Hacettepe Üniversitesi - Türkiye

Doç. Dr. Güçlü ŞEKERCİOĞLU
Akdeniz Üniversitesi - Türkiye

Doç. Dr. Gülşen TAŞDELEN TEKER
Hacettepe Üniversitesi - Türkiye

Doç. Dr. Halil İbrahim SARI
Texas A&M University - ABD

Doç. Dr. Inacio BO
University of Macau - Makao

Doç. Dr. İbrahim UYSAL
Abant İzzet Baysal Üniversitesi - Türkiye

Doç. Dr. İlker KALENDER
Bilkent Üniversitesi - Türkiye

Doç. Dr. Kağan BÜYÜKKARCI
Süleyman Demirel Üniversitesi - Türkiye

Doç. Dr. Kübra ATALAY KABASAKAL
Hacettepe Üniversitesi - Türkiye

Doç. Dr. Mahmut Sami KOYUNCU
Afyon Kocatepe Üniversitesi - Türkiye

Doç. Dr. Melek Gülşah ŞAHİN
Gazi Üniversitesi - Türkiye

Doç. Dr. M. Yalçın YILMAZ
İstanbul Üniversitesi - Türkiye

Doç. Dr. Mustafa İLHAN
Dicle Üniversitesi - Türkiye

Doç. Dr. Okan BULUT
University of Alberta - Kanada

Doç. Dr. Peter TROYAN
University of Virginia - ABD

Doç. Dr. Safiye BİLİCAN DEMİR
Kırıkkale Üniversitesi - Türkiye

Doç. Dr. Sakine GÖÇER ŞAHİN
MetaMetrics - ABD

Doç. Dr. Salim RAZI
Çanakkale 18 Mart Üniversitesi - Türkiye

Doç. Dr. Samson ALVA
University of Texas at San Antonio - ABD

Doç. Dr. Sedat ŞEN
Harran Üniversitesi - Türkiye

Doç. Dr. Seher YALÇIN
Ankara Üniversitesi - Türkiye

Doç. Dr. Selma ŞENEL
Balıkesir Üniversitesi - Türkiye

Doç. Dr. Serkan ARIKAN
Boğaziçi Üniversitesi - Türkiye

Doç. Dr. Stefanie WIND
University of Alabama - ABD

Doç. Dr. Szilvia PAPAI
Concordia University - Kanada

Doç. Dr. Tuncer CAN
İstanbul Üniversitesi / Cerrahpaşa - Türkiye

Doç. Dr. Qiwei B. HE
Georgetown University - ABD

Doç. Dr. Will PHAN
North Carolina State University - ABD

Dr. Christine COOMBE
Dubai Men's College - BAE

Dr. Öğr. Üyesi Derya AKBAŞ
Aydın Adnan Menderes Üniversitesi - Türkiye

Dr. Ergün Cihat ÇORBACI
Sinop Üniversitesi - Türkiye

Dr. Hanan KHALIFA
Cambridge University - Birleşik Krallık

Dr. Mehmet KAPLAN
New Jersey Institute of Technology – ABD

Dr. Merve YILDIRIM SEHERYELİ
Hasan Kalyoncu Üniversitesi - Türkiye

Dr. Nathan THOMPSON
Assessment Systems Corporation – ABD

Dr. Sebahat GÖREN
Kütahya Dumlupınar Üniversitesi - Türkiye

Dr. Sinan ERTEMEL
Rice University – ABD

Dr. Sungur GÜREL
Siirt Üniversitesi - Türkiye

Dr. Süleyman Abanoz
Selçuk Üniversitesi - Türkiye

Dr. Taner YAPAR
Türkiye Odalar ve Borsalar Birliği Üniversitesi - Türkiye

SEMPOZYUM SEKRETERYASI

Dr. Yusuf KÖMÜR
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Ahmet Murat SOYUĞUR
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Cem AKGÜR
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Ceren ABABAY
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Eylem KUBİLAY
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Gizem TÜRKYILMAZ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Hülya ATEŞ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Murat DEĞİRMENCİ
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Nisa GENDE
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

Zeynep KAZAK
Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı

İÇİNDEKİLER

TAKDİM ..... **I**
SUNUŞ ..... **III**
KURULLAR ..... **V**
BİLDİRİ METİNLERİ **1**

Konuşma Performansının Değerlendirilmesinde ChatGPT- 4o'nun Sağladığı Olanaklar Üzerine Bir Pilot Çalışma 24
A Pilot Study on the Affordances of ChatGPT-4o in Assessing Speaking Performance 25
CEFR düzeylerine göre metinlerin dilbilimsel özellikleri nedir ve ChatGPT bunları ne kadar yansıtabilir? 26
What are the linguistic features of texts according to the CEFR levels, and to what extent can ChatGPT reflect them? 27
Rater Performance Categorization System (RPCS): A new practical way to monitor rater performance 28
University Strategies in Centralized Admissions 29
Bilişsel Tanısal Modellemenin e-UTYS Testine Uygulanması 30
Application of Cognitive Diagnostic Modeling to e-UTYS 31
Yabancı Dil Okuma ve Dinleme Becerileri Arasındaki İlişki 32
The Relationship Between Foreign Language Reading and Listening Skills 33
Sözdiziminden Puanlara: Büyük Ölçekli Bir Çevrimiçi Yazışma Görevindeki Puan Düzeylerinin Sözdizimsel Göstergeleri 34
From Syntax to Scores: Syntactic Indicators of Writing Quality in a Large-Scale Online Exchange Writing Task 35
Beceri Temelli İngilizce Başarı Testinin Psikometrik Özelliklerinin Madde Tepki Kuramı Çerçevesinde İncelenmesi 36
Examining the Psychometric Properties of the Skill-Based English Achievement Test within the Framework of Item Response Theory 37
Avrupa Dilleri Ortak Çerçevesi Seviyeleri Madde Tepki Kuramı Karşılıkları: Bir Standart Belirleme Çalışması 38
Item Response Theory Correspondences of Common European Framework of Reference for Languages Levels: A Standard-Setting Study 39
Düşük ve Yüksek Okuma Başarısı Gösteren Öğrenci Özelliklerinin Kestirilmesi 40
Predicting Characteristics of Low and High Reading Achievers 41

Bayes Ağlarına Dayalı Kişiselleştirilmiş Geri Bildirim Tasarımı: Feedup, Feedback ve Feedforward Üçgeni 42
Personalized Feedback Design Based on Bayesian Networks: The Triangle of Feedup, Feedback, and Feedforward 43
Öğrencilerin Üst Bilişsel Düşünme Becerilerinin Geliştirilmesinde Örnek Bir Uygulama: SOT 44
An Exemplary Application in Developing Students' Metacognitive Thinking Skills: QFT 45
Kavram Karikatürlü Plicker Uygulamasının Eleştirel Düşünme Eğilimine Etkisi 46
The Effect of Plicker Application with Concept Cartoons on 47
Critical Thinking Disposition 47
Üstün Zekâlı İlkokul Öğrencilerin Bilimsel Yaratıcılık Becerilerinin CHAID Analizi ile Değerlendirilmesi 48
Eğitimde Ayrık Seçenekli Çoktan Seçmeli Madde Kullanımı Üzerine Bir İnceleme 50
An Examination of the Use of Discrete Option Multiple Choice Items in Education 51
Geleneksel ve Öncüllü Çoktan Seçmeli Maddelerin 52
Psikometrik Açıdan İncelenmesi 52
Psychometric Analysis of Traditional and Complex Multiple Choice Items 53
Şans Başarısı, Üst Bilişsel Beceri ve Ölçme Sonuçlarının İlişkisinin İncelenmesi 54
An Investigation of the Relationship between Chance Success, Metacognitive Skills and Measurement Results 55
Çoklu Cevaplı Madde Türünün Sınıf İçi Ölçme Uygulamalarında ve Geniş Ölçekli Sınavlarda Kullanımına İlişkin Öğrenci ve Öğretmen Görüşlerinin İncelenmesi 56
Investigation of Student and Teacher Opinions on the Use of Multiple- Response Item Types in Classroom Assessment Applications and Large-Scale Exams 57
Çoktan Seçmeli Formatta Bağlam Temelli Soruların Hazırlanması 58
Closing the Gender Gap in STEM: Evidence from the 2019 Taiwanese College Admission Reform 60
Market Incentives and College Admissions in the Absence of Standardized Tests 61
Okul Tercihinde Eşitlikçi Pozitif Ayrımcılık 62
Market Design of College-Major Switches for Reducing Student-Major Mismatch 64
Alternative Assessment in High School EFL Teaching: Practices and Perspectives 65
Yabancı Dilde Kelime Bilgisi Ölçme ve Değerlendirme için Fibonacci Tabanlı Algoritma Tasarımı: Rusça Örneği 66
Fibonacci-Based Algorithm Design for Vocabulary Measurement and Evaluation in a Foreign Language: A Russian Language Example 67
Rusça Yabancı Dil Bilgisi Seviye Tespit Sınavlarında (YDS) Güvenirlik ve Geçerliğin İncelenmesi 68

An Examination of Reliability and Validity in the Russian Foreign Language Proficiency Exams (YDS) 69
CEFR'e Göre Anadili Arapça Olmayanlar İçin Arapça Metinlerin Sınıflandırma Standartlarının Belirlenmesi 70
Non-Native Speakers According to CEFR 71
İlk, Orta ve Yükseköğretimde Okuma Sıklığına Göre Birleşik Sözcüklerin Anlamsal Saydamlığının İncelenmesi* 72
Investigation of Semantic Transparency of Compound Words According to Reading Frequency of Middle School, High School and University Students* 73
A Tool for Selecting and Writing Assessment Texts: Reader- Friendly Informative Text Rubric (RFITER) 75
Relationship Between Language Assessment Literacy and Teacher Agency of Pre-Service English Language Teachers 76
Eğitimde Ölçme ve Değerlendirme Alanında Yapay Zekâ: Sistematik Bir Derleme 77
Alıcı Beceriler Üretken Şekilde Değerlendirilebilir mi? Alternatif Bir Değerlendirme Aracı Olarak Hikâye Haritalama 78
Can Receptive Skills Be Assessed Productively? Story Mapping as an Alternative Assessment Tool 79
Dil Öğreniminde Alternatif Ölçme-Değerlendirme: Öğretmen Görüşleri ve Sınıf Uygulamaları 80
Alternative Assessments in Language Learning: Teacher Voices and Practical Applications 81
Kelime Denizinde Bir Yelkenli: Yapılandırılmış Grid Yöntemiyle Dil Alanında Yeni Ufuklar 82
A Sailboat in the Sea of Words: Fresh Horizons in the Realm of Language Via the Structure Grid Method 83
ChatGPT Destekli Ölçme Sürecinin Ortaokul Öğrencilerinin Öğrenme Eksikliklerini Giderme Üzerindeki Etkisi 84
The Effect of ChatGPT Supported Assessment Process on Eliminating Learning Deficiencies of Middle Students 85
Geniş Dil Modelleriyle Q Matrisi Üretimi: GPT-4o'nun Performans Analizi 86
Q Matrix Generation with Large Language Models: Performance Analysis of GPT-4o 87
PISA 2022 Matematik Okuryazarlığı Şekil Formatındaki Maddelerin Türkçe ve İngilizce Olarak Cevaplanmasında ChatGPT'nin Performansı 88
ChatGPT's Performance in Answering PISA 2022 Mathematics Literacy Questions in Image Format in Both Turkish and English 89
Okuduğunu Anlama Becerisini Yordayan Değişkenlerin Makine Öğrenme Algoritmalarıyla İncelenmesi: PIRLS 2021 Örneği 90

Investigation of Variables Predicting Reading Literacy Skills with Machine Learning Algorithms: The Case of PIRLS 2021 91
Psikomotor Becerilerin Ölçülmesinde Yapay Zekâ Tespit Araçlarından Elde Edilen Sonuçların Doğruluğunun İncelenmesi 92
Examining the Accuracy of Results Obtained from Artificial Intelligence Detection Tools in Measuring Psychomotor Skills 93
Yapay Zekâ ve Nitel Veri: Geleceğin Analiz Yöntemleri 94
Artificial Intelligence and Qualitative Data: Future Analysis Methods 95
Yükseköğretim Kurumları Sınavı (YKS)'nda Yapay Zekâ Kullanımına İlişkin Öğrenci Görüşleri 96
Student Opinions on the Use of Artificial Intelligence in Higher Education Institutions Examination (YKS) 97
Yükseköğretimde Üretken Yapay Zekanın Öğrenciler Tarafından Ödev ve Performans Görevlerinde Kullanımı: Avantajlar ve Riskler 98
Use of Generative Artificial Intelligence by Students in Assignments and Performance Tasks in Higher Education: Advantages and Risks 99
Seçmeli Sorularda Cevap Değiştirme Davranışlarının Kopya İstatistiklerine Etkisinin İncelenmesi 100
Investigation of the Effect of Answer Changing Behaviors on Cheating Statistics in Selective Questions 101
Türkiye Yüzyılı Maarif Modeli Öğretim Programları Ortak Metninin Kapsayıcı Ölçme ve Değerlendirme İlkeleri Açısından İncelenmesi 102
Examination of the Türkiye Century Education Model Curriculum Common Text in term of Inclusive Measurement and Evaluation Principles 103
Nasıl Ölçüyor? Nasıl Değerlendiriyoruz? 4-7. Sınıf Sosyal Bilgiler Ders Kitaplarında Ölçme ve Değerlendirme Durumlarının İncelenmesi 104
How Do We Measure? How Do We Evaluate? Examination of Measurement and Evaluatio in 4th-7th Grade Social Studies Textbooks 105
ÖABT-Din Kültürü ve Ahlak Bilgisi Öğretmenliği Testi'nde Farklılaşan Madde Fonksiyonu 106
Differential Item Function in the TFKT-Religious Culture and Moral Knowledge Teaching Test 107
Okul Öncesi Eğitimde Yapay Zekâ: "Yapay Zekâ ile Kişiselleştirilmiş Ölçme ve Değerlendirme" 108
NLP Tabanlı Açık Uçlu Soruların Otomatik Değerlendirme Sistemi: "T.C. İnkılap Tarihi ve Atatürkçülük Dersi için Bir Uygulama" 110
An NLP-Based Automatic Evaluation System for Open-Ended Questions: An Application for the Course on the T.C. İnkılap Tarihi ve Atatürkçülük 111

Yükseköğretim Programlarına Yerleşen Öğrencilerin Yeterlik Düzeylerinin Tanımlanması için Model ve Yazılım Geliştirilmesi 112
Development of a Model and Software for Defining Proficiency Levels of Students Admitted to Higher Education 113
Açık Uçlu Maddelerin Farklı Branşlardaki Puanlayıcılar Tarafından Puanlanmasına Karışan Hatalar ve Puanlayıcılar Arasındaki Uyumun İncelenmesi 114
Errors in the Scoring of Open-Ended Items by Raters in Different Disciplines and Examining the Agreement Between Raters 115
Madde Takımlarından Oluşan Testlerde Parametre Kestirimi, Sınıflama Doğruluğu ve Değişen Madde Grubu Fonksiyonunun İncelenmesi 116
Examination of Parameter Estimation, Classification Accuracy and Differential Item Bundle Functioning in Testlet Based Tests 117
Karşılaştırma Yargılarıyla Puanlamada Uyarlamalı Eşleme ve Standart Hata Durdurma Kuralı: Bir Monte Carlo Simülasyonu 118
Adaptive Matching and Standard Error Stopping Rule in Comparative Judgement: A Monte Carlo Simulation 119
Büyük Gruplarda Karşılaştırma Yargılarıyla Puanlama: Referans Set Temelli Bir Yaklaşım 120
Using Comparative Judgement in Large Groups: A Reference Set- Based Approach 121
İkinci/Yabancı Dil Olarak Türkçe Öğreticilerinin Ölçme ve Değerlendirme Okuryazarlıkları: Bir Durum Çalışması 122
Assessment Literacies of Teachers of Turkish as a Second/Foreign Language: A Case Study 123
Yabancı Dilde Anlam Tercihi Birlikteliklerinin İşlemlenmesi: Örtük Anlamsal Öğrenmenin Değerlendirilmesi 124
Processing Meaning Preference Associations in a Foreign Language: Assessing Implicit Semantic Learning 125
Kontrol Listelerinin Yabancı/İkinci Dil Olarak Türkçe Konuşma Becerisinin Değerlendirilmesinde Kullanımı 126
Use of Checklists in the Assessment of Speaking Turkish as a Foreign/Second Language 127
Konuşma ve Dinleme Becerilerine Yönelik Uygulama Sınavlarının Değerlendirilmesi 128
Yüksek Ölçekli Bir Dil Sınavında Yazma Bölümünün Değerlendirilmesine Yönelik Eğitim İçeriklerinin Planlanması 130
Planning Training Contents for the Evaluation of the Writing Section in a Large-Scale Language Exam 131
Merkezi Yerleştirme Sistemlerinde Stratejik Koordinasyonun Tercih Sıralamaları Üzerindeki Etkisi 132

The Impact of Strategic Coordination on Preference Rankings in Centralized Placement Systems 133
ÖSYM Sınav Puanlamaları İçin Alternatif Yaklaşımlar 134
Alternative Approaches for ÖSYM Exam Scoring 135
Rastgele Arama ile Merkezi Olmayan Çoktan Bire Eşleştirme 136
Decentralized Many-to-One Matching with Random Search 137
Admissions Game and Matching with Limited Resources 138
Preventing Academic Misconduct in Online Assessments 139
Otomatik Madde Üretiminde Görsellerin Işlevleri Ve Dinamik Olarak Oluşturulması 142
Dynamically Creation of Visuals And Their Functions In Automatic Item Generation 143
Üretken Yapay Zekâ Modellerinin Çoktan Seçmeli Soru Üretme Performanslarının Karşılaştırması 144
Geniş Dil Modeliyle Otomatik Madde Üretimi: 5. Sınıf Matematik Uygulaması 146
Automatic Item Generation with the Large Language Model: Grade 5 Mathematics Application 147
Klinik Akıl Yürütme Becerisinin Değerlendirmesinde Otomatik Soru Üretim Yöntemlerinin Karşılaştırılması 148
Comparison of Automatic Question Generation Methods in Assessing Clinical Reasoning Skills 149
Öğrenci ve Okul Özellikleri, Öğretme ve Öğrenme Süreci ile İlgili Değişkenler ve Bilişsel Olmayan Öğrenme Çıktılarının PISA 2015 Fen Okuryazarlığı Üzerindeki Etkilerinin İncelenmesi* 150
Investigation of the Effect of Student and School Background Variables, Teaching and Learning Variables and Non-Cognitive Outcomes on the Components of Scientific Literac in PISA 2015* 151
Test Eşitlemede Ortak Test Desenine Karşı Ortak Değişken Deseni Yaklaşımları 152
Anchor Test Design Versus Covariates Design Approaches in Test Equating 153
Using Artificial Intelligence in Standard Setting 155
KTK, MTK ve D-Puanlama Yöntemlerinin Karşılaştırılması ve Standart Belirleme Süreçleri Üzerine Bir Uygulama 156
Comparison of KTK, MTK and D-Scoring Methods and an Application on Standard Setting Processes 157
İngilizce Yeterlik Sınavı Dinleme Alt Testinin Geçerlik Kanıtlarının İncelenmesi 158
The Validation of English Proficiency Exam- Listening Subtest 159
Yapay Zekâ ile C1 Seviyesi Türkçe Okuma Sınavı Oluşturma ve Uygulama 160
C1 Level Turkish Reading Test with Artificial Intelligence Creation and Implementation 161

Okuduğunu Anlama Seviyesinin Ölçümünde Boşluk Tamamlama Testleri 162
Cloze Tests in Measuring Reading Comprehension Levels 163
Aynı Özelliği Ölçen Ortak Köklü ve Bağımsız Maddelerden Oluşan Yabancı Dil Olarak Türkçe Dinleme Testlerinin Psikometrik Özelliklerinin Karşılaştırılması 164
A Comparison of the Psychometric Properties of Common Stem and Independent Items in Turkish as a Foreign Language Listening Tests 165
Yabancı Dil Olarak Türkçe Öğretiminde Kriter Bazlı Soru Yazımı Eğitimi Tasarımı ve Değerlendirmesi 166
Design and Evaluation of Criteria-Based Question Writing Training in Teaching Turkish as a Foreign Language 167
Türkçenin Yabancı/İkinci Dil Olarak Öğretiminde Yenilenmiş Bloom Taksonomisi'ne Göre Hazırlanmış Soru ve Etkinliklerin Öğretici Farkındalıkları Açısından Değerlendirilmesi 168
Evaluation of Questions and Activities Prepared According to the Revised Bloom's Taxonomy in Teaching Turkish as a Foreign/Second Language in Terms of Teachers' Awareness 169
Üniversite Eğitimlerini İngilizce Alan Öğrencilerin Yabancı Dil Yazma Becerilerinin Yapay Zekâ Tabanlı Puanlama ve İnsan Tarafından Değerlendirilmesinin Karşılaştırmalı Bir Analizi 170
Transforming Speaking Assessments with AI-based Virtual Characters in the Metaverse 172
The practical effectiveness of ChatGPT as an automated scoring tool for writing assessment 173
Topluluk Önünde Konuşma Becerisi Değerlendirmesinin Öğrenen Özerkliği Süreci (ÖÖS) ile Geliştirilmesi 174
Enhancing Public Speaking Skill Assessment with Autonomous Learning Process (ALP) 175
Yazılı Sınavlar için Yapay Zekâ Destekli Web Tabanlı Ön Değerlendirme Sistemi Tasarımı: Rusça Dil Sınavları Örneği 176
Ai-Supported Web-Based Pre-Assessment System Design for Written Exams: Example of Russian Language Exams 177
Öğrencilerin Test Çözme Davranışlarının eTIMSS 2019 Süreç Verileri Kullanılarak İncelenmesi: Türkiye Verisi için Bulgular 178
Exploring Students' Test-Taking Behaviors Using Process Data in the eTIMSS 2019: Findings for Türkiye Data 179
Tepki Sürelerinin Karma Modellemeye Dahil Edilmesi, Matematik Okuryazarlığı Örtük Grupların Belirlenmesini Nasıl Etkiler? 180
How Integrating Response Times into Mixture Modeling Effect the Identification of Latent Groups in Mathematical Literacy? 181
Yapay Zeka Tarafından Oluşturulan Olgu Temelli Çoktan Seçmeli Maddelerin Değerlendirilmesi: Pediatri Örneği 182

Evaluation of Case-Based Multiple-Choice Items Generated by Artificial Intelligence: A Pediatrics Example 183
Geri Bildirimin Başarıya Etkisi: Bibliyometrik Bir Analiz 184
The Effect of Feedback on Achievement: A Bibliometric Analysis 185
Baskın Maddeler Faktör Sayısı Belirleme Yöntemlerinin Performansını Nasıl Etkiler? 186
Madde Parametrelerinin Tahmininde Ortak Madde Bağlantısı Yönteminin Geçerliliği 188
The Validity of Common Item Linking Method in Estimating Item Parameters 189
Katılımcıların Test Sürecine Dahil Olma Durumlarının İncelenmesi: Bir Yöntem Önerisi ve Farklı Yöntemlerin Karşılaştırılması 190
Are Participants Fully Engaged During a Test Process? A Sequential Argument and Comparison with Different Methods 191
Matlab Programında Destek Vektör Makinesi İle Sınıflandırma Doğrulunun İncelenmesi 192
Investigation of Classification Accuracy with Support Vector Machine in MATLAB Program 193
Uluslararası Okuma Becerilerinde Gelişim Araştırması (PIRLS) 2021 Türkiye Verisinin Bilişsel Tanı Modellemesi İle Analizi 194
Analysis Of Progress in International Reading Literacy Study (PIRLS) 2021 Türkiye Data Using Cognitive Diagnostic Modeling 195
Türkiye ve Amerika'da Uygulanan Merkezî Sınavların Dil Becerileri Bakımından Karşılaştırılması 196
Comparison of High-Stakes Test in Turkey and the United States in terms of Language Skills 197
PISA2018 Okuma Okuryazarlığı Temel Testinin Ölçme Değişmezliğinin İncelenmesi 198
Examining the Measurement Invariance of the PISA2018 Reading Literacy Core Test 199
Sürdürülebilir Kapsayıcı Başarı: Öğretmenlerle İlgili Faktörlerin Rolü* 200
Sustainable Inclusive Success: The Role of Teacher-Related Factors* 201
Özel Amaçlı İngilizce Öğretmenlerinin Mesleki Bilgilerinin Ölçülmesi 202
Assessing Professional Knowledge of English For Specific Purposes (ESP) Teachers 203
Dil Becerilerinin Ölçülmesi Alanında Yapılan Çalışmaların Konu Modellemesi ile İncelenmesi 204
Investigation of Research on Measurement of Language Skills with Topic Modeling 205
Akıllı Teknolojiler, Dijital Araçlar ve YZ Uygulamaları Üzerinden Konuşma Becerisinin Değerlendirilmesi 206
Evaluation of Speaking Skill Through Smart Technologies, Digital Tools and AI Apps 207
Geleneksel Dört Beceri Değerlendirmesi ile Alternatif Portföy Değerlendirmesinin B2 Seviye İngilizce Hazırlık Öğrencilerinde Karşılaştırılması 208
Comparing Traditional Four-Skill Assessment and Alternative Portfolio Evaluation in B2 Level English Preparatory Students 209
Kısa Yanıtlı Soruları Özdevinimli Puanlanma Platformu: Cevaplab 210

An Automatic Short Answer Grading Platform: Cevaplab 211
Açık Uçlu Maddelerin Puanlanmasında Yapay Zekâ Performansının İncelenmesi 212
Investigation of Artificial Intelligence Performance in Scoring Open-Ended Items 213
Boyutlar Arasındaki İlişki Düzeyinin Çok Boyutlu Test Eşitlemeye Etkisi * 214
Effect of the Level of Relationship Between Dimensions on Multidimensional Test Equalization * 215
Açık Uçlu Sınavların Otomatik Değerlendirilmesi: Görüntü İşleme ve Yapay Zekâ Destekli iOS Tabanlı Mobil Uygulama Geliştirme Örneği 216
Automatic Evaluation of Open-Ended Exams: A Case Study of Developing an iOS-Based Mobile Application Supported by Image Processing and AI 217
PISA 2022 Matematik Okuryazarlığı Performansı ve Okul Etkililiği ile İlişkili Faktörlerin Belirlenmesi: Kapsamlı Bir Değerlendirme 218
Determining the Factors Associated with PISA 2022 Mathematics Literacy Performance and School Effectiveness: A Comprehensive Evaluation 219
Yapay Zekâ İngilizce Yazma Becerilerini Puanlamada Ne Kadar Başarılı? 220
How Successful Is Artificial Intelligence in Scoring English Writing Skills? 221
Yapay Zeka ile Sınav Kağıdının Otomatik Değerlendirilmesi 222
Automatic Evaluation of Exam Paper with Artificial Intelligence 223
Ölçme ve Değerlendirme Uzmanlarının Yeni Öğretim Programındaki Ölçme Değerlendirme Anlayışı Açısından Görüşleri 224
Measurement and Evaluation Experts' Views on the Measurement and Evaluation Approach in the New Curriculum 225
Görme Engelli YKS Adaylarının Sınav Uyarlamaları Hakkındaki Bilgi Düzeyi: Mevcut Durum ve Geliştirme Önerileri 226
Knowledge Level of Visually Impaired YKS Candidates About Exam Accommodations: Current Status and İmprovement Recommendations 227
Uyumsuz Madde Puan Örüntülerinin Belirlenmesinde Kullanılan Tekniklerin Sistematik İncelemesi 228
A Systematic Review of Techniques Used to Identify Aberrant Item Score Patterns 229
Yükseköğretim Kurumlarında Test Adaletinin İncelenmesine Yönelik Bir Kontrol Listesi Geliştirme Çalışması 230
The Checklist Development Study to Investigate the Test Fairness in The Higher Education 231
Testlerde Adillik ve Test Düzenlemeleri: Bibliyometrik Analiz ve Sistematik İnceleme Çalışması ... 232
Test Fairness and Test Accommodations: A Bibliometric Analysis and Systematic Literature Review 233

Kayıp Veri Bağlamında Bayesyen Rasch ve Örtük Uzay Madde Tepki Modellerinin Karşılaştırılması 234
Comparison of Bayesian Rasch and Latent Space Item Response Models in the Context of Missing Data..... 235
Birleştirilmiş Ağaç BBT ve Ağaç BBT performansının Ağacın Büyümesini Kontrol Eden Ana Parametrelere Göre İncelenmesi 236
Investigation of Merged Tree CAT and Tree CAT Performance According to the Main Parameters Controlling the Growth of the Tree..... 237
Lisansüstü Programlara Başvuruda Kullanılan Sınavların Karşılaştırmalı İncelenmesi..... 238
Comparative Examination of Exams Used for Graduate Program Applications..... 239
Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) Başarısını Makine Öğrenmesi Algoritmalarıyla Yordama 240
Predicting Academic Personnel and Postgraduate Education Entrance Exam (ALES) Success with Machine Learning Algorithms 241
Afet Dönemlerinde e-Sınav Uygulaması için AUZEF Örneği 242
AUZEF Example for e-Exam Application in Disaster Periods..... 243
Aday Gözünden Çevrimiçi Sınavların Kullanılabilirliği: Göz Hareketleri Izleme Çalışması 244
Usability of Online Exams from The Candidate's Eye: An Eye Tracking Study..... 245
Performans Değerlendirmede Puanlayıcılar Arası Uyum İle Göz İzleme Sonuçları Arasındaki İlişkinin İncelenmesi 246
Investigation Of Interrater Agreement And Eye Tracking in Performance Evaluation 247
Evaluation of Eye Movements of Primary School Students with Attention Disorders During Reading with Artificial Intelligence Supported Application..... 249
Test Geliştirmede Paradigma Değişimi: Yapay Zekâ Destekli Çeldiricilerin Etkililiği..... 250
Paradigm Shift in Test Development: The Effectiveness of AI- Supported Distractors 251
Çoktan Seçmeli Testlerde Çeldiriciler için Puan Düzeltme Uygulamaları..... 252
The multilingual education and its assessment-evaluation systems in schools for immigrants 254

BİLDİRİ ÖZETLERİ

Konuşma Performansının Değerlendirilmesinde ChatGPT- 4o'nun Sağladığı Olanaklar Üzerine Bir Pilot Çalışma

Mehmet GAZAN [1]

İkinci veya yabancı dil öğrencilerinin konuşma becerilerini değerlendirmedeki zorluk, sınırlı zaman, mali kaynaklar ve eğitimli insan değerlendiricilerin eksikliği gibi kısıtlamalardan kaynaklanmaktadır (Hayashi vd., 2023). Ancak, konuşma tanıma, doğal dil işleme gibi ilerlemeler ve son zamanlarda piyasaya sürülen ChatGPT-4o gibi yapay zeka araçları, bu zorluklara potansiyel çözümler sunma ihtimalini ortaya koyuyor (Timpelaughin vd., 2022). Otomatik konuşma değerlendirmelerinin artan popüleritesine ve bu aracın kullanımına rağmen, ChatGPT-4o üzerine araştırmalar hala erken aşamadadır. Bu çalışma, ChatGPT-4o'nun (plus versiyonu) İngilizce konuşma performansını değerlendirme yeteneğini araştırmayı ve test katılımcılarının algılarını insan değerlendiricilerle karşılaştırmayı amaçlamıştır. Bu çalışmada karma yöntem yaklaşımı kullanılarak, insan değerlendiricilerin puanları ile ChatGPT-4o'nun puanları karşılaştırılmış, ayrıca 25 orta düzeyde İngilizce hazırlık öğrencisiyle yapılan nitel görüşmelerle desteklenmiştir. Katılımcılar başlangıçta iki görevi tamamlamışlar (yani, resim yardımlı hikaye anlatma ve görüş paylaşma) ve ardından görüşme sorularını cevaplamışlardır. İstatistiksel analiz, kurumsal sınav puanları ile ChatGPT-4o'nun puanları arasında anlamlı bir fark olduğunu göstermiştir. Görüşmeler, test katılımcılarının ChatGPT-4o puanlarını adil bulduklarını ve en çok anlık geri bildirim özelliğine değer verdiklerini ortaya koymuştur. Bu pilot çalışma, bu aracı sürekli geri bildirimlerle iyileştirmeyi ve yüksek riskli ve büyük ölçekli test bağlamlarına daha iyi uyarlamayı önermektedir. Bu aracın kullanımı ve optimal görevleri belirlemek için ileri araştırmalar da önerilmektedir.

Anahtar Kelimeler: *ChatGPT-4o, Konuşma değerlendirmesi, Otomatik puanlama.*

Kaynakça

Hayashi, Y., Kondo, Y., & Ishii, Y. (2023). Automated speech scoring of dialogue response by Japanese learners of English as a foreign language. *Innovation in Language Learning and Teaching, 18*(1), 32–46. https://doi.org/10.1080/17501229.2023.2217181

OpenAI. (2024). ChatGPT-4o [Large language model]. https://www.openai.com/chatgpt-4

Timpe-Laughlin V., Sydorenko T., Dombi J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. C*omputer Assisted Language Learning*. *37*(1-2), 149-178.https://doi.org/10.1080/09588221.2022.2032184

[1] mehmet.gazan@metu.edu.tr, ORCID: 0000-0002-7233-9033, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

A Pilot Study on the Affordances of ChatGPT-4o in Assessing Speaking Performance

Mehmet GAZAN [1]

Evaluating speaking skills in second or foreign language learners is challenging due to constraints such as limited time, financial resources, and a shortage of trained human evaluators (Hayashi et al., 2023). However, advancements in speech recognition, natural language processing, and recently launched AI tools like ChatGPT-4o (OpenAI, 2024) suggest potential solutions to these challenges (Timpelaughin et al., 2022). Despite the increasing popularity of automated speaking assessments and this tool, research on ChatGPT-4o is still in its early stages. Thus, this study aimed to assess ChatGPT-4o's (plus version) capability in evaluating L2 English speaking performance and explore the perceptions of test-takers (n= 25) compared to assessment by human raters. Using a mixed-method approach, it compared scores from human assessors with those from ChatGPT-4o, alongside qualitative interviews with 25 intermediate-level English prep students. Participants completed two tasks (i.e., picture-cued storytelling and opinion sharing) initially and then answered interview questions. Statistical analysis revealed a significant difference between institutional exam scores and scores from ChatGPT-4o. Interviews revealed that test-takers found ChatGPT-4o scores fair and valued its immediate feedback feature the most. This pilot study suggests refining ChatGPT-4o through ongoing feedback to better suit high-stakes and large-scale testing contexts. Further research is recommended to identify optimal tasks for use with this tool.

Keywords: *ChatGPT4-o, Speaking assessment, Automated scoring.*

References

Hayashi, Y., Kondo, Y., & Ishii, Y. (2023). Automated speech scoring of dialogue response by Japanese learners of English as a foreign language. *Innovation in Language Learning and Teaching, 18*(1), 32–46. https://doi.org/10.1080/17501229.2023.2217181

OpenAI. (2024). ChatGPT-4o [Large language model]. https://www.openai.com/chatgpt-4

Timpe-Laughlin V., Sydorenko T., Dombi J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. C*omputer Assisted Language Learning. 37*(1-2), 149-178.https://doi.org/10.1080/09588221.2022.2032184

[1] mehmet.gazan@metu.edu.tr, ORCID: 0000-0002-7233-9033, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

CEFR düzeylerine göre metinlerin dilbilimsel özellikleri nedir ve ChatGPT bunları ne kadar yansıtabilir?

Mehmet AKINCI[1], Aylin ÜNALDI[2]

İkinci dilde (L2) okuduğunu anlama becerisini ölçen sınavlar geliştirirken dikkat edilmesi gereken en önemli konulardan biri okuma metinlerin, sınava girecek öğrenci grubundan beklenen yeterlilik düzeyini doğru yansıtmasıdır. Bir okuma metni, sınavı alan öğrenci grubu için fazla zor veya kolay olduğunda, o sınavın geçerliği tartışmalı hale gelir. Ortak Avrupa Dil Referans Çerçevesi (CEFR) altı yeterlilik düzeyinde (A1-C2) metin özelliklerini genel olarak tanımlar ama metinlerin dilbilimsel özelliklerini detaylı olarak belirlemez. Bu konuda kapsamlı bir bilimsel çalışma henüz yapılmamıştır. Okuma metinlerinin öğrenci düzeyine uygunluğunu belirlemek için otomatik metin analizi araçlarının kullanılması önerilmektedir (Green vd., 2010). Ayrıca, günümüzde, sınav geliştirme konusunda yapay zekâ kullanımı hızlı bir biçimde artmaktadır. Buna rağmen, ChatGPT gibi araçların metinlerde olması beklenen dilbilimsel özellikleri ne ölçüde doğru yansıttığı araştırılmamıştır. Bu faktöriyel araştırma desenindeki çalışma, CEFR düzeylerine uygun olan metinlerin hangi dilbilimsel özellikleri taşıdığını ve ChatGPT kullanılarak üretilen metinlerin bu özellikleri ne kadar yansıttığını belirlemek amacıyla yapılmıştır. Beş farklı yayınevi tarafından yayımlanan A1-B2 seviyelerinde sekiz İngilizce öğretim (ELT) kitap serisinden yaklaşık 20,000 kelimelik 521 metin seçilerek bir derlem oluşturulmuş, metin özellikleri Coh-Metrix (McNamara vd., 2014) ile analiz edilmiştir. Ayrıca, benzer sayı, düzey ve konuda üretilen ChatGPT metinlerinden bir derlem daha oluşturulmuştur. Aynı biçimde analiz edilen metinler, dilbilimsel özellikleri bakımından ELT metinleriyle, ANOVA ve takip eden analizlerle karşılaştırılmıştır. Sonuçta, değişik CEFR düzeylerindeki metinlerinin dilbilimsel özellikleri belirlenmiş ve ChatGPT metinlerinin bazı yönlerden ELT kitap metinleriyle benzerlikler gösterdiği ancak düşük düzeyler için önemli ölçüde farklılıklar olduğu bulunmuştur. L2 okuma sınavı hazırlanırken bu çalışmada elde edilen bulgular metin seçme ve/veya oluşturma konusunda rehberlik edebilir.

Anahtar Kelimeler: İkinci *dil (L2) okuma becerisi değerlendirme, Yapay zeka metinleri, CEFR düzeyleri*

Kaynakça

Green, A., Ünaldi, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, *27*(2), 191-211. https://doi.org/10.1177/0265532209349471

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

[1] makinci@29mayis.edu.tr: ORCID: 0000-0002-7296-156X, İstanbul 29 Mayıs Üniversitesi, İstanbul, Türkiye

[2] aunaldi@yahoo.com: ORCID: 0000-0003-4119-6700, University of Huddersfield, Huddersfield, United Kingdom

What are the linguistic features of texts according to the CEFR levels, and to what extent can ChatGPT reflect them?

Mehmet AKINCI[1], Aylin ÜNALDI[2]

In developing second language (L2) reading assessment, it is essential that the reading texts accurately reflect the expected proficiency levels of the test-takers. When a reading text is excessively challenging or overly simplistic for the test-takers, the validity of the assessment is called into question. The Common European Framework of Reference for Languages (CEFR) gives a broad description of text features across six proficiency levels (A1-C2), yet it does not specify the linguistic features of texts in detail. To date, no comprehensive scientific investigation has been made on this subject. Automated text analysis tools are proposed to be efficient in evaluating the appropriateness of reading texts to student levels (Green et al., 2010). Moreover, the adoption of artificial intelligence in exam development has been on the rise in recent years. However, it has not been investigated to what extent tools like ChatGPT can accurately reflect the expected linguistic features of texts at various proficiency levels. This study with a factorial research design aims to determine the linguistic features of the texts aligned with CEFR levels and assess the extent to which texts generated by ChatGPT reflect these features. A corpus of texts from eight English Language Teaching (ELT) book series at A1-B2 levels published by five different publishing houses, totaling approximately 20,000 words, was formed. The text features were analyzed using Coh-Metrix (McNamara et al., 2014). Additionally, a parallel corpus of ChatGPT-generated texts, equivalent in number, level, and subject matter, was created. These texts were analyzed in the same manner and compared with the ELT book texts in terms of linguistic features using a series of mixed-ANOVAs and follow-up analyses. The results delineated the linguistic features across CEFR levels and indicated that while ChatGPT-generated texts exhibited similarities to ELT book texts in certain aspects, there were significant differences at lower proficiency levels. The findings obtained from this study would provide valuable guidance for text selection and/or creation when preparing second language reading comprehension assessments.

Keywords: *Second language (L2) reading assessment, AI-generated texts, CEFR levels*

Reference

Green, A., Ünaldi, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, *27*(2), 191-211. https://doi.org/10.1177/0265532209349471

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

[1] makinci@29mayis.edu.tr: ORCID: 0000-0002-7296-156X, İstanbul 29 Mayıs Üniversitesi, İstanbul, Türkiye

[2] aunaldi@yahoo.com: ORCID: 0000-0003-4119-6700, University of Huddersfield, Huddersfield, United Kingdom

Rater Performance Categorization System (RPCS): A new practical way to monitor rater performance

Alper Şahin[1]

The subjective nature of rubric-based scoring while assessing the writing and speaking skills of test takers constitutes one of the long-standing problems in English language testing. The accuracy of these scores is critically important, especially in situations where high-stakes decisions about test takers are made based on them. Therefore, maintaining accuracy is critical, and monitoring rater performance is one of the ways to achieve it. Besides, regular standardization training can also be beneficial. However, organizing such training for all raters is challenging, especially in organizations with a large number of raters. Therefore, identifying who needs training requires a practical rater performance monitoring system. However, existing rater performance monitoring systems involve facets like Rasch measurement, Bayesian approaches, the Rasch partial credit model, hierarchical rater models, and automated rater scoring. These systems also require specialized software, which limits their utility for practitioners. To overcome this issue, a practical Rater Performance Categorization System (RPCS) was developed, and a sample study was designed. For this purpose, data was first collected from 101 instructors using nine sample student performances across three types of tasks (3 paragraphs, 3 essays, and 3 speaking performances). Furthermore, the Rater Performance Score (RPS) and Rater Performance Category (RPC) for the instructors were calculated using MS Excel. The scores were shared with the instructors, and a standardization training of up to two hours was conducted for each performance task type. Then the instructors were asked to score another set of nine sample performance tasks. Eighty-six instructors scored the papers at this post-training stage. The root mean squared difference (RMSD) of the average RPS was calculated. It was found that most of the raters improved their RPC averages for paragraph tasks (60%), speaking tasks (66%), and overall (65%). However, only 45% of the raters improved their RPC averages. When the RMSD values were analyzed, it was found that the highest change occurred in the low-level paragraph graded by the raters for the paragraph task. Moreover, the speaking task was found to be the one from which raters benefited the most from the training and the RPCS. The findings indicated that, thanks to the RPCS, raters' performance was identified, classified, and the magnitude of their performance change after the training was measured successfully. As a result, this sample study indicated that the RPCS may provide the test administrators with easy-to-calculate, practical, and fair performance indicators about the raters' performance when used appropriately.

Keywords: Rater Performance Monitoring, assessing speaking and assessing writing, rater performance categorization system

[1] e-posta: alpersahin2@yahoo.com, ORCID: 0000-0001-7750-4408, Atılım University, Ankara, Türkiye

University Strategies in Centralized Admissions

Samson Alva[1], Hayri Alper Arslan[2], Umut Dur[3]

Türkiye operates a centralized college admissions system where students are placed in undergraduate programs based on their nationwide college entrance exam scores, high school GPAs, and submitted preference lists. Unlike a standard centralized system, Türkiye's admissions are influenced by the presence of different university types: State and Foundation (i.e., Private). State universities are fully funded by the government and do not require tuition, whereas foundation universities charge tuition but offer various financial aid opportunities. These distinctions influence student preferences and admission outcomes, as foundation universities adjust their tuition, funding levels, and the number of funded students to attract applicants.

In this study, we analyze the strategic behavior of foundation universities in Türkiye from 2003 to 2013, using program-level data on centralized college admissions outcomes and tuition rates for four-year departments. Our findings indicate that higher tuition and larger enrollment capacities increase actual revenue but lower the rank of the last admitted student. Conversely, offering more funded positions improves the rank of the last admitted student but reduces departmental revenue. Additionally, increasing capacity and financial aid options leads to more unfilled seats, while offering more funded positions reduces the likelihood of these unfilled seats. Furthermore, outcomes from previous admissions cycles significantly influence current strategies.

These results suggest that private universities have different objectives, balancing the need to increase demand by lowering tuition or offering more financial aid against the goal of maximizing revenue by raising tuition, which may reduce demand from higher-scoring students. Our regression analysis reveals that a department's capacity decisions are primarily influenced by the previous year's outcomes. Departments that experienced unfilled seats or lower-than-expected revenue typically reduced their overall capacity and increased the proportion of funded positions. Additionally, the local competitive landscape significantly affects tuition and funding policies; in cities with more universities, tuition tends to decrease, while in areas with more private competitors, tuition is more likely to increase. Departments also reduce tuition if they experienced a lower actual-to-expected revenue ratio in the previous year. Finally, we find that graduation rates decline when departments offer more funding types and admit more funded students. Programs with lower cutoff scores also tend to have lower graduation rates, suggesting that increased diversity in the admitted student body may lead to reduced graduation outcomes.

Keywords: *Centralized admissions, Private universities, Tuition and scholarship strategies*

[1] samson.alva@utsa.edu:, ORCID: 0000-0002-9660-3628, University of Texas at San Antonio, San Antonio, United States

[2] hayri.arslan@utsa.edu:, ORCID: 0000-0002-2359-7773, University of Texas at San Antonio, San Antonio, United States

[3] umutdur@gmail.com:, ORCID: 0000-0002-0553-8382, North Carolina State University, Raleigh, United States

Bilişsel Tanısal Modellemenin e-UTYS Testine Uygulanması

Fahri Yılmaz [1], Seçil Alaca [2], Gohar Sayedi [3]

Bu çalışma, Uluslararası Türkçe Yeterlik sınavı pilot uygulamalarına katılan yabancı dil olarak Türkçe öğrenen ve Türkiye'de yükseköğretim kurumlarında öğrenim görmekte olan öğrenicilerin okuma ve dinleme becerilerindeki bilişsel yeterlik performanslarını ortaya koymak için Bilişsel Tanısal Modelleme (BTM) uygulanmasını amaçlamaktadır. Birincil amaç, bu maddeleri yanıtlamak için gereken belirli bilişsel becerileri belirleyerek haritalamak ve öğrenciler için ayrıntılı bilişsel tanı profilleri oluşturmaktır. Bu profilleri inceleyerek, okuduğunu anlama ve dinleme becerileri için gerekli olan bilişsel becerilerin ediniminin karmaşık modellerini araştırmayı, öğrenci performansının ayrıntılı bir görünümünü sağlamayı amaçlıyoruz. Model karşılaştırmasına ek olarak, maddeleri bilişsel yeterlik tanılarına ve zorluk seviyelerine göre sınıflandırmak için derinlemesine bir madde analizi yapılacaktır. Bu analiz, madde özelliklerinin ve maddelerin bilişsel süreçlerle uyumunun kapsamlı bir incelemesini içerecek ve bilişsel becerilerin belirli test maddeleriyle ayrıntılı bir şekilde eşleştirilmesini sağlayacaktır. Araştırma, çeşitli analitik teknikleri kullanarak ve çok çeşitli değişkenleri göz önünde bulundurarak yabancı dil olarak Türkçe öğretimi alanında uygulanan eğitim programlarında alımlama becerilerinde bilişsel yeteneklerin geliştirilmesine yönelik geliştirilecek uygulamalar için önerilerde bulunacaktır.

Anahtar Kelimeler: *bilişsel tanısal modelleme, yabancılara Türkçe* öğretimi*, bilişsel profil oluşturma*

Kaynakça

de la Torre, J., & Chiu, C. Y. (2016). A general method of cognitive diagnosis. *Journal of Educational Measurement, 53*(4), 418-440.

Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non-diagnostic reading comprehension assessment: Application of the G-DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly*, 16, 294–311.

Kim, H. (2014). Application of cognitive diagnostic model for achievement profile analysis. *KAERA Research Forum, 1*(1), 15–25.

Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment, 18*(1), 1–25.

Liu, R., & Shi, D. (2020). Using diagnostic classification models in psychological rating scales. *The Quantitative Methods for Psychology, 16*(5), 442-456.

Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2017). The impact of Q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement, 77*(2), 220-240.

Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software, 93*(14), 1-26.

[1] fahri.yilmaz_01@metu.edu.tr, ORCID: 0000-0001-5919-8611, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[2] secil.alaca@osym.gov.tr, ORCID: 0000-0002-8731-4197, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, Türkiye

[3] gohar.sayedi@lokmanhekim.edu.tr, ORCID: 0000-0001-9324-7245, Lokman Hekim Üniversitesi, Ankara, Türkiye

Application of Cognitive Diagnostic Modeling to e-UTYS

Fahri Yılmaz [1], Seçil Alaca [2], Gohar Sayedi [3]

This study aims to apply Cognitive Diagnostic Modeling (CDM) to reveal the cognitive competence performance in reading and listening skills of Turkish as a foreign language learners and learners studying in higher education institutions in Türkiye who participated in the piloting of the International Turkish Proficiency (e-UTYS) test. The primary aim is to identify and map the specific cognitive skills required to answer these items and to create detailed cognitive diagnostic profiles for students. The aim is to investigate the complex patterns of acquisition of the cognitive skills required for reading comprehension and listening skills, providing a detailed view of student performance. Also, an in-depth item analysis will be conducted to classify items according to cognitive abilities and difficulty levels. This analysis will include a thorough examination of item characteristics and the alignment of items with cognitive processes and will provide a detailed mapping of cognitive skills to specific test items. By using various analytical techniques and taking into account a wide range of variables, the study aims to provide recommendations for the development of cognitive abilities in receptive skills in educational programs implemented in the field of teaching Turkish as a foreign language.

Keywords: *cognitive diagnostic modeling, teaching Turkish as a foreign language, cognitive profiling*

References

de la Torre, J., & Chiu, C. Y. (2016). A general method of cognitive diagnosis. *Journal of Educational Measurement, 53*(4), 418-440.

Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non-diagnostic reading comprehension assessment: Application of the G-DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly*, 16, 294–311.

Kim, H. (2014). Application of cognitive diagnostic model for achievement profile analysis. *KAERA Research Forum, 1*(1), 15–25.

Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment, 18*(1), 1–25.

Liu, R., & Shi, D. (2020). Using diagnostic classification models in psychological rating scales. *The Quantitative Methods for Psychology, 16*(5), 442-456.

Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2017). The impact of Q-matrix designs on diagnostic classification accuracy in the presence of attribute hierarchies. *Educational and Psychological Measurement, 77*(2), 220-240.

Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software, 93*(14), 1-26.

[1] fahri.yilmaz_01@metu.edu.tr, ORCID: 0000-0001-5919-8611, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[2] secil.alaca@osym.gov.tr, ORCID: 0000-0002-8731-4197, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, Türkiye

[3] gohar.sayedi@lokmanhekim.edu.tr, ORCID: 0000-0001-9324-7245, Lokman Hekim Üniversitesi, Ankara, Türkiye

Yabancı Dil Okuma ve Dinleme Becerileri Arasındaki İlişki

Levent Yakar [1], Elif Kantarcıoğlu[2], Tuba Arabacı Atlamaz[3], Erkan Hasan Atalmış[4],Reyhan Ağçam[5]

Dil edinimi okuma, dinleme, yazma ve konuşma olmak üzere dört temel becerinin bileşiminden oluşmaktadır. Avrupa Dilleri Ortak Çerçevesine göre okuma ve dinleme becerisi birlikte değerlendirilerek alımlama alanını oluşturmaktadır (Avrupa Konseyi, 2020). Birbirleriyle ile kuramsal boyutta yakından ilişkili olduğu görülen bu becerilerden okuma beceresi diğer becerilere göre kolay sınanması sayesinde en sık sınanan dil becerilerinin başında gelmektedir. Bu araştırmada bir İngilizce yeterlik sınavı sonucu elden edilen okuma ve dinleme yetenek parametrelerinin ilişki düzeyinin incelenmesi amaçlanmıştır. Araştırma verileri üçü vakıf, ikisi devlet üniversitesinin yabancı diller yüksekokulu İngilizce hazırlık programında öğrenim gören toplam 1462 öğrenciden toplanmıştır. Veri toplama aracı olarak araştırma ekibi tarafından hazırlanan 256 okuma ve 128 dinleme alanına yönelik toplam 384 çoktan seçmeli soru kullanılmıştır. Yapılan MTK analizleri sonucunda çok boyutlu 3PL modelinin en iyi uyuma sahip olduğu görülmüştür. EAP yöntemi ile açığa çıkartılan okuma ve dinleme yetenek parametreleri arasında.89 düzeyinde yüksek bir ilişki tespit edilmiştir. Okuma ve dinleme alanlarına ait doğru sayıları arasında ise.43 değerinde, orta düzeyde bir ilişki tespit edilmiştir. Elde edilen sonuçlar farklı yetenek düzeyleri için araştırılacak ve benzer çalışma sonuçları (Cheng ve Matthews, 2016) eşliğinde sonuçlar tartışılacaktır.

Anahtar Kelimeler: *okuma becerisi, dinleme becerisi, madde tepki kuramı.*

Kaynakça

Avrupa Konseyi, (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. Strasbourg: Council of Europe.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3-25.

[1] leventyakar@gmail.com:, ORCID: 0000-0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[2] kutevu@bilkent.edu.tr:, ORCID: XXXX-XXXX-XXXX-XXXX, Bilkent Üniversitesi, Ankara, Türkiye

[3] tubaarabaci@gmail.com:, ORCID: 0000-0002-6332-7727, Boğaziçi Üniversitesi, İstanbul, Türkiye

[4] erkanatalmis@gmail.com:, 0000-0001-9610-491X, Manisa Celal Bayar Üniversitesi. Manisa, Türkiye

[5] reyhanagcam@gmail.com:, ORCID: 0000-0002-5445-9031, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

The Relationship Between Foreign Language Reading and Listening Skills

Levent Yakar [1], Elif Kantarcıoğlu[2], Tuba Arabacı Atlamaz[3], Erkan Hasan Atalmış[4],Reyhan Ağçam[5]

Language acquisition is composed of four basic skills: reading, listening, writing and speaking. According to the Common European Framework of Reference for Languages, reading and listening skills are evaluated together to form the domain of reception (Council of Europe, 2020). Among these skills, which are seen to be closely related to each other in theoretical terms, reading skill is one of the most frequently tested language skills due to its ease of testing compared to other skills. The aim of this study is to examine the relationship level of reading and listening skill parameters obtained as a result of an English proficiency exam. The research data were collected from a total of 1462 students studying in the English preparatory program of the foreign languages school of three foundation universities and two state universities. A total of 384 multiple choice questions prepared by the research team for 256 reading and 128 listening areas were used as data collection tools. As a result of the MTK analyses, it was seen that the multidimensional 3PL model had the best fit. A high relationship of.89 was determined between the reading and listening ability parameters revealed by the EAP method. A moderate correlation of.43 was found between the correct answers in the reading and listening areas. The results obtained will be investigated for different ability levels and the results will be discussed in the context of similar studies.

Keywords: *reading comprehension, listening comprehension, item response theory*

Reference

Council of Europe, (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. Strasbourg: Council of Europe.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3-25.

[1] leventyakar@gmail.com:, ORCID: 0000-0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[2] kutevu@bilkent.edu.tr:, ORCID: XXXX-XXXX-XXXX-XXXX, Bilkent Ünversitesi., Ankara, Türkiye

[3] tubaarabaci@gmail.com:, ORCID: 0000-0002-6332-7727, Boğaziçi Üniversitesi, İstanbul, Türkiye

[4] erkanatalmis@gmail.com:, 0000-0001-9610-491X, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye

[5] reyhanagcam@gmail.com:, ORCID: 0000-0002-5445-9031, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

Sözdiziminden Puanlara: Büyük Ölçekli Bir Çevrimiçi Yazışma Görevindeki Puan Düzeylerinin Sözdizimsel Göstergeleri

Aysel Saricaoglu[1], Mustafa Çimrin[2], Ece Selva Küçükoglu[3], Aliye Evin Yörüdü[4]

Bu çalışma, İngilizce 'deki dört beceriyi ölçmek üzere Türkiye'de geliştirilmekte olan e-TEP sınavındaki yazılı anlatım bölümünün bir parçası olan çevrimiçi yazışma görevine verilen yanıtlardan alınan üç farklı puan düzeyindeki sözdizimsel özelikleri incelemektedir. Yazılı anlatım becerisinin ölçülmesinde çevrimiçi yazışma görevleri önem kazanarak, büyük ölçekli İngilizce Yeterlik sınavlarındaki (TOEFL'daki Çevrimiçi Yazılı Akademik Tartışma bölümü gibi) diğer yazma görevlerinin yerini aşamalı olarak almaya başlamıştır. İkinci dilde yazmanın değerlendirilmesine yönelik alan yazında okumaya dayalı karma yazma görevlerinin geçerlik ve güvenirliğine bir iki çalışmada değinilmiş olsa da, okumaya dayalı çevrimiçi yazma görevlerinin geçerlik ve güvenirliği araştırmacıların dikkatini çevirmesi gereken bir açık olarak alandaki yerini korumaktadır. Yeni yazılı anlatım görevleri "puana dayalı çıkarımların netlik kazanabilmesi için araştırmalar yoluyla kanıta" gereksinim duyduğundan bu çalışma yeni geliştirilmekte olan e-TEP sınavının geçerlik ve güvenirliğine yazılı anlatım bölümünün bir parçası olan çevrimiçi yazışma görevine adaylar tarafından verilmiş yanıtlardaki sözdizimi özelliklerinin incelenmesi yolu ile katkıda bulunmayı hedeflemektedir. Bu doğrultuda, çalışmamızın araştırma sorusu şöyledir: Yabancı dil olarak İngilizce test katılımcılarının yazışma görevine verdikleri yanıtlardaki hangi sözdizimsel özellikler puanlama ölçeğinin gramer boyutundan aldıkları üç farklı puan düzeyi ile (ileri, yüksek ve orta) örtüşmektedir?

Araştırma sorumuzu yanıtlamak üzere, 2024 yılında gerçekleştirilen üç e-TEP pilot testinde uygulanan aynı çevrimiçi yazışma görevine katılan toplam 189 test katılımcısının (her düzeyden 63 katılımcısı) yazılı performanslarını analiz ettik. Sözdizimi analizi için erişimi ücretsiz olan Lu (2010)'un Otomatik Sözdizimsel Karmaşıklık Analizi programını ve yine erişimi ücretsiz olan Kyle (2016)'nın Otomatik Sözdizimsel Karmaşıklık Ölçüm Aracı'nı kullandık. Analiz sonucunda üç farklı puan düzeyinde sözdizimsel farklılıklar gözlemlemeyi öngörüyoruz. Bu da, e-TEP Çevrimiçi Yazışma Görevinin ikinci dil yazılı anlatım becerileri açısından değişik puan düzeylerini farklı sözdizimi özelliklerine göre ayırt edebileceğine yönelik kanıt sağlanmış olacaktır. Ayrıca çalışma sonuçları son yıllarda önem kazanmaya başlayan okumaya dayalı yazma karma görevleri ve bu görevler yoluyla elde edilen puanlar açısından önemli çıkarımlar yapılmasını sağlayarak ikinci dilde yazılı anlatımın değerlendirilmesi alanındaki uygulamalara ve araştırmalara katkı sağlayacaktır.

Anahtar Kelimeler: Çevrimiçi *yazışma görevi, e-TEP, Sözdizimsel* özellikler*, Puan düzeyleri*

[1] aysel.saricaogluaygan@asbu.edu.tr, ORCID: 0000-0002-5315-018X, Sosyal Bilimler Üniversitesi, Ankara, Türkiye

[2] Mustafa.cimrin@osym.gov.tr, ORCID: 0009-0000-6333-564X, ÖSYM, Ankara, Türkiye

[3] aliyevin@metu.edu.tr, ORCID: 0000-0002-2473-2690, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[4] eceselva@metu.edu.tr, ORCID: 0000-0002-1856-8839, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

From Syntax to Scores: Syntactic Indicators of Writing Quality in a Large-Scale Online Exchange Writing Task

Aysel Saricaoglu[1], Mustafa Çimrin[2], Ece Selva Küçükoglu[3], Aliye Evin Yörüdü[4]

This study examines the syntactic features of written responses from three score levels of the Grammar dimension in an Online Exchange Writing Task in e-TEP, a recently developed four-skills English language proficiency exam in Türkiye. Online exchange writing tasks are gaining popularity in writing assessment. They progressively replace other prompt-based task types in high-stakes language proficiency exams (e.g., TOEFL's Writing for An Academic Discussion task). While the literature on L2 writing assessment contains several research articles addressing the validity of reading-to-write integrated tasks, the validation of online exchange writing tasks remains an area that requires attention from researchers. As new writing test tasks "warrant investigation to clarify score interpretations" (Plakans et al., 2019, p. 161), this study seeks to gather evidence to contribute to the validation of the newly developed e-TEP Online Exchange Writing Task by examining test-takers' written responses with regard to syntactic features. To this end, the following specific research question guides our study: What syntactic features in EFL test-takers' written responses to an online exchange writing task are associated with three different score levels (advanced, high, and intermediate) received from the grammar dimension of the scoring rubric?

To answer our research question, we analyzed the written performances of a total 189 test-takers (63 test takers from each level) to the same online exchange writing task implemented in three e-TEP pilot tests conducted in 2024. For the syntactic analysis, we used Lu's (2010) freely accessible automated L2 Syntactic Complexity Analyzer (L2SCA) and Kyle's (2016) freely-accessible Tool for the Automatic analysis of Syntactic Sophistication and Complexity. We expect to observe syntactic differences across the score levels, which will provide evidence that the e-TEP Online Exchange Writing Task elicits different syntactic features that effectively distinguish score levels in L2 writing performance. Our results will also have important implications for new reading-to-write integrated task types and score interpretations, thereby contributing to L2 writing assessment practices and research.

Keywords: *Online exchange writing task, e-TEP, Syntactic features, Score levels*

[1] aysel.saricaogluaygan@asbu.edu.tr, ORCID: 0000-0002-5315-018X, Sosyal Bilimler Üniversitesi, Ankara, Türkiye

[2] Mustafa.cimrin@osym.gov.tr, ORCID: 0009-0000-6333-564X, ÖSYM, Ankara, Türkiye

[3] aliyevin@metu.edu.tr, ORCID: 0000-0002-2473-2690, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[4] eceselva@metu.edu.tr, ORCID: 0000-0002-1856-8839, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

Beceri Temelli İngilizce Başarı Testinin Psikometrik Özelliklerinin Madde Tepki Kuramı Çerçevesinde İncelenmesi

Veli Atalay [1], Neşe ÖZTÜRK GÜBEŞ [2]

Bu araştırmanın amacı, Millî Eğitim Bakanlığı (MEB) tarafından yayınlanan İngilizce dersi beceri temelli başarı testinin psikometrik özelliklerini Madde Tepki Kuramı (MTK) çerçevesinde incelemektir. Araştırmanın çalışma grubunu, Burdur ilinde bir ortaokulda 5. Sınıfta öğrenim görmekte olan 189 öğrenci oluşturmaktadır. Öğrencilere, MEB tarafından "Movies" ünitesi kapsamında yayınlanan 20 çoktan seçmeli maddeden oluşan İngilizce başarı testi 2023-2024 eğitim-öğretim yılı bahar döneminde uygulanmıştır. Veri analizinin ilk aşamasında MTK'nın varsayımları incelenmiş ve testin güvenirliği MTK marjinal güvenirlik katsayısı ile hesaplanmıştır. Veri analizinin ikinci aşamasında testin madde ve yetenek parametreleri bir-, iki- ve üç parametreli lojistik (PL) modeller ile kestirilmiştir. Daha sonra veri setinin hangi modele uyum sağladığını belirlemek için modeller ANOVA yapılarak karşılaştırılmıştır. Araştırmanın sonucunda, tek boyutluluk varsayımını sağlayan ve güvenirliği yüksek olan İngilizce başarı testi puanlarının 2PL modele uyum sağladığı görülmüştür. Madde güçlük parametreleri -1.89 ile 0.53 arasında değerler alırken, madde ayırt edicilik parametreleri 0.69 ile 2.70 arasında değerler almıştır. Öğrencilerin yetenek parametrelerinin ise -1.80 ile 1.82 arasında değiştiği görülmüştür.

Anahtar Kelimeler: *Beceri temelli* İngilizce *başarı testi, madde tepki kuramı, 2 PL model*

[1] 2330425009@ogr.mehmetakif.edu.tr, ORCID: 0000-0001-2345-6789, MEB, Burdur, Türkiye

[2] nozturk@mehmetakif.edu.tr, ORCID: 0000-0003-0179-1986, Burdur Mehmet Akif Ersoy Üniversitesi, Burdur,Türkiye

Examining the Psychometric Properties of the Skill-Based English Achievement Test within the Framework of Item Response Theory

Veli ATALAY[1], Neşe ÖZTÜRK GÜBEŞ[2]

The purpose of this research is to examine the psychometric properties of the English course skill- based achievement test published by the Ministry of National Education (MoNE) within the framework of Item Response Theory (IRT). The study group of the research consists of 189 students studying in the 5th grade at a secondary school in Burdur. The English achievement test, consisting of 20 multiple- choice items published by the MoNE within the scope of the "Movies" unit, was administered to the students in the spring semester of the 2023-2024 academic year. At the first step of the data analyses assumptions of IRT were checked and and the IRT marginal reliability coefficient was calculated to provide evidence for its reliability. At the secon step, the item and ability parameters of the test were estimated with one-, two-, and three-parameter logistic (PL) models. Then, the models were compared by ANOVA to determine which model the data set fits. As a result of the research, it was seen that English achievement test scores, which meet the one-dimensionality assumption and have high reliability, fit the 2PL model. While item difficulty parameters took values between -1.89 and 0.53, item discrimination parameters took values between 0.69 and 2.70. It was observed that the ability parameters of the students varied between -1.80 and 1.82.

Keywords: *Skills-based English achievement test, item response theory, 2PL model*

[1] 2330425009@ogr.mehmetakif.edu.tr, ORCID: 0000-0001-2345-6789, MEB, Burdur, Türkiye

[2] nozturk@mehmetakif.edu.tr, ORCID: 0000-0003-0179-1986, Burdur Mehmet Akif Ersoy Üniversitesi, Burdur,Türkiye

Avrupa Dilleri Ortak Çerçevesi Seviyeleri Madde Tepki Kuramı Karşılıkları: Bir Standart Belirleme Çalışması

Levent Yakar [1], Elif Kantarcıoğlu[2], Tuba Arabacı Atlamaz[3], Erkan Hasan Atalmış[4], Reyhan Ağçam[5]

Avrupa Dilleri Ortak Çerçevesi (ADOÇ) Avrupa Konseyinin vatandaşları için nitelikli ve kapsayıcı bir eğitim alma hakkını sağlamaya amaçlayan ve dil öğrenimine odaklanan öncü çalışmalarındandır. Hâlihazırda 40 dilde mevcut ADOÇ yeterlilikleri halen güncellemekte ve kapsamını genişletmektedir (Avrupa Konseyi, 2020). ADOÇ ile dil öğreniminde önemli standartları beraberinde getirmiştir. Madde tepki kuramının (MTK) gruptan bağımsız olarak elde edilen madde parametreleri sayesinde herkes için geçerli stabil bir yapısı en belirgin özelliklerindendir. Bu araştırmada bir İngilizce yeterlik sınavından elden edilen zorluk parametresinin standart belirleme çalışmasıyla ADOÇ seviyelerine göre sınıflandırılması amaçlanmıştır. Araştırma verileri üçü vakıf, ikisi devlet üniversitesinin yabancı diller yüksekokulu İngilizce hazırlık programında öğrenim gören toplam 1462 öğrenciden toplanmıştır. Veri toplama aracı olarak araştırma ekibi tarafından hazırlanan 256 okuma ve 128 dinleme alanına yönelik toplam 384 çoktan seçmeli soru kullanılmıştır. Yapılan MTK analizleri sonucunda 3PL modelinin en iyi uyuma sahip olduğu görülmüştür. Maddelerden istenen madde uyuma elendikten sonra geriye kalan 218 okuma sorusu iki ve 109 dinleme sorusu için bir olmak üzere üç farklı soru takımı standart belirleme çalışması gerçekleştirilmiştir. Standart belirleme yöntemlerinden MTK'ya uyumlu olan işaretleme yöntemi (Mitzel vd., 2001) kullanılmış ve 2/3 yanıtlama olasılığı tercih edilmiştir. Her bir grupta 10'ar İngilizce alan uzmanından alınan ilk görüşten sonraki her iki tartışma oturumu sonunda da kesme noktası hakkında görüş alınarak yöntem gerçekleştirilmiştir. Son turda elde edilen sonuçlara göre okuma alanındaki kesme puanları A2 düzeyi için -,63, B1 seviyesi için,03 ve B2 seviyesi için 1,18 olarak belirlenmiştir. Dinleme alanındaki kesme puanları ise A1 düzeyi için -,21, A2 düzeyi için -,17, B1 seviyesi için,71 ve B2 seviyesi için 1,64 olarak belirlenmiştir. Ortaya çıkan sonuçlar önceki çalışmalar eşliğinde tartışılacaktır.

Anahtar Kelimeler: *Avrupa dilleri ortak* çerçevesi*, madde tepki kuramı, standart belirleme.*

Kaynakça

Avrupa Konseyi, (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume.* Strasbourg: Council of Europe.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2013). The bookmark procedure: Psychological perspectives. In Setting performance standards (pp. 263-296). Routledge.

[1] leventyakar@gmail.com:, ORCID: 0000-0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[2] kutevu@bilkent.edu.tr:, ORCID: XXXX-XXXX-XXXX-XXXX, Bilkent Üniversitesi, Ankara, Türkiye

[3] tubaarabaci@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Boğaziçi Üniversitesi, İstanbul, Türkiye

[4] erkanatalmis@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Manisa Celal Bayar Üniversitesi. Manisa, Türkiye

[5] reyhanagcam@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

Item Response Theory Correspondences of Common European Framework of Reference for Languages Levels: A Standard-Setting Study

Levent Yakar [1], Elif Kantarcıoğlu[2], Tuba Arabacı Atlamaz[3], Erkan Hasan Atalmış[4], Reyhan Ağçam[5]

The Common European Framework of Reference for Languages (CEFR) is one of the pioneering studies of the Council of Europe that aims to ensure the right to a qualified and inclusive education for its citizens and focuses on language learning. The CEFR qualifications currently available in 40 languages are currently being updated and expanded (Council of Europe, 2020). It has brought important standards in language learning with CEFR. The most prominent feature of the item response theory (CRT) is its stable structure that is valid for everyone thanks to the item parameters obtained independently of the group. In this study, it was aimed to classify the difficulty parameter obtained from an English proficiency exam according to the CEFR levels with a standard-setting study. The research data were collected from a total of 1462 students studying in the English preparatory program of the foreign languages school of three foundation universities and two state universities. A total of 384 multiple-choice questions prepared by the research team for 256 reading and 128 listening areas were used as data collection tools. As a result of the CRT analyses, it was seen that the 3PL model had the best fit. After the required items were eliminated from the items, three different question sets were standardized, two for the remaining 218 reading questions and one for the 109 listening questions. The marking method compatible with MTK (Mitzel et al., 2001) was used among the standardized methods and a 2/3 response probability was preferred. The method was implemented by obtaining opinions about the cut-off point at the end of both discussion sessions after the first opinion received from 10 English field experts in each group. According to the results obtained in the last round, the cut-off scores in the reading area were determined as -.63 for level A2,.03 for level B1 and 1.18 for level B2. The cut-off scores in the listening area were determined as -.21 for level A1, -.17 for A2,.71 for level B1 and 1.64 for level B2. The results will be discussed in light of previous studies.

Keywords: *Common European Framework of Reference, Item Response Theory, Standard Setting*

Reference

Council of Europe, (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume.* Strasbourg: Council of Europe.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2013). The bookmark procedure: Psychological perspectives. In Setting performance standards (pp. 263-296). Routledge.

[1] leventyakar@gmail.com:, ORCID: 0000-0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[2] kutevu@bilkent.edu.tr:, ORCID: XXXX-XXXX-XXXX-XXXX, Bilkent Üniversitesi, Ankara, Türkiye

[3] tubaarabaci@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Boğaziçi Üniversitesi, İstanbul, Türkiye

[4] erkanatalmis@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Manisa Celal Bayar Üniversitesi. Manisa, Türkiye

[5] reyhanagcam@gmail.com:, ORCID: XXXX-XXXX-XXXX-XXXX, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

Düşük ve Yüksek Okuma Başarısı Gösteren Öğrenci Özelliklerinin Kestirilmesi

Ömer Kutlu [1], Neslihan Tuğçe Özyeter [2], Dilek Tokhay [3], Kutlutekin Baş [4]

Düşük öğrenci başarısı alanyazında *gelişme sürecinin daha gerisinde* öğrenme *gerçekleştiren* öğrenciler olarak (Vann ve Abraham, 1990), öğrenmek *için daha fazla zaman ayırmasına rağmen yaşıtlarından daha düşük puan elde* öğrenciler olarak (Wen ve Jonhson, 1997), *temel* öğrenme *düzeyinin altında performans gösterme* olarak tanımlanmaktadır (OECD, 2016). PISA, düşük okuma başarısı gösteren öğrencileri gelecekte, doğaya ve topluma karşı sorumluluklarını bilmeyen, ekonomik gelişime katkı sunamayan, bireysel ya da toplumsal değişime direnen yetişkinler olacağına dikkat çekmektedir.

Bu araştırma, "Ortaokul 5. sınıfta düşük ve yüksek okuma başarısı gösteren öğrencilerin okuma başarısını yordayan değişkenler nelerdir?" sorusuna yanıt aramaktadır. Araştırma korelasyonel araştırmadır ve Mersin ilinde, Mersin İl Milli Eğitim Müdürlüğünün desteğiyle yürütülmüştür. Araştırmada MEB Türkçe ders kitabında yer alan okuma metinleri kullanılmıştır.

Öğrenci özelliklerinin yordayıcı rollerinin araştırılmasında çoklu doğrusal regresyon modeli kullanılacaktır. Bu çalışma kapsamında düşük ve yüksek başarı grubundaki öğrenciler başarı puanları açısından dar bir varyansa sahip olacakları için Klasik Test Kuramına dayalı yöntemlerin bir sonuç vermemesi durumunda Madde Tepki Kuramı modellerinden Birey Açıklayıcı Madde Tepki Kuramı kullanılarak birey yetenekleri kestirilecektir. Öğrenci özelliklerinin yeteneği kestirmedeki rolleri belirlenecektir.

Anahtar Kelimeler: *Okuduğunu anlama, okuma okuryazarlığı, düşük okuma başarısı, düşük başarı,* ölçme *ve durum belirleme, regresyon* çözümlemesi.

Kaynakça

Organisation for Economic Co-Operation and Development. (2016). *Low-performing students: Why they fall behind and how to help them succeed*. Organisation for Economic Co-Operation and Development Publishing.

Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly,*

24(2) 177-198. https://doi.org/10.2307/3586898

Wen, Q., & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary- level English majors in China. *Applied Linguistics, 18*(1), 27-48. https://doi.org/10.1093/applin/18.1.27

[1] omerkutlu@ankara.edu.tr, ORCID: 0000-0003-4364-5629, Ankara Üniversitesi, Ankara, Türkiye.

[2] neslihan.ozyeter@kocaeli.edu.tr, ORCID: 0000-0003-1558-1293, Kocaeli Üniversitesi, Kocaeli, Türkiye.

[3] dilektokhay@gmail.com, ORCID: 0009-0000-4019-5592, Mersin İl Milli Eğitim Müdürlüğü, Mersin, Türkiye.

[4] kutlutekinbas.ilmem@gmail.com, ORCID: 0009-0003-6548-806X, Mersin İl Milli Eğitim Müdürlüğü, Mersin, Türkiye.

Predicting Characteristics of Low and High Reading Achievers

Ömer Kutlu[1], Neslihan Tuğçe Özyeter[2], Dilek Tokhay[3], Kutlutekin Baş[4]

Low student achievement is defined in the literature as students who are lagging behind in their developmental process (Vann & Abraham, 1990), students who earn lower scores than their peers despite spending more time learning (Wen & Johnson, 1997), and students who perform below the basic learning level (OECD, 2016). PISA highlights that students with low reading achievement are likely to become adults who are unaware of their responsibilities toward nature and society, who cannot contribute to economic development, and who resist individual or social change.

This research aims to answer the question, "What are the variables that predict the reading success of students with low and high reading achievement in the 5th grade of secondary school?" The research employs a correlational design and was conducted in Mersin province with support from the Mersin Provincial Directorate of National Education. Reading texts from the MEB Turkish coursebook were used in the study.

A multiple linear regression model will be used to investigate the predictive roles of student characteristics. Given that students in the low and high achievement groups will have a narrow range of achievement scores within the scope of this study, if methods based on Classical Test Theory do not yield satisfactory results, individual abilities will be estimated using the Individual Explanatory Item Response Theory from Item Response Theory models. The roles of student characteristics in estimating ability will be examined.

Keywords: *Reading comprehension, reading literacy, low reading achievement, low achievement, measurement and assessment, regression analysis.*

References

Organisation for Economic Co-Operation and Development. (2016). *Low-performing students: Why they fall behind and how to help them succeed*. Organisation for Economic Co-Operation and Development Publishing.

Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly, 24*(2) 177-198. https://doi.org/10.2307/3586898

Wen, Q., & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary- level English majors in China. *Applied Linguistics, 18*(1), 27-48. https://doi.org/10.1093/applin/18.1.27

[1] omerkutlu@ankara.edu.tr, ORCID: 0000-0003-4364-5629, Ankara Üniversitesi, Ankara, Türkiye.

[2] neslihan.ozyeter@kocaeli.edu.tr, ORCID: 0000-0003-1558-1293, Kocaeli Üniversitesi, Kocaeli, Türkiye.

[3] dilektokhay@gmail.com, ORCID: 0009-0000-4019-5592, Mersin İl Milli Eğitim Müdürlüğü, Mersin, Türkiye.

[4] kutlutekinbas.ilmem@gmail.com, ORCID: 0009-0003-6548-806X, Mersin İl Milli Eğitim Müdürlüğü, Mersin, Türkiye.

Bayes Ağlarına Dayalı Kişiselleştirilmiş Geri Bildirim Tasarımı: Feedup, Feedback ve Feedforward Üçgeni

Cennet Terzi Müftüoğlu[1], Halil Yurdugül[2]

Bu çalışma, e-öğrenme ortamlarında öğrenme süreçlerini desteklemek ve daha etkin kılmak amacıyla, öğrenenlerin bireysel ihtiyaçlarına uygun geri bildirimlerin zamanında ve etkili bir şekilde sağlanmasını amaçlayan yapay zeka temelli bir geri bildirim sisteminin tasarımını ve uygulanmasını ele almaktadır. E-öğrenme ortamlarında öğrenenlerin bıraktığı dijital izler sayesinde öğrenme deneyimleri izlenebilir, modellenebilir ve öğrenmeyi iyileştirmek amacıyla tasarlanabilir hale gelmiştir. Eğitsel veri madenciliği, öğrenme analitikleri ve yapay zeka, öğrenme deneyimini oluşturan verilerdeki örüntüleri keşfetme ve örüntüleri işe koşma süreçlerinde metodolojik yaklaşımlara dayalı uygulamalar üretilmesine olanak sağlamaktadır. Yapay zeka, öğrenme analitikleri aracılığıyla her öğrenen için kişiselleştirilmiş geri bildirimler üretilmesini ve bu geri bildirimlerin öğrenenlerin ihtiyaç duyduğu anlarda sunulmasını mümkün kılmaktadır. Öğrenmeyi iyileştirmede en güçlü stratejilerimizden biri geri bildirimlerdir (Hattie & Timperley, 2007; Pardo, 2018). Eğitim sistemini de devingen (dinamik) bir sisteme dönüştüren unsur da geri bildirim sistemidir. Bu araştırma kapsamında geri bildirim kavramı içerdiği bilgi bağlamına göre ele alınmıştır. İdeal bir öğrenme ortamının "Nereye gidiyorum? (amaca yönelik bildirim)", "Nasıl gidiyorum? (doğrulamaya yönelik bildirim)", "Sırada ne var? (ayrıntılandırmaya yönelik bildirim)" olmak üzere üç soruya yanıt vermesi gerekmektedir. Bu geri bildirim tasarımı aynı zamanda ideal bir geri bildirim deneyimi olarak kavramsallaştırılmaktadır (Hattie ve Timperley, 2007). Araştırmada kullanılan yapay zeka temelli geri bildirim sistemi bayes ağları yaklaşımına göre modellenmiş, 89 lisans öğrencisi ve 5 öğretim üyesinin katılımıyla yürütülen uygulamalarda test edilmiştir. Öğrencilerin öğrenme süreçleri, başarı yönelimleri ve yardım arama eğilimleri açısından incelenmiş ve sistemin, öğrenenlerin performanslarını olumlu yönde etkilediği sonucuna varılmıştır. Bu çalışmanın bulguları, e- öğrenme ortamlarında kişiselleştirilmiş geri bildirim sistemlerinin öğrenme süreçlerini iyileştirmede kritik bir rol oynadığını göstermektedir. Özellikle yapay zeka tabanlı geri bildirim sistemleri, öğrenenlere daha esnek, dinamik ve etkili bir öğrenme deneyimi sunarak, eğitim teknolojilerinde yeni bir dönemin kapılarını aralamaktadır.

Anahtar Kelimeler: *e-*öğrenme*, yapay zeka, bayes ağları, geri bildirim tasarımı, feedup, feedback, feedforward*

Kaynakça

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81- 112. https://doi.org/10.3102/003465430298487

Pardo, A. (2018). A feedback model for data-rich learning experiences. A*ssessment & Evaluation in Higher Education, 43*(3), 428-438. https://doi.org/10.1080/02602938.2017.1356905

[1] cennet.terzi@gmail.com:, ORCID: 0000-0002-0875-4806, Hacettepe Üniversitesi, Ankara, Türkiye

[2] yurdugul@hacettepe.edu.tr, ORCID: 0000-0001-7856-4664, Hacettepe Üniversitesi, Ankara, Türkiye

Personalized Feedback Design Based on Bayesian Networks: The Triangle of Feedup, Feedback, and Feedforward

Cennet Terzi Müftüoğlu[1], Halil Yurdugül[2]

This study focuses on designing an artificial intelligence-based feedback system to support and enhance learning processes in e-learning environments by providing timely and effective feedback according to individual learners' needs. Through the digital traces left by learners in e-learning environments, learning experiences can be monitored, modeled, and designed to improve learning. Educational data mining, learning analytics, and artificial intelligence enable the discovery of patterns in the data that shape the learning experience and the development of methodological approaches for implementing these patterns. Artificial intelligence allows for the generation of personalized feedback for each learner through learning analytics and the delivery of this feedback when learners need it. Feedback is considered one of the most powerful strategies in improving learning. The feedback system transforms the educational system into a dynamic system. In this research, the concept of feedback is discussed within the context of the information it contains. An ideal learning environment should answer three questions: "Where am I going? (goal-oriented feedback)", "How am I doing? (verification-oriented feedback)", and "What's next? (detail-oriented feedback)". This feedback design is conceptualized as an ideal feedback experience. The artificial intelligence-based feedback system used in the research was modeled according to Bayesian networks, tested in applications involving 89 undergraduate students and 5 faculty members. The students' learning processes, achievement orientations, and help-seeking tendencies were examined, and it was concluded that the system positively affected learners' performance. The findings of this study demonstrate that personalized feedback systems in e-learning environments play a critical role in improving learning processes. Specifically, artificial intelligence-based feedback systems open the doors to a new era in educational technologies by providing learners with a more flexible, dynamic, and effective learning experience.

Keywords: *E-learning, artificial intelligence, Bayesian networks, feedback design, feed up, feedback, feedforward.*

Reference

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81- 112. https://doi.org/10.3102/003465430298487

Pardo, A. (2018). A feedback model for data-rich learning experiences. A*ssessment & Evaluation in Higher Education, 43*(3), 428-438. https://doi.org/10.1080/02602938.2017.1356905

[1] cennet.terzi@gmail.com:, ORCID: 0000-0002-0875-4806, Hacettepe Üniversitesi, Ankara, Türkiye

[2] yurdugul@hacettepe.edu.tr, ORCID: 0000-0001-7856-4664, Hacettepe Üniversitesi, Ankara, Türkiye

Öğrencilerin Üst Bilişsel Düşünme Becerilerinin Geliştirilmesinde Örnek Bir Uygulama: SOT

Seda YILMAZ [1]

Öğrenme, merak etme ve soru sorma ile başlayan uçsuz bucaksız bir süreçtir. Soru sormak, insanın doğasında bulunan ve varoluş nedenlerinin başında gelen önemli bir özellik olarak görülürken, hem kendisi hem de çevresi için değişim yaratma çabaşı olarak da nitelendirilmektedir (Fisher, 2022). Eğitim perspektifinden bakıldığında, yapılan bilimsel çalışmalarda öğrencilerin düşünme becerileri doğrultusunda soru sorma becerilerinin de geliştiği ve sesli veya sessiz olarak soru soran öğrencilerin daha üst düzey düşünme becerilerine sahip oldukları belirtilmektedir (Giordan, 2008). Bu bağlamda soru sormanın, hem öğretmenler hem de öğrenciler için bilgiyi genişletme ve yeni fikirler üretmek bakımından değerli olduğu vurgulanmaktadır (Büyükalan, 2007). Bununla birlikte soru sorma becerisi gelişmeyen öğrencilerde öğrenmeye karşı isteksizlik ve çekingenlik gözlemlenmektedir. Bu durum da, soru sormaktan vazgeçen ve sorgulama isteğinde bulunmayan bireylerin ortaya çıkmasına neden olmaktadır (Right Question Institute, 2024). Bu çalışmanın amacı, eğitim ortamlarında ve ölçme değerlendirme süreçlerinde madde yazımında Soru Oluşturma Tekniği'ni (SOT) kullanan öğretmenlerin bu teknikle ilgili görüşlerini almak ve bu tekniğin eğitim ve ölçme değerlendirme faaliyetleri üzerindeki etkisini belirlemektir. Çalışmanın ayrıca SOT tekniğinin ölçme süreçlerinde kullanımı ile ilgili öğretmenlere farklı bir bakış açısı ve alternatif sunacağı düşünülmektedir. Çalışmada nitel araştırma yöntemi kullanılmış, ölçme değerlendirme çalışmalarında ve sınıf içi etkinliklerde SOT tekniğini kullanan 12 öğretmen ile yarı yapılandırılmış görüşme formu aracılığıyla görüşmeler yapılmıştır. Elde edilen veriler içerik analizi yöntemi ile analiz edilmiştir. SOT tekniğini sınıf içi etkinliklerde kullanan öğretmenler, tekniğin öğrencilerin soru sorma becerisini geliştirdiğini, öğrenmeye daha istekli olduklarını ve derse katılımlarını büyük oranda artırdığını belirtmişlerdir. Tekniği ölçme değerlendirme çalışmalarında madde yazımında kullanan öğretmenler ise, kazanımı doğrudan ölçen daha isabetli sorular sorabildiklerini ve bu tekniği kullanarak hazırladıkları soruların öğrencilerin yakınsak, ıraksak, eleştirel düşünme gibi üst bilişsel düşünme becerilerini geliştirmeye fırsat tanıdığını vurgulamışlardır. Bu bağlamda, SOT tekniğinin daha fazla öğretmen tarafından bilinmesi, hem ölçme değerlendirme süreçlerinde hem de eğitim ortamlarında öğrencilerin yaş grubu ve gereksinimlerine göre kullanılabilmesi için öğretmenlere yönelik hizmet içi eğitimlerin düzenlenmesi önerilmektedir.

Anahtar Kelimeler: üst *bilişsel düşünme, soru sorma becerisi, SOT.*

Kaynakça

Büyükalan, F. (2007). Soru Sorma Sanatı (2. Baskı). *Ankara: Nobel Yayın Dağıtım.*

Fisher, R. (2022). Çocuklara Düşünmeyi Öğretmek (2. Baskıdan çeviri). Z. Nurdan Baysal & E. Sarıcan (Çev. Ed.). *Ankara: Pegem Akademi Yayıncılık.*

Giordan, A. (2008). Öğrenme. (Çev. M. Baştürk, E. Bozavlı & M. Tulan). *Ankara: De Ki Basım Yayın.*
Right Question Institute, (2024). https://rightquestion.org/what-is-the-qft/ (Erişim Tarihi: 11.07.2024).

[1]sedaipek48@gmail.com, ORCID: 0000-0002-6255-061X, Muğla İl Milli Eğitim Müdürlüğü, Muğla, Türkiye.

An Exemplary Application in Developing Students' Metacognitive Thinking Skills: QFT

Seda YILMAZ[1]

Learning is a vast process that begins with curiosity and asking questions. While asking questions is seen as an important feature found in human nature and one of the main reasons for existence, it is also described as an effort to create change for both oneself and the environment (Fisher, 2022). From an educational perspective, scientific studies indicate that students' questioning skills improve in line with their thinking skills, and that students who ask questions aloud or silently have higher-level thinking skills (Giordan, 2008). In this context, it is emphasized that asking questions is valuable for both teachers and students in terms of expanding knowledge and generating new ideas (Büyükalan, 2007). However, reluctance and shyness towards learning are observed in students who do not develop the ability to ask questions. This situation leads to the emergence of individuals who give up asking questions and do not desire to question (Right Question Institute, 2024). The purpose of this study is to get the opinions of teachers who use the Question Formation Technique (SOT) in item writing in educational environments and measurement and evaluation processes, about this technique and to determine the effect of this technique on education and evaluation activities. In addition, it is thought that this study will offer teachers a different perspective and alternative regarding the use of the QFT technique in measurement processes. Qualitative research method was used in the study, and interviews were conducted through a semi-structured interview form with 12 teachers who used the QFT technique in measurement and evaluation studies and in-class activities. The data obtained was analyzed using the content analysis method. Teachers who used the QFT technique in classroom activities stated that the technique improved students' questioning skills, made them more willing to learn, and greatly increased their participation in the lesson. Teachers who use the technique in writing items in assessment and evaluation studies have emphasized that they can ask more accurate questions that directly measure the outcome and that the questions they prepare using this technique provide students with the opportunity to develop metacognitive thinking skills such as convergent, divergent and critical thinking. In this context, it is suggested that in-service training, courses and activities can be organized on provincial and district basis in order to make the QFT technique known by more teachers and to popularize its use both in measurement and evaluation processes.

Keywords: *metacognitive thinking, questioning skills, QFT.*

Reference

Büyükalan, F. (2007). *The art of asking questions* (2nd Edition). Nobel.

Fisher, R. (2022). *Teaching children to think* (Translation from the 2nd edition). Z. Nurdan Baysal & E. Sarıcan (Trans. ed.). Pegem.

Giordan, A. (2008). *Learning.* (Trans. M. Baştürk, E. Bozavlı & M. Tulan). De Ki.

Right Question Institute, (2024).https://rightquestion.org/what-is-the-qft/(Date of access:11.07.2024)

[1] e-mail: sedaipek48@gmail.com, ORCID: 0000-0002-6255-061X, Muğla Provincial Directorate of National Education, Muğla, Türkiye.

Kavram Karikatürlü Plicker Uygulamasının Eleştirel Düşünme Eğilimine Etkisi

Gizem Sarıkaya [1], Elif Selvi [2], Ömer Faruk Şen[3]

Öğretmenler sınıf içi uygulamalarda, öğrencilerin öğrenmelerine yönelik planlı ve plansız birçok veri toplamaktadırlar. Öğrenci öğrenme süreçlerine yönelik planlı ve sistematik veri toplamanın bir yolu olan hızlı cevap sistemleri, bu bağlamda Plicker uygulaması ile öne çıkmaktadır. Plicker, öğretmenlerin öğrenci yanıtlarını anında toplamasına ve değerlendirmesine olanak tanıyan, teknolojik altyapıya ihtiyaç duymayan bir mobil uygulamadır. Hızlı cevap sistemi sayesinde, öğretmenler geri bildirimleri anında alarak öğretim stratejilerini daha etkin şekilde geliştirebilir (Van Laar vd., 2020). Ancak, alan yazın incelendiğinde, Plicker'ın genellikle sadece veri toplama amacıyla kullanıldığı ve kullanılan problemin niteliğine dair sınırlı bilgi bulunduğu tespit edilmiştir (Tuncer & Şimşek, 2019). Weiss ve diğerleri (2020) anlık cevap sistemlerinin tek başına yeterli olmadığını, kullanılan problemin üst düzey düşünme becerilerini teşvik edici nitelikte olması gerektiğini vurgulamışlardır. Bu bağlamda kavram karikatürleri, farklı bakış açıları sunarak ve öğrencilerin argümanlarını açıklamalarına fırsat vererek eleştirel düşünme becerilerinin gelişimini desteklemektedir. Bu çalışma, Plicker destekli kavram karikatürlerinin öğrencilerin eleştirel düşünme eğilimlerine etkisini araştırmayı amaçlamaktadır. Zayıf deneysel desen kullanılarak 6. sınıf öğrencileri ile ısı ve sıcaklık ünitesinde yürütülen çalışmada, öğrencilerin eleştirel düşünme eğilimleri "Ortaokul Öğrencileri İçin Eleştirel Düşünme Eğilimi Ölçeği" ile ölçülmüştür (Yıldırım Döner & Demir, 2021). Sonuç olarak, Plicker uygulamalarının, kavram karikatürleri ile kullanıldığında öğrencilerin eleştirel düşünme eğilimlerine ve sınıf içi katılımlarına olumlu etkiler sağladığı belirlenmiştir.

Anahtar Kelimeler: *hızlı cevap sistemleri, plicker, kavram karikatürü, eleştirel düşünme eğilimi*

Kaynakça

Tuncer, M., & Şimşek, M. (2019). Ortaokul 5. sınıf matematik dersi bölme işlemi konusunda plickers uygulamasının matematik kaygısına ve matematik başarısına etkisi. *OPUS International Journal of Society Researches, 13*(19), 281-310.

van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. Sage Open.

Weiss, D. J., McGuire, P., Clouse, W., & Sandoval, R. (2020). Clickers are not enough: Results of a decade-long study investigating instructional strategies in chemistry. *Journal of College Science Teaching, 49*(3), 58- 65.

Yıldırım Döner, S., & Demir, S. (2021). Ortaokul öğrencileri için eleştirel düşünme eğilimi ölçeği'nin geliştirilmesi. *Pamukkale* Üniversitesi *Eğitim Fakültesi Dergisi, 54*, 99-129.

[1] g.sarikaya25@gmail.com, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

[2] elifselvi06@gmail.com, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

[3] ofaruksen@kku.edu.tr, 0000-0002-2103-8791, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

The Effect of Plicker Application with Concept Cartoons on Critical Thinking Disposition

Gizem Sarıkaya [1], Elif Selvi [2], Ömer Faruk Şen[3]

Teachers collect both planned and unplanned data regarding students' learning during classroom activities. Quick response systems, a method of collecting planned and systematic data on student learning processes, stand out with the Plicker application in this context. Plicker is a mobile application that allows teachers to instantly collect and evaluate student responses without requiring technological infrastructure. Through the quick response system, teachers can receive immediate feedback and develop their teaching strategies more effectively (Van Laar et al., 2020). However, a review of the literature reveals that Plicker is generally used only for data collection, with limited information available on the nature of the problems used (Tuncer & Şimşek, 2019). Weiss and others (2020) emphasized that quick response systems alone are not sufficient; the problems used should be of a nature that encourages higher-order thinking skills. In this context, concept cartoons support the development of critical thinking skills by presenting different perspectives and allowing students to explain their arguments. This study aims to investigate the effect of Plicker-supported concept cartoons on students' critical thinking disposition. Using a weak experimental design, the study was conducted with 6th-grade students in the unit on heat and temperature, and students' critical thinking dispositions were measured with the "Critical Thinking Disposition Scale for Middle School Students" (Yıldırım Döner & Demir, 2021). As a result, it was determined that the use of Plicker applications with concept cartoons positively affected students' critical thinking dispositions and their participation in classroom activities..

Keywords: *quick response systems, Plicker, concept cartoons, critical thinking disposition*

Reference

Tuncer, M., & Şimşek, M. (2019). Ortaokul 5. sınıf matematik dersi bölme işlemi konusunda plickers uygulamasının matematik kaygısına ve matematik başarısına etkisi. *OPUS International Journal of Society Researches, 13*(19), 281-310.

van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. Sage Open.

Weiss, D. J., McGuire, P., Clouse, W., & Sandoval, R. (2020). Clickers are not enough: Results of a decade-long study investigating instructional strategies in chemistry. *Journal of College Science Teaching, 49*(3), 58- 65.

Yıldırım Döner, S. ve Demir, S. (2021). Ortaokul öğrencileri için eleştirel düşünme eğilimi ölçeği'nin geliştirilmesi. *Pamukkale* Üniversitesi *Eğitim Fakültesi Dergisi, 54*, 99-129

[1] g.sarikaya25@gmail.com, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

[2] elifselvi06@gmail.com, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

[3] ofaruksen@kku.edu.tr, 0000-0002-2103-8791, Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale, Türkiye

Üstün Zekâlı İlkokul Öğrencilerin Bilimsel Yaratıcılık Becerilerinin CHAID Analizi ile Değerlendirilmesi

Niymet Demirci [1], Burcu Sel [2]

Üstün zekâlı ilkokul öğrencilerinin bilimsel yaratıcılık becerilerinin değerlendirilmesi, öğrencilerin bilimsel düşünme yeteneklerini ve yenilikçi çözümler geliştirme kapasitelerini anlamak açısından kritik bir öneme sahiptir. Bilimsel yaratıcılık, çeşitli bileşenlerden oluşur ve her bir bileşen farklı bilişsel süreçleri içerir. Yaratıcılık literatüründe, bu bileşenlerin bireyler arasında farklılık gösterebileceği ve bu farklılıkların yaratıcılık düzeylerini etkilediği vurgulanmaktadır (Asal, 2020). Bu bağlamda, araştırmada üstün zekâlı öğrencilerin sahip oldukları bilimsel yaratıcılıklarının düzeyleri ve bilimsel yaratıcılıklarını yordayan değişkenlerin incelenmesi amaçlanmıştır. Bu çalışmada ilişkisel tarama modeli kullanılmıştır. İlişkisel tarama modeli, mevcut durumu belirlemek ve öğrencilerin bilimsel yaratıcılık becerilerini kapsamlı bir şekilde incelemek için uygun bir yöntem olarak seçilmiştir. Bu araştırmada, ilkokul dördüncü sınıf seviyesindeki 251 üstün zekâlı öğrencinin bilimsel yaratıcılık düzeylerini belirlemek amacıyla Hu ve Adey (2002) tarafından geliştirilen Bilimsel Yaratıcılık Ölçeği kullanılmıştır. Ölçeğin Türkçe uyarlaması Asal (2020) tarafından yapılmış ve güvenilirliği bu çalışmada tekrar test edilerek doğrulanmıştır. Ölçek, öğrencilerin bilimsel düşünme süreçleri, yaratıcı fikir üretme becerileri ve bilimsel problemlere özgün çözümler geliştirme yeteneklerini değerlendirmek üzere yedi açık uçlu soru içermektedir. Verilerin analizinde CHAID analizi kullanılmıştır. Normallik varsayımına bağlı kalmadan, farklı veri türlerini bir arada analiz edebilen CHAID analizi, öğrencilerin bilimsel yaratıcılığı gibi çok değişkenli kavramların altında yatan faktörlerin karmaşık yapısını daha iyi anlaşılmasını sağlamıştır. Araştırma sonuçlarına göre üstün zekâlı ilkokul öğrencilerinin özgün deneyler tasarlama ve hipotez kurma becerilerinin beklenenden düşük çıkmıştır. Öğrencilerin bilimsel yaratıcılığını ise Kitap okuma (özellikle macera kitapları), belgesel izleme ve anne-baba eğitim düzeyi gibi faktörler önemli ölçüde etkilemektedir. Bu bulgular, öğrencilerin bilimsel meraklarını desteklemek için evde ve okulda yapılabilecek çeşitli etkinliklerin önemini vurgulamaktadır.

Anahtar Kelimeler: Üstün *zekâ,* İlkokul*, Bilimsel yaratıcılık*

Kaynakça

Asal, R. (2020). *The impact of engineering design-based science teaching on fourth-grade elementary students' scientific creativity and critical thinking skills] (Master's thesis).* Gazi University, Institute of Educational Sciences, Ankara.

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education, 24*(4), 389-403.

[1] niymetdemirci@gmail.com, ORCID: 0000-0002-2763-1879, Torbalı Bilim ve Sanat Merkezi, İzmir, Türkiye

[2] burcusel3@gmail.com, ORCID: 0000-0002-7663-0434, Emine Rasim Özkaya İlkokul, Ankara, Türkiye

Assessment of Scientific Creativity Skills in Gifted Elementary School Students Using CHAID Analysis

Niymet Demirci [1], Burcu Sel[2]

The assessment of scientific creativity skills in gifted elementary school students is essential for comprehending their scientific reasoning abilities and potential to generate novel solutions. Scientific creativity is a multifaceted construct encompassing various components that rely on distinct cognitive processes. The creativity literature highlights that these components can exhibit individual differences, thereby influencing creativity levels (Asal, 2020). This study aimed to investigate the scientific creativity levels of gifted students and identify the variables that predict their scientific creativity.A correlational research design was adopted to investigate the scientific creativity skills of gifted fourth-grade students. This design was deemed suitable for comprehensively assessing the current state of scientific creativity among these students. The study involved 251 gifted fourth-graders and employed the Scientific Creativity Scale developed by Hu and Adey (2002). The scale, previously validated in England with a Cronbach's alpha of 0.89, demonstrated a reliability coefficient of 0.71 in this study. The scale, comprising seven open-ended questions, measures various dimensions of scientific creativity, including process, character, and product. CHAID analysis was utilized to analyze the data, enabling the exploration of complex relationships among variables without the constraints of normality assumptions. This method is particularly well-suited for understanding multifaceted constructs like scientific creativity. The study findings indicated that gifted elementary school students demonstrated lower than expected abilities in designing original experiments and formulating hypotheses. These results underscore the necessity for tailored instructional approaches and supportive interventions to cultivate scientific creativity among gifted learners. Furthermore, the study revealed that reading habits, particularly a preference for adventure books, documentary viewing, and parental education level significantly influenced the scientific creativity levels of gifted students.

Keywords: *Giftedness, Elementary Education, Scientific Creativity*

References

Asal, R. (2020). *The impact of engineering design-based science teaching on fourth-grade elementary students' scientific creativity and critical thinking skills] (Master's thesis).* Gazi University, Institute of Educational Sciences, Ankara.

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education, 24*(4), 389-403.

[1] niymetdemirci@gmail.com, ORCID: 0000-0002-2763-1879, Torbalı Science and Art Center, İzmir, Turkey

[2]burcusel3@gmail.com, ORCID: 0000-0002-7663-0434, Emine Rasim Özkaya Primary School, Ankara, Turkey

Eğitimde Ayrık Seçenekli Çoktan Seçmeli Madde Kullanımı Üzerine Bir İnceleme

Atilla ÖZDEMİR [1], Selahattin GELBAL [2]

Öğrenci başarısının doğru ve güvenilir bir şekilde ölçülmesi, eğitim sistemlerinin etkinliğini değerlendirmek ve iyileştirmek için kritik bir öneme sahiptir (Özdemir ve Gelbal, 2016). Doğru ölçme ve değerlendirme araçları, öğrenci performansını objektif bir şekilde değerlendirerek eğitim politikalarının ve uygulamalarının geliştirilmesine katkı sağlar. Bu bağlamda, ölçme ve değerlendirme alanında yenilikçi yöntemlerin araştırılması ve uygulanması gerekmektedir. Bu yenilikçi çalışmalardan biri olarak Ayrık Seçenekli Çoktan Seçmeli (ASÇS) maddelere yönelik artan ilgi dikkat çekmektedir (Özdemir, 2022). ASÇS maddeleri, geleneksel çoktan seçmeli maddelerden farklı olarak, seçeneklerin tek tek sunulduğu ve öğrencinin her bir seçeneğe ayrı ayrı yanıt verdiği bir format sunar (Foster, 2016). Bu madde formatı, testlerin güvenilirliğini ve geçerliliğini artırma potansiyeline sahiptir (Foster ve Miller, 2009). Kaynakların etkili ve verimli kullanılabilmesi için kanıta dayalı politikalar ve uygulamalar geliştirilmesi gerekmektedir. Bu doğrultuda, bilimsel araştırmaların güvenilirliğini sağlamak adına alanyazında yapılan çalışmaların dikkatlice incelenmesi önem taşımaktadır. Bu çalışma, ASÇS maddelerinin ölçme ve değerlendirme alanındaki yerini ve etkisini incelemeyi amaçlamaktadır. Bu amaca ulaşmak için, Scopus, Web of Science, Google Akademik, Ulakbim ve YÖK Tez veri tabanlarında yer alan ASÇS ile ilgili makaleler ve tezler PRISMA modeline göre belirlenmiştir. Elde edilen veriler, araştırmacılar tarafından geliştirilen inceleme formu kullanılarak analiz edilecektir. Analiz süreci devam etmekte olup, çalışmadan elde edilecek olası sonuçlarla araştırmacıların ASÇS maddeleri konusunda daha fazla bilgi edinmeleri ve istatistiksel analizler için yeterli teorik bilgiye sahip olmaları beklenmektedir. Ayrıca, bu konuda çalışmak isteyen araştırmacılar için yeni konu başlıkları ve araştırmalara ilişkin sınırlılıklar ortaya konulacaktır.

Anahtar Kelimeler: *Ayrık seçenekli* çoktan *seçmeli madde, Geleneksel* çoktan *seçmeli madde,* Öğrenci *başarısı.*

Kaynakça

Foster, D. (2016). *The discrete option multiple choice test item.* Utah: Caveon.

Foster, D. F., & Miller, H. L., Jr. (2009). A new format for multiple-choice testing: Discrete option multiple- choice. Results from early studies. *Psychology Science Quarterly, 51*, 355-369.

Özdemir, A. (2022). *Ayrık seçenekli* çoktan *seçmeli testlerin uygulanabilirliği.* Yayınlanmamış Doktora Tezi) Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Özdemir, A., & Gelbal, S. (2016). İlköğretim ve ortaöğretim başarı ölçülerinin yükseköğretime geçiş sınav puanlarını yordama gücü. *Eğitimde ve Psikolojide* Ölçme *ve Değerlendirme Dergisi, 7*(2), 309-334. DOI: https://doi.org/10.21031/epod.277887

[1] atillaozdemir@sdu.edu.tr, ORCID: 0000-0003-4775-4435, Süleyman Demirel Üniversitesi, Isparta, Türkiye

[2] gelbal@hacettepe.edu.tr, ORCID: 0000-0001-5181-7262, Hacettepe Üniversitesi, Ankara, Türkiye

An Examination of the Use of Discrete Option Multiple Choice Items in Education

Atilla ÖZDEMİR [1], Selahattin GELBAL [2]

Accurately and reliably measuring student achievement is critical for evaluating and improving the effectiveness of educational systems (Özdemir & Gelbal, 2016). Accurate assessment tools contribute to the development of educational policies and practices by objectively evaluating student performance. In this context, it is essential to explore and implement innovative methods in the field of measurement and evaluation. One such innovative approach is the increasing interest in Discrete Option Multiple Choice (DOMC) items (Özdemir, 2022). Unlike traditional multiple-choice items, DOMC items present options individually, requiring students to respond to each option separately (Foster, 2016). This item format has the potential to enhance the reliability and validity of tests (Foster & Miller, 2009). To utilize resources effectively and efficiently, evidence-based policies and practices need to be developed. Therefore, it is crucial to carefully examine studies in literature to ensure the reliability of scientific research. This study aims to investigate the role and impact of DOMC items in the field of measurement and evaluation. To achieve this goal, articles and theses related to DOMC items in the Scopus, Web of Science, Google Scholar, National Academic Network and Information Center, and National Thesis Centre databases were identified using the PRISMA model. The obtained data will be analyzed using a review form developed by the researchers. The analysis process is ongoing, and it is anticipated that the study will provide researchers with more information about DOMC items and sufficient theoretical knowledge for statistical analyses. Additionally, new research topics and limitations will be identified for researchers interested in this field.

Keywords: *Discrete option multiple choice, Traditional multiple-choice item, Student achievement*

Reference

Foster, D. (2016). *The discrete option multiple choice test item*. Utah: Caveon.

Foster, D. F., & Miller, H. L., Jr. (2009). A new format for multiple-choice testing: Discrete option multiple- choice. Results from early studies. *Psychology Science Quarterly, 51*, 355-369.

Özdemir, A. (2022). *Ayrık seçenekli* çoktan *seçmeli testlerin uygulanabilirliği*. Yayınlanmamış Doktora Tezi) Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Özdemir, A., & Gelbal, S. (2016). İlköğretim ve ortaöğretim başarı ölçülerinin yükseköğretime geçiş sınav puanlarını yordama gücü. *Eğitimde ve Psikolojide* Ölçme *ve Değerlendirme Dergisi, 7*(2), 309-334. DOI: https://doi.org/10.21031/epod.277887

[1] e-mail:atillaozdemir@sdu.edu.tr, ORCID: 0000-0003-4775-4435, Süleyman Demirel University, Isparta, Türkiye

[2] e-mail:gelbal@hacettepe.edu.tr, ORCID: 0000-0001-5181-7262, Hacettepe University, Ankara, Türkiye

Geleneksel ve Öncüllü Çoktan Seçmeli Maddelerin Psikometrik Açıdan İncelenmesi

Erkan Hasan Atalmış[1], Levent Yakar[2], Abdullah Çevik[3]

Çoktan seçmeli madde türleri yüksek güvenirlik ve geçerlik nedeniyle ulusal sınavlarda yaygın bir şekilde kullanılmaktadır. Ancak alan yazı incelendiğinde öncüllü çoktan seçmeli madde türlerinin test güvenirliği ve geçerliğini düşürdüğü gerekçesiyle kullanılmaması gerektiği savunulmaktadır (Haladyna, Downing, & Rodriguez, 2002; Haladyna & Rodriguez, 2013). Bu çalışmanın amacı geleneksel çoktan seçmeli ve öncüllü çoktan seçmeli soruların madde güçlük ve madde ayırt edicilik indekslerine göre karşılaştırmaktır. Bunun veri seti olarak Türkiye'deki bir ilde 2023-204 eğitim-öğretim döneminde her biri yaklaşık 15000 olmak üzere 8. sınıf öğrencisine uygulanan iki deneme sınavının fen bilimleri, inkılap tarihi ve din kültürü ve ahlak derslerinde yer alan 70 çoktan seçmeli madde kullanılmıştır. Bu maddelerin 47'u geleneksel 23'ü öncüllü çoktan seçmeli madde olarak gruplandırılmış ve her bir maddenin madde güçlük (p) ve madde ayırt edicilik indeksleri (r) hesaplanmıştır. Elde edilen sonuçlara göre madde gerek *p* değeri gerekse *r* değeri bu iki madde türü arasında istatistiksel olarak anlamlı bir farklılık göstermediği bulunmuştur. Ancak öncüllü çoktan seçmeli maddelerin seçenekleri kombinasyon türlerine göre sınıflandırıldığında, sonuçların istatistiksel olarak anlamlı farklılık gösterdiği bulunmuştur. Seçeneklerde daha fazla kombinasyon içeren öncüllü çoktan seçmeli maddelerin daha az kombinasyon içeren öncüllü çoktan seçmeli ve geleneksel çoktan seçmeli madde türlerine göre istatiksel olarak daha zor ve daha az ayırt edici olduğu bulunmuştur.

Anahtar Kelimeler: Öncüllü çoktan *seçmeli maddeler, madde güçlük indeksi, madde ayırt edicilik indeksi.*

Kaynakça

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. New York, NY: Routledge.

[1] erkanatalmis@gmail.com, ORCID: 0000-0001-9610-491X, Manisa Celal Bayar Ünversitesi, Manisa, Türkiye

[2] l_yakar@hotmail.com, ORCID: 0000- 0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[3] cevik-abdullah@hotmail.com, ORCID: 0000-0003-0659-480X, Milli Eğitim Bakanlığı, Kahramanmaraş, Türkiye

Psychometric Analysis of Traditional and Complex Multiple Choice Items

Erkan Hasan Atalmış[1], Levent Yakar[2], Abdullah Çevik[3]

Multiple-choice item types are widely used in national exams due to their high reliability and validity. However, when the literature is examined, it is argued that complex multiple-choice item types should not be used because they reduce test reliability and validity (Haladyna, Downing, & Rodriguez, 2002; Haladyna & Rodriguez, 2013). The aim of this study is to compare traditional multiple-choice and complex multiple-choice items in terms of item difficulty and item discrimination indices. The dataset for this study was 70 multiple-choice items from science, history of revolution, and religious culture and ethics courses of two trial exams administered to approximately 15,000 8^{th} grade students each in a province in Turkey during the 2023-2024 academic year. 47 of these items were grouped as traditional and 23 as complex multiple-choice items, and the item difficulty (p) and item discrimination indices (r) for each item were calculated. According to the results obtained, it was found that neither the *p* value nor the *r* value of the item showed a statistically significant difference between these two item types. However, when the options of the complex multiple choice items were classified according to their combination types, it was found that the results showed a statistically significant difference. It was found that multiple choice items with more combinations in the options were statistically more difficult and less discriminatory than multiple choice and traditional multiple choice items compared to fewer combinations.

Keywords: *Complex multiple choice item, Item difficulty, Item Discrimination*

Reference

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. New York, NY: Routledge.

[1] erkanatalmis@gmail.com, ORCID: 0000-0001-9610-491X, Manisa Celal Bayar Ünversitesi, Manisa, Türkiye

[2] l_yakar@hotmail.com, ORCID: 0000- 0001-7856-6926, Kahramanmaraş Sütçü İmam Üniversitesi, Kahramanmaraş, Türkiye

[3] cevik-abdullah@hotmail.com, ORCID: 0000-0003-0659-480X, Milli Eğitim Bakanlığı, Kahramanmaraş, Türkiye

Şans Başarısı, Üst Bilişsel Beceri ve Ölçme Sonuçlarının İlişkisinin İncelenmesi

Bedri Nejat YÜZÜAK [1], Nuri DOĞAN [2], Funda NALBANTOĞLU YILMAZ[3]

Günümüzde eğitimciler güvenirliğin yüksek olması, kullanışlılığın arttırılması ve objektif puanlamanın sağlanması amacıyla akademik sınavlarda ve başarı testlerinde çoktan seçmeli sorulara sıklıkla başvurmaktadır (Bereby-Meyer, Meyer, & Flascher, 2002). Ancak çoktan seçmeli sorularda bireyin tahminde bulunması, kopya çekiminin kolaylığı ya da kısmi bilgileriyle birlikte doğru yanıta ulaşması mümkündür (Kurz, 1999). Araştırmada çoktan seçmeli testlerde şans hatasının test puanına etkisini azaltmak için üst bilişsel beceri temelli şans başarısı düzeltmesi kullanılmış olup üst bilişsel beceri temelli şans başarısı düzeltmesinden elde edilen puanlar ile ham puanların ROC analizi ile incelenmesi amaçlanmıştır. Araştırma, 2023-2024 eğitim öğretim yılında bir devlet üniversitesinin eğitim fakültesindeki ikinci, üçüncü ve dördüncü sınıf öğrencilerinden 238 öğrenci ile yürütülecektir. Araştırma verilerinin toplanmasında ÖSYM tarafından 2012-2021 yılları arasında hazırlanmış ve halka açık erişimi olan ALES kapsamında bulunan Türkçe ve Matematik soruları arasından uzman görüşleri ile seçilenler kullanılacaktır. Verilerin toplanmasında öğrenciler her bir soruyu cevapladıktan sonra ilgili sorudaki cevapları için 1-5 arasında eminlik derecesi belirtilecektir. Elde edilen verilere eminlik derecesi temelli şans başarısından arındırma formülü uygulanacak olup veri analizi sürecinde SPSS

26.0 programıyla betimsel istatistik ve ROC analizi uygulamaları gerçekleştirilecektir. Araştırma sonucunda, ham puan ve şanstan arındırılmış puanla ROC analizine yönelik pozitif-negatif öngörü, AUC değeri, ROC eğrisi saptanacak olup değerler arasında anlamlı bir farklılık gözükmesi ve düzeltme formülü uygulanmış yapıların tüm değerlerde artışa sebep olması beklenmektedir.

Anahtar Kelimeler: Şans *başarısı,* üstbilişsel *değerlendirme, ROC analizi.*

Kaynakça

Bereby-Meyer, Y., Meyer, J., & Flascher, O. M. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, *15*(4), 313-327.

Kurz, T.B. (1999). “A review of scoring algorithms for multiple-choice tests”. Paper presented at the annual meeting of the Southwest Education Research Association. San Antonio, TX. Available at: https://eric.ed.gov/?q=%22A+review+of+scoring+algorithms+for+multiplechoice+tests%22&id=ED428076

[1] bnyuzuak@nevsehir.edu.tr, ORCID: 0009-0006-8397-1030, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, Türkiye

[2] nurid@hacettepe.edu.tr, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

[3] fundan@nevsehir.edu.tr, ORCID: 0000-0002-3228-4605, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, Türkiye

An Investigation of the Relationship between Chance Success, Metacognitive Skills and Measurement Results

Bedri Nejat YÜZÜAK [1], Nuri DOĞAN [2], Funda NALBANTOĞLU YILMAZ[3]

Today, educators frequently use multiple-choice questions in various structures such as academic exams and achievement tests in order to ensure high reliability, usefulness, and objective scoring (Bereby-Meyer, Meyer, & Flascher, 2002). However, in multiple-choice questions, it is possible for the individual to guess, to reach the correct answer with the ease of cheating or partial information (Kurz, 1999). In this study, metacognitive skill-based chance achievement correction was used in order to reduce the effect of chance error on the test score in multiple-choice tests, and it was aimed to examine the scores obtained from the metacognitive skill-based chance achievement correction and raw scores by ROC analysis. The research will be conducted in the 2023-2024 academic year with 238 students from the second-, third-, and fourth-year students of the faculty of education of a state university. In the collection of research data, those selected with expert opinions among the Turkish and Mathematics questions within the scope of ALES, which were prepared by OSYM between 2012- 2021 and have public access, will be used. In the collection of data, after the students answer each question, a degree of certainty between 1-5 will be indicated for their answers to the relevant question. A luck success formula based on the degree of certainty will be applied to the data obtained, and descriptive statistics and ROC analysis applications will be carried out with the SPSS

26.0 program in the data analysis process. As a result of the research, a positive-negative prediction, AUC value, ROC curve for ROC analysis will be determined with raw score and chance-adjusted score, and a significant difference is expected to appear between the values and the structures with the correction formula applied are expected to cause an increase in all values and a significant difference is expected.

Keywords: *Chance achievement, metacognitive assessment, ROC analysis.*

References

Bereby-Meyer, Y., Meyer, J., & Flascher, O. M. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, *15*(4), 313-327.

Kurz, T.B. (1999). “A review of scoring algorithms for multiple-choice tests”. Paper presented at the annual meeting of the Southwest Education Research Association. San Antonio, TX. Available at: https://eric.ed.gov/?q=%22A+review+of+scoring+algorithms+for+multiplechoice+tests%22&id=ED428076

[1] bnyuzuak@nevsehir.edu.tr, ORCID: 0009-0006-8397-1030, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, Türkiye

[2] nurid@hacettepe.edu.tr, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

[3] fundan@nevsehir.edu.tr, ORCID: 0000-0002-3228-4605, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, Türkiye

Çoklu Cevaplı Madde Türünün Sınıf İçi Ölçme Uygulamalarında ve Geniş Ölçekli Sınavlarda Kullanımına İlişkin Öğrenci ve Öğretmen Görüşlerinin İncelenmesi

İsmail KARAKAYA [1], Gökhan GÜMÜŞ [2], Hanifi ÖZDEMİR[3]

Eğitim ve öğretim süreçlerinin çıktılarını ölçme sürecinde çeşitli madde türleri kullanılabilir. Bu madde türlerinden birisi ise her soru formunda birden çok seçeneğin doğru olabileceği çoklu cevaplı madde türleridir. Orleans ve Sealy (1928) tarafından önerilen ve öğrencilerin düşünme becerilerini geliştirme açısından önemli olduğu belirtilmiştir. Duncan ve Milton (1978), Petersen ve diğerleri (2016), Muckle ve diğerleri (2011) gibi araştırmacılar çoklu cevaplı maddelerin öğrenci becerilerinin ölçülme sürecinde olumlu etkisine vurgu yapmaktadır. Bu kapsamda bu araştırma sürecinde çoklu cevaplı maddelerin okul uygulamalarında ve geniş ölçekli sınav uygulamalarında kullanımı ile ilgili öğretmen ve öğrenci görüşlerini almak amacıyla nitel araştırma yöntemlerinden durum çalışması kullanılmıştır. Bu amaçla bir devlet ortaokulunda öğrencilerin çoklu cevaplı madde türünü deneyimlemesi amacıyla bir sunum ve sınav uygulaması yapılmıştır. Sonrasında öğrencilerden gönüllü olanlar ile araştırmacılar tarafından hazırlanan yarı yapılandırılmış görüşme formu kullanılarak görüşme yapılmıştır. Öğretmenlerden ise öğrencileri ile sınav hakkında görüşlerini değerlendirmeleri istenmiştir. Sonrasında öğretmenlerle de araştırmacılar tarafından hazırlanan yarı yapılandırılmış görüşme formu kullanılarak görüş alınmıştır. Katılımcı grubu toplamda 4 öğrenci ve 3 öğretmenden oluşmaktadır. Uygulama sonrasında öğrenci ve öğretmenler ile görüşmeler yapılarak elde edilen sonuçlar tematik analizi ile incelenmiştir. Sonuçlar katılımcılar çoklu cevaplı maddelerin sırası ile derinlemesine öğrenme sağlama, farklı düşünme biçimlerini geliştirme, ayırt edici olma gibi olumlu; şans başarısını arttırma, zorlayıcı olma gibi olumsuz sayılabilecek yönlerine vurgu yapmaktadır.

Anahtar Kelimeler: Çoklu *cevaplı madde, sınıf içi ve geniş* ölçekli ölçme *uygulamaları*

Kaynakça

Duncan, G. T. (1974). An empirical Bayes approach to scoring multiple-choice tests in the misinformation model. *Journal of the American Statistical Association, 69,* 50-57.

Muckle, T. J., Becker, K. A. & Wu, B. (2011, April). Investigating the multiple answer multiple choice item format. *In annual conference of the National Council on Measurement in Education,* New Orleans, LA.

Orleans, J. S. & Sealy, G. A. (1928). *Objective tests.* Yonkers, New York: World Book Company. Parshall, C. G., Harmes, J. C., Davey, T. & Pashley, P. (2003). *Innovative Items for Computerized*

Testing. In W. J. van der Linden & C. A. W. Glas (Eds.), Computerized Adaptive Testing: Theory and Practice, Norwell, MA: Kluwer Academic Publishers.

[1] ikarakaya2002@gmail.com, ORCID: 0000-0003-4308-6919, Gazi Üniversitesi, Ankara, Türkiye

[2] gokhangumus2357@gmail.com, ORCID: 0000-0002-4161-0425, Hatay Mustafa Kemal Üniversitesi, Hatay, Türkiye

[3] hanifi_ozdemir@yahoo.com, ORCID: 0009-0005-4072-0197, Milli Eğitim Bakanlığı, Malatya, Türkiye

Investigation of Student and Teacher Opinions on the Use of Multiple-Response Item Types in Classroom Assessment Applications and Large-Scale Exams

İsmail KARAKAYA [1], Gökhan GÜMÜŞ [2], Hanifi ÖZDEMİR[3]

Various item types can be used in the process of measuring the outcomes of education and training processes. One such item type is the multiple-response item type, where more than one option can be correct in each question. Proposed by Orleans and Sealy (1928), it is noted for its importance in enhancing students' thinking skills. Researchers such as Duncan and Milton (1978), Petersen et al. (2016), and Muckle et al. (2011) have highlighted the positive impact of multiple-response items on the process of measuring student skills. In this context, a case study, one of the qualitative research methods, was employed to gather teachers' and students' opinions on the use of multiple-response items in school practices and large-scale exam applications. For this purpose, a presentation and exam were conducted at a public middle school to provide students with the experience of multiple- response items. Subsequently, interviews were conducted with students who volunteered, using a semi-structured interview form prepared by the researchers. Teachers were also asked to evaluate their views on the exam with their students. Following this, interviews were conducted with the teachers using a semi-structured interview form. The participant group consisted of a total of 4 students and 3 teachers. After the application, the results obtained from interviews with students and teachers were analyzed using thematic analysis. The findings revealed that participants emphasized positive aspects of multiple-response items, such as facilitating deep learning, developing different ways of thinking, and being distinctive; while also highlighting negative aspects, such as increasing success by chance and being challenging.

Keywords: *Multiple-response items, classroom and large-scale assessment applications*

Kaynakça

Duncan, G. T. (1974). An empirical Bayes approach to scoring multiple-choice tests in the misinformation model. *Journal of the American Statistical Association, 69,* 50-57.

Muckle, T. J., Becker, K. A. & Wu, B. (2011, April). Investigating the multiple answer multiple choice item format. *In annual conference of the National Council on Measurement in Education,* New Orleans, LA.

Orleans, J. S. & Sealy, G. A. (1928). *Objective tests.* Yonkers, New York: World Book Company. Parshall, C. G., Harmes, J. C., Davey, T. & Pashley, P. (2003). *Innovative Items for Computerized*

Testing. In W. J. van der Linden & C. A. W. Glas (Eds.), Computerized Adaptive Testing: Theory and Practice, Norwell, MA: Kluwer Academic Publishers.

[1] ikarakaya2002@gmail.com, ORCID: 0000-0003-4308-6919, Gazi Üniversitesi, Ankara, Türkiye

[2] gokhangumus2357@gmail.com, ORCID: 0000-0002-4161-0425, Hatay Mustafa Kemal Üniversitesi, Hatay, Türkiye

[3] hanifi_ozdemir@yahoo.com, ORCID: 0009-0005-4072-0197, Milli Eğitim Bakanlığı, Malatya, Türkiye

Çoktan Seçmeli Formatta Bağlam Temelli Soruların Hazırlanması

Erhan KARATAŞ [1], Bilal GÜNEŞ[2]

Başlangıcında teorik bilgilerin verildiği daha sonra bu bilgiler kullanılarak idealize edilmiş problemlerin çözülmeye çalışıldığı fizik derslerinde, öğrencilerin fiziğin yaşamla olan ilişkisini görmesi kolay değildir. Bu sebeple öğrencinin fizik ile yaşam arasındaki ilişkiyi öğrenimden sonra kurmasını beklemeden öğrenim sırasında oluşturması sağlanmalıdır (Whitelegg ve Parry, 1999). Bu durum için geliştirilen ve önerilen yaklaşım *bağlam temelli yaklaşım*dır. Bir yaklaşımda başarı sağlanabilmesi için öğretimin ölçme ve değerlendirme kısmı da benzer anlayışla kurgulanmalıdır. Fizik başarısını ölçen testlerde bağlam temelli sorular kullanılmalıdır. Alan yazında birçok farklı tanımı bulunsa da bu çalışmada bağlam temelli soru için *"Yaşamdan olgu ve olayların konu edindiği, yoklanan bilgi ve beceriler kullanılarak* çözümüne *ulaşılabilecek, ilgi uyandırıcı ve eksik tanımlanmış bir sorun, bir problem durumu."* tanımı esas alınmıştır (Karataş, 2022). Türkiye'de yapılan üniversiteye giriş sınavlarında yer alan fizik testlerinde hiçbir sorunun bağlam temelli sorulara ait "*Soruda, sorunun* çözümü *için gerekli olmayan veriler, sabitler, bilgiler vb. bulunmaktadır.*" ve "*Soruda, sorunun* çözümü *için ihtiyaç duyulan bazı veriler, sabitler, bilgiler vb.* öğrencinin *araştırıp elde etmesi için kasten verilmemiştir."* kriterlerine sahip olmadığı görülmektedir. Bu çalışmada, çoktan seçmeli soru formatında ve sınav güvenliği kaygıları giderilmiş ortamlarda uygulanabilir bağlam temelli soruların üretilmesi amaçlanmıştır. Hedefe ulaşabilme adına, içerisinde problem durumlarının öğrenciye sesli ve görüntülü olarak aktarıldığı animasyonlar ile öğrencilerin sorunun çözümünde ihtiyaç duyduğu bazı bilgileri elde etmesi için değişkenleri manipüle edebildiği ve ayrıca uzunluk ve zaman ölçümleri yapabildiği simülasyonları barındıran sorular geliştirilmiştir. Hazırlanan sorular uzman görüşüne sunulmuştur. Soruların bağlam temelli soru tanımına uygun hazırlanıp hazırlanmadığını belirlemek için Karataş (2022) tarafından geliştirilen bağlam temelli soru kriterleri kontrol listesi kullanılmıştır. Sonuç olarak bağlam temelli soru kriterlerinin tamamına (13 kriter) sahip olan, çoktan seçmeli soru formatında ve sınav güvenliği kaygılarının karşılandığı fizik soruları üretilmiştir. Bağlam temelli soru kriterlerinin tamamına sahip soruların üniversiteye giriş sınavlarında yer alması için bu sınavların bilgisayar ortamında yapılması önerilmektedir.

Anahtar Kelimeler: *Bağlam temelli soru,* üniversiteye *giriş sınavı.*

Kaynakça

Whitelegg, E., & Parry, M. (1999). Real-life contexts for learning physics: Meanings, issues and practice. *Physics Education*, *34*(2), 68–72.

Karataş, E. (2022). *Kuvvet ve hareket konusu ile ilgili bağlam temelli bir başarı testi geliştirilmesi*. Yüksek lisans tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

[1] e-posta: erhan_krts@hotmail.com, ORCID: 0000-0002-9612-1821, Gazi Üniversitesi, Ankara, Türkiye

[2] e-posta: bgunes@gazi.edu.tr, ORCID: 0000-0002-3911-6778, Gazi University, Ankara, Türkiye

Preparing Context-Based Questions in Multiple-Choice Format

Erhan KARATAŞ [1], Bilal GÜNEŞ[2]

It is often difficult for students to see the connection between physics and real life in physics classes, where theoretical information is presented first, followed by solving idealized problems. Therefore, it is crucial to make the connection between physics and real life during the classroom, rather than expecting it to be realized afterwards (Whitelegg and Parry, 1999). To achieve this, *a context-based approach* has been developed and recommended. For this approach to be successful, the measurement and evaluation components of teaching should be structured with a similar understanding. Therefore, contextual questions should be included in physics tests. Despite the various definitions in the literature, this study adopts the definition of contextual questions as: "*A fascinating and ill-defined problem situation based on facts and events from life, the solution of which can be achieved by using the knowledge and skills to be measured*" (Karataş, 2022). In the physics tests for university entrance exams in Türkiye, none of the questions meet the criteria for contextual questions. These criteria include elements such as "*The question contains data, constants, information, etc., that students do not need to use to solve the problem*" and "*The question contains necessary data, constants, information, etc., that are intentionally not provided, requiring students to recall or find them*". The goal of this study was to generate contextual questions in a multiple-choice format suitable for testing security environments. The questions incorporate animations that present problem scenarios to students both audibly and visually. Students can manipulate variables and measure length and time in simulations to obtain the information necessary to solve the problems. The prepared questions were reviewed by experts using the Checklist for the Criteria of Context-based Questions developed by Karataş (2022) to ensure that they met the standards for context-based questions. As a result, physics questions were created that meet all 13 criteria of context-based questions and are suitable for a multiple-choice format while addressing exam security concerns. It is recommended that these questions, which meet all of the criteria for context-based questions, be included in university entrance exams administered in computer-based environments.

Keywords: *Context-based question, university entrance exams.*

Reference

Whitelegg, E., & Parry, M. (1999). Real-life contexts for learning physics: Meanings, issues and practice. *Physics Education*, *34*(2), 68–72.

Karataş, E. (2022). *Development of a context-based achievement test for the subject of force and motion.* Master's thesis, Gazi University Institute of Education Sciences, Ankara.

[1] erhan_krts@hotmail.com, ORCID: 0000-0002-9612-1821, Gazi University, Ankara, Türkiye

[2] bgunes@gazi.edu.tr, ORCID: 0000-0002-3911-6778, Gazi University, Ankara, Türkiye

Closing the Gender Gap in STEM: Evidence from the 2019 Taiwanese College Admission Reform

Danny Po-Hsien Kang [1], Joseph Tao-yi Wang[2], Yi-Cheng Kao[3], Wei-Cheng Chen[4]

In this study, we examine the impact of the 2019 reform on Taiwan's college application channels. The reform restricts the number of subjects from the General Scholastic Ability Test (GSAT) that a program can consider and allows students to focus on fewer GSAT subjects. Our empirical findings indicate a significant rise in the number of female students who passed at least one program during the first stage of screening. Also, there is a universal increase in the share of female students admitted to both STEM and non-STEM programs post-reform. The most significant effect is observed in top-tier STEM programs, which exhibit a substantial 6.6% increase in the second stage of the application process, which introduced non-gender-blind evaluations. These results suggest that female students have larger performance improvement when focusing on fewer GSAT subjects, highlighting the potential of such reforms to close the gender gap in STEM.

Keywords: *Gender, College Admission, STEM*

[1] b09201020@ntu.edu.tw, Department of Economics, National Taiwan University, City, Taipei, Taiwan

[2] josephw@ntu.edu.tw, ORCID: 0000-0001-6786-8099, Department of Economics, National Taiwan University, City, Taipei, Taiwan

[3] 1976kao@gmail.com, ORCID:0000-0002-8877-7361, Department of Economics, Fu Jen Catholic University, New Taipei City, Taiwan

[4] weicheng.c@gmail.com, ORCID: 0000-0002-8237-0683, Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

Market Incentives and College Admissions in the Absence of Standardized Tests

A. Arda Gitmez [1], Román Andrés Zárate[2]

The debate surrounding the role of standardized tests for college admissions remains a contentious issue. While advocates argue that these tests are related to merit and capture an objective measure of ability, critics contend that standardized tests can perpetuate inequalities, as they may not accurately reflect potential and are influenced by socioeconomic factors. Despite some progress in recent research, the aggregate consequences of standardized tests remain unclear. Such consequences may depend on the presence of private education providers. In particular, when there is no standardized information about applicants' abilities, private institutions may respond to market incentives by changing their admission criteria to favor students from advantageous socioeconomic backgrounds. This project aims to explore this phenomenon in Peru, a country without a high school exit standardized test and a mixed system with public and private institutions at the secondary and higher education levels. In particular, we quantify the consequences of introducing a high school exit standardized exam on equity in higher education access by estimating a model that quantifies the private institutions' market incentives to change their admission policies. We use comprehensive data characterizing students' transition from high school to higher education between 2015 and 2023 with information on all college applications and admissions in Peru. While some applicants are admitted via a university-specific admission exam, others benefit from special admission criteria that allow them to waive the admission test. The model estimates allow us to quantify the gains of the special admission mode to private universities and the consequences of various hypothetical counterfactual policies. These counterfactual policies will mirror the ones of other countries, such as Türkiye, where there is a standardized test and a centralized college admission mechanism, or Colombia, where a standardized test is available but admission decisions are decentralized.

Keywords: *Standardized Exams, Special Admission Modes, Decentralized Markets*

[1] arda.gitmez@bilkent.edu.tr, ORCID: 0000-0003-0944-7477, Bilkent University, Ankara, Türkiye

[2] ra.zarate@utoronto.ca, University of Toronto, Toronto, Canada

Okul Tercihinde Eşitlikçi Pozitif Ayrımcılık

Furkan Doğan[1], Özgür Yılmaz[2]

Bu makale, rezerve kategoriler altında birimlerin rastgele dağılımını inceleyerek, bu tür dağıtımlarda ortaya çıkan eşitlik etkilerini ele almaktadır. Modelimiz, kategoriler içinde bağımsız öncelik sıralamasına izin vererek, kategori üyeleri arasındaki temel sıralamaya dayalı önceliklendirmeyi esnetir. Makale eşitliği ajanların fayda profili arasında Lorenz hâkimiyeti kriterleriyle tanımlanan eşitlikçi bir dağılım olarak tanımlar. Tüm deterministik dağılımların kümesini karakterize eden kesme kümeleri kavramını tanıtarak önceliklere saygı gösterme ilkesini ihlal etmeden eşitlikçi bir dağılımın mevcut olduğu koşulları ortaya koyar. Chicago'nun okul kabul politikaları ve Hindistan'ın eyalet işleri ve devlet okulu kontenjanlarını tahsis etmek için benimsenen temel sıralamaya bağımlı öncelik yapıları üzerinde derinleşen araştırma, böyle kısıtlı bir model altında, önceliklere saygı gösteren eşitlikçi bir dağılımın her zaman mevcut olduğunu kanıtlar.

Anahtar Kelimeler: *Piyasa Tasarımı, Eşleşme, Pozitif Ayrımcılık, Lorenz Baskınlığı*

[1] fdogan19@ku.edu.tr, ORCID: 0000-0001-8848-8810, Koç Üniversitesi, İstanbul, Türkiye

[2] ozyilmaz@ku.edu.tr, ORCID: XXXX-XXXX-XXXX-XXXX, Koç Üniversitesi, İstanbul, Türkiye

Equitable Affirmative Action in School Choice

Furkan Doğan[1], Özgür Yılmaz[2]

This paper investigates the random allocation of units under reserve categories, addressing the equity implications inherent in such distributions. Our model allows independent priority ordering within categories, relaxing the prioritizing agents based on baseline ordering within the category members. We define equity, an egalitarian allocation, through the criteria of Lorenz dominance among the utility profile of the agents. By introducing the concept of cutoff sets, which characterizes the set of all deterministic allocations, we identify conditions under which an egalitarian allocation exists without jeopardizing respecting priorities. Our analysis delves into baseline ordering-dependent priority structures adopted by Chicago for school admission policies and India to allocate state jobs and public school seats. An egalitarian allocation that respecting priorities always exists under such a restricted model.

Keywords: *Market Design, Matching, Affirmative Action, Lorenz Dominance*

[1] fdogan19@ku.edu.tr, ORCID: 0000-0001-8848-8810, Koç Üniversitesi, İstanbul, Türkiye

[2] ozyilmaz@ku.edu.tr, ORCID: XXXX-XXXX-XXXX-XXXX, Koç Üniversitesi, İstanbul, Türkiye

Market Design of College-Major Switches for Reducing Student-Major Mismatch

Umut Dur[1], Yi-Cheng Kao[2], Scott Paiement[3]

In many countries, students must choose college-major pairs jointly in college admissions. It is not unusual for students to feel that their major is not a good fit for them. When such a mismatch between students and majors occurs, there exists a demand for major switching. In practice, the number of major switches is usually constrained by the total number of initially vacant seats at each major. This constraint limits the number of successful major switches, preventing potential welfare gains for students. We propose a new mechanism to address the problem and demonstrate how the mechanism could increase the success rate of major switching using data provided by National Taiwan University.

Keywords: *market design, two-sided matching, major switching, college admissions*

[1] umutdur@gmail.com, Department of Economics, North Carolina State University, Raleigh, NC, USA.

[2] 1976kao@gmail.com, ORCID: 0000-0002-8877-7361. Department of Economics, Fu Jen Catholic University, Taipei, Taiwan, R.O.C.

[3] smpaieme@ncsu.edu, Department of Economics, North Carolina State University, Raleigh, NC, USA.

Alternative Assessment in High School EFL Teaching: Practices and Perspectives

Ayşe Gençer [1], Şeyda Selen Çimen[2]

This research focuses on exploring alternative assessment practices in language classrooms and the awareness of English as a Foreign Language (EFL) teachers and learners in the context of public high schools in Turkiye. Primary objectives of the study are (1) to investigate the perspectives of EFL teachers and their learners regarding the use of alternative assessment methods; (2) identify the commonly employed alternative assessment methods; and (3) identify which methods are perceived as the most effective by the learners. The study employs qualitative research design, incorporating open-ended teacher and learner questionnaires, document analyses of assessment tools, and interviews with teachers. The participant groups consist of EFL teachers who work at public high schools and high school students who take EFL classes. The study mainly concludes that the teachers have a low level of awareness on alternative language assessment; most of them do not employ alternative assessment methods, and the learners are unaware of the possible advantages of the use of alternative assessment methods for their language development. Both the teachers and the learners are in need of raising their awareness about the effectiveness of alternative assessment in producing better results when used with traditional assessment practices.

Keywords: *Alternative language assessment, practices, EFL teachers and learners*

[1] aysegencer2@posta.mu.edu.tr, ORCID: 0009-0008-6427-0268, Muğla Sıtkı Koçman University, Muğla, Türkiye

[2] seydaselen@mu.edu.tr, ORCID: 0000-0001-6840-6558, Muğla Sıtkı Koçman University, Muğla, Türkiye

Yabancı Dilde Kelime Bilgisi Ölçme ve Değerlendirme için Fibonacci Tabanlı Algoritma Tasarımı: Rusça Örneği

Cihan Kara [1]

Bu çalışmada yabancı dil olarak Rusça örneğinde kelimelerin kullanım sıklığına göre sıralandığı sıklık sözlüğü (Ляшевская & Шаров, 2009) referans alınarak yabancı dilde kelime bilgisini daha hızlı ve etkin bir şekilde ölçmek amacıyla Fibonacci tabanlı bir algoritma geliştirilmiştir. Kendinden öndeki iki sayının toplamı şeklinde ilerleyen Fibonacci dizisinde başlangıç aralıkları birbirine yakın iken sonraki aşamalarda aralıklar giderek artmaktadır. Sıklık sözlüklerinde ise kelimeler en çok kullanılandan en az kullanılana göre sıralanmaktadır. Sıklık sözlüğündeki kelimeler, Fibonacci dizisinin doğal artış örüntüsüne bağlı olarak sorulduğunda testin zorluk düzeyi adaptif olarak ayarlanabilmektedir.

Algoritmanın temel çalışma prensibi şu şekildedir: Öğrenciye sıklık listesinde ilk sırada yer alan kelime sorulur, öğrenci kelimeyi biliyorsa Fibonacci dizisinin artış örüntüsüne bağlı olarak "+1, +2, +3, +5, +8..." şeklinde sıralama artırılarak listedeki kelimeler sorulur. Öğrenci doğru cevap vermeye devam ederse 23. soruda listede 46368.sırada yer alan Rusçada en nadir kullanılan kelimelerden biriyle karşılaşacaktır. Öğrenci yanlış cevap verdiğinde Fibonacci dizisinin artış örüntüsü tersine çevrilir ve "-1, -2, -3, -5, -8..." şeklinde sıralama azaltılarak listede karşılık gelen kelimeler sorulur. Öğrenci yanlış cevap vermeye devam ederse daha sıklıkla kullanılan kelimelere doğru hızla ilerleyecektir. Sorulan kelimeler bilinen ve bilinmeyen kelimeler şeklinde liste haline getirilir. Bu listelerdeki kelimeler öğrenciye tekrar sorulmaz ve nihai değerlendirme için veri olarak kullanılır. Bu şekilde öğrencinin tahmini olarak bildiği kelime sayısı aralığı hızlı ve etkin bir şekilde ölçülebilir. Örneğin öğrenci 20. soru sonrasında kelime listesinde 7500-10000 arasında sıralamalarda yer alıyorsa bu aralık öğrencinin tahmini kelime bilgisi sayısı olacaktır. Bu ölçüm nihai değerlendirme için kesin bir sonuç olarak değil, öğrencinin kelime bilgisi hakkında önemli bir veri olarak düşünülebilir. Bu veriler kullanılarak öğretim elemanı tarafından özel bir değerlendirme yapılabilir. Çalışma kapsamında, bu algoritmanın uygulanması için PHP ve MySQL tabanlı bir web uygulaması geliştirilmiştir. Bu algoritma tasarımı yalnızca Rusça için değil, kelime sıklık listesi oluşturulan diğer yabancı diller için de kullanılabilir. Çalışmanın teorik temelleri, dil eğitiminde teknoloji destekli kişiselleştirilmiş öğrenme yaklaşımları için de yararlı olacağı düşünülmektedir.

Anahtar Kelimeler: *Kelime bilgisi,* ölçme *ve değerlendirme, Fibonacci dizisi, Rusça*

Kaynakça

О. Н. Ляшевская & С. А. Шаров. (2009). Новый частотный словарь русской лексики. http://dict.ruslang.ru/freq.php. (Erişim Tarihi: 05 Ağustos 2024).

[1] e-posta: cihan_kara@anadolu.edu.tr, ORCID: 0000-0002-2638-7237, Anadolu Üniversitesi, Eskişehir, Türkiye

Fibonacci-Based Algorithm Design for Vocabulary Measurement and Evaluation in a Foreign Language: A Russian Language Example

Cihan Kara [1]

This study develops a Fibonacci-based algorithm to measure foreign language vocabulary knowledge more quickly and effectively, using Russian as an example and referencing a frequency dictionary (Ляшевская & Шаров, 2009) where words are ranked according to their usage frequency. In the Fibonacci sequence, which progresses as the sum of the two preceding numbers, the initial intervals are close to each other, while in later stages, the intervals gradually increase. In frequency dictionaries, words are ordered from most frequently used to least used. When the words in the frequency dictionary are asked based on the natural increase pattern of the Fibonacci sequence, the difficulty level of the test can be adaptively adjusted. The basic working principle of the algorithm is as follows: The student is asked the word that ranks first in the frequency list. If the student knows the word, words from the list are asked by increasing the ranking according to the Fibonacci sequence's increase pattern: "+1, +2, +3, +5, +8...". If the student continues to answer correctly, they will encounter one of the rarest used words in Russian, ranked 46,368th in the list, at the 23rd question. When the student gives an incorrect answer, the increase pattern of the Fibonacci sequence is reversed, and words corresponding to the list are asked by decreasing the ranking as "-1, -2, -3, -5, -8...". If the student continues to give incorrect answers, they will quickly progress towards more frequently used words. The asked words are listed as known and unknown words. The words in these lists are not asked again to the student and are used as data for final evaluation. In this way, the range of the estimated number of words known by the student can be measured quickly and effectively. For example, if the student is ranked between 7500-10000 in the word list after the 20th question, this range will be the student's estimated vocabulary knowledge. This measurement should be considered not as a definitive result for final evaluation, but as important data about the student's vocabulary knowledge. Using this data, a special evaluation can be made by the instructor. Within the scope of the study, a PHP and MySQL- based web application was developed for the implementation of this algorithm. This algorithm design can be used not only for Russian but also for other foreign languages for which word frequency lists have been created. The theoretical foundations of the study are thought to be useful for technology- supported personalized learning approaches in language education.

Keywords: *Vocabulary, measurement and evaluation, Fibonacci sequence, Russian language*

Reference

О. Н. Ляшевская & С. А. Шаров. (2009). Новый частотный словарь русской лексики. http://dict.ruslang.ru/freq.php. (Retrieved August 5, 2024).

[1] cihan_kara@anadolu.edu.tr, ORCID: 0000-0002-2638-7237, Anadolu University, Eskişehir, Türkiye

Rusça Yabancı Dil Bilgisi Seviye Tespit Sınavlarında (YDS) Güvenirlik ve Geçerliğin İncelenmesi[1]

Hülya ATEŞ [2]

Küreselleşme süreciyle birlikte dil yeterliliklerinin ölçülmesi, eğitim ve sınav sistemlerinde önemli bir yer edinmiştir. Ülkemizde bu amaçla uygulanan sınavlardan biri olan Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS), dil yeterliliklerini değerlendirmede kullanılan önemli bir ölçme aracıdır. Bu çalışmanın amacı, Rusça YDS'nin adayların dil yeterliliklerini yeterli seviyede ölçüp ölçmediğini güvenilirlik ve geçerlik yönünden incelemektir. Çalışma, Rusça YDS'nin geçerlik ve güvenilirliğini değerlendirmeyi ve mevcut literatürdeki eksikliklere ışık tutmayı hedeflemektedir. Mevcut literatürde Rusça YDS sınavının güvenilirliği ve geçerliliği üzerine herhangi bir çalışma bulunmamakta olup bu durum sınavın adayların dil yeterliliklerini ne derece doğru ve tutarlı bir şekilde ölçtüğü hakkında belirsizlikler yaratmaktadır.

2021 yıllarında yapılan YDS/1 ve YDS/2 testleri bu çalışmanın örneklemi olarak seçilmiştir. Güvenilirlik analizi, ÖSYM'den alınan resmi veriler ışığında madde güçlük ve ayırt edicilik indeksleri ile Cronbach alfa katsayısı kullanılarak gerçekleştirilmiştir. Madde güçlük indeksi test maddelerinin adaylar tarafından ne kadar zor veya kolay bulunduğunu, madde ayırt edicilik indeksi ise test maddelerinin başarılı ve başarısız adayları ne derece ayırt edebildiğini belirlerken, Cronbach alfa katsayısı testin iç tutarlılığını değerlendirmiştir (Güler, 2013). Çalışmada verilerin elde edilme ve incelenme tekniği bakımından nitel bir veri analizi yöntemi olan doküman analizi yönteminden yararlanılmıştır. Doküman analizi, yazılı belgelerin içeriğini sistematik bir şekilde analiz etmek amacıyla faydalanılan nitel bir araştırma yöntemidir (Wach & Ward, 2013, s. 3).

Araştırma sonucunda, Rusça YDS testinin genel olarak yüksek bir güvenirlik ve geçerlik düzeyine sahip olduğu görülmüştür. Özellikle, testin Cronbach alfa iç tutarlılık katsayısının çok yüksek çıkması, testin iç tutarlılığı ve güvenilirliğinin oldukça yüksek olduğunu göstermektedir. Bu bulgular, Rusça YDS'nin adayların dil yeterliliklerini doğru ve tutarlı bir şekilde ölçme kapasitesine sahip olduğunu ortaya koymaktadır. Sınavın mevcut formunun dil yeterliliklerini değerlendirmede etkili bir araç olduğunu ve diğer dil yeterlilik sınavları için bir referans noktası oluşturabileceğini göstermektedir.

Anahtar Kelimeler: *Rusça YDS, güvenirlik, geçerlik.*

Kaynakça

Wach, E. & Ward, R. (2013). Learning about qualitative document analysis. *IDS Practice Paper in Brief, 13,* 1-11.

Güler, N. (2013). *Eğitimde* Ölçme *ve Değerlendirme (4. Baskı).* Pegem Akademi.

---

[1] Bu bildiri özeti, "Rusça Yabancı Dil Bilgisi Seviye Tespit Sınavlarında (YDS) Güvenirlik ve Geçerliğin İncelenmesi" başlıklı yüksek lisans tezimin özetine dayanılarak hazırlanmış ve genişletilmiştir.

[2] hulyaturhan07@gmail.com, ORCID: https:0009-0005-5466-2987, ÖSYM, Ankara, Türkiye

An Examination of Reliability and Validity in the Russian Foreign Language Proficiency Exams (YDS)1

Hülya ATEŞ [2]

With the process of globalization, the assessment of language proficiencies has gained significant importance in education and examination systems. One of the exams implemented for this purpose in our country is the Foreign Language Proficiency Exam (YDS), which is an important measurement tool used to evaluate language proficiency. The aim of this study is to examine whether the Russian YDS adequately measures candidates' language proficiencies in terms of reliability and validity. The study aims to evaluate the validity and reliability of the Russian YDS and shed light on the gaps in the existing literature. There is no existing research on the reliability and validity of the Russian YDS, which creates uncertainties about how accurately and consistently the exam measures candidates' language proficiencies.

The YDS/1 and YDS/2 tests conducted in 2021 have been selected as the sample for this study. The reliability analysis was carried out using item difficulty and discrimination indices along with the Cronbach alpha coefficient, based on official data obtained from ÖSYM. The item difficulty index determines how difficult or easy the test items are for the candidates, while the item discrimination index indicates how well the test items can distinguish between successful and unsuccessful candidates. The Cronbach alpha coefficient assesses the internal consistency of the test (Güler, 2013). The study employed a qualitative data analysis method known as document analysis for data collection and examination. Document analysis is a qualitative research method used to systematically analyze the content of written documents (Wach & Ward, 2013, p. 3).

The research results indicate that the Russian YDS test generally has a high level of reliability and validity. Specifically, the high Cronbach alpha internal consistency coefficient of the test demonstrates the high internal consistency and reliability of the test. These findings suggest that the Russian YDS has the capacity to accurately and consistently measure candidates' language proficiencies. The current form of the exam is an effective tool for evaluating language proficiencies and can serve as a reference point for other language proficiency exams.

Keywords: *Russian YDS, Reliability, Validity*

Reference

Wach, E. & Ward, R. (2013). Learning about qualitative document analysis. *IDS Practice Paper in Brief, 13,* 1-11.

Güler, N. (2013). *Eğitimde* Ölçme *ve Değerlendirme (4. Baskı).* Pegem Akademi.

---

1 This abstract is based on and has been expanded from the summary of my master's thesis titled "An Examination of Reliability and Validity in the Russian Foreign Language Proficiency Exams (YDS)"

2 hulyaturhan07@gmail.com, ORCID: 0009-0005-5466-2987, ÖSYM, ANKARA, TÜRKİYE

CEFR'e Göre Anadili Arapça Olmayanlar İçin Arapça Metinlerin Sınıflandırma Standartlarının Belirlenmesi[1]

İlsmail RAMAZAN[2]

Okunabilir konuların genel olarak dillerin, özel olarak Arapçanın edinilmesinde önemli olduğu hipotezinden yola çıkarak okuma; yabancı dil öğretim programlarında yönlendirici bir unsur olarak kabul edilmektedir. Şu şekilde tanımlanmaktadır; "İçeriği anlamak için yazılı bir metni algılamak. (Richard ve Schmidt, 2002)". Öğretmenler ve müfredat geliştiriciler, özellikle bir rehber olarak bilimsel bir standardın yokluğunda, öğrencilere sunulabilecek metinleri seçmede gerçek bir zorlukla karşı karşıyadırlar. (Brown, 2007) denklemine göre; ana dili İngilizce olmayan öğrenciye sunulan metnin zorluğu = öğrencinin düzeyi. Bu demektir ki; ne (+1) ne (+2) ne de (+0) yani metin kolay olmamalıdır. Gunning (2013), bu nedenle öğrenci için bir zorluk olmadığını, aynı zamanda öğrencinin yeteneklerinin ötesinde zorlanmaması gerektiğini böylelikle onun için ciddi bir engele dönüşmesinin engellenmesini önermiştir. Kendileri için hiçbir zorluk teşkil etmeyen kolay konuları çalışmak dilin fosilleşmesiyle sonuçlanabilirken, yeteneklerini aşan zor konuları inceleyen kişiler için ise dilsel düzeyde ilerleme kaydedememelerine neden olmaktadır. Öğrencinin dil düzeyine göre uygun metni seçmeye karar vermede faydalı olabilecek şey bilimsel standartlardır. Bu araştırma; Avrupa çerçevesinde Arapça metinlerin okunabilirliğini dilsel okuma becerisi düzeylerine göre sınıflandırmak için kriterlerin belirlenmesi üzerinde çalışarak bu sorunu çözmeyi amaçlamaktadır. Bu araştırmanın, öğrencinin dil düzeyine göre uygun metinlerin seçilmesinde eğitim kurumları ve müfredat geliştiricilerin yanı sıra öğretmenler ve öğrenciler için de yararlı olabilecek bilimsel standartların sağlanmasına yardımcı olması beklenmektedir. Ayrıca yayınlarını dil düzeyine göre şekillendirebilen uzman yayınevleri de bulunmaktadır. Araştırma, veri toplama ve analiz etmede "Delphi" metodolojisine dayanacaktır.

Anahtar Kelimeler: *Arapça, okuma becerisi, Arapça metin, standart.*

Kaynakça

Richards, J. C. & Schmidt, R. 2002. Longman Dictionary of Language Teaching and Applied Linguistics (3ed). London: Longman.

Brown, H. D. 2007. Principles of Language and Teaching. New York: Pearson Education.

Gunning, T. 2003. "The Role of Readability in Today's Classrooms", Topics in Language Disorders, 23/3, 175–189.

Ramadan, İ. 2020. "Arap dilinin anadili Arapça olmayanlar için okunabilirliği: Bilimsel Bir Standart Oluşturmaya Doğru", II. Uluslararası Ana Dili Arapça Olmayanlar İçin Günümüzde ve Gelecekte Arapça Sempozyumu 12-13 Ağustos 2020 Bildiriler Kitabı. 188-207

[1] Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenen 223K736 numaralı projenin çıktılarından biridir. Projeye verdiği destekten ötürü TÜBİTAK'a teşekkürlerimizi sunarız.

[2] hany.ramadan@giresun.edu.tr:, ORCID: 0000-0001-5935-811X, Giresun Üniversitesi, Giresun, Türkiye

Determination of Classification Standards of Arabic Texts for Non-Native Speakers According to CEFR[1]

İsmail RAMAZAN[2]

Stemming from the hypothesis of the importance of the readable subjects in acquiring languages in general and Arabic in particular, reading is considered a governing element in foreign languages teaching programs. It is defined as; "perceiving a written text in order to understand its contents. (Richard &Schmidt, 2002). Teachers and curriculum developers face a real challenge in selecting texts that can be presented to learners, especially in the absence of a scientific standard as a guide. According to (Brown, 2007) equation; the difficulty of the text presented to the non-native speaker learner = the level of the learner (+1) neither (+2) nor (+0). Meaning that the text should not be easy, so it is not a challenge for the learner, and at the same time, it should not be difficult beyond the learner's abilities and capabilities, so it turns into a serious obstacle for him/her. Gunning (2013) suggested that the learners who study easy subjects that represent no challenge for them could end with linguistic fossilization, while those who study difficult subjects that are beyond their abilities fail to advance in their linguistic level. In a previous study (Ramadan, 2020) the researcher noticed the absence of a scientific standard that can be useful in making the decision to select the appropriate text according to the linguistic level of the learner. This research aims to address this problem by working on determining criteria for classifying the readability of Arabic texts according to the linguistic levels of reading skill in the European framework. It is expected that this research will help provide scientific standards that may be useful for teachers and learners, in addition to educational institutions and curriculum developers, in selecting appropriate texts according to the linguistic level of the learner. As well as the specialized publishing houses that can label their publications according to the linguistic level. The research will depend on "Delphi" methodology for collecting and analyzing data.

Keywords: *Arabic language, reading skill, Arabic text, standard.*

Reference

Richards, J. C. & Schmidt, R. 2002. Longman Dictionary of Language Teaching and Applied Linguistics (3ed). London: Longman. Brown, H. D. 2007. Principles of Language and Teaching. New York: Pearson Education.

Gunning, T. 2003. "The Role of Readability in Today's Classrooms", Topics in Language Disorders, 23/3, 175–189.

Ramadan, İ. 2020. "Arap dilinin anadili Arapça olmayanlar için okunabilirliği: Bilimsel Bir Standart Oluşturmaya Doğru", II. Uluslararası Ana Dili Arapça Olmayanlar İçin Günümüzde ve Gelecekte Arapça Sempozyumu 12-13 Ağustos 2020 Bildiriler Kitabı. 188-207

[1]This research is one of the outputs of project No. 223K736 sponsored by the Scientific and Technological Research Council of Turkey (TÜBİTAK).

[2]hany.ramadan@giresun.edu.tr:, ORCID: 0000-0001-5935-811X, Giresun University, Giresun, Türkiye

İlk, Orta ve Yükseköğretimde Okuma Sıklığına Göre Birleşik Sözcüklerin Anlamsal Saydamlığının İncelenmesi*

Nermin YAZICI[1], Leyla ÇİMEN[2], Mehmet Ali BAHAR[3]

Birleşik sözcükler, farklı çeşitlilikte bütünleşmeler sergilemesi nedeniyle tam çözümleme yaklaşımının (Taft, 1981) ileri sürdüğünden farklı işlemleme ve geri çağırma süreçlerini ortaya koyar. Türkçe birleşik sözcüklerin işlemlemesini ele alan çalışmaların sınırlılığı ve var olan yazım kurallarını belirlenirken sözcük işlemleme yaklaşımlarının göz önüne alınmaması bir problem olarak görülmektedir. Bu problem birleşik sözcüklerin yazımına ilişkin soru yazımını da etkilemektedir. Bu nedenle çalışmada, üç eğitim düzeyindeki anadili konuşucularının saydam, yarı saydam ve opak kategorideki birleşik sözcükleri işlemlemeleri, tepki süresi üzerinden incelenmektedir. Çalışmada Türkçe birleşik sözcüklere erişimde, zihinde temsil edilme şekillerinde sözcüğün yazımı, birinci ve ikinci bileşeninin saydamlık ve sıklık açısından rolü olup olmadığının belirlenmesi ve bunların okuma sıklığıyla ilişkisinin incelenmesi amaçlanmıştır. Çalışma ilişkisel araştırma yöntemi ile gerçekleştirilmiştir. 124 birleşik sözcük; sıklıkları, uzunlukları, kodlayıcı uyumu, anadili koşucularının ölçeklendirmeleri sonucunda seçilmiştir. 1.Deneyde, 184 katılımcının tepki süreleri, sözcüklerin saydamlıklarına ve sıklıklarına göre karşılaştırılmış; saydam, yarı saydam, ve opak birleşik sözcük kategorileri arasındaki tepki süresi farklılıklarının anlamlı olduğu; kategoriler arasında madde analizinde yalnızca opak ile saydam; saydam ile opak-saydam karşılaştırmasında; katılımcı analizinde ise yalnızca opak başlangıçlı kategorilerin karşılaştırmasında anlamlı farklılığın olmadığı görülmüştür. Toplam 340 katılımcının katıldığı, bitişik yazma faktörünün incelendiği 2. ve 3. Deneyde, madde analizi sonuçları boşluklu yazımın tepki sürelerini kısalttığı ortaya koymuştur. Katılımcılar için yapılan analizler ise boşluklu yazımın ortaokul ve lise öğrencilerinin tepki sürelerini azalttığını ve sözcük tanımayı kolaylaştırdığını; öte yandan yetişkinlerde aralıklı yazmanın tepki sürelerini etkilemediğini ortaya koymuştur. Okuma sıklığı verileri birinci, ikinci ve üçüncü deneyde anlamlı sonuçlar göstermemiştir. Araştırmanın genel varsayımından hareketle ulaşılan bulgular sonucunda, saydamlık kategorilerine ve bitişik-ayrı yazıma verilen tepki sürelerindeki değişimlere bağlı olarak işlemleme farklılıkları tartışılmıştır. Bu proje, alanyazını destekler biçimde saydamlığın sözcük işlemlemeye etki ettiğini; bitişik yazımın ise tepki süresinde gecikmelere neden olduğunu önermektedir.

Anahtar Kelimeler: *Birleşik sözcük, anlamsal saydamlık, işlemleme, tepki süresi.*

Kaynakça

Taft, M. (1981). Prefix stripping revisited. *Journal of verbal learning and verbal behavior*, 20(3), 289- 297.

* Bu çalışma TÜBİTAK -2100A Hızlı Destek Modülünce desteklenen ""Birleşik Sözcüklerin Saydamlıklarının Sözcük İşleme ile İlişkisinin İncelenmesi" başlıklı proje kapsamında gerçekleştirilmiştir.

[1] e-posta: nerminyazici@gmail.com, ORCID: 0000-0003-0145-9772, Hacettepe Üniversitesi, Ankara, Türkiye

[2] e-posta: leylacimen@erciyes.edu.tr, ORCID: 0000-0003-0692-6421, Erciyes Üniversitesi, Kayseri, Türkiye

[3] e-posta: mehmetalibahar@erciyes.edu.tr, ORCID: 0000-0003-4193-4515, Erciyes Üniversitesi, Kayseri, Türkiye

Investigation of Semantic Transparency of Compound Words According to Reading Frequency of Middle School, High School and University Students*

Nermin YAZICI[1], Leyla ÇİMEN[2], Mehmet Ali BAHAR[3]

Compound words exhibit different processing and retrieval processes than those suggested by the full listing approach (Taft, 1981) due to their different types of compounding. The limited number of studies addressing the processing of Turkish compound words and the fact that word processing approaches are not considered when determining existing spelling rules are seen as a problem. This problem also affects the writing of questions regarding the spelling of compound words. Therefore, in the study, the processing of compound words in transparent, semi-transparent and opaque categories by native speakers at three education levels is examined in terms of reaction time. The aim of the study is to determine whether the spelling rule, transparency and frequency of the first and second components of a compound word play a role in accessing Turkish compound words and the ways these words are represented in lexicon, and to examine their relationship with reading frequency. The study is carried out with the relational research method. 124 compound words were selected because of their frequencies, lengths, coder agreement and the rating of native speakers. In Exp.1, the reaction times of 184 participants were compared according to the transparency and frequency of the words; it was seen that the differences in reaction times between transparent, semitransparent, and opaque compound word categories were significant. In Exp. 2 and 3, in which a total of 340 participants participated and the spelling style factor was examined, the results of the item analysis revealed that spaced orthography shortened reaction times. Analyses conducted for the participants revealed that spaced orthography reduced the reaction times of middle and high school students and facilitated word recognition; on the other hand, it did not affect reaction times of adults. The reading frequency data did not show significant results in the first, second, and third experiments. As a result of the findings reached based on the general assumption of the research, processing differences were discussed depending on the changes in reaction times. This project, supporting the literature, proposes that transparency affects word processing; suggests that spaced writing causes delays in response time.

Keywords: *compounds, semantic transparency, word processing, reaction time*

Reference

Taft, M. (1981). Prefix stripping revisited. *Journal of verbal learning and verbal behavior*, 20(3), 289- 297.

---

* This study was carried out within the scope of the project titled " Investigating the Relation between Transparency on Turkish Compounds and Word Processing" supported by the TÜBİTAK 1002-A Rapid Support Module.

[1] e-posta: nerminyazici@gmail.com, ORCID: 0000-0003-0145-9772, Hacettepe Üniversitesi, Ankara, Türkiye

[2] e-posta: leylacimen@erciyes.edu.tr, ORCID: 0000-0003-0692-6421, Erciyes Üniversitesi, Kayseri, Türkiye

[3] e-posta: mehmetalibahar@erciyes.edu.tr, ORCID: 0000-0003-4193-4515, Erciyes Üniversitesi, Kayseri, Türkiye

Ölçme Değerlendirme Metinlerinin Seçimi ve Yazımına Yönelik Bir Araç: Okur Dostu Bilgilendirici Metin Değerlendirme Rubriği (ODBMDR)

Dilek Fidan[1], Sercan Demirgüneş[2], Tuğba Çelik Korat [3], Nida Kırömeroğlu[4]

Konusu okur dostu metin değerlendirme rubrikleri olan bu çalışmada, bilgilendirici metinlerin ne kadar okur dostu olduğuna dair bir değerlendirme yapılabilmesi için alanyazına dayandırılarak oluşturulan *Okur Dostu Bilgilendirici Metin Değerlendirme Rubriği*'nin (ODBMDR) tanıtılması amaçlanmıştır. Böylece MEB, ÖSYM gibi sınav uygulamaları yapan kurumların sınavlarında kullanılan ya da eğitim ortamlarında okutulan kitaplarda yer alan bilgilendirici metinlerin ne kadar okur dostu olduğu sayısal olarak ifade edilebilecek ve metnin sınavda ya da derste kullanılıp kullanılmamasına karar verilebilecektir. Alanyazında (Özdemir ve Kıroğlu, 2017, Çelik ve ark., 2015) Okur Dostu Metin Değerlendirme Rubriği- ODMDR (Çelik ve ark., 2014) kullanılarak ders kitabı metinlerinin öğrenciler için uygunluğu, çeşitli açılardan değerlendirilmiştir. Tüm metin türleri için kullanılan ODMDR, bu çalışmada salt bilgilendirici metinler için kullanılmak üzere "Okur Dostu Bilgilendirici Metin Değerlendirme Rubriği" olarak geliştirilmiştir. ODBMDR için toplam 8 boyut belirlenmiştir. Doküman analizi yöntemi kapsamında, rubrikte kullanılacak olan boyutlar 5,6,7 ve 8. sınıf düzeyindeki ders kitaplarından seçilen bilgilendirici metinler üzerinde değerlendirilmiştir. Rubriğin geçerlik güvenirliğini sağlamak amacıyla, metinlerdeki boyutların değerlendirilmesinde, birbirinden bağımsız 2 puanlayıcının puanlarının ağırlıklı Kappa katsayısına bakılmış ve ağırlıklı ortalama Kappa sayısı 0,59 olarak belirlenerek geçerlik güvenirlik açısından anlamlı bir sonuca ulaşılmıştır.

Anahtar Kelimeler: *Okur dostu bilgilendirici metin değerlendirme rubriği,* ölçme *değerlendirme, metin değerlendirme,* ölçek

Kaynakça

Özdemir, Y, & Kıroğlu, K. (2017). İlkokul dördüncü sınıf öğrencilerinin bilgilendirici metinlerde okuduğunu anlama düzeyleri, *Küreselleşen dünyada eğitim,* Pegem, 496-502. doi: 10.14527/9786053188407.33

Çelik, T., Demirgüneş, S., & Fidan, D. (2015). Okur dostu metin olma özelliği ile okuduğunu anlama başarısı arasındaki ilişkinin incelenmesi, *Başkent University Journal of Education,* 2,1,119–127.

Çelik,T, Demirgüneş,S., & Baştuğ, M. (2014). Metin oluşturma ve değerlendirme çalışmalarına yönelik okur dostu metin değerlendirme rubriği geliştirme çalışması. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, 5, 14, 65–82.

[1] dilek.fidan@kocaeli.edu.tr, ORCID: 0000-0003-1349-580X, Kocaeli Üniversitesi, Kocaeli, Türkiye

[2] sdemirgunes@gmail.com, ORCID: 0000-0002-5695-5071, Niğde Ömer Halisdemir Üniversitesi, Niğde, Türkiye.

[3] tugba.celikkorat@bau.edu.tr, ORCID: 0000-0002-2211- 9243, Bahçeşehir Üniversitesi, İstanbul, Türkiye

[4] nidakiromeroglu@gmail.com, ORCID: 0000-0001-8252-979X, İstanbul Üniversitesi, İstanbul, Türkiye

A Tool for Selecting and Writing Assessment Texts: Reader- Friendly Informative Text Rubric (RFITER)

Dilek Fidan[1], Sercan Demirgüneş[2], Tuğba Çelik Korat[3], Nida Kırömeroğlu[4]

Reader-friendly texts are easily perceived and processed by reader(s)/listener(s). This study, which is about reader-friendly text rubrics, aims to introduce the Reader-Friendly Informative Text Rubric (RFITR). The rubric is created according to the standards determined in the basis of the literature in order to decide how reader-friendly informative texts should be textualized. Thus, it will be possible to express numerically whether or not the informative texts used in the exams of institutions such as MEB and ÖSYM or taught in educational environments are reader-friendly. It will be possible to decide whether or not these texts will be used in the exam or in the course. In the literature (Özdemir and Kıroğlu, 2017, Çelik et al., 2015), the readbility of textbook texts for students has been evaluated from various perspectives using the Reader-Friendly Text Rubric-RFTR (Çelik et al., 2014). RFTER, which is used for all text types, has been developed as the "Reader-Friendly Informative Text Rubric" (RFITR) to be used only for informative texts in this study. A total of 8 dimensions have been determined for RFITR. Within the scope of the document analysis method, the dimensions to be used in the rubric were evaluated on informative texts selected from 5th, 6th, 7th and 8th grade textbooks. In order to ensure the validity and reliability of the rubric, the weighted Kappa coefficient of the scores of two independent raters was examined in the evaluation of the dimensions in the texts and the weighted average Kappa number was determined as 0.59, thus achieving a significant result in terms of validity and reliability.

Keywords: *Reader-friendly informative text evaluation rubric, measurement and assessment, scale*

Reference

Özdemir, Y, & Kıroğlu, K. (2017). İlkokul dördüncü sınıf öğrencilerinin bilgilendirici metinlerde okuduğunu anlama düzeyleri, *Küreselleşen dünyada eğitim,* Pegem, 496-502. doi: 10.14527/9786053188407.33

Çelik, T., Demirgüneş, S., & Fidan, D. (2015). Okur dostu metin olma özelliği ile okuduğunu anlama başarısı arasındaki ilişkinin incelenmesi, *Başkent University Journal of Education,* 2,1,119–127.

Çelik,T, Demirgüneş,S., & Baştuğ, M. (2014). Metin oluşturma ve değerlendirme çalışmalarına yönelik okur dostu metin değerlendirme rubriği geliştirme çalışması. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, 5, 14, 65–82.

[1] dilek.fidan@kocaeli.edu.tr, ORCID: 0000-0003-1349-580X, Kocaeli Üniversitesi, Kocaeli, Türkiye

[2] sdemirgunes@gmail.com, ORCID: 0000-0002-5695-5071, Niğde Ömer Halisdemir Üniversitesi, Niğde, Türkiye.

[3] tugba.celikkorat@bau.edu.tr, ORCID: 0000-0002-2211- 9243, Bahçeşehir Üniversitesi, İstanbul, Türkiye Cumhuriyeti

[4] nidakiromeroglu@gmail.com, ORCID: 0000-0001-8252-979X, İstanbul Üniversitesi, İstanbul, Türkiye

Relationship Between Language Assessment Literacy and Teacher Agency of Pre-Service English Language Teachers.

Mehmet Abi [1], Şevki Kömür[2],

Language assessment is an important factor influencing teaching and learning practices in L2 classrooms, and it provides invaluable data for practitioners. In recent years, researchers seeking ways to understand language assessment procedures better have shifted their focus to the assessment literacy field, as it plays a crucial role. Thus, this paper aims to reveal how knowledgeable the pre- service English language teachers are regarding several aspects of language assessment. Additionally, as another factor, teacher agency is known to have a central role in professional teaching practices and L2 education. In this vein, the current paper also aims to determine the perceptions of pre-service English teachers affecting their language classroom practices.

This paper employs a convergent parallel mixed method design in which qualitative and quantitative data are collected and analyzed together. With purposeful sampling, senior pre-service English language teachers attending a state university are selected as the sampling group. The Assessment Literacy Survey (Kremmel, and Harding, 2019) and EFL Teachers' Assessment Agency Questionnaire (Ghiasvand, Jahanbakhsh, and Sharifpour, 2023) are employed to collect quantitative data, and descriptive and inferential statistical tests are applied. Reflection papers to collect qualitative data and content analysis are applied. Potentially, it is expected to figure out what participants think about the Language Assessment Course as a part of their education at the university. In the second phase, conclusions from statistical analyses are expected to provide data consistent with the reflections of participants. At the final stage, it is hoped to understand how teacher agency plays a role in language assessment practices. Finally, it is also anticipated to obtain data to understand the potential relationship between language assessment literacy and language assessment agencies.

Keywords: Language Assessment Literacy*, Language Assessment agencyPre-service English language education. etkinliği, değerlendirme okuryazarlığı, hizmet* öncesi İngilizce öğretmenliği *eğitimi*

Reference

Ghiasvand, F., Jahanbakhsh, A.A. & Sharifpour, P. Designing and validating an assessment agency questionnaire for EFL teachers: an ecological perspective. *Lang Test Asia* **13**, 41 (2023). https://doi.org/10.1186/s40468-023-00255-z

Kremmel, B., & Harding, L. (2019). Towards a Comprehensive, Empirical Model of Language Assessment Literacy across Stakeholder Groups: Developing the Language Assessment Literacy Survey. *Language Assessment Quarterly*, *17*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

[1] mehmetabi@mu.edu.tr, ORCID: 0000-0002-4976-5173, Muğla Sıtkı Koçman University, Muğla, Türkiye

[2] coal@mu.edu.tr, ORCID: 0000-0001-6739-7864, Muğla Sıtkı Koçman University, Muğla, Türkiye

Eğitimde Ölçme ve Değerlendirme Alanında Yapay Zekâ: Sistematik Bir Derleme

F. Gökçen Ayva Yörü[1], Serap Büyükkıdık[2]

Yapay zekânın etkin kullanılmaya başlanmasıyla günlük yaşamdan eğitime birçok alanda hayatımız kolaylaşmıştır. Bu durum eğitim araştırmalarına da yansımaktadır. 2010'dan beri Web of Science ve Google akademikte yapay zekâ ve eğitimle ilgili yapılan araştırmaların sayısı artmakla birlikte, son yıllarda bu araştırmalar popülerliğini iyice arttırmıştır (Chen, Chen, & Lin, 2020). Eğitimde ölçme ve değerlendirme araştırmalarında da yapay zekâ dönüşümü yaşanmıştır. Bu araştırmaların sistematik bir derlemesini yapmak ise araştırmacılara, politika yapıcılara, öğretmenlere ve eğitime gönül verenlere yol gösterici olacaktır. Bu araştırmanın amacı Web of Science'ta eğitimde ölçme ve değerlendirme alanında yapay zekâ üzerine yapılmış olan makalelerin sistematik bir derlemesini yapmaktır. Bu sayede eğitimde ölçme ve değerlendirme yapay zekâ çalışmalarıyla ilgili var olan durumu ortaya çıkararak gelecekteki çalışmalara ışık tutmak amaçlanmaktadır. Bu araştırma, Web of Science'ta eğitimde ölçme ve değerlendirme alanında yapay zekâ üzerine yapılmış olan makalelerin derinlemesine betimlenebilmesi ve var olan durumun ortaya çıkarılabilmesi için sistematik derleme yapıldığı nitel bir çalışmadır. İncelemeye alınacak makaleler belirlenirken Web of Science'da eğitim alanında ve İngilizce dilinde yayımlanmış olan çalışmalar dikkate alınmış ve tüm alanlarda artificial intelligence in educational measurement and evaluation, "artificial intelligence in education" ve "AI in education" kelimeleri taranmış ve kriterlere uygun olan eğitimde ölçme ve değerlendirme alanındaki makaleler araştırmaya dâhil edilmiştir. Eğitimde ölçme ve değerlendirme alanında yapay zekâ üzerine yapılmış çalışmaların incelenebilmesi için araştırmacılar tarafından geliştirilen "makale inceleme formu" kullanılmıştır. Form iki bölümden oluşmaktadır. Formun birinci bölümünde makalelere ilişkin genel bilgiler (dergi adı, dergi indeksi, yayınlanma yılı, yazar sayısı vb.) ve ikinci bölümde ise tematik bilgileri (anahtar kelimeler, araştırmanın türü, ele alınan konular vb.) elde etmeye yönelik sorular yer almaktadır. Araştırmanın veri analizi devam etmektedir.

Anahtar Kelimeler: *Yapay zekâ, Eğitim, Eğitimde* Ölçme *ve Değerlendirme, Sistematik derleme.*

Kaynakça

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8,* 75264- 75278. https://doi.org/10.1109/ACCESS.2020.2988510

[1] korelasyon_@hotmail.com, ORCID: 0000-0002-4555-1987, Afyon Kocatepe Üniversitesi, Afyonkarahisar, Türkiye

[2] serapbuyukkidik@gmail.com, ORCID: 0000-0003-4335-2949, İstanbul Üniversitesi - Cerrahpaşa, İstanbul, Türkiye

Alıcı Beceriler Üretken Şekilde Değerlendirilebilir mi? Alternatif Bir Değerlendirme Aracı Olarak Hikâye Haritalama

Hatice İğsen [1]

Değerlendirme, öğrenme ve öğretme sürecinin temel dayanaklarından biridir. Geleneksel değerlendirmenin eksiklikleri ve dil öğretiminde öğretmen merkezlilikten öğrenci merkezliliğe doğru yaşanan paradigma değişikliği, alternatif değerlendirmeyi gündeme getirmiştir (Moqbel, 2016). Dört dil becerisinden biri olan okuma, yazılı bilgiyi anlamanın temelini oluşturduğundan genellikle ilk öğrenilmesi gereken beceridir. Alıcı bir beceri olarak okuma, çoğunlukla alıcı test formatlarıyla değerlendirilir; ancak alternatif değerlendirme türleri uygulanarak okuma değerlendirmesinin kalitesi artırılabilir. Okuma becerisini değerlendirmenin kalitesini artırmak ve öğrencilerin iletişimsel yeterliliklerini geliştirmek amacıyla araştırmacı-öğretmen hikâye haritalama tekniğini kullanmıştır. Bu teknik, metnin önemli kısımlarını bulmak için görsel veya grafik çerçeveler kullanarak öğrencilerin metindeki temel bilgileri hatırlamalarını ve metinle ilgili soruları yanıtlamalarını sağlar. Hikâye haritalamada amaç, tema, olay örgüsü, karakterler gibi anlatı metninin önemli unsurlarına odaklanmak için öğrencilerin zihinsel şemalarını harekete geçirmek ve hikâyede kullanılan kronolojik örüntüyü netleştirmek ve anlamaktır (Ibnian, 2010). Bir proje okulunun 10. sınıfında öğrenim gören 30 öğrencinin katıldığı bu çalışma hikâye haritalama tekniğinin uygulanışını ve öğrencilerin hikâye haritalarının öğretmen tarafından hazırlanan bir dereceli puanlama anahtarı kullanılarak nasıl değerlendirildiğini anlatmaktadır. Öğrencilere hikâye haritalama tekniği tanıtıldıktan ve hikâye haritalarını değerlendirmek için dereceli puanlama anahtarı verildikten sonra, seviyelerine uygun üç hikâye seçilmiş ve her ay bunlardan birinin okunması sağlanmıştır. Öğrencilerden okudukları hikâyeleri bu tekniği kullanarak özetlemeleri istenmiştir. Öğrencilerin haritaları kullanılarak her ay sınıfta gerçekleştirilen okuma çemberi etkinliklerinde öğrenciler olay örgüsünü tartışmış, karakterleri analiz etmiş ve "Bu hikâyenin yazarı ben olsaydım..." gibi yaratıcı yorumlar yapmışlardır. Sürecin sonunda öğrenciler favori karakterlerini belirterek düşüncelerini paylaşmış ve hikâye haritaları sınıf panosunda sergilenmiştir. Öğrencilerin hikâye haritaları dil, hikâye unsurları, tasarım ve yaratıcılık gibi çeşitli kriterlere göre değerlendirilmiştir. Araştırmacı öğretmen süreç boyunca saha notları alarak gözlemler yapmış ve hikâye haritalamanın alternatif bir değerlendirme aracı olarak kullanılmasının öğrencilerin okuduğunu anlamasını, motivasyonunu ve aktif katılımını artırdığını tespit etmiştir.

Anahtar Kelimeler: *Hik*âye *haritalama, Alternatif değerlendirme, Okuduğunu kavrama.*

Kaynakça

Ibnian, S. S. K. (2010). The Effect of Using the Story-Mapping Technique on Developing Tenth Grade Students' Short Story Writing Skills in EFL. *English Language Teaching*, *3*(4), 181-194.

Moqbel, M. S. S. (2016). Alternative assessment in EFL classrooms: Why and how to implement it.

In *The proceedings of the 19th and 20th Nile TESOL/AUC Conference* (pp. 56-79).

[1] higsen@outlook.com, ORCID: 0000-0002-9864-4028, Teacher (M.A.), Batman, Turkey

Can Receptive Skills Be Assessed Productively? Story Mapping as an Alternative Assessment Tool

Hatice İğsen[1]

Assessment is one of the fundamental pillars of the learning and teaching process. The shortcomings of traditional assessment and the paradigm change in ELT instruction from teacher-centered to student centered has brought the proposition of the alternative assessment (Moqbel, 2016). In ELT, among the four language skills, reading is generally the first to master as it is fundamental to understand written information. As a receptive skill, reading is mostly assessed by receptive test formats, however; the quality of reading assessment can be increased by implementing alternative assessment types. Therefore, to increase the quality of assessing reading and enhance learners' communicative competence, the teacher researcher implemented story mapping. This technique utilises visual or graphic frames to find the important parts of the text, enabling students to remember the basic information in the text and to answer questions about the text. The aim is to activate students' mental schemata to focus on important elements of the narrative text such as theme, plot, characters and to clarify and understand the chronological pattern used in the story (Ibnian, 2010). This study describes the implementation of the story mapping technique for reading comprehension in a 10th-grade class with 30 students at a project high school, and how the students' story maps were evaluated using a teacher-made rubric. After introducing the story mapping technique to the students and providing them with the rubric for assessing their story maps, three stories appropriate to their level were selected and one of them was given to be read each month. Students were asked to summarise the stories they read using this technique. Through the reading circle activities carried out in the classroom every month with students' maps, students discussed the plot, analysed the characters and made creative comments such as "If I were the author of the story...". At the end of the process, students shared their thoughts by indicating their favourite characters and the story maps were displayed on the class board. Students' story maps were evaluated with several criteria: language, story elements, design and creativity. The researcher teacher made observations throughout the process by taking field notes and found that using story mapping as an alternative assessment tool increased students' reading comprehension, their motivation and active participation.

Keywords: *Story mapping, Alternative assessment, Reading comprehension*

Reference

Ibnian, S. S. K. (2010). The Effect of Using the Story-Mapping Technique on Developing Tenth Grade Students' Short Story Writing Skills in EFL. *English Language Teaching*, *3*(4), 181-194.

Moqbel, M. S. S. (2016). Alternative assessment in EFL classrooms: Why and how to implement it.

In *The proceedings of the 19th and 20th Nile TESOL/AUC Conference* (pp. 56-79).

[1] higsen@outlook.com, ORCID: 0000-0002-9864-4028, Teacher (M.A.), Batman, Turkey

Dil Öğreniminde Alternatif Ölçme-Değerlendirme: Öğretmen Görüşleri ve Sınıf Uygulamaları

Esin Çağlayan [1]

Bu çalışma, Türkiye'deki yükseköğretim kurumlarında İngilizceyi Yabancı Dil olarak Öğreten (EFL) öğretmenlerin alternatif e-değerlendirmelerle ilgili görüşlerini incelemektedir. Türkiye genelindeki 17 üniversiteden EFL öğretim elemanlarıyla yapılan yarı yapılandırılmış görüşmelere dayanan çalışma, tematik analiz yöntemiyle altı ana tema ortaya koymuştur: öğrenmeye etkisi, iş yükü, akademik dürüstlük, dijital teknoloji kullanımı, not verme ve geri bildirim süreçleri ve ölçme-değerlendirme yetkinliği. Bulgular, öğretmenlerin farklı bakış açılarını yansıtarak çözülmesi gereken temel sorunlara dikkat çekmektedir.

Çalışmada, bulguların yanı sıra, e-portfolyo, öz/akran değerlendirmeleri, proje bazlı değerlendirmeler, tartışma forumları ve bloglar gibi çeşitli alternatif e-değerlendirme teknikleri ele alınacaktır. Bu tekniklerin uygulama örnekleri üzerinden pratik bilgiler sunularak, katılımcıların alternatif e- değerlendirme tekniklerini sınıflarında nasıl kullanabileceklerine dair somut ipuçları verilecektir. Ayrıca, öğretmenlerin alternatif e-değerlendirme tekniklerini etkili bir şekilde uygulayabilmeleri için ihtiyaç duydukları becerileri kazandırmayı amaçlayan mesleki gelişim programlarının nasıl tasarlanabileceği tartışılacaktır.

Çalışmanın ortaya koyduğu sonuçlar, dijital çağda eğitim uygulamalarını daha da geliştirmek için dijital ölçme-değerlendirme politikaları ve prosedürlerinin oluşturulmasına katkı sağlayacaktır. Katılımcılar, alternatif ölçme-değerlendirmelerin sağladığı fırsatlar ve uygulanmasında karşılaşılan zorluklar hakkında daha derin bir anlayış edinerek, yabancı dil eğitiminde yenilikçi ölçme-değerlendirme yöntemleri üzerine yapılan tartışmalara katkıda bulunacaklardır.

Anahtar Kelimeler: *Alternatif ölçme-değerlendirme, e-değerlendirme, EFL, yükseköğretim, nitel çalışma.*

[1] esin.caglayan@ieu.edu.tr, ORCID: 0000-0003-4516-2743, İzmir Ekonomi Üniversitesi, İzmir, Türkiye

Alternative Assessments in Language Learning: Teacher Voices and Practical Applications

Esin Çağlayan[1]

This presentation explores the perceptions of English as a Foreign Language (EFL) teachers regarding alternative e-assessments within the context of Turkish higher education. Through semi-structured interviews with EFL instructors from 17 universities in Türkiye, this study utilized thematic analysis uncovering six primary themes: the impact on learning, workload, academic integrity, digital technology, grading and feedback, and assessment competence. Our findings reveal diverse perceptions and highlight key areas of concern that need to be addressed.

In addition to discussing the findings, the presentation will describe various alternative assessment techniques such as e-portfolios, self/peer assessments, project-based assessments, discussion forums and blogs. Examples of implementation of these techniques will be showcased, providing practical insights into their application. The presentation will also discuss the implications of the findings for professional development programs aimed at equipping teachers with the necessary skills to effectively implement alternative e-assessments.

Furthermore, the study's insights will inform the development of assessment policies and procedures that enhance educational practices in the digital age. Attendees will gain a deeper understanding of the challenges and opportunities associated with alternative e-assessments, contributing to the broader discourse on innovative assessment methods in foreign language education.

Keywords: *Alternative assessments, e-assessments, EFL, higher education, qualitative study*

[1] esin.caglayan@ieu.edu.tr, ORCID: 0000-0003-4516-2743, Izmir Ekonomi Üniversitesi, İzmir, Türkiye

Kelime Denizinde Bir Yelkenli: Yapılandırılmış Grid Yöntemiyle Dil Alanında Yeni Ufuklar

Gülten AKGÜL[1]

Günümüz eğitim ikliminde alternatif ölçme ve değerlendirme, öğrencinin gerçek hayatta karşılaşabileceği durumları simüle ederek dil becerilerini değerlendiren bir yaklaşımdır. Alternatif ölçme ve değerlendirme yöntemlerinden biri olan yapılandırılmış grid yöntemi, öğrencilerin dil becerilerini çok yönlü bir şekilde değerlendirmek için etkili bir araçtır. Bu bağlamda, bu makalenin amacı, tanınmış bir Amerikalı yazar olan O. Henry'nin (1862–1910) "The Gift of the Magi" adlı kısa öyküsü üzerinden yapılandırılmış grid yönteminin öğrencilerin dil becerilerini ölçme ve değerlendirmedeki etkinliğini analiz etmektir. Bir ay süren araştırmanın örneklem grubunu, 2022-2023 eğitim öğretim yılı güz döneminde haftada 2 saat İngilizce öğretim programı olan bir meslek lisesinin iki ayrı 11. sınıf şubesinde öğrenim gören 24 öğrenci oluşturmuştur. Nitel ve nicel verilerin bir arada kullanılarak verilerin elde edildiği karma desen araştırma yöntemi kullanılmıştır. Araştırmanın deneysel deseninde, veri toplama aşamasında deney (12) ve kontrol (12) gruplarına ön test ve son testler uygulanmıştır. Buna ek olarak, deney grubuna yapılandırılmış grid yöntemi ile hazırlanmış bir ara sınav da uygulanmıştır. Her bir dil becerisi için ayrı bir grid oluşturulmuş ve bu gridlerde belirlenen gereksinimleri içeren sütunlar eklenmiştir. Nitel veriler, görüşme teknikleri ve gözlem yoluyla toplanmıştır. Bulgular, yapılandırılmış grid yönteminin ara sınav olarak uygulandığı deney grubu öğrencileri (%93) ile hiçbir yöntemle ara sınav ile değerlendirilmeyen kontrol grubu öğrencileri (%67) arasında akademik başarı açısından anlamlı bir fark olduğunu ortaya koymuştur. Sonuç olarak, yapılandırılmış grid yöntemi öğrencilerin dil becerilerini değerlendirmek için etkin bir şekilde uygulanabilir.

Anahtar Kelimeler: *Dil becerileri,* ölçme *ve değerlendirme, yapılandırılmış grid yöntemi*

Kaynakça

Bachman, L. F., & Clark, J. L. (1987). The measurement of foreign/second language proficiency. *The Annals of the American Academy of Political and Social Science*, *490(*1), 20-33.

Can, T. (2009). Learning and teaching languages online: A constructivist approach. *Novitas-ROYAL (Research on Youth and Language), 3*(1).

Gultom, E. (2016). Assessment and evaluation in EFL teaching and learning. *Proceedings of ISELT FBS Universitas Negeri Padang, 4*(1), 190-198.

[1] gulakgul5@gmail.com, ORCID: 0000-0002-0476-6826, Milli Eğitim Bakanlığı, Kütahya, Türkiye

A Sailboat in the Sea of Words: Fresh Horizons in the Realm of Language Via the Structured Grid Method

Gülten AKGÜL[1]

In today's educational climate, alternative measurement and assessment is an approach that evaluates language skills by simulating situations that the student may encounter in real-life. The structured grid method, one of the alternative measurement and assessment methods, is an effective tool to evaluate students' language skills in a versatile way. In this context, the purpose of this paper is to analyze the efficacy of the structured grid method in measuring and assessing students' language skills through the short story "The Gift of the Magi" by a well-known American writer, O. Henry (1862– 1910). The sample group of the research, which lasted one month, consisted of 24 students from two separate 11th grade branches at a vocational high school in the fall semester of the 2022-2023 academic year, with an English teaching program of 2 hours per week. A mixed-design research method in which qualitative and quantitative data were utilized together to obtain the data. During the data collection phase, pre- and post-tests were applied to the experimental (12) and control groups (12) in the experimental design of the research. In addition to this, a mid-term exam prepared with the structured grid method was also applied to the experimental group. A separate grid was created for each language skill, and columns containing the requirements determined in these grids were added. Qualitative data was collected through interview techniques and observation. The findings revealed that there was a significant difference between the academic achievements of the experimental group students (%93), to whom the structured grid method was applied as a midterm exam, and the control group students (%67), who were not assessed with a midterm exam by any method. As a result, the structured grid method can be applied flourishingly to evaluate students' language skills.

Keywords: *Language skills, measurement and evaluation, structured grid method*

References

Bachman, L. F., & Clark, J. L. (1987). The measurement of foreign/second language proficiency. *The Annals of the American Academy of Political and Social Science*, *490(*1), 20-33.

Can, T. (2009). Learning and teaching languages online: A constructivist approach. *Novitas-ROYAL (Research on Youth and Language), 3*(1).

Gultom, E. (2016). Assessment and evaluation in EFL teaching and learning. *Proceedings of ISELT FBS Universitas Negeri Padang, 4*(1), 190-198.

[1] gulakgul5@gmail.com, ORCID: 0000-0002-0476-6826, The Ministry of National Education, Kütahya, Türkiye

ChatGPT Destekli Ölçme Sürecinin Ortaokul Öğrencilerinin Öğrenme Eksikliklerini Giderme Üzerindeki Etkisi

Gülşah GÜRKAN [1], Mehmet Akif ARDUÇ [2]

Bu çalışmanın amacı yapay zekâ dil modelleri kullanılarak yapılan ölçme işleminin 7. sınıfta okuyan öğrencilerdeki öğrenme eksikliklerini gidermedeki etkisini araştırmaktır. Nicel metodoloji kapsamında yarı deneysel deseninin kullanıldığı çalışmada her iki grupta (deney ve kontrol grubu) da "Güneş Sistemi ve Ötesi" ünitesi işlenmesi sonucunda başarı testleri uygulanacaktır. Kontrol grubunda bulunan öğrencilere geçerli ve güvenilir başarı testi geleneksel kâğıt-kalem yöntemi kullanılarak uygulanırken deney grubundaki öğrencilere aynı sorular yapay zekâ dil araçlarından olan ChatGPT aracılığıyla uygulanacaktır. Deney grubundaki öğrencilere ChatGPT yapay zekâ dil aracı, öğrencilerin doğru cevabı bulacağı şekilde yönlendirmeler yaparak öğrencilerin ilerlemelerini sağlayacaktır. Deney ve kontrol grubuna yapılan uygulamalardan sonra ChatGPT'nin öğrenmedeki kalıcılık etkisini analiz etmek amacıyla kalıcılık testi uygulanacaktır.

Bu bağlamda çalışma, öğrencilerin yanlış cevapladığı sorularda doğruyu bulana kadar yapay zekâ aracı tarafından verilen ipuçları ile öğrencilerin öğrenme eksiklikleri ortaya konacak ve deney grubu ile kontrol grubu arasındaki başarı farkı karşılaştırılarak, yapay zekâ desteğinin öğrenme üzerindeki etkisi analiz edilecektir. Elde edilecek bulgular, gelecekteki ölçme değerlendirme uygulamalarında yapay zekâ teknolojilerinin daha etkin kullanılmasına katkı sağlayabilir.

Anahtar Kelimeler: *Yapay z lığı.*

[1] gulsah.gurkan@ozal.edu.tr: ORCID: 0000-0003-0297-3060, Malatya Turgut Özal Üniversitesi, Malatya, Türkiye

[2] arducakif@gmail.com: ORCID: 0000-0002-1026-2843, *Millî Eğitim Bakanlığı, Malatya, Türkiye*

The Effect of ChatGPT Supported Assessment Process on Eliminating Learning Deficiencies of Middle Students

Gülşah GÜRKAN [1], Mehmet Akif ARDUÇ [2]

The aim of this study is to investigate the effect of measurement using artificial intelligence language models in eliminating the learning deficiencies in 7th grade students. In the study where quasi- experimental design is used within the scope of quantitative methodology, achievement tests will be applied to both groups (experimental and control groups) as a result of the "Solar System and Beyond" unit. While the valid and reliable achievement test will be applied to the students in the control group using the traditional paper-and-pencil method, the same questions will be applied to the students in the experimental group through ChatGPT, an artificial intelligence language tool.

ChatGPT artificial intelligence language tool will help students in the experimental group progress by providing guidance in a way that they will find the correct answer. After the applications to the experimental and control groups, a permanence test will be applied to analyze the permanence effect of ChatGPT on learning. In this context, the study will reveal the learning deficiencies of the students with the clues given by the artificial intelligence tool until the students find the answer to the questions they answered incorrectly, and the effect of artificial intelligence support on learning will be analyzed by comparing the difference in achievement between the experimental group and the control group. The findings may contribute to the more effective use of artificial intelligence technologies in future measurement and evaluation applications.

Keywords: *Artificial intelligence, ChatGPT, learning retention.*

[1] gulsah.gurkan@ozal.edu.tr: ORCID: 0000-0003-0297-3060, Malatya Turgut Özal Üniversitesi, Malatya, Türkiye

[2] arducakif@gmail.com: ORCID: 0000-0002-1026-2843, *Millî Eğitim Bakanlığı, Malatya, Türkiye*

Geniş Dil Modelleriyle Q Matrisi Üretimi: GPT-4o'nun Performans Analizi

Semih AŞİRET [1], Seçil ÖMÜR SÜNBÜL [2]

Bilişsel tanı modelleri (BTM), genel bir test puanı yerine, her bir bireyin beceri ve nitelikleri hakkında ayrıntılı bir profil sunmaktadır. Tüm BTM'nde, maddeler ile nitelikler arasındaki ilişkileri gösteren bir Q-matrisi gerektirmektedir. Günümüzde Q-matrisi oluşturma süreci, genellikle alan uzmanlarının deneyimlerine dayalı olarak gerçekleştirilmektedir. Ancak, bu sürecin sağlam bir teorik temele dayanması zorunludur. Son yıllarda, geniş dil modellerinin gelişimiyle birlikte, bu modellerin eğitim alanındaki kullanımı yaygınlaşmaktadır. Bu çalışmada, geniş dil modellerinden biri olan GPT-4o tarafından belirtilen nitelik ve maddelere ilişkin otomatik Q matrisi oluşturma performansının incelenmesi amaçlanmaktadır. Bu çalışmada, de la Torre (2011) tarafından geliştirilen kesirlerde çıkarma, Park ve Lee (2014) tarafından geliştirilen TIMSS 2007 4. Sınıf matematik ve Chen ve Chen (2016) tarafından geliştirilen PISA 2000 okuma testleri için Q matrisleri ve üç farklı veri seti, *CDM* paketinden alınarak kullanılmıştır. Verilerin analizi R programında yer alan *Qval* ve *GDINA* paketleri kullanılarak gerçekleştirilmiştir. Q matrisi oluşturma sürecinde, öncelikle GPT-4o üzerinde özel bir model geliştirilmiş ve Q matrisinin yapısına ilişkin bilgiler bu modele aktarılmıştır. Tüm maddelerin *JSON* kodları üretilmiş ve her çalışmada yer alan nitelikler ve açıklamaları modele tanımlanmıştır. Son olarak, her bir madde tek tek GPT-4o'ya aktarılarak, ilgili teste ilişkin Q matrisi oluşturulmuştur. Oluşturulan Q matrislerinin geçerliği, MLR-B ve Hull yöntemleri kullanılarak değerlendirilmiş ve GPT-4o tarafından üretilen ile araştırmacılar tarafından oluşturulan Q matrislerinin geçerlik düzeyleri karşılaştırılmıştır. Ayrıca farklı Q matrislerinden oluşan modellerin model veri uyumları incelenmiş ve GPT-4o'nun ürettiği Q matrisleri ile araştırmacıların ürettiği matrisler arasındaki model veri uyumu karşılaştırılmıştır. Çalışmanın sonucunda, matematik testlerinde GPT-4o tarafından üretilen Q matrislerinin, araştırmacılar tarafından belirlenen Q matrislerine kıyasla daha yüksek geçerlik ve model veri uyumu sergileyeceği; ancak okuma temelli testlerde bu oranın daha düşük olacağı öngörülmektedir.

Anahtar Kelimeler: *Geniş Dil Modelleri, Bilişsel Tanı Modeli, Q-matrix, Eğitsel* Ölçme

Kaynakça

Chen, H., & Chen, J. (2016). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, 36(6), 1049-1064.

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. Applied Psychological Measurement, 37, 419-437.

Park, Y. S., & Lee, Y. S. (2014). An extension of the DINA model using covariates: Examining factors affecting response probability and latent classification. *Applied Psychological Measurement*, 38(5), 376-390.

[1] semihasiret@gmail.com:, ORCID ID: 0000-0002-0577-2603, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

[2] secilomur@gmail.com:, ORCID ID: 0000-0001-9442-1516, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

Q Matrix Generation with Large Language Models: Performance Analysis of GPT-4o

Semih AŞİRET [1], Seçil ÖMÜR SÜNBÜL [2]

Cognitive diagnostic models (CDM) provide a detailed profile of each individual's skills and attributes, rather than an overall test score. All CTMs require a Q-matrix showing the relationships between items and attributes. Today, the process of constructing a Q-matrix is often based on the experience of domain experts. However, it is imperative that this process is based on a solid theoretical foundation. In recent years, with the development of large language models, the use of these models in the field of education has become widespread. In this study, we aim to examine the performance of automatic Q matrix generation for the attributes and items specified by GPT-4o, which is one of the large language models. In this study, Q matrices and three different data sets for the subtraction of fractions tests developed by de la Torre (2011), TIMSS 2007 4th grade mathematics tests developed by Park and Lee (2014), and PISA 2000 reading tests developed by Chen and Chen (2016) were obtained from the *CDM* package. The data were analyzed using the *Qval* and *GDINA* packages in the R program. In the process of creating the Q matrix, a special model was first developed on GPT-4o and information about the structure of the Q matrix was transferred to this model. JSON codes of all items were generated and the attributes and their descriptions in each study were defined in the model. Finally, each item was transferred one by one to GPT-4o and the Q matrix for the related test was created. The validity of the Q matrices was evaluated using MLR-B and Hull methods, and the validity levels of the Q matrices generated by GPT-4o and those generated by the researchers were compared. In addition, the model data fit of the models consisting of different Q matrices were examined and the model data fit between the Q matrices produced by GPT-4o and the matrices produced by the researchers were compared. As a result of the study, it is predicted that the Q matrices produced by GPT-4o will exhibit higher validity and model data fit in mathematics tests compared to the Q matrices determined by the researchers; however, this ratio will be lower in reading-based tests.

Keywords: *Large Language Models, Cognitive Diagnostic Model, Q matrix, Educational Measurement*

Reference

Chen, H., & Chen, J. (2016). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, 36(6), 1049-1064.

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. Applied Psychological Measurement, 37, 419-437.

Park, Y. S., & Lee, Y. S. (2014). An extension of the DINA model using covariates: Examining factors affecting response probability and latent classification. *Applied Psychological Measurement*, 38(5), 376-390.

[1] semihasiret@gmail.com:, ORCID ID: 0000-0002-0577-2603, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

[2] secilomur@gmail.com:, ORCID ID: 0000-0001-9442-1516, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

PISA 2022 Matematik Okuryazarlığı Şekil Formatındaki Maddelerin Türkçe ve İngilizce Olarak Cevaplanmasında ChatGPT'nin Performansı

Sümeyra Soysal [1], Elif Özlem Ardıç Kemerkaya [2]

Bu çalışma kapsamında, ağırlıklı alanın matematik okuryazarlığı olduğu PISA 2022 uygulamasında yer alan ve OECD tarafından yayımlanan 4 çoktan seçmeli madde, İngilizce ve Türkçe dillerinde ChatGPT 4o'ya sorulmuş ve cevapların doğruluğu karşılaştırılmıştır. Çalışmanın sonucunda, ChatGPT'nin güneş sistemi ünitesiyle ilgili maddeyi hem Türkçe hem de İngilizce olarak sorulduğunda doğru cevapladığı bulunmuştur. Üçgen deseni ile ilgili, çokluk konu alanındaki madde Türkçe ve İngilizce olarak sorulduğunda ChatGPT yanlış cevap vermiştir. Ancak cevabını gözden geçirmesi istenildiğinde, Türkçe olarak yanlış cevap verirken, İngilizce olarak doğru cevabı bulmuştur. Üçgen deseni ünitesine ilişkin değişim ve ilişkiler içerik alanındaki madde İngilizce olarak sorulduğunda doğru cevaplanmış, Türkçe olarak sorulduğunda ise yanlış cevap verilmiştir. Programdan cevabını gözden geçirmesi istenildiğinde de doğru cevabı bulamamıştır. Son olarak, ormanlık alan ünitesi ile ilgili maddeyi Türkçe ve İngilizce olarak doğru cevaplamıştır. Çalışmanın sonuçları ChatGPT'nin Türkçe ve İngilizce olarak sorulan maddelere verdiği cevaplar arasında farklılıklar olduğunu, programın İngilizce dilinde daha iyi performans gösterdiğini ortaya koymuştur.

Anahtar Kelimeler: *ChatGPT, matematik okuryazarlığı, PISA.*

[1] sumeyrasoysal@hotmail.com, ORCID: 0000-0002-7304-1722, Necmettin Erbakan Üniversitesi, Konya, Türkiye

[2] ardic@trabzon.edu.tr, ORCID: 0000-0003-2983-2970, Trabzon Üniversitesi, Trabzon, Türkiye

ChatGPT's Performance in Answering PISA 2022 Mathematics Literacy Questions in Image Format in Both Turkish and English

Sümeyra Soysal [1], Elif Özlem Ardıç Kemerkaya [2]

In this study, four multiple-choice items from the PISA 2022 assessment, which focuses primarily on mathematics literacy and was published by the OECD, were presented to ChatGPT 4o in both English and Turkish, and the accuracy of the responses was compared. The results showed that ChatGPT correctly answered the item related to the solar system unit in both English and Turkish. However, for the item related to triangle patterns in the domain of quantity, ChatGPT provided incorrect answers in both languages. When asked to review its responses, it still answered incorrectly in Turkish but correctly in English. For the item related to the unit of triangle patterns in the domain of change and relationships, ChatGPT answered correctly in English but incorrectly in Turkish, and it could not provide the correct answer when asked to review its response. Finally, ChatGPT answered the item related to the forest unit correctly in both Turkish and English. The results of the study indicate that there are differences in the answers ChatGPT provides to items asked in Turkish and English, with the program performing better in English.

Keywords: *ChatGPT, mathematics literacy, PISA.*

[1] sumeyrasoysal@hotmail.com, ORCID: 0000-0002-7304-1722, Necmettin Erbakan Üniversitesi, Konya, Türkiye

[2] ardic@trabzon.edu.tr, ORCID: 0000-0003-2983-2970, Trabzon Üniversitesi, Trabzon, Türkiye

Okuduğunu Anlama Becerisini Yordayan Değişkenlerin Makine Öğrenme Algoritmalarıyla İncelenmesi: PIRLS 2021 Örneği

Fatma Nur AYDIN [1], Kübra ATALAY KABASAKAL[2]

Yapay zekâ ve veri bilimindeki gelişmelerle beraber çeşitli kararların karmaşık sistemler aracılığıyla alınması söz konusudur. Bu bağlamda çeşitli verilerden yararlanarak bir kararın/sonucun tahmin edilmesi prensine dayanan makine öğrenme modellerinin çalışma prensiplerinin anlaşılması giderek daha önemli hale gelmektedir (Redelmeier ve ark., 2020). Bunlardan topluluk öğrenme yöntemlerinde, daha basit modellerden oluşan pek çok ağacın güçlü yönlerinin birleştirilmesiyle daha isabetli tahmin modelleri oluşturulur (Hastie ve ark., 2009). Bu bağlamda çalışmada, PIRLS 2021 okuduğunu anlama becerisini yordayan değişkenlerin torbalama, rastgele orman ve gradyan artırma algoritmalarıyla incelenerek, bu yöntemlerin tahmin performanslarının kıyaslanması amaçlanmıştır. Çalışmada PIRLS 2021 döngüsüne Türkiye'den katılan 5022 kişinin verileri kullanılmıştır. Elde edilen sonuçlara göre tüm modeller benzer metrik değerlerine sahip olmakla birlikte, gradyan artırma daha iyi tahmin performansı (MAE=52.009, MAPE=0.110, MSE=4218.942, RMSE=64.953, yanlılık=0.657, R=0.634,

R^2=0.401) sergilemiştir. Tüm algoritmalarda okumaya yönelik özgüven düzeyi başarıyı yordayan en önemli değişken olmuştur. Gradyan artırmadan elde edilen değişken önem sıralamasına göre en iyi yordayıcılar sırasıyla okumaya yönelik özgüven düzeyi, evin sosyoekonomik durumu ve öğrenmeye yönelik evdeki kaynak miktarı olmuştur. Bu üç değişkene ilişkin kısmi bağımlılık grafikleri incelendiğinde, tümünün okuduğunu anlama ile arasında pozitif yönde ve neredeyse doğrusal bir ilişki olduğu dikkat çekmiştir.

Anahtar Kelimeler: *Makine* öğrenmesi*, topluluk* öğrenme *algoritmaları, PIRLS 2021*

Kaynakça

Redelmeier, A., Jullum, M., & Aas, K. (2020). Explaining predictive models with mixed features using shapley values and conditional inference trees. In Holzinger, A., Kieseberg, P., Tjoa, A., & Weippl, E. (Eds.). *Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4* (pp. 117-137). Springer International Publishing. https://doi.org/10.1007/978-3-030-57321-8_7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

[1] fatmanuraydin.2012@gmail.com: ORCID: 0000-0003-0887-395X, Milli Eğitim Bakanlığı, Ankara, Türkiye

[2] kkatalay@gmail.com: ORCID: 0000- 0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

Investigation of Variables Predicting Reading Literacy Skills with Machine Learning Algorithms: The Case of PIRLS 2021

Fatma Nur AYDIN [1], Kübra ATALAY KABASAKAL[2]

With the advances in artificial intelligence and data science, various decisions are being made through complex systems. In this context, it is becoming increasingly important to understand the working principles of machine learning models based on the principle of predicting a decision/outcome by utilizing various data (Redelmeier et al., 2020). In ensemble learning methods, more accurate prediction models are created by combining the strengths of many trees consisting of simpler models (Hastie et al., 2009). In this context, this study aims to examine the variables predicting PIRLS 2021 reading literacy skills with bagging, random forest and gradient boosting algorithms and to compare the prediction performance of these methods. The study uses data from 5022 participants from Türkiye in the PIRLS 2021 cycle. According to the results obtained, although all models have similar metric values, gradient boosting has better prediction performance (MAE=52.009, MAPE=0.110, MSE=4218.942, RMSE=64.953, bias=0.657, R=0.634, R^2=0.401). In all algorithms, the level of self- confidence towards reading was the most important variable predicting success. According to the variable importance ranking obtained from gradient boosting, the best predictors were the level of self-confidence towards reading, the socioeconomic status of the home, and the amount of learning resources at home, respectively. When the partial dependency graphs of these three variables are analyzed, it is noteworthy that all of them have a positive and almost linear relationship with reading literacy.

Keywords: *Machine learning, ensemble learning algorithms, PIRLS 2021*

Reference

Redelmeier, A., Jullum, M., & Aas, K. (2020). Explaining predictive models with mixed features using shapley values and conditional inference trees. In Holzinger, A., Kieseberg, P., Tjoa, A., & Weippl, E. (Eds.). *Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4* (pp. 117-137). Springer International Publishing. https://doi.org/10.1007/978-3-030-57321-8_7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

[1] fatmanuraydin.2012@gmail.com, ORCID: 0000-0003-0887-395X, Republic of Türkiye Ministry of National Education, Ankara, Türkiye

[2] kkatalay@gmail.com, ORCID: 0000- 0002-3580-5568, Hacettepe University, Ankara, Türkiye

Psikomotor Becerilerin Ölçülmesinde Yapay Zekâ Tespit Araçlarından Elde Edilen Sonuçların Doğruluğunun İncelenmesi

Emre KUCAM[1], İsmet MALAKCIOĞLU[2]

Yapay zekâ dedektörü/tespit araçları olarak ifade edilen araçlar özellikle ödevler, çevrimiçi olarak gerçekleştirilen sınavlar ve proje ödevleri gibi süreçlerde, hile ve kopya gibi girişimleri belirleyebilmek için kullanılabilmektedir (Aktay, 2024). Bu araştırmanın genel amacı, yapay zekâ tespit araçlarından elde edilen sonuçların doğruluğunun belirlenmesidir. Bu bağlamda, bir özel okulun görsel sanatlar, müzik, beden eğitimi ve bilişim teknolojileri öğretmeni alımı için uyguladığı yazılı sınavda sorulan ölçme ve değerlendirmeye yönelik bir soruya verilen 17'si gerçek ve 17'si yapay zekâ tarafından yazdırılan yanıtlar, iki farklı yapay zekâ tespit aracı (Scribb ve ZeroGPT) ve bir uzman tarafından incelenmiş, benzerlik yüzdeleri elde edilmiştir. Soruya verilen gerçek yanıtların ortalaması en düşük ZeroGPT aracı ile en yüksek de uzman değerlendirmesi ile elde edilmiştir. Scribbr uygulaması ise uzman değerlendirmesine daha yakın bir ortalamada bulunmaktadır. İlgili soruya verilen oluşturulmuş yanıtların ortalaması ise en düşük Scribb aracı ile, en yüksek de zeroGPT aracı ile belirlenmiştir. Uzman değerlendirmesi ise zeroGPT aracına daha yakın bir ortalamada bulunmaktadır. Oluşturulmuş yanıtlar arasında yapay zekâ yüzdelerini tespit eden iki araç ile bir uzmandan elde edilen verilere uygulanan Wilcoxon Sıra toplamları testinde zeroGPT ve Scribb tespit yüzdelerinin istatistiksel olarak birbirinden farklı olmadığı, ancak Scribb ve uzman ile zeroGPT ve uzman tespit yüzdelerinin birbirinden manidar düzeyde farklılaştığı belirlenmiştir. Verilere uygulanan Wilcoxon Sıra toplamları testinde zeroGPT ve Scribb ile zeroGPT ve uzman tespit yüzdelerinin istatistiksel olarak birbirinden farklılaşmadığı, scribb ve uzman tespit yüzdelerinin istatistiksel olarak birbirinden farklılaştığı belirlenmiştir. Sonuçlar incelendiğinde; tüm verilerin zeroGPT ve Scribb ile zeroGPT ve uzman tespitlerinin birbirinden farklılaşmadığı ifade edilebilir. Oluşturulmuş yanıtların her iki tespit aracı arasında birbirinden manidar düzeyde farklılaşmadığı ifade edilebilir. Uzman ile zeroGPT ve uzman ile Scribb tespitlerinin ise birbirinden farklılaşmadığı belirlenmiştir.

Anahtar Kelimeler: *zeroGPT, Scribb, dedektör.*

Kaynakça

Aktay, S. (2024). Eğitimde yapay zeka tehlikesi ve yapay zeka tespit araçları. *EU 4. Uluslararası Sosyal ve Beşeri Bilimler Kongresi içinde (ss. 230-237).* Paris, Fransa: Academy Global Publishing House.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever. (2018). *Improving Language Understanding by Generative Pre-Training*. OpenAI Blog. https://openai.com/research/lan guage- unsupervised

Zielinski, C., M. A. Winker, R. Aggarwal, L. E. Ferris, M. Heinemann, S. A. P. Jose Florencio Lapeña Jr, I., Edsel, et al. (2023). *"Chatbots, Generative AI, and Scholarly Manuscripts: WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications."* WAME Revised May 31, 2023. https://wame.org/page3.php?id= 106.

[1] emrekucam@gmail.com, ORCID: 0000-0002-4283-7103, TOBB ETÜ Koleji, Ankara, Türkiye

[2] ismetkurt@gmail.com, ORCID: 0000-0002-5332-0046, TOBB ETÜ Koleji, Ankara, Türkiye

Examining the Accuracy of Results Obtained from Artificial Intelligence Detection Tools in Measuring Psychomotor Skills

Emre KUCAM[1], İsmet MALAKCIOĞLU[2]

The tools referred to as artificial intelligence detector/detection tools can be used to identify attempts such as cheating and cheating, especially in processes such as homework, online exams and project assignments (Aktay, 2024). The general purpose of this research is to determine the accuracy of the results obtained from artificial intelligence detection tools. In this context, 17 real and 17 artificial intelligence-written answers given to a question about measurement and evaluation in a written exam applied by a private school for the recruitment of visual arts, music, physical education and information technologies teachers were examined by two different artificial intelligence detection tools (Scribb and ZeroGPT) and an expert, and similarity percentages were obtained. The average of the real answers given to the question was obtained with the lowest ZeroGPT tool and the highest with expert evaluation. The Scribbr application is at an average closer to expert evaluation. The average of the generated answers given to the relevant question was determined with the lowest Scribb tool and the highest with the zeroGPT tool. The expert assessment is closer to the zeroGPT tool on average. In the Wilcoxon Rank Sum test applied to the data obtained from two tools that detect artificial intelligence percentages among the generated responses and an expert, it was determined that the zeroGPT and Scribb detection percentages were not statistically different from each other, but the Scribb and expert and zeroGPT and expert detection percentages differed significantly from each other. In the Wilcoxon Rank Sum test applied to the data, it was determined that the zeroGPT and Scribb and zeroGPT and expert detection percentages did not differ statistically from each other, and the scribb and expert detection percentages differed statistically from each other. When the results were examined; it can be stated that all data did not differ between zeroGPT and Scribb and zeroGPT and expert detections. It can be stated that the generated responses did not differ significantly from each other between both detection tools. It was determined that the expert and zeroGPT and expert and Scribb detections did not differ from each other.

Keywords: *zeroGPT, Scribbr, detector.*

Reference

Aktay, S. (2024). Eğitimde yapay zeka tehlikesi ve yapay zeka tespit araçları. *EU 4. Uluslararası Sosyal ve Beşeri Bilimler Kongresi içinde (ss. 230-237).* Paris, Fransa: Academy Global Publishing House.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever. (2018). *Improving Language Understanding by Generative Pre-Training.* OpenAI Blog. https://openai.com/research/lan guage-unsupervised

Zielinski, C., M. A. Winker, R. Aggarwal, L. E. Ferris, M. Heinemann, S. A. P. Jose Florencio Lapeña Jr, I., Edsel, et al. (2023). *"Chatbots, Generative AI, and Scholarly Manuscripts: WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications."* WAME Revised May 31, 2023. https://wame.org/page3.php?id= 106.

[1] emrekucam@gmail.com, ORCID: 0000-0002-4283-7103, TOBB ETÜ Koleji, Ankara, Türkiye

[2] ismetkurt@gmail.com, ORCID: 0000-0002-5332-0046, TOBB ETÜ Koleji, Ankara, Türkiye

Yapay Zekâ ve Nitel Veri: Geleceğin Analiz Yöntemleri

İbrahim Hakkı TEZCİ [1], Bayram BIÇAK [2],

Bu çalışma, ChatGPT gibi ileri düzey yapay zekâ modellerinin, nitel verilerin analizinde etkili bir şekilde nasıl kullanılabileceğine odaklanmaktadır. Çalışmada bir araştırma kapsamında toplanan görüşme verileri ChatGPT aracılığıyla analiz edilmiştir. İlk adımda, görüşme transkriptlerinin ChatGPT'ye yüklenmesi ve metinlerin otomatik olarak özetlenmesi sağlanmıştır. Daha sonra, metinlerin belirli temalar ve kodlar çerçevesinde analiz edilmesi için doğal dil işleme yeteneklerinden faydalanılmıştır. ChatGPT, metinleri analiz ederek sıkça tekrar eden temaları ve önemli ifadeleri (kodları) belirlemiştir. Bu süreçte araştırmacı, yapay zekanın önerdiği temaları inceleyip kendi uzmanlığıyla doğrulamalar yapmış ve gerekli düzeltmeleri gerçekleştirmiştir. Son olarak, elde edilen bulguların görselleştirilmesi ve raporlanması aşamasına geçilmiştir. Söz konusu bu çalışma, ilgili adımların ve nitel veri analizinde kullanılan promptların nasıl olması gerektiği üzerine odaklanmaktadır. Bu adımlar, nitel veri analizinin daha sistematik ve objektif bir şekilde gerçekleştirildiğini göstermektedir.

Elde edilen sonuçlardan hareketle yapay zekâ destekli nitel veri analizi, akademik araştırmalarda birçok avantaj sunmaktadır. İlk olarak, büyük veri setlerinin analiz edilmesinde zaman ve iş gücü tasarrufu sağlar. Otomatik kodlama ve tema oluşturma gibi süreçler, araştırmacıların veri analizine harcadıkları süreyi önemli ölçüde azaltır. İkinci olarak, yapay zekâ, daha objektif ve tekrarlanabilir analizler sunar. İnsan hatası ve önyargıların en aza indirildiği bu analizler, güvenilirliği artırır. Üçüncü olarak, yapay zekâ, karmaşık veri setlerinde gizli kalıpları ve ilişkileri tespit etme yeteneği sayesinde daha derinlemesine içgörüler elde edilmesine olanak tanır. Bu durum, akademik çalışmalarda daha zengin ve detaylı sonuçların elde edilmesini sağlar.

Yapay zekâ destekli nitel veri analizinin birçok avantajı olmasına rağmen, bazı olumsuzluklar ve etik ihlaller de göz önünde bulundurulmalıdır. Yapay zekâ algoritmalarının şeffaf olmaması, analiz sonuçlarının nasıl elde edildiğinin anlaşılmasını zorlaştırabilir ve bu da güven sorunlarına yol açabilir. Ayrıca, verilerin gizliliği ve anonimlik konularında hassasiyet gösterilmemesi durumunda, katılımcıların mahremiyeti tehlikeye girebilir. Etik ihlallerin önlenmesi için, yapay zekâ kullanımında veri gizliliğine ve anonimlik ilkelerine sıkı bir şekilde uyulmalı ve algoritmaların şeffaflığı artırılmalıdır. Bunun yanı sıra, yapay zekanın önerdiği analiz sonuçlarının mutlaka uzman araştırmacılar tarafından gözden geçirilmesi ve doğrulanması gerekmektedir.

Anahtar Kelimeler: *Yapay zeka destekli nitel veri analizi, ChatGPT, nitel veri analizi*

[1] ihtezci@akdeniz.edu.tr:, ORCID: 0000-0003-0273-8853, Akdeniz Üniversitesi, Antalya, Türkiye

[2] bayrambicak@yahoo.com:, ORCID: 0000-0003-0860-9374, Akdeniz Üniversitesi, Antalya, Türkiye

Artificial Intelligence and Qualitative Data: Future Analysis Methods

İbrahim Hakkı TEZCİ [1], Bayram BIÇAK[2]

This study focuses on how advanced AI models, such as ChatGPT, can be effectively utilized in qualitative data analysis. Within the scope of the research, interview data collected were analyzed using ChatGPT. Initially, interview transcripts were uploaded to ChatGPT, allowing for automatic summarization of the texts. Subsequently, the natural language processing capabilities were utilized to analyze the texts within the framework of specific themes and codes. ChatGPT analyzed the texts, identifying frequently recurring themes and significant phrases (codes). During this process, the researcher reviewed the AI-suggested themes, validated them with their expertise, and made necessary adjustments. Finally, the findings were visualized and reported. This study focuses on these steps and how prompts used in qualitative data analysis should be constructed. These steps demonstrate that qualitative data analysis can be conducted more systematically and objectively.

Drawing from the results obtained, AI-supported qualitative data analysis offers several advantages in academic research. First, it saves time and labor in analyzing large data sets. Processes such as automatic coding and theme development significantly reduce the time researchers spend on data analysis. Second, AI provides more objective and reproducible analyses. These analyses, which minimize human error and biases, enhance reliability. Third, AI enables the identification of hidden patterns and relationships in complex data sets, allowing for deeper insights. This leads to richer and more detailed results in academic studies.

Despite the numerous advantages of AI-supported qualitative data analysis, certain drawbacks and ethical violations must also be considered. The lack of transparency in AI algorithms can make it difficult to understand how analysis results are obtained, potentially leading to trust issues. Additionally, if sensitivity to data privacy and anonymity is not maintained, participant confidentiality may be compromised. To prevent ethical violations, strict adherence to data privacy and anonymity principles in AI usage is essential, and the transparency of algorithms should be increased. Moreover, the analysis results suggested by AI must always be reviewed and validated by expert researchers.

Keywords: *AI-Supported Qualitative Data Analysis, ChatGPT, Qualitative Data Analysis*

[1] ihtezci@akdeniz.edu.tr, ORCID: 0000-0003-0273-8853, Akdeniz University, Antalya, Türkiye

[2] bayrambicak@yahoo.com, ORCID: 0000-0003-0860-9374, Akdeniz University, Antalya, Türkiye

Yükseköğretim Kurumları Sınavı (YKS)'nda Yapay Zekâ Kullanımına İlişkin Öğrenci Görüşleri

Şeref Yiğit AKPINAR [1], Bayram ÇETİN [2]

Bu araştırmanın amacı, ortaöğretim 11. Sınıfta öğrenim görmekte olan öğrencilerin Yükseköğretim Kurumları Sınavı (YKS)'nda yapay zekâ kullanımına ilişkin görüş ve düşüncelerinin ne yönde olduğunu belirlemektir. Araştırma nitel araştırma yöntemlerinden fenomenoloji yaklaşıma göre oluşturulmuş bir çalışmadır. Araştırmada elde edilen verilerin toplanması, analiz ve yorumlanmasında nitel araştırma yöntemlerinden görüşme yöntemi kullanılmıştır. Araştırmanın çalışma grubu Kahramanmaraş ili Göksun ilçesindeki Fen Lisesi, Anadolu Lisesi ve İmam Hatip Lisesi olmak üzere toplam 3 farklı ortaöğretim okulunda 11. sınıfta öğrenim görmekte olan 12 ortaöğretim öğrencisi oluşmaktadır. Araştırmanın amacına uygun olarak maksimum çeşitleme örneklemesi ile çalışma grupları belirlenmiştir. Araştırmada nitel veri toplama aracı olarak "Yarı Yapılandırılmış Görüşme Formu" kullanılmıştır. Elde edilen verilerin analizi ise içerik analizi ile gerçekleştirilmiştir. Araştırmadan elde edilen sonuçlara göre, öğrencilerin yapay zekânın eğitimdeki rolüne ilişkin, yapay zekânın YKS gibi merkezi sınavlarda kullanımına ilişkin ve açık uçlu soruların cevaplarının okunmasının ve puanlanmasının yapay zekâ tarafından yapılmasına ilişkin genel olarak olumlu görüşlere sahip oldukları, bir kısım öğrencilerin ise olumsuz görüşlere sahip oldukları tespit edilmiştir. Öğrenciler yapay zekânın YKS gibi merkezi sınavlarda soru oluşturmada kullanılması durumunda özellikle soru konu dağılımlarının dengeli olacağını ifade etmişlerdir. Ayrıca soru zorluk seviyelerinin de dengeli olacağını belirtilmişlerdir. Açık uçlu soruların cevaplarının okunmasının ve puanlanmasının yapay zekâ tarafından yapması halinde ise objektif değerlendirmenin olacağını ifade etmişlerdir.

Anahtar Kelimeler: *Yükseköğretim Kurumları Sınavı,Yapay Zeka, Açık Uçlu Soru.*

[1] serefyigitakpinar@gmail.com, ORCID: 0009-0001-0482-0108, Milli Eğitim Bakanlığı, Kahramanmaraş, Türkiye

[2] bcetin27@gmail.com, ORCID ID: 0000-0001- 5321-8028, Gaziantep Üniversitesi, Gaziantep, Türkiye

Student Opinions on the Use of Artificial Intelligence in Higher Education Institutions Examination (YKS)

Şeref Yiğit AKPINAR [1], Bayram ÇETİN [2]

The aim of this study is to determine the opinions and thoughts of 11th grade secondary school students about the use of artificial intelligence in the Higher Education Institutions Examination (YKS). The research is a study designed according to the phenomenology approach, one of the qualitative research methods. Interview method, one of the qualitative research methods, was used to collect, analyze and interpret the data obtained in the study. The study group of the research consists of 12 secondary school students studying in the 11th grade in 3 different secondary schools, namely Science High School, Anatolian High School and Imam Hatip High School in Göksun district of Kahramanmaraş province. In accordance with the purpose of the study, the study groups were determined by maximum variation sampling. "Semi-structured Interview Form" was used as a qualitative data collection tool in the study. The analysis of the data obtained was carried out by content analysis. According to the results obtained from the research, it was determined that students generally had positive opinions about the role of artificial intelligence in education, the use of artificial intelligence in central exams such as YKS and the use of artificial intelligence in reading and scoring the answers of open-ended questions, while some students had negative opinions. Students stated that if artificial intelligence is used to create questions in central exams such as YKS, especially the distribution of question topics will be balanced. They also stated that question difficulty levels would be balanced. They also stated that if artificial intelligence is used to read and score the answers to open-ended questions, objective evaluation will be possible.

Keywords: *Higher Education Institutions Exam, Artificial Intelligence, Open-Ended Question.*

[1] serefyigitakpinar@gmail.com, ORCID: 0009-0001-0482-0108, Ministry of National Education, Kahramanmaraş, Turkey

[2] bcetin27@gmail.com, ORCID ID: 0000-0001- 5321-8028, Gaziantep University, Gaziantep, Turkey

Yükseköğretimde Üretken Yapay Zekanın Öğrenciler Tarafından Ödev ve Performans Görevlerinde Kullanımı: Avantajlar ve Riskler

Duygu Koçak [1], Ahmet Salih Şimşek[2]

Bu araştırmanın amacı, yükseköğretim kurumlarının öğrencilerin ödev, performans görevi ve araştırma raporu gibi ölçme ve değerlendirme uygulamalarında üretken yapay zeka kullanımının avantaj ve risklerini belirlemektir. Bu doğrultuda, Times Higher Education World University Rankings 2024 sıralamasına göre belirlenen eğitim kategorisindeki ilk 10 üniversitenin (University of California, Berkeley; Stanford University; University of Oxford; University of Cambridge; Harvard University; University of Hong Kong; Tsinghua University; University of Michigan-Ann Arbor; University of Toronto ve University of California, Los Angeles) öğretim üyeleri ve öğrenciler için yayınladığı üretken yapay zeka yönergeleri incelenmiştir. Yönergelerin incelenmesi sonucunda, tüm üniversitelerin yapay zekanın araştırma ve ödevlerde kullanımının etik boyutuna vurgu yaptığı gözlemlenmiştir. Bazı yönergelerde, yapay zekanın kullanımının kaçınılmaz olduğu belirtilerek, öğrencilerin bu konuda bilinçli bir şekilde eğitilmesi gerektiği vurgulanmıştır. Diğer bazı yönergelerde ise öğrencilerin yapay zeka kullanmaya özellikle teşvik edilmesi gerektiği ifade edilmiştir. Tüm yönergeler, öğrencilerin yapay zeka kullanımının önüne geçilemeyeceğini kabul etmekte ve bu doğrultuda, öğrencilerin becerilerini geliştirecek şekilde görevler verilmesi gerektiğine değinmektedir. Örneğin, sınıf içi ölçme ve değerlendirme etkinliklerinde yapay zeka kullanımına izin verilmesi, ancak ne ölçüde kullanılabileceğinin net sınırlarla belirlenmesi önerilmektedir. Ayrıca, öğrencilerin bütünüyle yapay zeka kullanamayacağı şekilde kurgulanan ödev ve görevler verilmesi tavsiye edilmektedir. Buna göre yazılı bir rapor yerine bir sunum ya da ses kaydı gibi ürünler oluşturulması, öğrencinin ödevini bütünüyle yapay zekaya yaptırmasının önüne geçebilecektir. Sonuç olarak, ödevlerde yapay zeka kullanımının önüne geçmek ve kullanılıp kullanılmadığını tespit etmek şu an için mümkün görünmemektedir. Bu nedenle incelenen yönergeler, öğrencileri yapay zekayı kullanırken etik ilkelere uymaları, yapay zekayı yardımcı kaynak, öz değerlendirme aracı ve planlayıcı olarak kullanmaları konusunda bilinçlendirilmesi gerektiğini savunmaktadır.

Anahtar Kelimeler: Üretken *yapay zeka, yüksek* öğretim*,* öğrenci ödevleri*, performans görevi.*

[1] duygu.kocak@alanya.edu.tr:, ORCID: 0000-0003-3211-0426, Alanya Alaaddin Keykubat Üniversitesi, Antalya, Türkiye

[2] asalihsimsek@gmail.com:, ORCID: 0000-0002-9764-3285, Kırşehir Ahi Evran Üniversitesi, Kırşehir, Türkiye

Use of Generative Artificial Intelligence by Students in Assignments and Performance Tasks in Higher Education: Advantages and Risks

Duygu Koçak [1], Ahmet Salih Şimşek[2],

The aim of this research is to determine the advantages and risks of using generative artificial intelligence in assessment practices such as assignments, performance tasks, and research reports in higher education institutions. In this context, the guidelines on the use of generative artificial intelligence published for faculty and students by the top 10 universities in the education category, as determined by the Times Higher Education World University Rankings 2024, were examined. These universities include the University of California, Berkeley; Stanford University; University of Oxford; University of Cambridge; Harvard University; University of Hong Kong; Tsinghua University; University of Michigan-Ann Arbor; University of Toronto; and the University of California, Los Angeles. The analysis of the guidelines revealed that all universities emphasize the ethical dimension of using artificial intelligence in research and assignments. Some guidelines highlight the inevitability of AI usage, stressing the need for students to be educated on its conscious use. Other guidelines advocate for actively encouraging students to use artificial intelligence. All guidelines acknowledge that it is impossible to prevent students from using artificial intelligence and thus recommend assigning tasks that enhance students' skills. For example, allowing the use of AI in classroom assessment activities but setting clear boundaries on the extent of its use is suggested. Additionally, it is recommended to design assignments and tasks in such a way that students cannot rely entirely on artificial intelligence. For instance, instead of a written report, assignments could involve creating a presentation or an audio recording, which would prevent students from having AI complete their work entirely. In conclusion, it currently seems impossible to prevent the use of AI in assignments and to detect whether it has been used. Therefore, the examined guidelines advocate for educating students on adhering to ethical principles while using AI, and for raising awareness about using AI as an auxiliary resource, a self- assessment tool, and a planner.[3]

Keywords: *Generative artificial intelligence, higher education, student assignments, performance tasks*

[1] duygu.kocak@alanya.edu.tr:, ORCID: 0000-0003-3211-0426, Alanya Alaaddin Keykubat University, Antalya, Turkey

[2] asalihsimsek@gmail.com:, ORCID: 0000-0002-9764-3285, Kırşehir Ahi Evran University, Kırşehir, Türkiye

Seçmeli Sorularda Cevap Değiştirme Davranışlarının Kopya İstatistiklerine Etkisinin İncelenmesi

Nuri DOĞAN [1], Meltem Yurtçu [2], Özlem Doğan[3]

Çoktan seçmeli test uygulamalarında öğrencilerin cevaplarını değiştirmesiyle ilgili çok sayıda araştırma bulunmaktadır (Reile, ve Briggs, 1952; Baştürk, 2011; Merry, Elenchin ve Surma, 2021). Bu çalışmaların neredeyse tamamı cevap değiştirmenin birey puanlarına ve testin psikometrik özelliklerine etkisi üzerinedir. Diğer yandan bu çalışmalara bakıldığında Türkiye kökenli araştırmanın yok denecek kadar sınırlı olduğu, sadece bir araştırma bulunduğu görülmüştür. Öğrencilerin cevaplarını değiştirme çalışmalarında genellikle "ilk verilen cevap genelde doğrudur" düşüncesini test etmek üzerinde odaklanılmış, değiştirilen cevapların birey veya testi alan grubun puanında artış sağlayıp sağlamadığı incelenmiş ya da test ve madde istatistiklerine bakılmıştır. Ancak öğrencilerin cevaplarını değiştirmede kopya çekme davranışının etkili olup olmadığı dolayısıyla cevap değiştirmenin kopya indekslerine etkisi incelenmemiştir.

Bu araştırmanın amacı öğrencilerin cevaplarını değiştirme davranışının kopya indekslerine (istatistiklerine) etkisini araştırmaktır.

Araştırma betimseldir. Araştırma Eğitimde Ölçme ve Değerlendirme dersini alan 5 farklı şubede kayıtlı eğitim fakültesi dışındaki fakültelerden gelen 220 formasyon öğrencisi ile gerçekleştirilmiştir. Araştırmada 27 soruluk çoktan seçmeli test kullanılmıştır. Öğrencilerin optik formları incelenerek değiştirilen cevaplar olası üç durum dikkat alınarak Kodlamalar doğrudan yanlışa, yanlıştan doğruya ve yanlıştan yanlışa şeklinde yeniden yapılmıştır.

Verilerin analizinde kopya indekslerinden standartlaştırılmış cevap benzerlik indeksi, Wollac'ın Omegası, Wesolowsy Z değeri, Belezza'nın hata benzerlik indeksi sonuçlarına etkisi dikkate alınmıştır.

Analiz sonuçlarının öğrencilerin cevaplarını değiştirmelerinin kopya indekslerini etkilemesi beklenmektedir.

Anahtar Kelimeler: *Seçmeli Sorular, kopya, cevap verme alışkanlığı*

Kaynakça

Baştürk, R. (2011). Impact of Answer-Switching Behavior on Multiple-Choice Test Scores in Higher Education. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 2*(1), 114-120.

Merry, J.W., Elenchin, M. K., & Surma, R.N. (2021). Should students change their answers on multiple choice questions? *Journal of Physiology Education*,, 45, 182-190.

Reile, P. J., & Briggs, L.J. (1952). Should students change their initial answers on objective-type tests? More evidence regarding an old problem. *Journal of Eductional Psychology*, 43, 110-115.

[1] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

[2] meltem.yurtcu@gmail.com, ORCID: 0000-0003-3303-5093, İnönü Üniversitesi, Malatya, Türkiye

[3] sudoganozlem@gmail.com, Rehber Öğretmen, MEV, Ankara, Türkiye

Investigation of the Effect of Answer Changing Behaviors on Cheating Statistics in Selective Questions

Nuri DOĞAN[1], Meltem Yurtçu[2], Özlem Doğan[3]

There are many studies on students changing their answers in multiple-choice test applications (Reile, & Briggs, 1952; Baştürk, 2011; Merry, Elenchin, & Surma, 2021). Almost all of these studies focus on the effect of changing answers on individual scores and psychometric properties of the test. On the other hand, when these studies are examined, it is seen that there is almost no research originating from Turkey and there is only one study. The studies on changing students' answers generally focused on testing the idea that "the first answer given is usually correct", examining whether the changed answers increased the scores of the individual or the group taking the test, or looking at test and item statistics. However, whether cheating behavior was effective in changing students' answers and thus the effect of changing answers on cheating indices was not examined.

The purpose of this study is to investigate the effect of changing students' answers on cheating indices (statistics).

The research is descriptive. The study was conducted with 220 formation students from faculties other than the faculty of education enrolled in 5 different branches taking the Measurement and Evaluation in Education course. A 27-question multiple-choice test was used in the study. The optical forms of the students were examined and the changed answers were re-coded by taking into consideration three possible situations: direct to false, false to true and false to false.

In the analysis of the data, the effects of copying indices such as standardized answer similarity index, Wollac's Omega, Wesolowsy Z value, Belezza's error similarity index on the results were taken into consideration.

The results of the analysis are expected to affect the cheating indices of students' changing their answers.

Keywords: *Selective Questions, cheating, answering habits*

Reference

Baştürk, R. (2011). Impact of Answer-Switching Behavior on Multiple-Choice Test Scores in Higher Education. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 2*(1), 114-120.

Merry, J.W., Elenchin, M. K., & Surma, R.N. (2021). Should students change their answers on multiple choice questions? *Journal of Physiology Education*,, 45, 182-190.

Reile, P. J., & Briggs, L.J. (1952). Should students change their initial answers on objective-type tests? More evidence regarding an old problem. *Journal of Eductional Psychology*, 43, 110-115.

[1] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

[2] meltem.yurtcu@gmail.com, ORCID: 0000-0003-3303-5093, İnönü Üniversitesi, Malatya, Türkiye

[3] sudoganozlem@gmail.com, Rehber Öğretmen, MEV, Ankara, Türkiye

Türkiye Yüzyılı Maarif Modeli Öğretim Programları Ortak Metninin Kapsayıcı Ölçme ve Değerlendirme İlkeleri Açısından İncelenmesi

Süleyman TEMUR [1]

Eğitim sistemlerinin güncelliğini koruması, sürekli yenilenen ve zamanın ruhunu ve ihtiyaçlarını yansıtan öğretim programlarıyla sağlanabilir (Önlen, Tatan & İbret, 2020). Bu doğrultuda, dinamik ve zengin toplumsal çeşitliliğe sahip olan Türkiye'nin eğitim sistemini yeniden yapılandırması ve güncel ihtiyaçlara cevap verecek şekilde uyarlanması kritik önem taşımaktadır (Ülçay, 2024). Bu bağlamda Milli Eğitim Bakanlığı (MEB), "Türkiye Yüzyılı Maarif Modeli" adı altında öğretim programlarını yenilemiş ve 2024-2025 eğitim-öğretim yılından itibaren kademeli olarak uygulamaya konulacağı kamuoyuna duyurmuştur (MEB, 2024). Bu nedenle araştırma, 2024 yılında yayınlanan "Türkiye Yüzyılı Maarif Modeli" "Öğretim Programları Ortak Metni"nin kapsayıcı ölçme ve değerlendirme ilkelerine uygunluğunu değerlendirmeyi amaçlamaktadır. Çünkü öğretim programlarının kapsayıcı olarak nitelendirilebilmesi ancak ölçme ve değerlendirme yaklaşımının da kapsayıcı olması ile mümkündür (Taneri, 2019). Araştırmada, nitel araştırma yöntemi kullanılmış ve doküman incelemesi yöntemiyle veriler toplanmıştır. Elde edilen veriler içerik analizine tabi tutulmuştur. Araştırma sonucunda "Öğretim Programları Ortak Metni"nde geçen "bireysel farklılıklar, "süreç odaklı değerlendirme", "iş birliği", "dijital teknolojilerin kullanımı", "geri bildirim, şeffaf ölçme ve değerlendirme süreci", "çeşitli ölçme ve değerlendirme araçları", "öğretmen desteği" ve "öğrenci katılımı" gibi ifadelerin doğrudan kapsayıcı ölçme ve değerlendirme ilkeleriyle örtüştüğü tespit edilmiştir. Bu sonuç, "Öğretim programları Ortak Metni"nin eğitimde fırsat eşitliği, eşit ve adil bir eğitim süreci ve tüm öğrencilerin ihtiyaçlarına cevap verme hedefleriyle kapsayıcı ölçme ve değerlendirme anlayışının örtüştüğünü göstermektedir.

Anahtar Kelimeler: *Türkiye yüzyılı maarif modeli,* öğretim *programları ortak metni, kapsayıcı* ölçme *ve değerlendirme.*

Kaynakça

Önlen, M., Tatan, M., & İbret, B. Ü. (2020). 2005-2018 sosyal bilgiler öğretim programı 5, 6 ve 7 sınıf kazanımlarının yenilenen bloom taksonomisine göre karşılaştırmalı analizi. *Türkiye Bilimsel Araştırmalar Dergisi*, *5*(1), 1-13.

Ülçay, O. (2024). Türkiye yüzyılı maarif modeli değerlendirmesi. *Ulusal Eğitim, Toplum ve Dünya Dergisi*, *1*(2), 70-75. https://doi.org/10.5281/zenodo.11097248

Milli Eğitim Bakanlığı. (2024). *Türkiye yüzyılı maarif modeli:* Öğretim *programları ortak metni*. MEB Yayınları.

Taneri, P. O. (2019). Kapsayıcı değerlendirme. İçinde P. O. Taneri (Ed.), *Kuramdan uygulamaya kapsayıcı eğitim* (s. 93-123). Pegem Akademi.

[1] temursuleyman19@gmail.com, ORCID: 0000-0002-5203-6553, Niğde Ömer Halisdemir Üniversitesi/MEB, Niğde, Türkiye.

Examination of the Türkiye Century Education Model Curriculum Common Text in terms of Inclusive Measurement and Evaluation Principles

Süleyman TEMUR [1]

Keeping education systems up-to-date can be ensured through curricula that are constantly renewed and reflect the spirit and needs of the time (Önlen, Tatan & İbret, 2020). Accordingly, it is critical for Turkey, a dynamic and richly diverse society, to restructure its education system and adapt it to meet current needs (Ülçay, 2024). In this context, the Ministry of National Education (MoNE) has renewed the curricula under the name of "Turkey Century Education Model" and announced to the public that it will be implemented gradually starting from the 2024-2025 academic year (MoNE, 2024). For this reason, this study aims to evaluate the conformity of the "Turkish Century Education Model" "Curriculum Common Text" published in 2024 with the principles of inclusive measurement and evaluation. Because curricula can only be described as inclusive if the measurement and evaluation approach is also inclusive (Taneri, 2019). In the study, qualitative research method was used and data were collected through document analysis. The data obtained were subjected to content analysis. As a result of the research, it was concluded that the expressions such as "individual differences", "process-oriented assessment", "collaboration", "use of digital technologies", "feedback, transparent assessment and evaluation process", "various assessment and evaluation tools", "teacher support" and "student participation" in the "Common Text of Curricula" are directly related to inclusive assessment and evaluation. This result shows that inclusive measurement and evaluation approach overlaps with the goals of the "Common Text of Curricula", which are equality of opportunity in education, an equal and fair education process and responding to the needs of all students.

Keywords: *Türkiye century education model, common text of curricula, inclusive assessment and evaluation.*

Reference

Önlen, M., Tatan, M., & İbret, B. Ü. (2020). 2005-2018 sosyal bilgiler öğretim programı 5, 6 ve 7 sınıf kazanımlarının yenilenen bloom taksonomisine göre karşılaştırmalı analizi. *Türkiye Bilimsel Araştırmalar Dergisi*, *5*(1), 1-13.

Ülçay, O. (2024). Türkiye yüzyılı maarif modeli değerlendirmesi. *Ulusal Eğitim, Toplum ve Dünya Dergisi*, *1*(2), 70-75. https://doi.org/10.5281/zenodo.11097248

Milli Eğitim Bakanlığı. (2024). *Türkiye yüzyılı maarif modeli:* Öğretim *programları ortak metni.* MEB Yayınları.

Taneri, P. O. (2019). Kapsayıcı değerlendirme. İçinde P. O. Taneri (Ed.), *Kuramdan uygulamaya kapsayıcı eğitim* (s. 93-123). Pegem Akademi.

[1] temursuleyman19@gmail.com, ORCID: 0000-0002-5203-6553, Niğde Ömer Halisdemir Üniversitesi/MEB, Niğde, Türkiye.

Nasıl Ölçüyor? Nasıl Değerlendiriyoruz? 4-7. Sınıf Sosyal Bilgiler Ders Kitaplarında Ölçme ve Değerlendirme Durumlarının İncelenmesi

Kadir Can ERBUDAK [1]

Eğitim öğretim sürecinde her ne kadar birçok değişken etkili olsa da en temel materyal ya da ders aracı olarak kabul edilen ders kitaplarının çağdaş ve yeterli olması oldukça önemlidir. Ders kitabı herkes için her yer de ve resmî programla başat biçimde kullanılan, belirlenen öğrenme çıktılarının aktarılmasını sağlayan önemli bir işlevi üstlenmektedir. Bu bakımdan ders kitaplarının; çağdaş, güncel, bireyi araştırmaya sevk eden, görsel olarak zengin, farklı etkinliklerle tüm bireylere hitap edici şekilde oluşturulması, yapılan eğitim öğretim faaliyetlerin bireylerin öğrenme çıktılarına ne ölçüde ulaşıp ulaşılmadığının belirlenmesinde kullanılacak öğrenme alanı sonu ve konu sonu ölçme ve değerlendirme bölümlerinin bulunması gibi niteliklere sahip olması gerekmektedir. Bu çalışmanın amacı 4-7. Sınıf sosyal bilgiler ders kitaplarında yer alan konu sonu ve öğrenme alanının sonunda yer alan ölçme ve değerlendirme durumların incelenmesi olacaktır. Çalışmada doküman incelemesi yönteminden yararlanılacak ve elde edilecek veriler içerik analizine tabii tutulacaktır. Çalışmanın verileri 2023-2024 yılında ilk ve ortaokullarda okutulan 4-7. sınıf sosyal bilgiler ders kitaplarından oluşacaktır. Çalışma; konu sonu ve öğrenme alanı sonunda yer alan ölçme ve değerlendirme kısımlarını; soruların çeşitliliği, Yenilenmiş Bloom Taksonomisine göre düzeyleri, farklı ölçme ve değerlendirme araçlarının kullanılıp kullanılmadığı, çoktan seçmeli testlerin yapıları, öğrenme alanına yönelik becerileri içerip içermediği, ölçme ve değerlendirme durumlarının yoğunluk durumu ve ölçme ve değerlendirme bölümlerinde yer alan etkinliklerin cevap anahtarı içerip içermediği ele alınıp ortaya konulacaktır. Böylelikle sosyal bilgiler ders kitaplarının ölçme ve değerlendirme durumları açıkça ortaya konulabilecektir.

Anahtar Kelimeler: *Sosyal Bilgiler Eğitimi, Sosyal Bilgiler Ders Kitapları,* Ölçme *ve Değerlendirme*

[1] kadircanerbudak@windowslive.com, ORCID: 0000-0003-4595-1179, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, TÜRKİYE

How Do We Measure? How Do We Evaluate? Examination of Measurement and Evaluation in 4th-7th Grade Social Studies Textbooks

Kadir Can ERBUDAK [1]

Although many variables are effective in the education and training process, it is very important that the textbooks, which are considered the most basic material or course tool, are modern and adequate. Textbooks undertake an important function that is used for everyone, everywhere and with the official program, and that enables the transfer of the determined learning outcomes. In this respect, textbooks should have qualities such as being contemporary, up-to-date, encouraging the individual to research, visually rich, appealing to all individuals with different activities, and having end-of-learning area and end-of-subject measurement and evaluation sections that will be used to determine the extent to which individuals have achieved the learning outcomes of the education and training activities. The aim of this study is to analyze the 4-7. The aim of this study is to examine the measurement and evaluation situations at the end of the subject and at the end of the learning area in the 4th-7th grade social studies textbooks. In the study, document analysis method will be used and the data to be obtained will be subjected to content analysis. The data of the study will consist of 4th-7th grade social studies textbooks taught in primary and secondary schools in 2023-2024. The study will examine and reveal the measurement and evaluation sections at the end of the subject and at the end of the learning area; the variety of questions, their levels according to the Revised Bloom Taxonomy, whether different measurement and evaluation tools are used, the structures of multiple-choice tests, whether they contain skills for the learning area, the intensity of measurement and evaluation situations, and whether the activities in the measurement and evaluation sections contain an answer key. In this way, the assessment and evaluation status of social studies textbooks will be clearly revealed.

Keywords: *Social Studies Education, Social Studies Textbooks, Assessment and Evaluation.*

[1] kadircanerbudak@windowslive.com, ORCID: 0000-0003-4595-1179, Center for Assessment, Selection and Placement, Ankara, TÜRKİYE

ÖABT-Din Kültürü ve Ahlak Bilgisi Öğretmenliği Testi'nde Farklılaşan Madde Fonksiyonu

Süleyman Abanoz [1]

Farklılaşan Madde Fonksiyonu-FMF, aynı yetenek seviyesinde olmakla birlikte farklı alt gruplarda yer alan bireylerin maddeyi cevaplama olasılığının farklılaşmasıdır. FMF içeren madde bireyin yeteneğinin yanlı ölçülmesine yol açabilir. Başka bir ifadeyle ölçmeyi amaçladığı değişken ile ilgisiz herhangi bir özellik nedeniyle yanıtlayıcılar arasında bulunan bir gruba, diğer gruplara göre avantaj veya dezavantaj sağlayabilir. Dolayısıyla testte FMF içeren maddelerin yer alması testin geçerliğini ve buna bağlı olarak güvenirliğini olumsuz yönde etkiler. Araştırma, KPSS-ÖABT Din Kültürü ve Ahlak Bilgisi Öğretmenliği testini 2013-2016 yılları arasında cevaplayan toplam 64810 aday ile yürütülmüştür. Araştırmanın amacı lisans eğitimi sırasında pedagojik formasyon alan ve almayan adaylar ile erkek ve kız adayların cevapladıkları maddeler içerisinde FMF içeren maddelerin olup olmadığını incelemektir. Yanlılık olduğunu tespit etmek için adayların sorulara verdikleri cevaplar "R" programı ile incelenmiş; aynı eğitim ve yetenek düzeyinde olduğu kabul edilen adaylar için farklılaşan madde olup olmadığına bakılmıştır. "Farklılaşan Madde Fonksiyonu"nun (FMF) belirlenmesi için Mantel-Haenszel, Logistic Regresyon ve Lord yöntemleri kullanılmıştır. Bu yöntemlerden en az ikisinin FMF gösterdiği maddeler araştırmaya dâhil edilmiştir. FMF dereceleri A (hafif düzey), B (orta düzey) ve C (üst düzey) olarak sıralanmıştır. Bunlardan A düzeyinde FMF gösterenler değerlendirmeye alınmamış; B ve C düzeyinde gösterenler araştırmaya dâhil edilmiştir. Cinsiyete göre FMF içeren maddelerin sayısı 8'i kızlar lehine, 5'i erkekler lehine olmak üzere toplam 13 tanedir. Pedagojik formasyon açısından FMF içeren maddeler incelendiğinde 22 maddenin lisans eğitimi sırasında pedagojik formasyon alanlar lehine, 25 maddenin ise almayanlar lehine FMF içerdiği tespit edilmiştir. Alan eğitimiyle ilgili maddelerin genellikle pedagojik formasyonu lisans eğitimi sırasında alanlar lehine FMF içermesi, pedagojik formasyonun lisans eğitimi sırasında verilmesinin daha faydalı olacağını göstermektedir. İlahiyat fakültelerinde pedagojik formasyonun lisans eğitimi sırasında verilmeye başlanması bu açıdan olumlu bir durumdur. Testlerde FMF içeren maddelerin inceler uzmanlar tarafından incelenmeli; bunların adaylar arasındaki gerçek farktan mı kaynaklandığı yoksa yanlılık mı içerdiği ortaya konmalıdır. Sonuç olarak yanlılık içeren durumlar tespit edilmeli ve bu tür maddelerin testte yer alması engellenmelidir.

Anahtar Kelimeler: *Din Kültürü ve Ahlak Bilgisi* Öğretmenliği*, pedagojik formasyon, farklılaşan madde fonksiyonu, Mantel-Haenszel, logistic regresyon.*

[1] abanoz_s@hotmail.com, ORCID: 0000-0002-4466-7080, Selçuk Üniversitesi İlahiyat Fakültesi, Konya, Türkiye

Differential Item Function in the TFKT-Religious Culture and Moral Knowledge Teaching Test

Süleyman Abanoz [1]

Differential Item Function-DIF is the difference in the probability of answering the item for individuals in different subgroups at the same ability level. An item containing DIF may lead to a biased measurement of an individual's ability. In other words, it may provide an advantage or disadvantage to one group among the respondents over the other groups due to any feature unrelated to the variable it aims to measure. Therefore, the inclusion of items containing DIF in the test negatively affects the validity and, accordingly, the reliability of the test. The research was conducted with a total of 64810 candidates who answered the PPSE-TFKT Religious Culture and Moral Knowledge Teaching Test between 2013 and 2016. The aim of the research is to examine whether there are any DIF in the items answered by male and female candidates and candidates who received and did not receive pedagogical formation during undergraduate education. In order to detect DIF, the answers given by the candidates to the questions were examined with the "R" program; it was checked whether there were any differentiating items for candidates who were considered to be at the same education and ability level. Mantel-Haenszel, Logistic Regression and Lord methods were used to determine the "Differential Item Functioning" (DIF). Items showing DIF by at least two of these methods were included in the study. The DIF grades are ranked as A (mild level), B (moderate level) and C (high level). Of these, those showing DIF at A level were not taken into consideration; those showing DIF at levels B and C were included in the study. The number of items containing DIF according to gender is 13 in total, 8 in favor of girls and 5 in favor of boys. When the items containing DIF in terms of pedagogical formation were examined, it was determined that 22 items contained DIF in favor of those who received pedagogical training during their undergraduate education and 25 items contained DIF in favor of those who did not. The fact that items related to field education are generally contain DIF in favor of those who received the pedagogical formation during undergraduate education shows that it would be more beneficial to give the pedagogical formation during undergraduate education. In this respect, it is a positive situation that pedagogical formation begins to be given during undergraduate education in faculties of theology. Items containing DIF in the tests should be examined by experts; it should be revealed whether they are due to real differences between the candidates or whether they are biased. As a result, cases containing bias should be detected and such items should be prevented from being included in the test.

Keywords: *Religious Culture and Moral Knowledge Teaching, pedagogical formation, differentiating item function, Mantel-Haenszel, logistic regression*

[1] abanoz_s@hotmail.com, ORCID: 0000-0002-4466-7080, Selçuk University Faculty of Theology, Konya, Türkiye

Okul Öncesi Eğitimde Yapay Zekâ: "Yapay Zekâ ile Kişiselleştirilmiş Ölçme ve Değerlendirme"

Gülçin Güven [1], Seher Yumugan [2]

Bu çalışmanın amacı, yapay zekanın (AI) okul öncesi eğitimde ölçme ve değerlendirme süreçlerini daha etkili, kişiselleştirilmiş ve veri odaklı hale getirme potansiyelini incelemektir. Yapay zekâ teknolojilerinin sunduğu avantajlardan yararlanarak, okul öncesi dönem çocuklarının dil gelişimi, sosyal etkileşim becerileri ve genel akademik performanslarını değerlendirmek ve bu değerlendirme süreçlerini iyileştirmek hedeflenmiştir. Çalışma, literatürdeki araştırmaları derleyerek ve analiz ederek okul öncesi eğitimde yapay zekâ destekli bir ölçme ve değerlendirme modelinin tasarımını yapmayı amaçlamaktadır.

Yapay zekâ teknolojilerinin eğitimdeki uygulamaları dünya genelinde artmasına rağmen, bu teknolojilerin okul öncesi eğitimdeki spesifik kullanım alanları ve potansiyel faydaları yeterince araştırılmadığı tespit edilmiştir. Bu çalışmanın özgünlüğü, yapay zekâ destekli değerlendirme araçlarının okul öncesi eğitime entegrasyonunu detaylandırarak, eğitimde kişiselleştirilmiş öğrenme deneyimlerinin nasıl geliştirilebileceğine dair yeni perspektifler sunmasıdır. Araştırma, nitel bir yaklaşımla yürütülmüştür ve veri toplama aracı olarak kapsamlı bir literatür taraması kullanılmıştır. Literatür taraması, yapay zekâ teknolojilerinin eğitimdeki uygulamaları, okul öncesi eğitimdeki değerlendirme yöntemleri ve yapay zekâ destekli sistemlerin potansiyel etkileri üzerine yapılan akademik çalışmaları içermektedir. Toplanan veriler, tematik analiz yöntemi ile analiz edilmiştir.

Araştırma sonuçlarının, yapay zekâ destekli ölçme ve değerlendirme modelinin, okul öncesi dönem çocuklarının dil gelişimi ve sosyal etkileşim becerilerinde anlamlı gelişmeler sağlayacağı düşünülmektedir. Ortaya konulan modelin uygulama süreci, veri toplama, veri analizi, kişiselleştirilmiş geri bildirim ve uyarlanabilir öğrenme platformlarının entegrasyonunu içermektedir. Örneğin, doğal dil işleme (NLP) teknolojisi, çocukların konuşmalarını ve yazılı çalışmalarını analiz ederek dil kullanımındaki gelişim ve becerileri tespit etmiş, görüntü tanıma teknolojisi ise, fiziksel aktiviteler ve sosyal etkileşimleri değerlendirilecektir. Analizler, makine öğrenimi algoritmaları tarafından değerlendirilecek ve her çocuğun bireysel öğrenme stilleri ve gelişim hızları belirlenecektir.

Araştırma, yapay zekâ destekli ölçme ve değerlendirme modelinin, okul öncesi eğitimde öğrencilerin bireysel ihtiyaçlarını karşılamada ve öğretim süreçlerini optimize etmede önemli bir araç olarak kullanılabileceğini örneklem üzerinden ortaya koyacaktır. Yapay zekâ teknolojilerinin eğitimde kullanımı, öğretim süreçlerinin kişiselleştirilmesi ve eğitim kalitesinin artırılması açısından büyük fırsatlar sunmaktadır.

Anahtar Kelimeler: *Yapay Zeka, Okul* Öncesi *Eğitim, Kişiselleştirilmiş Değerlendirme.*

[1] gulcinm@marmara.edu.tr:, ORCID: 0000-0002-9638-025X, Marmara Üniversitesi, İstanbul, Türkiye

[2] yumugans34@gmail.com:, ORCID: 0009-0004-8809-5012, Marmara Üniversitesi, İstanbul, Türkiye

Artificial Intelligence in Preschool Education: "Personalized Measurement and Assessment with AI"

Gülçin Güven [1], Seher Yumugan [2]

The purpose of this study is to examine the potential of artificial intelligence (AI) to make assessment and evaluation processes in preschool education more effective, personalized, and data- driven. By leveraging the advantages offered by AI technologies, the aim is to evaluate the language development, social interaction skills, and overall academic performance of preschool children and to improve these assessment processes. The study intends to design an AI-supported assessment and evaluation model in preschool education by compiling and analyzing existing research in the literature.

Despite the increasing applications of AI technologies in education worldwide, it has been found that the specific use cases and potential benefits of these technologies in preschool education have not been sufficiently researched. The originality of this study lies in detailing the integration of AI-supported assessment tools into preschool education and providing new perspectives on how personalized learning experiences in education can be developed. The research was conducted using a qualitative approach and a comprehensive literature review was used as the data collection tool. The literature review includes academic studies on the applications of AI technologies in education, assessment methods in preschool education, and the potential impacts of AI-supported systems. The collected data were analyzed using thematic analysis.

The results of the research suggest that the AI-supported assessment and evaluation model will provide significant improvements in the language development and social interaction skills of preschool children. The implementation process of the proposed model includes data collection, data analysis, personalized feedback, and the integration of adaptive learning platforms. For instance, natural language processing (NLP) technology will analyze children's speech and written works to identify developments and skills in language use, while image recognition technology will evaluate physical activities and social interactions. The analyses will be assessed by machine learning algorithms to determine each child's individual learning styles and development rates.

The research will demonstrate through a sample that the AI-supported assessment and evaluation model can be used as an important tool in meeting the individual needs of students and optimizing teaching processes in preschool education. The use of AI technologies in education offers great opportunities in terms of personalizing teaching processes and improving the quality of education.

Keywords: *Artificial Intelligence, Preschool Education, Personalized Assessment.*

[1] gulcinm@marmara.edu.tr:, ORCID: 0000-0002-9638-025X, Marmara Üniversitesi, İstanbul, Türkiye

[2] yumugans34@gmail.com:, ORCID: 0009-0004-8809-5012, Marmara Üniversitesi, İstanbul, Türkiye

NLP Tabanlı Açık Uçlu Soruların Otomatik Değerlendirme Sistemi: "T.C. İnkılap Tarihi ve Atatürkçülük Dersi için Bir Uygulama"

Oğuzhan Sakarya [1]

Günümüzde eğitimde yapay zeka ve doğal dil işleme (NLP) teknolojilerinin kullanımı giderek artmaktadır. Bu teknolojiler, öğretim ve değerlendirme süreçlerini daha verimli ve etkili hale getirme potansiyeline sahiptir. T.C. İnkılap Tarihi ve Atatürkçülük dersi, tarihsel olayların anlaşılması ve yorumlanması üzerine odaklandığı için açık uçlu soruların önemli olduğu bir derstir. Bu makale, T.C. İnkılap Tarihi ve Atatürkçülük dersinde NLP tabanlı bir sistemin, açık uçlu soruların otomatik olarak değerlendirilmesi için nasıl kullanılabileceğini incelemektedir.

Literatür taramasında, doğal dil işleme ve yapay zeka tabanlı sistemlerin eğitimde kullanımı üzerine yapılan çalışmalar incelenmiştir. Mevcut araştırmalar, bu teknolojilerin öğretmenlerin iş yükünü azaltma, değerlendirme süreçlerinde tutarlılığı artırma ve öğrencilere daha hızlı geri bildirim sağlama potansiyeline sahip olduğunu göstermektedir. Ancak, özellikle Türkçe dilinde NLP uygulamalarının sınırlı olması ve dilin karmaşıklığı gibi zorluklar da mevcuttur.

Bu çalışmada, NLP tabanlı sistemin uygulanabilirliği bir örnek soru ve cevap analizi ile gösterilmiştir. Örnek kod oluşturularak, öğrenci cevaplarını analiz etmek ve değerlendirmek için spaCy ve Hugging Face Transformers kütüphanelerini kullanılmıştır. Sistem, öğrenci cevaplarını doğru ve tutarlı bir şekilde değerlendirerek öğretmenlerin iş yükünü hafifletmekte ve öğrencilere hızlı geri bildirim sağlamak dışında ölçme ve değerlendirmede insan hatasını da minimuma indirmektedir. Mustafa Kemal Atatürk'ün liderlik özellikleri ve Millî Mücadele sürecindeki rolünü açıklayan bir öğrenci cevabı, doğruluk ve içerik uyumu, analiz ve yorumlama, dil kullanımı kriterlerine göre değerlendirilmiş ve puanlanmıştır.

NLP tabanlı değerlendirme sistemlerinin avantajları arasında öğretmenlerin değerlendirme sürecini hızlandırması, tutarlılık sağlaması, öğrencilere anlık geri bildirim sunması ve öğrencilerin gelişimlerini izleme imkanı sunması yer almaktadır. Ancak, Türkçe dilinde doğal dil işleme zorlukları, geniş ve çeşitli veri setlerine ihtiyaç duyulması ve cevapların doğru bağlamda ve anlamda değerlendirilmesi gibi zorluklar da bulunmaktadır.

Sonuç olarak, NLP tabanlı otomatik değerlendirme sistemleri, T.C. İnkılap Tarihi ve Atatürkçülük dersinde önemli avantajlar sunmaktadır. Bu sistemlerin gelecekte eğitimde daha yaygın bir şekilde kullanılması ve geliştirilmesi, öğretim süreçlerini daha verimli ve etkili hale getirebilir

Anahtar Kelimeler: *NLP tabanlı otomatik değerlendirme sistemleri, Tarih dersi,* Ölçme *ve değerlendirme*

[1] osaka.26060@gmail.com:, ORCID: 0009-0000-0841-4249, Anadolu Üniversitesi, Eskişehir, Türkiye

An NLP-Based Automatic Evaluation System for Open-Ended Questions: An Application for the Course on the T.C. İnkılap Tarihi ve Atatürkçülük

Oğuzhan Sakarya [1]

The use of artificial intelligence and natural language processing (NLP) technologies in education is increasing rapidly today. These technologies have the potential to make teaching and assessment processes more efficient and effective. The course on the History of the T.C. İnkılap Tarihi ve Atatürkçülük focuses on understanding and interpreting historical events, making open-ended questions crucial for this subject. This article examines how an NLP-based system can be used to automatically evaluate open-ended questions in this course.

A review of the literature on the use of natural language processing and artificial intelligence- based systems in education has been conducted. Existing studies show that these technologies have the potential to reduce teachers' workload, increase consistency in assessment processes, and provide faster feedback to students. However, challenges such as the limited applications of NLP in the Turkish language and the complexity of the language are also present.

In this study, the applicability of an NLP-based system is demonstrated through the analysis of a sample question and answer. Using the spaCy and Hugging Face Transformers libraries, a sample code was created to analyze and evaluate student responses. The system accurately and consistently evaluates student answers, reducing teachers' workload and providing quick feedback to students, while also minimizing human error in assessment. A student response explaining Mustafa Kemal Atatürk's leadership qualities and his role in the National Struggle was evaluated and scored based on criteria of accuracy and content alignment, analysis and interpretation, and language use.

The advantages of NLP-based evaluation systems include accelerating the assessment process for teachers, ensuring consistency, providing instant feedback to students, and offering the ability to monitor students' progress. However, challenges such as the difficulties of natural language processing in Turkish, the need for extensive and diverse datasets, and the correct contextual and semantic evaluation of responses are also noted.

NLP-based automatic evaluation systems offer significant advantages in the course on the T.C. İnkılap Tarihi ve Atatürkçülük. The future widespread use and development of these systems in education could make teaching processes more efficient and effective.

Keywords: *NLP-based automatic evaluation systems, History course, Assessment and evaluation*

[1] osaka.26060@gmail.com:, ORCID: 0009-0000-0841-4249, Anadolu Üniversitesi, Eskişehir, Türkiye

Yükseköğretim Programlarına Yerleşen Öğrencilerin Yeterlik Düzeylerinin Tanımlanması için Model ve Yazılım Geliştirilmesi

İlker Kalender [1]

Üniversite giriş sisteminde kullanılan alt testlere verilen yanıtlar ve öğrencilerin puanlar yükseköğretim programlarına yerleştirme bakımından başlıca kriter durumundadır. Fakat hesaplanan bu puanlar üstünde öğrencilerin sahip oldukları bilgi ve beceriler ile zihinsel süreçler hakkında ya da puanlara karşılık gelen gözlemlenebilir davranışlar cinsinden bir çözümleme yapılmamaktadır. Mevcut durumda, öğrenciler için hesaplanan puanlar öğrencileri sıralamak amacı ile kullanılmaktadır. Uluslararası geniş ölçekli çalışmalarda sıklıkla yeterlik düzeyleri tanımları kullanılmakta ve böylelikle öğrenciler hakkında puanlar yanında daha geniş kapsamlı bilgiler elde edilebilmektedir. Bu tür yeterlik düzey tanımlarının belirlenmesi hem yükseköğretim hem de ortaöğretim kurumları açısından önemli bilgiler sağlama potansiyeli taşımaktadır: (i) üniversiteler programlarına yerleşen adaylar hakkında sadece puanlar bazında değil, gözlemlenebilir davranışlar şeklinde tanımlanacak yeterlik düzeylerine sahip olabilecekler, (ii) üniversiteler öğretim programlarını tasarlarken ve/veya güncellerken ek bir bilgi kaynağı elde edecekler, (iii) yeterlik düzeyi tanımlanması farklı üniversitelerin aynı programlarına yerleşen öğrencilerin farklılıklarını ortaya koyacaktır, (iv) ÖSYM'nin ortaöğretimden yükseköğretime geçişteki rolü açısından çok daha anlamlı ve detaylı bilgiler ortaya konabilecektir. Bu çalışma kapsamında araştırmacıya daha önce ÖSYM tarafından sağlanan bir veri seti kullanılarak madde haritalama (Karantonis ve Sireci, 2006) yaklaşımları kullanılmıştır. Çalışmanın ilk aşamasında standart oluşturma çalışması yapılmıştır. Bu aşamadaki analizler Madde Tepki Kuramı modelleri kullanılarak yapılmıştır (Baker ve Kim, 2004). Standart tanımlama aşamasında yeterlik düzeylerini tanımlayan puan aralıkları ve kesme puanları belirlenmiş ve ardından tanımlanan puan aralıkları kapsamındaki yeterlik düzeyleri ortaya konmuştur. Sonrasında ise soruların ölçtükleri kazanımlar tanımlanmıştır. Bu aşamada sorular hedeflediği kazanımlar ve madde parametreleri de dikkate alınmıştır. Soruların için elde edilecek olasılık dağılımları kullanılarak soruların farklı bilgi, beceri ve zihinsel süreç ölçmeye bağladığı noktalar tespit edilmiş ve ardından yeterlikler tanımlanmıştır. Yeterlik tanımları yapılırken soruların içerikleri incelenerek her sorunun ölçmeyi hedeflediği kazanımlar ortaya konmuştur. Bu çalışma araştırmacı tarafından geliştirilmiş bir model ve bu modeli uygulamaya yönelik geliştirdiği yazılım ile yapılmıştır. Model ve yazılım yapılacak sunumda paylaşılacaktır.

Anahtar Kelimeler: *yeterlik düzeyleri, madde haritalama, yükseköğretime geçiş*

Kaynakça

Baker, B. B., & Kim, S. (2004). Item response theory: Parameter estimation techniques. New York, NY: Marcel-Dekker.

Karantonis, A & Sireci, S. G. (2006). The bookmark standard setting method: A literature review. *Educational Measurement: Issues and Practice, 25*(1),4-12

[1] kalenderi@bilkent.edu.tr, ORCID: 0000-0003-1282-4149, İhsan Doğramacı Bilkent Üniversitesi, Ankara, Türkiye

Development of a Model and Software for Defining Proficiency Levels of Students Admitted to Higher Education

İlker Kalender [1]

The answers given to the subtests used in the university entrance system and the scores of the students are the main criteria for placement in higher education programs. However, no analysis is made on these calculated scores regarding the knowledge and skills that the students have, their mental processes or the observable behaviors corresponding to the scores. Currently, the scores calculated for the students are only used to rank the students. In large-scale international studies, definitions of proficiency levels are frequently used, and thus, more comprehensive information about the students can be obtained in addition to the scores. Defining proficiency levels has the potential to provide important information for both higher education and secondary education institutions: (i) Candidates placed in university programs will have proficiency levels that will be defined not only on the basis of scores but also in terms of observable behaviors, (ii) universities may obtain an additional source of information when designing and/or updating their curriculum, (iii) definition of proficiency levels will reveal the differences between students placed in the same programs of different universities, (iv) more meaningful and detailed information can be revealed in terms of the role of ÖSYM in the transition from secondary education to higher education. Within the scope of this study, item mapping (Karantonis and Sireci, 2006) approaches were used using a data set previously provided to the researcher by ÖSYM. In the first stage of the study, a standardization analysis was conducted. The analyses at this stage were conducted using Item Response Theory models (Baker and Kim, 2004). In the standard definition stage, the score ranges and cut-off scores defining the proficiency levels were determined, and then the proficiency levels within the defined score ranges were revealed. Then, the gains measured by the questions will be defined. At this stage, the outcomes targeted by the questions and the item parameters were also taken into consideration. Using the probability distributions to be obtained for the questions, the points that the questions connected to measuring different knowledge, skills, and mental processes were determined, and then the competencies were defined. While defining the competencies, the contents of the questions were examined, and the outcomes that each question aimed to measure were revealed.

This study was conducted based on a model developed by the researcher and the software developed to implement this model. The model and software will be shared in the presentation to be made.

Keywords: *Proficiency levels, Item mapping, transition to higher edcuation*

References

Baker, B. B., & Kim, S. (2004). Item response theory: Parameter estimation techniques. New York, NY: Marcel-Dekker.

Karantonis, A & Sireci, S.G. (2006). The bookmark standard setting method: A literature review. *Educational Measurement: Issues and Practice, 25*(1),4-12.

[1] kalenderi@bilkent.edu.tr, ORCID0000-0003-1282-4149, İhsan Doğramacı Bilkent University, Ankara, Türkiye

Açık Uçlu Maddelerin Farklı Branşlardaki Puanlayıcılar Tarafından Puanlanmasına Karışan Hatalar ve Puanlayıcılar Arasındaki Uyumun İncelenmesi

İsmail KARAKAYA [1], Hanifi ÖZDEMİR [2], Gökhan GÜMÜŞ[3]

Puanlama sürecinde puanlayıcılardan kaynaklanan çeşitli hatalar ölçmeyi dolayısıyla değerlendirmeyi etkileyebilir. Bu durumda da puanlayıcılar tarafından elde edilen sonuçlar öğrencilerin gerçek yetenek düzeyini yansıtma açısından sınırlı olabilir. Puanlama sürecinde puanlayıcı katılığı-cömertliği, halo etkisi, merkezi yığılma etkisi gibi hata kaynakları ortaya çıkabilir (Saal, Downey & Lahey, 1980). Puanlama sürecinde rubrik olarak bilinen dereceli puanlama anahtarları kullanılarak oluşabilecek hataların azaltılabileceği söylenilebilir (Kutlu, Doğan & Karakaya, 2010). Bu çalışmada puanlayıcı hatalarının incelenmesi amacıyla Türkiye'de 2023 – 2024 Eğitim ve Öğretim Yılı sürecinde uygulanmaya başlanan ortak yazılı sınavlar üzerinden ele alınması elde edilen öğrenci puanlarının geçerliliği ve güvenilirliği konusunda önemli bir çıkarımda bulunulmasını sağlayabilir. Bu açıdan puanlama hatalarının azaltılması sürecinde puanlayıcıların niteliklerinin puanlama sürecine olan etkisinin ortaya konması önemlidir. Bu kapsamda temel araştırma yöntemiyle sürdürülen bu araştırmada puanlayıcı davranışlarına yönelik literatüre katkı sunulması hedeflenmiştir. Bu kapsamda Milli Eğitim Bakanlığına bağlı bir devlet ortaokulunda 6. sınıfta öğrenim gören 39 öğrencinin ikinci dönem matematik MEB ortak yazılı sınavı matematik ve fen bilimleri öğretmenleriyle; ilköğretim matematik öğretmenliği 4. sınıfında öğrenim gören aday öğretmenlere sınavlar puanlatılmıştır. Elde edilen sonuçlar Krippendorf Alfa Katsayısı ve Çok Yüzeyli Rasch Modelleri kullanılarak incelenmiştir. Elde edilen sonuçlara göre puanlama sürecinde katılımcıların çeşitli puanlama hataları ile karşılaştıkları görülmektedir.

Anahtar Kelimeler: *Açık uçlu soruların puanlanması, puanlayıcı davranışları, puanlayıcı hataları*

Kaynakça

Kutlu, Ö., Doğan, C.D. & Karakaya, İ. (2010). Öğrenci *başarısının belirlenmesi performansa ve portfolyoya dayalı durum belirleme.* Ankara: Pegem Akademi Yayınları.

Murphy, K. R. & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of applied psychology*, *74*(4), 619-624

Saal, F. E., Downey, R. G. & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological bulletin*, *88*(2), 413–428. https://doi.org/10.1037/0033- 2909.88.2.413

Wang, B. (2010). On Rater Agreement and Rater Training. *English language teaching*, *3*(1), 108-112.

[1] ikarakaya2002@gmail.com, ORCID: 0000-0003-4308-6919, Gazi Üniversitesi, Ankara, Türkiye

[2] hanifi_ozdemir@yahoo.com, ORCID: 0009-0005-4072-0197, Milli Eğitim Bakanlığı, Malatya, Türkiye

[3] gokhangumus2357@gmail.com, ORCID: 0000-0002-4161-0425, Hatay Mustafa Kemal Üniversitesi, Hatay, Türkiye

Errors in the Scoring of Open-Ended Items by Raters in Different Disciplines and Examining the Agreement Between Raters

İsmail KARAKAYA [1], Hanifi ÖZDEMİR [2], Gökhan GÜMÜŞ[3]

Various errors caused by raters during the scoring process can affect measurement, and consequently, evaluation. In such cases, the results obtained by raters may be limited in accurately reflecting the true ability levels of students. Sources of error such as rater severity-leniency, halo effect, and central tendency bias can emerge during the scoring process (Saal, Downey & Lahey, 1980). It can be suggested that the use of rubrics, or scoring guides, can reduce potential errors during the scoring process (Kutlu, Doğan & Karakaya, 2010). In this study, the examination of rater errors through the common written exams implemented in Turkey during the 2023–2024 academic year could provide significant insights into the validity and reliability of student scores. In this context, highlighting the impact of rater qualifications on reducing scoring errors is crucial. This study, conducted using a basic research method, aims to contribute to the literature on rater behavior. In this context, the second-term mathematics common written exam of 39 sixth-grade students in a public middle school affiliated with the Ministry of National Education was scored by mathematics and science teachers, as well as by pre- service teachers in their fourth year of elementary mathematics teaching. The results were analyzed using Krippendorff's Alpha Coefficient and Multi-Faceted Rasch Models. The findings indicate that participants encountered various scoring errors during the scoring process.

Keywords: *Scoring of Open-Ended Questions, Scorer Behaviors, Scorer Errors*

Kaynakça

Kutlu, Ö., Doğan, C.D. & Karakaya, İ. (2010). Öğrenci *başarısının belirlenmesi performansa ve portfolyoya dayalı durum belirleme.* Ankara: Pegem Akademi Yayınları.

Murphy, K. R. & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of applied psychology*, *74*(4), 619-624

Saal, F. E., Downey, R. G. & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological bulletin*, *88*(2), 413–428. https://doi.org/10.1037/0033- 2909.88.2.413

Wang, B. (2010). On Rater Agreement and Rater Training. *English language teaching*, *3*(1), 108-112.

[1] ikarakaya2002@gmail.com, ORCID: 0000-0003-4308-6919, Gazi Üniversitesi, Ankara, Türkiye

[2] hanifi_ozdemir@yahoo.com, ORCID: 0009-0005-4072-0197, Milli Eğitim Bakanlığı, Malatya, Türkiye

[3] gokhangumus2357@gmail.com, ORCID: 0000-0002-4161-0425, Hatay Mustafa Kemal Üniversitesi, Hatay, Türkiye

Madde Takımlarından Oluşan Testlerde Parametre Kestirimi, Sınıflama Doğruluğu ve Değişen Madde Grubu Fonksiyonunun İncelenmesi

Kübra ATALAY KABASAKAL [1], Sebahat GÖREN [2]

Madde takımları; metin, grafik, tablo gibi ortak bir uyarana bağlı birden çok maddenin yer aldığı madde grupları/kümeleridir (Wainer & Kiely, 1987). Madde takımlarının bu yapısı sayesinde ortak bir içeriğe bağlı birden fazla soru sorularak zaman ve çaba açısından ekonomiklik sağlanabilir (Ho & Dodd, 2012; Wainer & Wang, 2000). Fakat ortak bir uyaran paylaşmalarından dolayı madde yanıtlarının birbiriyle ilişkili olma ihtimali oldukça yüksektir ve bu durum Madde Tepki Kuramı (MTK) 'nın yerel bağımsızlık varsayımını ihlal ederek madde takımlarında yer alan maddeler arası yerel bağımlılığa neden olur (Baghaei & Ravand, 2016; Wainer vd., 2007). Bu nedenle bu çalışmada madde takımlarından kaynaklı yerel bağımlılığın madde ve birey parametre kestirimi, sınıflama doğruluğu ve değişen madde/madde demeti fonksiyonu üzerindeki etkilerini değerlendirmek için MTK ve Madde Takımı Tepki Kuramı (MTTK) modelleri kullanılmış ve elde edilen sonuçlar karşılaştırmalı olarak incelenmiştir. TIMSS 2019 uygulamasında 3. ve 4. kitapçıkta yer alan 14. matematik madde kümesindeki madde takımlarına verilen cevaplar R yazılımındaki "mirt" paketi kullanılarak analiz edilmiştir. Madde takımlarındaki yerel bağımlılık derecesine bağlı olarak (yüksek-orta-düşük) MTTK modelinin daha iyi ya da MTK modeline eşdeğer bir performans göstermesi beklenmektedir.

Anahtar Kelimeler: *Madde takımı, madde takımı tepki kuramı, parametre kestirimi, sınıflama doğruluğu, değişen madde demeti fonksiyonu.*

Kaynakça

Baghaei, P. & Ravand, H. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicológica, 37*(1), 85-104.

Ho, T. H., & Dodd, B. G. (2012). Item selection and ability estimation procedures for a mixed-format adaptive test. *Applied Measurement in Education*, *25*(4), 305-326.

https://doi.org/10.1080/08957347.2012.714686

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*.

Cambridge University Press.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets.

Journal of Educational Measurement, 24, 185–201

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement, 37*(3), 203-220.

[1] kkatalay@gmail.com, ORCID: 0000-0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

[2] sebahatgoren@gmail.com, ORCID: https://orcid.org/0000-0002-6453-3258, Hacettepe Üniversitesi,Ankara, Türkiye

Examination of Parameter Estimation, Classification Accuracy and Differential Item Bundle Functioning in Testlet Based Tests

Kübra ATALAY KABASAKAL [1], Sebahat GÖREN [2]

Testlets are groups/clusters of items that contain multiple items related to a common stimulus such as text, graph, or table (Wainer & Kiely, 1987). Thanks to this structure of testlets, more than one question related to a common content can be asked, thus providing economy in terms of time and effort (Ho & Dodd, 2012; Wainer & Wang, 2000). However, since they share a common stimulus, the probability of item responses being related to each other is quite high, and this violates the local independence assumption of Item Response Theory (IRT), causing local dependency among the items in the testlets (Baghaei & Ravand, 2016; Wainer et al., 2007). Therefore, in this study, IRT and Testlet Response Theory (TRT) models were used to evaluate the effects of local dependency originating from testlets on item and individual parameter estimation, classification accuracy, and differential item/item bundle functioning and the obtained results were examined comparatively. The responses to the testlet in the 14th mathematics block in the TIMSS 2019 3rd and 4th booklets were analyzed using the "mirt" packages in the R software. Depending on the degree of local dependency in the testlets (high-moderate-low), the TRT model is expected to perform better or equivalent to the IRT model.

Keywords: *Testlet, testlet response theory, parameter estimation, classification accuracy, differential item bundle function.*

References

Baghaei, P., & Ravand, H. (2016). Modeling local item dependence in cloze and reading comprehension test items using testlet response theory. *Psicológica, 37*(1), 85-104.

Ho, T. H., & Dodd, B. G. (2012). Item selection and ability estimation procedures for a mixed-format adaptive test. *Applied Measurement in Education*, *25*(4), 305-326.

https://doi.org/10.1080/08957347.2012.714686

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*.

Cambridge University Press.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets.

Journal of Educational Measurement, 24, 185–201

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement, 37*(3), 203-220.

[1] kkatalay@gmail.com, ORCID: 0000-0002-3580-5568, Hacettepe University, Ankara, Türkiye

2 sebahatgoren@gmail.com, ORCID: https://orcid.org/0000-0002-6453-3258, Hacettepe University, Ankara, Türkiye

Karşılaştırma Yargılarıyla Puanlamada Uyarlamalı Eşleme ve Standart Hata Durdurma Kuralı: Bir Monte Carlo Simülasyonu

Sungur Gürel[1], Murat Doğan Şahin[2], İbrahim Uysal[3], Ali İhsan İbileme[4], Tuba Gündüz[5]

Performansa dayalı becerilerin ölçülmesinde geleneksel anlayış, bireyin performansının iki ya da daha fazla puanlayıcı tarafından bir rubrik referans alınarak puanlanmasına dayanır. Ancak puanlayıcılar arası uyumu sağlamak zordur. Özellikle performansın bütünsel şekilde ölçülmesi gerektiğinde karşılaştırma yargılarıyla puanlama (KYP) metodundan yararlanılabilir. KYP'de puanlayıcının karşısına belirli bir kurala göre eşlenmiş iki farklı performans gelir; puanlayıcının yapması gereken ise bu iki performanstan hangisinin daha iyi olduğuna karar vermektir (Pollit, 2012). Bir dizi ikili karşılaştırmaya tabi tutulan performansların başarı seviyeleri BTL model (Bradley & Terry, 1952; Luce, 1959) ile belirlenir. Karşılaştırılacak performans çiftlerinin nasıl belirleneceği ve karşılaştırmaların ne zaman sonlanacağı KYP'nin önemli bileşenleri arasındadır. Bu araştırmanın amacı KYP'de eşlemenin uyarlamalı, sonlandırmanın ise standart hataya (SH) dayalı olarak gerçekleştirildiği durumda performansın ne kadar iyi puanlandığını belirlemektir. 250, 500 ve 1000 örneklem büyüklüklerinde uyarlamalı eşleme metodu ile ve SH'ye (.30,.35 ve.40) dayalı sonlandırma kuralının kullanıldığı toplam 9 koşul simüle edilecek ve koşulların karşılaştırılmasında güvenirlik (SSR), sıra korelasyonu, gerçek ve kestirilen yetenekler arasındaki korelasyon ve performans başına düşen karşılaştırma sayısı esas alınacaktır. Tüm koşullarda 50 replikasyon yapılması planlanmaktadır. Örneklemin büyümesiyle aynı SH koşulunu sağlamak için daha fazla karşılaştırma yapılması, aynı örneklem büyüklüğünde ise daha küçük SH koşulunu sağlamak için daha fazla ikili karşılaştırma yapılması beklenen durumlardır. Bununla birlikte simülasyon çalışması sonucunda temel olarak, her bir örneklem büyüklüğünde farklı SH koşulunu sağlamak için gerekli olan karşılaştırma sayısı ve elde edilmesi olası güvenirlik istatistiklerinin belirlenmesiyle, gerçek uygulamaların planlanmasına rehberlik edilmesi amaçlanmaktadır.

Anahtar Kelimeler: *Karşılaştırma yargılarıyla puanlama, Monte Carlo simülasyon, Uyarlamalı eşleme*

Kaynakça

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: 1. The method of paired comparisons. *Biometrika, 39*, 324–345.

Luce, R. D. (1959). *Individual choice behaviours: A theoretical analysis*. New York, NY: Wiley.

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice, 19* (3), 281-30

1 s.gurel@siirt.edu.tr, ORCID: 0000-0003-3425-858X, Siirt Üniversitesi, Siirt, Türkiye

2 mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu Üniversitesi, Eskişehir, Türkiye

3 ibrahimuysal@ibu.edu.tr, ORCID: 0000-0002-6767-0362, Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

4 aliihsan@eskisehir.edu.tr, ORCID: 0000-0002-4705-3973, Eskişehir Teknik Üniversitesi, Eskişehir, Türkiye

5 tuba.gunduz@osym.gov.tr, ORCID: 0000-0002-0921-9290, ÖSYM, Ankara, Türkiye

Adaptive Matching and Standard Error Stopping Rule in Comparative Judgement: A Monte Carlo Simulation

Sungur Gürel[1], Murat Doğan Şahin[2], İbrahim Uysal[3], Ali İhsan İbileme[4], Tuba Gündüz[5]

In the measurement of performance-based skills, the traditional approach relies on scoring an individual's performance by two or more raters using a rubric as a reference. However, ensuring inter-rater reliability is challenging. Particularly when performance needs to be assessed holistically, the comparative judgement (CJ) method can be utilized. In CJ, two different performances are paired according to a specific rule and presented to the rater; the rater's task is to decide which of the two performances is better (Pollit, 2012). The achievement levels of performances subjected to a series of pairwise comparisons are determined using the BTL model (Bradley & Terry, 1952; Luce, 1959). How the pairs of performances to be compared are determined and when the comparisons should be terminated are among the important components of CJ. The purpose of this research is to determine how well performance is scored when matching in CJ is adaptive, and the termination is based on the standard error (SE). A total of 9 conditions will be simulated using the adaptive matching method and a termination rule based on SE (.30,.35, and.40) with sample sizes of 250, 500, and 1000. Reliability (SSR), rank correlation, correlation between true and estimated abilities, and the number of comparisons per performance will be considered for the comparison of conditions. It is planned to conduct 50 replications in all conditions. As the sample size increases, more comparisons are expected to be required to achieve the same SE condition, and more pairwise comparisons are expected to be necessary to achieve a smaller SE condition within the same sample size. Additionally, the primary goal of the simulation study is to guide the planning of real applications by determining the number of comparisons required to meet different SE conditions at each sample size and the reliability statistics likely to be obtained.

Keywords: *Comparative judgement, Monte Carlo simulation, Adaptive matching*

References

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: 1. The method of paired comparisons. *Biometrika, 39*, 324–345.

Luce, R. D. (1959). *Individual choice behaviours: A theoretical analysis.* New York, NY: Wiley.

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice, 19* (3), 281-300.

1 s.gurel@siirt.edu.tr, ORCID: 0000-0003-3425-858X, Siirt University, Siirt, Türkiye

2 mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu University, Eskişehir, Türkiye

3 ibrahimuysal@ibu.edu.tr, ORCID: 0000-0002-6767-0362, Bolu Abant İzzet Baysal University, Bolu, Türkiye

4 aliihsan@eskisehir.edu.tr, ORCID: 0000-0002-4705-3973, Eskişehir Technical University, Eskişehir, Türkiye

5 tuba.gunduz@osym.gov.tr, ORCID: 0000-0002-0921-9290, ÖSYM, Ankara, Türkiye

Büyük Gruplarda Karşılaştırma Yargılarıyla Puanlama: Referans Set Temelli Bir Yaklaşım

Murat Doğan Şahin[1], Sungur Gürel[2], İbrahim Uysal[3], Ali İhsan İbileme[4]

Karşılaştırma yargılarıyla puanlama (KYP), puanlayıcının eşleştirilmiş iki performanstan hangisinin daha üstün olduğunu belirlediği bir yöntem olup, bireysel performansların bütünsel değerlendirilmesinde giderek daha fazla kullanılmaktadır. (Bramley & Vitello, 2018). KYP ile puanlamada amaç, mümkün olduğunca düşük karşılaştırma sayısıyla yüksek güvenirlik (SSR) değerleri elde edebilmektir. Nispeten küçük örneklemlerde bile.85 üzerinde bir SSR için her bir performansın 30 civarında karşılaştırmaya dahil olması gerekmektedir (Crompvoets et al., 2020); bu durum, büyük örneklemlerde KYP uygulamasının kullanışlılığını düşürmektedir. Bu araştırma kapsamında büyük örneklemlerde, daha az karşılaştırma ile yüksek güvenirlik değerlerini mümkün kılabilecek referans set temelli KYP (RST-KYP) modeli önerilmektedir. Gerçekleştirilecek olan Monte Carlo simülasyon çalışması ile önerilen modelin etkililiği belirlenecektir. KYP'de performanslar birbiriyle belirli bir kurala göre eşlenir. Mevcut çalışmada kullanılacak RST-KYP'de, KYP'lerin uygulamalarına kıyasla büyük bir gruptan (2500), rassal olarak 200'ü seçilerek referans set oluşturulacaktır. Bu referans set KYP'lerde kullanılan uyarlamalı eşleme ve standart hataya dayalı (.30) sonlandırma kuralına göre puanlanacak, sonuçta referans set için elde edilen kestirimlerin.90 üzeri güvenirlikte olması sağlanacaktır. Sonraki aşamada ise ana set (kalan 2300 performans) sürekli olarak referans set (200 performans) ile eşlenerek puanlanacaktır. Bu basamakta da uyarlamalı eşleme ve standart hataya dayalı (.30) sonlandırma kuralı uygulanacaktır. Elde edilen sonuçlar, 2500 performansta geleneksel KYP'nin uygulanmasıyla elde edilen sonuçlarla karşılaştırılacaktır. KYP ile RST-KYP uygulamasının performanslarının karşılaştırılmasında güvenirlik (SSR), sıra korelasyonu, gerçek ve kestirilen yetenekler arasındaki korelasyon ve performans başına düşen karşılaştırma sayısı esas alınacaktır. RST-KYP'nin, KYP ile benzer güvenirlik düzeyine daha az ikili karşılaştırma sonucunda ulaşması beklenmektedir.

Anahtar Kelimeler: *Karşılaştırma yargılarıyla puanlama, Monte Carlo simülasyon, Bütünsel değerlendirme*

Kaynakça

Bramley, T., & Vitello, S. (2018). The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 1–16. doi:10.1080/0969594X.2017.1418734.

Crompvoets, E. A. V., Béguin, A. A., & Sijtsma, K. (2020). Adaptive Pairwise Comparison for Educational Measurement. J. Educ. Behav. Stat. 45, 316–338. doi:10.3102/1076998619890589

[1] mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu Üniversitesi, Eskişehir, Türkiye

[2] s.gurel@siirt.edu.tr, ORCID: 0000-0003-3425-858X, Siirt Üniversitesi, Siirt, Türkiye

[3] ibrahimuysal@ibu.edu.tr, ORCID: 0000-0002-6767-0362, Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

[4] aliihsan@eskisehir.edu.tr, ORCID: 0000-0002-4705-3973, Eskişehir Teknik Üniversitesi, Eskişehir, Türkiye

Using Comparative Judgement in Large Groups: A Reference Set-Based Approach

Murat Doğan Şahin[1], Sungur Gürel[2], İbrahim Uysal[3], Ali İhsan İbileme[4]

Comparative Judgement (CJ), where the rater determines the superior of two paired performances, has increasingly been used for holistic evaluation of individual performances (Bramley & Vitello, 2018). The goal of CJ is to achieve high reliability (SSR) with as few comparisons as possible. Even in relatively small samples, achieving an SSR above.85 requires each performance to be included in around 30 comparisons (Crompvoets et al., 2020); this reduces the feasibility of using CJ in large samples. This research proposes a reference set-based CJ (RSB-CJ) model that can achieve high reliability with fewer comparisons in large samples. A Monte Carlo simulation study will be conducted to determine the effectiveness of the proposed model. In CJ, performances are paired according to a specific rule. In the RSB-CJ to be used in this study, a reference set will be created by randomly selecting 200 performances from what is considered a large group (2500) in the context of CJ applications. This reference set will be scored using the adaptive matching and standard error-based (.30) stopping rule used in CJ applications, ensuring that the estimates for the reference set achieve reliability above.90. In the next phase, the main set (the remaining 2300 performances) will be continuously paired and scored against the reference set (200 performances). At this stage, the adaptive matching and standard error-based (.30) stopping rule will also be applied. The results obtained will be compared with the results from applying traditional CJ to the 2500 performances. The comparison of CJ and RSB-CJ applications will be based on reliability (SSR), rank correlation, correlation between true and estimated abilities, and the number of comparisons per performance. It is expected that RSB-CJ will achieve similar reliability levels as CJ with fewer pairwise comparisons.

Keywords*: Comparative Judgement, Monte Carlo simulation, Holistic assessment*

Reference

Bramley, T., & Vitello, S. (2018). The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 1–16. doi:10.1080/0969594X.2017.1418734.

Crompvoets, E. A. V., Béguin, A. A., & Sijtsma, K. (2020). Adaptive Pairwise Comparison for Educational Measurement. J. Educ. Behav. Stat. 45, 316–338. doi:10.3102/1076998619890589

[1] mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu Üniversitesi, Eskişehir, Türkiye

[2] s.gurel@siirt.edu.tr, ORCID: 0000-0003-3425-858X, Siirt Üniversitesi, Siirt, Türkiye

[3] ibrahimuysal@ibu.edu.tr, ORCID: 0000-0002-6767-0362, Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

[4] aliihsan@eskisehir.edu.tr, ORCID: 0000-0002-4705-3973, Eskişehir Teknik Üniversitesi, Eskişehir, Türkiye

İkinci/Yabancı Dil Olarak Türkçe Öğreticilerinin Ölçme ve Değerlendirme Okuryazarlıkları: Bir Durum Çalışması

İsa Koyuncu [1], Eda Gürlen[2]

Bu çalışmanın amacı Türkçeyi ikinci/yabancı dil olarak öğreten öğreticilerin dil değerlendirme okuryazarlıkları ile ölçme ve değerlendirme hakkındaki algılarını belirlemektir. Çalışma, teorik olarak akademik ve profesyonel başarının bir parçası olarak dil öğreticisi/öğretmeni eğitiminde dil değerlendirme okuryazarlığı (DDO) gelişiminin önemini göstermeyi amaçlamaktadır. Bu amaç doğrultusunda, çalışma, nitel araştırma yöntemlerinden durum çalışması (case study) ile yürütülmüştür. Durum çalışması yöntemiyle Türkçeyi ikinci/yabancı dil olarak öğreten öğreticilerin ölçme ve değerlendirmeye ilişkin mevcut durumlarını; yaptıkları uygulamaları, deneyimlerini, görüşlerini ve ihtiyaçlarını daha iyi anlayabilmek için gerçek bağlamdan veriler ortaya konularak dil değerlendirme okuryazarlıkları açıklanmaya çalışılmıştır. Araştırmanın çalışma grubu, amaçlı örnekleme (purposeful sampling) yöntemiyle belirlenen ve Türkiye'deki üniversitelerin Türkçe hazırlık birimlerinde (TÖMER, DİLMER, TÖBİR, TÜRMER vb.) görev yapan öğreticilerden oluşmaktadır. Araştırma için gerekli Etik Kurul İzinleri Hacettepe Üniversitesi Sosyal ve Beşeri Bilimler Araştırma Etik Kurulundan alınmıştır (Belge No: E-66777842-300-00003671940).

Çalışmada veri toplama aracı olarak araştırmacılar tarafından konuyla ilgili mevcut alanyazın da taranarak geliştirilen yarı yapılandırılmış görüşme formu kullanılmıştır. 9 sorudan oluşan yarı yapılandırılmış görüşme formu ile öğreticilerle birebir yarı yapılandırılmış görüşmeler yapılmıştır. Elde edilen nitel verilerin analizi MAXQDA veri analizi yazılımı ile gerçekleştirilmiştir. Verilerin analizinde ayrıca katılımcıların ölçme ve değerlendirmeye yönelik algılarını, mevcut okuryazarlık düzeylerini ve ihtiyaçlarını belirginleştirmek amacıyla durum kodları oluşturulmuştur.

Araştırmada, öğreticilerin; özellikle geçerlik ve güvenirlik başta olmak üzere değerlendirme prosedürlerini kısmen uygulamaya koyabildikleri ve kavramsal farkındalıklarının düşük olduğu bununla birlikte dilde ölçme ve değerlendirmeye yönelik bir eğitim ihtiyaçlarının bulunduğu, gerek dil becerilerinin değerlendirilmesinde kullanılan standart testlerdeki madde tiplerinin gerekse biçimlendirici/tamamlayıcı ölçme ve değerlendirme tekniklerinin dil düzeylerine göre değişiklik gösterdiği, yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirme sürecinin bazı avantajlı ve dezavantajlı yönlerinin olduğu sonuçlarına ulaşılmıştır.

Anahtar Kelimeler: Ölçme *ve Değerlendirme Okuryazarlığı, Dil Değerlendirmesi, Dil Değerlendirme Okuryazarlığı,* İkinci*/Yabancı Dil Olarak Türkçe* Öğretimi.

[1] ikoyuncu@hacettepe.edu.tr, ORCID: 0000-0003-4796-7086, Hacettepe Üniversitesi, Ankara, Türkiye

[2] edaerdem@hacettepe.edu.tr, ORCID: 0000-0002-1719-9840, Hacettepe Üniversitesi, Ankara, Türkiye

Assessment Literacies of Teachers of Turkish as a Second/Foreign Language: A Case Study

İsa Koyuncu [1], Eda Gürlen[2]

The aim of this study is to determine the language assessment literacies of Turkish as a second/foreign language teachers and their perceptions about assessment and measurement. Theoretically, the study aims to demonstrate the importance of developing language assessment literacy (LAL) as a part of academic and professional success in language instructor/teacher education. In line with this purpose, the study was conducted with a case study, one of the qualitative research methods. Through the case study method, language assessment literacies of Turkish as a second/foreign language teachers were tried to be explained by revealing data from the real context in order to better understand their current situation regarding measurement and evaluation; their practices, experiences, opinions and needs. The focal group of the research consists of instructors working in Turkish language preparation units (TÖMER, DİLMER, TÖBİR, TÜRMER, etc.) of universities in Turkey. Ethics Committee Permission for the research was granted from Hacettepe University Social Sciences and Humanities Research Ethics Committee (Document No: E-66777842-300-00003671940).

In the study, a semi-structured interview form developed by the researchers by reviewing the existing literature on the subject was used as a data collection tool. With a 9-question semi-structured interview form, one-on-one semi-structured interviews were conducted with the instructors. The qualitative data were analyzed with MAXQDA data analysis software. To analyze the data, case codes were created to clarify the participants' perceptions of measurement and evaluation, their current literacy levels and their needs.

Results of the study revealed that the instructors were partially able to implement the assessment procedures, especially validity and reliability, and that their conceptual awareness was low, however, they needed training on measurement and evaluation in the language, that both the item types in standardized tests used in the evaluation of language skills and formative/complementary measurement and evaluation techniques vary according to language levels, and that the measurement and evaluation process in teaching Turkish as a foreign language has some advantages and disadvantages aspects.

Keywords: *Assessment Literacy, Language Assessment, Language Assessment Literacy, Teaching Turkish as a Second/Foreign Language.*

[1] ikoyuncu@hacettepe.edu.tr, ORCID: 0000-0003-4796-7086, Hacettepe Üniversitesi, Ankara, Türkiye

[2] edaerdem@hacettepe.edu.tr, ORCID: 0000-0002-1719-9840, Hacettepe Üniversitesi, Ankara, Türkiye

Yabancı Dilde Anlam Tercihi Birlikteliklerinin İşlemlenmesi: Örtük Anlamsal Öğrenmenin Değerlendirilmesi

Nermin Yazıcı[1], Günçe Günduğdu[2], Leyla Çimen,[3] Çağrı Kaygısız[4]

Örtük anlamsal öğrenme, düzenli anlamsal birlikteliklere maruz kalmayla gerçekleşen öğrenme türüdür ve anlamsal genellemelerle ortaya çıkan bu türden öğrenmeler, sözcüklerin anlam tercihleriyle ilişkilendirilir. Anlam tercihi, sözcüklerin belirli bir rolde, belirli anlamsal kategorilerdeki sözcüklerle birlikte bulunma eğilimidir ve çağrışımsal öğrenme ilkesiyle açıklanır. Buna göre kullanım sıklığı yüksek eşdizimlilikler bellekte kodlanarak uzun süreli temsilleri güçlendirilir. Ancak anlamsal açıdan eşdeğer kimi sözcüklerin anlam birlikteliklerine ilişkin bilgi, bu sözcüklerin pragmatik özelliklerine dönük temsilleri de motive eder. Dil öğreniminin önemli bir boyutunu oluşturan bu sürecin dil öğrenicisinin maruz kaldığı iletişim bağlamlarında örtük olarak gerçekleşebileceği varsayımı, kullanım temelli dil öğrenimi çerçevesinde tartışılır. Özellikle yabancı dilde sözcüklerin pragmatik uzanımlarının tanınması, dil kullanıcılarının bağlama uygun dilsel seçimler gerçekleştirmelerini desteklemektedir. Bu araştırmada da yabancı dil olarak Türkçe konuşucularının yakın anlamlı sözcüklerin anlam tercihi birlikteliklerinin bilişsel işlemlenme süreçleri incelenmektedir. Araştırmanın katılımcı grubu deney ve kontrol olmak üzere iki gruptan oluşmaktadır. Katılımcılar C1 dil düzeyindeki yabancı dil olarak Türkçe öğrenicileridir. Araştırma, deneysel yöntemle tasarlanmış olup PsychoPy programında gerçekleştirilmiştir. Deney eğitim ve test aşamasından oluşmaktadır. Deneyin eğitim aşamasında her hedef eylem kendi anlam tercihini yansıtan 8 tümce içinde sunulmuş ve deney grubunun dilsel olayı fark etmemeleri için katılımcılardan tümcenin zorluk düzeyini belirlemeleri istenmiş, böylece katılımcılar örtük öğrenme koşulu için hedef eylem birlikteliklerine maruz bırakılmışlardır. Deneyin test aşamasında ise hedef eylemler anlam tercihinin uyumlu olduğu birliktelikler ve manipüle edilen uyumsuz birliktelikler ile sunulmuş ve katılımcılardan ekrandaki sözcüklerin eğitim aşamasında okudukları tümcelerde yer alıp almadıklarını belirlemeleri istenerek tepki süreleri ölçülmüştür. Örtük anlamsal öğrenmenin gerçekleşip gerçekleşmediğini gözlemlemek amacıyla kontrol grubuna ise yalnızca test aşaması uygulanmış, katılımcılardan anlam tercihi uyumlu ve uyumsuz koşullu sözcük birlikteliklerıyle daha önce karşılaşıp karşılaşmadıklarını belirtmeleri istenerek tepki süresi ölçümü gerçekleştirilmiştir. Çalışma sonucunda uyumsuz dizgelerde işlemleme yükü meydana gelmesi beklenmektedir. Söz konusu etkilerin sözcük öğreniminde örtük anlamsal öğrenmeye dönük psikodilbilimsel çıkarımları olanaklı kılacağı varsayılmaktadır.

Anahtar Kelimeler: Örtük *anlamsal* öğrenme*, sözcük* öğrenimi*, anlam tercihi, eşdizim.*

[1] nerminyazici@hacettepe.edu.tr, ORCID: 0000-0003-0145-9772, Hacettepe Üniversitesi, Ankara, Türkiye

[2] guncegundugdu@hacettepe.edu.tr, ORCID: 0000-0002-4512-2170, Hacettepe Üniversitesi, Ankara, Türkiye

[3] leylacimen@erciyes.edu.tr, ORCID: 0000-0003-0692-6421, Erciyes Üniversitesi, Kayseri, Türkiye

[4] cgr.kaygisiz@gmail.com, ORCID: 0000-0002-9650-3889, Türk Hava Kurumu Üniversitesi, Ankara, Türkiye

Processing Meaning Preference Associations in a Foreign Language: Assessing Implicit Semantic Learning

Nermin Yazıcı[1], Günçe Günduğdu[2], Leyla Çimen,[3] Çağrı Kaygısız[4]

Implicit semantic learning is the type of learning that occurs through exposure to regular semantic associations, and this type of learning, which occurs through semantic generalizations, is associated with the semantic preferences of words. Semantic preference is the tendency of words to co-occur in a particular role with words in particular semantic categories and is explained by the principle of associative learning. Accordingly, collocations with high frequency of use are encoded in memory and their long-term representations are strengthened. However, information about the semantic associations of some semantically equivalent words also motivates the representations of the pragmatic features of these words. The assumption that this process, which constitutes an important dimension of language learning, can take place implicitly in the communication contexts to which the language learner is exposed is discussed within the framework of usage-based language learning. Recognizing the pragmatic extensions of words, especially in a foreign language, supports language users to make contextually appropriate linguistic choices. In this study, the cognitive processing processes of the meaning preference associations of near- synonym words of Turkish as a foreign language speakers are examined. The participants were Turkish as a foreign language learner at C1 level. The study was designed with experimental method and conducted in PsychoPy program. The experiment consisted of training and test phases. In the training phase of the experiment, each target word was presented in 8 sentences reflecting its semantic preference and the participants were asked to determine the difficulty level of the sentence so that the experimental group would not notice the linguistic event, thus the participants were exposed to target verb associations for the implicit learning condition. In the test phase of the experiment, the participants were presented with target verb associations that were congruent with their semantic preference and manipulated incongruent associations, and their reaction times were measured by asking them to determine whether the words on the screen appeared in the sentences they read in the training phase. In order to observe whether implicit semantic learning occurred or not, only the test phase was applied to the control group, and reaction times were measured by asking the participants to indicate whether they had encountered semantic preference compatible and incompatible conditional word associations before. As a result of the study, a cognitive load is expected to occur in incongruent strings. It is hypothesized that these effects will enable psycholinguistic inferences about implicit semantic learning in vocabulary learning.

Keywords: *Implicit semantic learning, vocabulary learning, semantic preference, collocation.*

[1] nerminyazici@hacettepe.edu.tr, ORCID: 0000-0003-0145-9772, Hacettepe Üniversitesi, Ankara, Türkiye

[2] guncegundugdu@hacettepe.edu.tr, ORCID: 0000-0002-4512-2170, Hacettepe Üniversitesi, Ankara, Türkiye

[3] leylacimen@erciyes.edu.tr, ORCID: 0000-0003-0692-6421, Erciyes Üniversitesi, Kayseri, Türkiye

[4] cgr.kaygisiz@gmail.com, ORCID: 0000-0002-9650-3889, Türk Hava Kurumu Üniversitesi, Ankara, Türkiye

Kontrol Listelerinin Yabancı/İkinci Dil Olarak Türkçe Konuşma Becerisinin Değerlendirilmesinde Kullanımı

Aybegüm AKSAK[1], Günsu Taşköprü [2]

Ölçme değerlendirme uygulamaları öğretimin etkililiğini yansıtma ve yol gösterici olma noktasında önem taşımaktadır. Eğitim öğretimin tüm alanlarında olduğu gibi yabancı dil öğretiminde de öğrenicilerin performanslarının iyi yapılandırılmış ölçme ve değerlendirme araçları ile nesnel şekilde değerlendirilmesi gerekmektedir. Özellikle üretim becerileri olan konuşma ve yazma becerilerinin ölçülmesi dil testleri ile değil, açık uçlu veya görev temelli sorular ile gerçekleştirildiğinden yabancı/ikinci dil öğretiminde nesnel değerlendirilmesi güç olan becerilerdir. Konuşma ve yazma becerileri performansa dayalı olarak ölçülür ve performans değerlendirmede kullanılan derecelendirme ölçekleri (analitik ve bütüncül puanlama anahtarları), kontrol listeleri (Mertler, 2001,s. 1) gibi bazı değerlendirme araçları söz konusudur. Konuşma becerisinin ölçülmesi ve değerlendirilmesi eş zamanlı olarak gerçekleştirilir, bu nedenle konuşma becerisinin değerlendirilmesinde kullanılacak değerlendirme aracının kullanılabilirliği ve pratikliği yüksek olmalıdır. Kontrol listeleri konuşma becerisinin hızlı bir şekilde değerlendirilmesine yardımcı olabilmesi söz konusu olduğundan kullanışlılığı yüksektir. Araştırma kapsamında konuşma becerisinin değerlendirilmesine yönelik sözlü anlatım ve karşılıklı konuşma becerileri için iki ayrı kontrol listesi hazırlanması amaçlanmaktadır.

Araştırmada nitel araştırma yönteminden yararlanılacaktır. Amaçlar doğrultusunda öncelikle literatür taraması yapılarak kontrol listesinde yer alması gereken maddeler için bir madde havuzu oluşturulacaktır. Ardından kontrol listesinde yer alacak maddeler için uzman görüşü alınacaktır. Gall, Borg ve Gall (1996)'in literatür taramasının özellikle yeni araştırma konularının yakalanabilmesine ve gelecek çalışmaların neler olabileceğinin belirlenmesine katkı sağlayacağı ifadelerinden hareketle bu yöntem tercih edilmiştir. Çünkü özellikle konuşma becerisinin değerlendirilmesi sürecinde hâlihazırda sınırların belirlenmesinde sorunlar yaşandığı ve daha önce uygulanan yöntemlere ek olarak içerik açısından farklılaşmaya ve yapılandırmaya ihtiyaç duyulduğu anlaşılmaktadır. Araştırma sonucunda konuşma becerisini nesnel ve tutarlı şekilde değerlendirecek iki farklı kontrol listesinin oluşacağı ve hem alan yazına hem de uygulamaya katkı sağlayacağı düşünülmektedir.

Anahtar Kelimeler: *yabancı dil olarak Türkçe,* ölçme *değerlendirme, konuşma becerisi, kontrol listesi*

Kaynakça

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Education research: An introduction (6th ed.). White Plains, NY: Longman.

Mertler, C. A. (2001). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research, and Evaluation*, *7*(25), 1–8.

[1] aybegumce@gmail.com:, ORCID: 0000-0002-5367-6608, Ankara Yıldırım Beyazıt Üniversitesi, Ankara, Türkiye

[2] gunsutaskopru@gmail.com:, ORCID: 0000-0002-9724-4698, Dokuz Eylül Üniversitesi, İzmir, Türkiye

Use of Checklists in the Assessment of Speaking Turkish as a Foreign/Second Language

Aybegüm AKSAK[1], Günsu Taşköprü [2]

Assessment and evaluation practices are important in terms of reflecting the effectiveness of teaching and providing guidance. As in all areas of education and training, in foreign language teaching, learners' performances need to be assessed objectively with well-structured assessment and evaluation tools. Speaking and writing skills, which are the production skills, are not assessed with language tests but with open-ended or task-based questions, so they are skills that are difficult to assessed objectively in foreign/second language teaching. Speaking and writing skills are assessed based on performance, and there are some assessment tools such as rating scales (analytical and holistic rubrics) and checklists (Mertler, 2001, p. 1) used in performance assessment. The assessment and evaluation of speaking skills are carried out simultaneously, therefore the usability and practicality of the evaluation tool to be used in the assessment of speaking skills should be high. Checklists are highly useful as they can help quickly assess speaking skills. Speaking skills are assessed as two skills, oral expression and conversation. Considering this situation, it is aimed to prepare two separate checklists for oral expression and conversation skills for the assessment of speaking skills within the scope of the research.

Qualitative research method will be used in the research. In line with the objectives, firstly a literature review will be conducted and an item pool will be created for the items that should be included in the checklist. Then, expert opinions will be obtained for the items to be included in the checklist. This method was preferred based on the statements of Gall, Borg and Gall (1996) that literature review will contribute to the detection of new research topics and determination of what future studies may be. Because it is understood that there are currently problems in determining the boundaries in the process of evaluating speaking skills in particular and that there is a need for differentiation and structuring in terms of content in addition to the previously applied methods. It is thought that two different checklists that will assess speaking skills objectively and consistently will be formed as a result of the research and will contribute to both the literature and the application.

Keywords: *Turkish as a foreign language, assessment and evaluation, speaking skills, checklist*

Reference

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Education research: An introduction (6th ed.). White Plains, NY: Longman.

Mertler, C. A. (2001). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research, and Evaluation*, *7*(25), 1–8.

[1] aybegumce@gmail.com:, ORCID: 0000-0002-5367-6608, Ankara Yıldırım Beyazıt Üniversitesi, Ankara, Türkiye

[2] gunsutaskopru@gmail.com:, ORCID: 0000-0002-9724-4698, Dokuz Eylül Üniversitesi, İzmir, Türkiye

Konuşma ve Dinleme Becerilerine Yönelik Uygulama Sınavlarının Değerlendirilmesi

Orhan Özen[1]

Türkçe öğretiminde temel hedef karşı tarafa bilgi aktarmak değil beceriler kazandırmaktır (Şahbaz, 2019). Bu beceriler ise Türkçe öğretiminin ana yapısını oluşturan dört temel beceridir. Türkçe Dersi Öğretim Programı bu dört temel beceriyi öğrencilerin hayat boyu kullanabilmeleri, sosyal ve bireysel yönden kendilerini geliştirmeleri, iletişimi etkili kullanabilmeleri doğrultusunda yapılandırılmıştır. MEB, beceri eğitimi kapsamında Türkçe, Türk dili ve edebiyatı ile yabancı dil derslerinin dinleme ve konuşma sınavlarının uygulamalı olarak yapılacağını belirtmiştir. Çalışmanın amacı Türkçe derslerinde kullanılan ölçme uygulamalarından konuşma ve dinleme/izleme sınavlarının öğrencilerin konuşma ve dinleme becerilerine etkisini belirlemektir. Aynı zamanda bu sınavların hazırlanmasında, uygulanmasında ve değerlendirilmesinde yaşanan sorunları tespit etmek ve bunlara yönelik çözüm önerileri getirmektir. Araştırma durum çalışması olarak planlanmıştır. Nitel araştırma desenlerinden biri olan durum çalışması bir veya birden fazla kaynak kullanarak gerçek hayat ya da sınırları belli bir durum hakkından derinlemesine ve ayrıntılı bilgi toplamaya olanak sağlamaktadır (Creswell, 2016). Derslerinde konuşma ve dinleme/izleme uygulama sınavlarını yapmakta olan Türkçe öğretmenleri araştırmanın katılımcılarını oluşturmaktadır. Araştırma sorularına yönelik katılımcı çeşitliliğinin fazla olması için maksimum çeşitlilik örneklem yöntemi kullanılmıştır (Yıldırım & Şimşek, 2008). Bu bağlamda MEB'de Türkçe öğretmeni olarak görev yapmakta olup derslerinde konuşma ve dinleme sınavlarını uygulamakta olan 30 Türkçe öğretmeniyle görüşmeler yapılmıştır. Araştırma verilerinin toplanmasında yarı yapılandırılmış görüşme formu kullanılmıştır. Bu görüşme yöntemi derinlemesine incelenmesi gereken durumlarda katılımcı görüşlerinden yola çıkarak konuyu daha iyi anlamayı sağlar. Araştırma verilerinin analizinde içerik analizi yöntemi kullanılmıştır. Araştırma sonuçlarının konuşma ve dinleme/izleme sınavlarının etkinliğine yönelik ayrıntılı bir bakış açısı sağlaması beklenmektedir. Aynı zamanda sınavların öğretmenler tarafından hazırlanması, uygulanması ve değerlendirilmesi süreçlerinde karşılaşılan sorunlara yönelik çözüm önerileri belirlenecektir.

Anahtar Kelimeler: *Beceri eğitimi, uygulama sınavları, konuşma ve dinleme.*

Kaynakça

Creswell, J. W. (2016). *Nitel araştırma yöntemleri. Beş yaklaşıma göre nitel araştırma ve araştırma deseni.* M. Bütün ve S. B. Demir (Çev.). (3. Baskı). Ankara: Siyasal Kitabevi.

Şahbaz, N. K. (2019). Türkçe öğretiminde kullanılan yöntem ve teknikler. A. Güzel ve H. Karatay (Ed.), *Türkçe* öğretimi *el kitabı* içinde (3. baskı, ss. 131-147). Ankara: Pegem Akademi.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri.* Ankara: Seçkin.

[1] orhanozen25@gmail.com:, ORCID: 0000-0002-3534-3084, Millî Eğitim Bakanlığı, Erzurum, Türkiye

Evaluation of Practice Exams for Speaking and Listening Skills

Orhan Özen[1]

The main goal in teaching Turkish is not to convey information to the other party but to provide skills (Şahbaz, 2019). These skills are the four basic skills that form the main structure of teaching Turkish. The Turkish Language and Literature Curriculum is structured so that students can use these four basic skills throughout their lives, develop themselves socially and individually, and use communication effectively. The Ministry of National Education has stated that the listening and speaking exams of Turkish, Turkish language and literature, and foreign language courses will be held practically within the scope of skills training. The aim of the study is to determine the effects of speaking and listening/monitoring exams, which are among the measurement applications used in Turkish lessons, on students' speaking and listening skills. At the same time, it is to identify the problems experienced in the preparation, implementation, and evaluation of these exams and to offer solution suggestions for them. The research was planned as a case study. A case study, which is one of the qualitative research designs, allows for in-depth and detailed information to be collected about real life or a situation with certain boundaries by using one or more sources (Creswell, 2016). The participants of the study are Turkish teachers who conduct speaking and listening/monitoring practice exams in their classes. In order to have a wide range of participants for the research questions, maximum diversity sampling method was used (Yıldırım & Şimşek, 2008). In this context, interviews were conducted with 30 Turkish teachers who work as Turkish teachers in the Ministry of National Education and who apply speaking and listening exams in their classes. A semi-structured interview form was used to collect the research data. This interview method provides a better understanding of the subject by starting from the participant views in cases where in-depth examination is required. Content analysis method was used in the analysis of the research data. It is expected that the research results will provide a detailed perspective on the effectiveness of speaking and listening/monitoring exams. At the same time, solution suggestions will be determined for the problems encountered in the preparation, implementation and evaluation processes of the exams by teachers.

Keywords: *Skills training, practice exams, speaking and listening.*

Reference

Creswell, J. W. (2016). *Nitel araştırma yöntemleri. Beş yaklaşıma göre nitel araştırma ve araştırma deseni.* M. Bütün ve S. B. Demir (Çev.). (3. Baskı). Ankara: Siyasal Kitabevi.

Şahbaz, N. K. (2019). Türkçe öğretiminde kullanılan yöntem ve teknikler. A. Güzel ve H. Karatay (Ed.), *Türkçe* öğretimi *el kitabı* içinde (3. baskı, ss. 131-147). Ankara: Pegem Akademi.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri.* Ankara: Seçkin.

[1] orhanozen25@gmail.com:, ORCID: 0000-0002-3534-3084, Millî Eğitim Bakanlığı, Erzurum, Türkiye

Yüksek Ölçekli Bir Dil Sınavında Yazma Bölümünün Değerlendirilmesine Yönelik Eğitim İçeriklerinin Planlanması

Eylem KUBİLAY[1], Dr. Seçil ALACA [2], Dr. Nida KIRÖMEROĞLU[3]

"Elektronik, Uluslararası Türkçe Yeterlik Sınavı"e-UTYS; yüksek ölçekli bir dil sınavı olarak ÖSYM tarafından okuma, dinleme, konuşma ve yazma becerilerini farklı düzeylerde derecelendirmek için geliştirilen bir sınavdır. ÖSYM tarafından bir proje kapsamında geliştirilen ve Türkçe dil yeterliğini ölçmeyi amaçlayan bu sınavın okuma ve dinleme bölümleri, otomatik olarak değerlendirilecek şekilde tasarlanmıştır. Üretim becerilerinden konuşma ve yazma çıktılarının değerlendirilmesi ise -belirli kurallar çerçevesinde belirlenen- değerlendiriciler tarafından gerçekleştirilecektir. Dil sınavlarında üretim becerilerinin puanlanması ve düzeylendirilmesinde değerlendiricilerin standartlaşması sınavın geçerlik ve güvenirliğini sağlayan en önemli aşamadır. Bunun yanı sıra yanıtların değerlendirmesinde kullanılacak ölçütlerin belirlenmesi, değerlendiricilerin ortak kriterlere uygun değerlendirme yapabilmesi için temel bir gerekliliktir. Bu çalışma; geliştirilen sınavın yazma bölümünün değerlendirilmesine başlanmadan önce planlanan eğitim içerikleri hakkında bilgi vermek için oluşturulmuştur. Çalışmanın birinci bölümünde yazma bölümü için belirlenen "dereceli puanlama anahtarı"nda bulunması gereken kriterler üzerinde durulmuştur. E-UTYS'nin bir derecelendirme sınavı olması dolayısıyla yazma becerisi için kullanılacak ölçekte D-AOBM tanımlayıcıları kullanılmıştır. Çalışmanın ikinci aşamasında, değerlendiricilerin oluşturulan puanlama anahtarına göre değerlendirme yapabilmesi için değerlendirici eğitimlerinde uygulanan yöntemlere yer verilmiştir. Çalışmanın üçüncü aşamasında değerlendiricilerin güvenirliğini sağlamak için düzenlenen standardizasyon eğitimlerinin önemi üzerinde durulmuştur. Böylece bu çalışma ile e-UTYS'den hareketle açık uçlu soruların standart yaklaşımla değerlendirmesine katkı sağlanmış olacaktır.

Anahtar Kelimeler: *e-UTYS, ÖSYM, yazma, değerlendirme, güvenirlik*

Kaynakça

ALTE, (2011). *Manual For Language Test Devolopment and Examing*: *For Use With The CEFR*.

Avrupa Konseyi, (2021). *Diller için Avrupa Ortak Başvuru Metni:* Öğrenme*,* Öğretme *ve Değerlendirme Tamamlayıcı Cilt*, (çev. Millî Eğitim Bakanlığı Komisyonu).

Council of Europe The Language Policy Division (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) A Manual*. Strasbourg.

[1] eylem1958@ hotmail.com, 0000-0000 3587-2629, ÖSYM, Ankara, Türkiye

[2] secilalaca@gmail.com, 0000-0002-8731-4197, İstanbul Üniversitesi, İstanbul, Türkiye

[3] nidakiromeroglu@gmail.com, 0000-0001-8252-979X, İstanbul Üniversitesi, İstanbul, Türkiye

Planning Training Contents for the Evaluation of the Writing Section in a Large-Scale Language Exam

Eylem KUBİLAY[1], Dr. Seçil ALACA [2], Dr. Nida KIRÖMEROĞLU[3]

"Electronic, International Turkish Proficiency Exam" e-UTYS is a high-scale language exam developed by ÖSYM to rate reading, listening, speaking and writing skills at different levels. The reading and listening sections of this exam, which was developed by ÖSYM within the scope of a project and aims to measure Turkish language proficiency, are designed to be evaluated automatically. The assessment of speaking and writing outputs, which are production skills, will be carried out by assessors who are determined within the framework of certain rules. The standardization of assessors in the scoring and leveling of production skills in language exams is the most important stage in ensuring the validity and reliability of the test. In addition to this, determining the criteria to be used in the evaluation of the answers is a basic requirement for the evaluators to make evaluations in accordance with common criteria. This study is designed to provide information about the planned training contents before the assessment of the writing section of the developed exam. In the first part of the study, the criteria that should be included in the "rubric" determined for the writing section were emphasized. Since E-UTYS is a grading test, D-AOBM descriptors were used in the scale to be used for writing skill. In the second stage of the study, the methods applied in the rater trainings were included in order for the raters to make evaluations according to the rubric created. In the third stage of the study, the importance of standardization trainings organized to ensure the reliability of evaluators was emphasized. Thus, this study will contribute to the evaluation of open-ended questions with a standardized approach based on e-UTYS.

Keywords: *e-UTYS, ÖSYM, writing, evaluation, reliability*

Reference

ALTE, (2011). *Manual For Language Test Devolopment and Examing*: *For Use With The CEFR*.

Avrupa Konseyi, (2021). *Diller için Avrupa Ortak Başvuru Metni:* Öğrenme*,* Öğretme *ve Değerlendirme Tamamlayıcı Cilt*, (çev. Millî Eğitim Bakanlığı Komisyonu).

Council of Europe the Language Policy Division (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) A Manual*. Strasbourg.

[1] eylem1958@ hotmail.com, 0000-0000 3587-2629, ÖSYM, Ankara, Türkiye

[2] secilalaca@gmail.com, 0000-0002-8731-4197, İstanbul Üniversitesi, İstanbul, Türkiye

[3] nidakiromeroglu@gmail.com, 0000-0001-8252-979X, İstanbul Üniversitesi, İstanbul Türkiye

Merkezi Yerleştirme Sistemlerinde Stratejik Koordinasyonun Tercih Sıralamaları Üzerindeki Etkisi

Muhammet Yorulmaz[1], Umut Mert Dur[2], Sinan Ertemel[3], İsmail Yıldız[4]

Yerleştirme araştırmaları uzun yıllara dayanmaktadır ve kullanılan algoritmaların iyileştirme çalışmaları araştırmacıların ilgisini çekmeye devam etmektedir (Gale & Shapley, 1962) (Balinski & Sönmez, 1999) (Abdülkadiroğlu & Sönmez, 2003). Özellikle puan eşitliği ve öncelikler yerleştirme iyileştirmelerinde ayrı bir yer tutmaktadır. Türkiye'de Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı (ÖSYM) tarafından yürütülen merkezi yerleştirme işlemlerinde, adaylar arasında puan eşitliği durumunda farklı öncelik politikaları uygulanmaktadır.

Bu çalışma, puan eşitliği durumunda uygulanan politikaların yerleştirme sonuçlarına etkilerini incelemektedir. Üretilen örnek veri setleri kullanılarak, bu politikaların merkezi yerleştirme sonuçlarına olan etkileri gösterilecektir. Bazı merkezi sınavlarda, eşit sınav puanına sahip adayların tercih sıralamalarını koordine ederek aynı programa aynı anda yerleşebildikleri gözlemlenmiştir. Bu tür durumlarda, adaylar arasında gerçekleşecek koordinasyonun bir Nash dengesi oluşturması olasıdır. Bu bağlamda, Nash dengesi ve diğer çözüm kavramlarını analiz ederek, adayların stratejik davranışlarının ve bu davranışların tercih sıralamalarına dayalı yerleştirme sonuçlarını nasıl etkilediğine dair daha derin bir araştırma yapılacaktır. Çalışma, politika yapıcılar ve karar vericiler için rehber niteliğinde önerilerle sonuçlanacaktır.

Anahtar Kelimeler: *merkezi yerleştirme, eşitlik bozma, tercih manipülasyonu, denge analizi.*

Kaynakça

Abdülkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, *93*(3), 729–747.

Balinski, M., & Sönmez, T. (1999). A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, *84*(1), 73–94. https://doi.org/10.1006/JETH.1998.2469

Gale, D., & Shapley, L. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

[1] muhammed@baskent.edu.tr, ORCID: 0000-0002-4904-9396, Başkent University, Ankara, Türkiye

[2] umutdur@gmail.com, ORCID: 0000-0002-0553-8382, North Carolina State University, North Carolina, USA

[3] ertemels@itu.edu.tr, ORCID: 0000-0003-0089-4641, İstanbul Technical University, İstanbul, Türkiye

[4] ismaily@kastamonu.edu.tr, ORCID: 0000-0003-3048-2840, Kastamonu University, Kastamonu, Türkiye

The Impact of Strategic Coordination on Preference Rankings in Centralized Placement Systems

Muhammet Yorulmaz[1], Umut Mert Dur [2], Sinan Ertemel[3], İsmail Yıldız[4]

Placement research has been around for many years, and studies on improving the algorithms used continue to attract the attention of researchers (Gale & Shapley, 1962) (Balinski & Sönmez, 1999) (Abdülkadiroğlu & Sönmez, 2003). In particular, score ties and priorities have a special place in placement improvements. In the centralized placement procedures conducted by the Center for Assessment, Selection, and Placement (ÖSYM) in Türkiye, various policies are implemented in cases of score ties between candidates.

This study examines the effects of these policies on placement outcomes in cases of score ties. Using generated sample data sets, the impact of these policies on centralized placement results will be demonstrated. It has been observed that in some centralized exams, candidates with equal exam scores can coordinate and rank their preferences in the same order and be admitted to the same program simultaneously. In such scenarios, coordination among candidates is likely to lead to a Nash equilibrium. By analyzing the Nash equilibrium and other solution concepts within this context, we can gain a deeper insight into the strategic behaviors of candidates and how these behaviors influence placement outcomes based on preference rankings. The study will conclude with recommendations aimed at guiding policymakers and decision-makers.

Keywords: *centralized admissions, tie-breaking, preference manipulation, equilibrium analysis.*

References

Abdülkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, *93*(3), 729–747.

Balinski, M., & Sönmez, T. (1999). A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, *84*(1), 73–94. https://doi.org/10.1006/JETH.1998.2469 Gale, D., & Shapley, L. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

[1] muhammed@baskent.edu.tr, ORCID: 0000-0002-4904-9396, Başkent University, Ankara, Türkiye

[2] umutdur@gmail.com, ORCID: 0000-0002-0553-8382, North Carolina State University, North Carolina, USA

[3] ertemels@itu.edu.tr, ORCID: 0000-0003-0089-4641, İstanbul Technical University, İstanbul, Türkiye

[4] ismaily@kastamonu.edu.tr, ORCID: 0000-0003-3048-2840, Kastamonu University, Kastamonu, Türkiye

ÖSYM Sınav Puanlamaları İçin Alternatif Yaklaşımlar

Mehmet Güven GÜNVER[1], Murat KİRİŞÇİ[2], Rahmet SAVAŞ[3], Duran EREN[4]

ÖSYM uyguladığı sınavların birçoğunda, adayların net cevap sayılarını dağılıma uydurarak puan hesaplamaktadır. Bu konuda, Dünya'da kabul görmüş standart uygulama olan t dönüşümü kullanılmaktadır. T dönüşümü uygulanabilmesi için adayların ürettiği net sayılarının normal dağılım varsayımı karşılaması gerekmektedir. Ancak, POC kapsamında çalıştığımız DGS 2023 sınavında ve diğer birçok sınavda normal dağılım varsayımının karşılanmadığı görülmüştür. Normal dağılmayan (çarpık) veride daha gerçeğe yakın sonuçlar üreten Golden Ratio in Statistics "GRiS" (1,2) yöntemiyle üretilen parametreleri kullanarak yapılan transformasyon sonuçları paylaşılacaktır.

ÖSYM'nin gerçekleştirdiği sınavlarda, ölçümü yapabilmek için farklı zorluk seviyelerinde sorular sormaktadır. Soruların zorluk derecesi birbirinden farklı olsa da, bu soruların standart puana katkısı eşittir. Soruya doğru cevap veren kişi sayısından hareketle, zor soruya "yüksek", kolay soruya "düşük" katkı katsayısı üreten bir yaklaşım üretilmiştir.

Anahtar Kelimeler: *normal dağılım,* çarpıklık*, zorluk derecesi.*

Kaynakça

Gunver, M. G, Senocak, M. S., Vehid, Suphi. (2018). To determine skewness, mean and deviation with a new approach on continuous data. Int. J, 10.

Gunver, M. G. (2022). Norm-Referenced Scoring on Real Data: A Comparative Study of GRiSTEN and STEN. SAGE Open, 12(2), 21582440221091253.

[1] guven.gunver@istanbul.edu.tr, 0000-0002-4628-8391, Istanbul University, Istanbul, Turkey

[2] murat.kirisci@iuc.edu.tr, 0000-0003-4938-5207, Istanbul University Cerrahpasa, Istanbul, Turkey

[3] rahmetsavas@gmail.com, 0000-0002-3670-622X, Haliç University, Istanbul Turkey

[4] duran.eren@osym.gov.tr, ---, OSYM, Ankara

Alternative Approaches for ÖSYM Exam Scoring

Mehmet Güven GÜNVER [1], Murat KİRİŞÇİ [2], Rahmet SAVAŞ[3], Duran EREN [4]

In many of the exams it administers, ÖSYM calculates scores by adapting the candidates' net answer numbers to the distribution. In this regard, t transformation, which is the accepted standard practice in the world, is used. In order for T transformation to be applied, the net numbers produced by the candidates must meet the normal distribution assumption. However, it was observed that the normal distribution assumption was not met in the DGS 2023 exam and many other exams that we studied within the scope of POC. Golden Ratio in Statistics "GRiS", which produces more realistic results in non- normally distributed (skewed) data. Transformation results using the parameters produced by the (1,2) method will be shared.

In the exams conducted by ÖSYM, questions at different difficulty levels are asked in order to make the measurement. Although the difficulty levels of the questions differ from each other, the contribution of these questions to the standard score is equal. Based on the number of people who answered the question correctly, an approach was produced that produced a "high" contribution coefficient for the difficult question and a "low" contribution coefficient for the easy question.

Key Words: *normal distribution, skewness, degree of difficulty.*

References

Gunver, M. G, Senocak, M. S., Vehid, Suphi. (2018). To determine skewness, mean and deviation with a new approach on continuous data. Int. J, 10.

Gunver, M. G. (2022). Norm-Referenced Scoring on Real Data: A Comparative Study of GRiSTEN and STEN. SAGE Open, 12(2), 21582440221091253.

[1] guven.gunver@istanbul.edu.tr, 0000-0002-4628-8391, Istanbul University, Istanbul, Turkey

[2] murat.kirisci@iuc.edu.tr, 0000-0003-4938-5207, Istanbul University Cerrahpasa, Istanbul, Turkey

[3] rahmetsavas@gmail.com, 0000-0002-3670-622X, Haliç University, Istanbul Turkey

[4] duran.eren@osym.gov.tr, ---, OSYM, Ankara

Rastgele Arama ile Merkezi Olmayan Çoktan Bire Eşleştirme

Günnur Ege BİLGİN[1]

Sonlu sayıda firma ve işçinin piyasa temizlenene kadar rastgele buluştuğu, friksiyonların olduğu merkezi olmayan bir arama modelinde kanonik bir çoktan bire eşleştirme piyasasını analiz etmekteyim. Temel piyasanın stabil eşleşmelerini ve zamanın neredeyse maliyetsiz olduğu dengenin sonuçlarını karşılaştırıyorum. Friksiyonlar ortadan kalktıkça arama modellerindeki sonuçların stabil sonuçlara yakınsaması, literatürde farklı ortamlar için çeşitli kereler test edilmiş bir varsayımdır (Roth & Sotomayor, 1992). Her firmanın sadece tek bir boş pozisyona sahip olduğu durumun aksine, firmalar çok boş pozisyona sahip olduğunda stabil eşleşmelerin o kadar kolay elde edilemediğini gösteriyorum. Özellikle, her alt oyunda hem işçi hem de firma açısından optimal stabil eşleşmeyi tekdüze olarak uygulayan bir Markovian dengesi olmayabilir. Buradaki zorluk, firmaların stratejik olarak kapasitelerini geri tutma kabiliyetinden kaynaklanmaktadır. Yine de, bir tarafta dikey tercihlerin olduğu piyasalar için durum böyle değildir ve neredeyse kesin olarak benzersiz stabil eşleşmeye yol açan denge strateji profilini inşa etmekteyim. Buna ek olarak, çoklu boş pozisyonlar, Markovian dengeler altında bile firmaların dolaylı olarak gizli anlaşma yapmasına ve stabil olmayan ancak firma tercihli eşleşmeler elde etmesine olanak tanır. Son olarak, bu tür fırsatları elemek için tercihler üzerinde yeterli bir koşul tanımlamaktayım. Makaledeki sonuçlar, Wu (2015) tarafından analiz edilen bire bir durumdan farklıdır ve bire çok ve bire bir modellerin arama ve eşleştirme çerçevesinde analiz edildiğinde birbirlerinden önemli ölçüde farklı olduğuna işaret etmektedir.

Anahtar Kelimeler: çoktan *bire eşleşme, arama, stabilite, denge*

Kaynakça

Wu, Q. (2015). A finite decentralized marriage market with bilateral search. *Journal of Economic Theory*, 160, 216-242.

Roth, A. E., & Sotomayor, M. (1992). Two-sided matching. *Handbook of game theory with economic applications*, 1, 485-541.

[1] ege.bilgin@tedu.edu.tr, ORCID: 0000-0002-2060-8342, TED Üniversitesi, Ankara, Türkiye

Decentralized Many-to-One Matching with Random Search

Günnur Ege BİLGİN[1]

I analyze a canonical many-to-one matching market within a decentralized search model with frictions, where a finite number of firms and workers meet randomly until the market clears. I compare the stable matchings of the underlying market and equilibrium outcomes when time is nearly costless. The outcomes in search models converging to stable outcomes as frictions vanish is a conjecture (Roth & Sotomayor, 1992), that has been tested various times in the literature for different settings. In contrast to the case where each firm has just a single vacancy, I show that stable matchings are not obtained as easily. In particular, there may be no Markovian equilibrium that uniformly implements either the worker- or the firm-optimal stable matching in every subgame. The challenge results from the firms' ability to withhold capacity strategically. Yet, this is not the case for markets with vertical preferences on one side, and I construct the equilibrium strategy profile that leads to the unique stable matching almost surely. Moreover, multiple vacancies enable firms to implicitly collude and achieve unstable but firm-preferred matchings, even under Markovian equilibria. Finally, I identify one sufficient condition on preferences to rule out such opportunities. The results in the paper are different from the one-to-one case analyzed by Wu (2015), hinting at many-to-one and one-to-one models being significantly different from each other when analyzed within a search and matching framework.

Keywords: *many-to-one matching, search, stability, equilibrium*

Reference

Wu, Q. (2015). A finite decentralized marriage market with bilateral search. *Journal of Economic Theory*, 160, 216-242.

Roth, A. E., & Sotomayor, M. (1992). Two-sided matching. *Handbook of game theory with economic applications*, 1, 485-541.

[1] ege.bilgin@tedu.edu.tr, ORCID: 0000-0002-2060-8342, TED University, Ankara, Türkiye

Admissions Game and Matching with Limited Resources

Vasily Gusev [1], Alexander Nesterov[2], Mikhail Reshetov[3], Alex Suzdaltsev[4]

In a variety of economic situations discrete agents choose one resource among several available resources and, once admitted to the resource of choice, divide it among fellow agents admitted there. The amount of the resource an agent gets is proportional to her relative ability to acquire this particular resource, what we refer to as an agent's weight at the resource. The relevant applications include students self-selecting into colleges, politicians self-selecting into races, and athletes self-selecting into teams. We find that this game has a pure-strategy Nash equilibrium in at least three special cases: 1) when agents have the same weight at each resource, 2) when all resources are the same, 3) when there are only two resources. We also show that this game always has an approximate Nash equilibrium when the number of players is large. Existence in the general case remains an open problem.

Keywords: *congestion games, potential games, pure Nash equilibrium, sorting into contests, college admissions*

[1] vgusev@hse.ru, HSE University, St. Petersburg, Russia

[2] nesterovu@gmail.com, HSE University, St. Petersburg, Russia

[3] ma_reshetov@mail.ru, HSE University, St. Petersburg, Russia

[4] asuzdaltsev@gmail.com, HSE University, St. Petersburg, Russia

Preventing Academic Misconduct in Online Assessments

Arum Perwitasari[1]

Traditional exams rely on human proctors to ensure fairness, whereas online exams require advanced proctoring techniques to verify identities, monitor behaviors, and protect test integrity (Langenfeld, 2020). Remote human monitoring via webcams includes identity verification and scanning the test environment for unauthorized materials (Atoum et al., 2017; Nigam et al., 2021). Turani et al. (2020) emphasized the critical importance of security in online exams, stressing the need for systems to verify test-taker identities and maintain overall exam integrity and security.

While motivations for cheating in online and offline exams are similar (Turner & Uludag, 2013), detecting and preventing online cheating poses unique challenges. This study conducts a narrative review on cheating in online exams, analyzing various proctoring methods and their effectiveness. It categorizes cheating into personal (e.g., using unauthorized exam tools, unlawful access to test content) and group cheating (e.g., impersonation, unauthorized communication with third parties). The study explores strategies to mitigate these behaviors and share best practices of effective prevention measures.

By synthesizing existing research, this study provides insights into online cheating behaviors and technologies aimed at combating them. These findings offer valuable guidance for educators, policymakers, and technology developers to enhance the integrity of online assessments.

Keywords: *test security, online assessment, remote proctoring*

Reference

Atoum, Y., Chen, L., Liu, A. X., Hsu, S., & Liu, X. (2017). Automated online exam proctoring. IEEE Transactions on Multimedia, 19(7), 1609-1624.

Langenfeld, T. (2020). Internet-based proctored assessment: Security and fairness issues. Educational Measurement: Issues and Practice, 39(3), 24-27. doi: 10.1111/emip.12359

Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A systematic review on AI-based proctoring systems: Past, present and future. Education and Information Technologies, 26(5), 6421– 6445. DOI: https://doi.org/10.1007/s10639-021-10597-x

Turani, A. A., Alkhateeb, J. H., & Alsewari, A. A. (2020). Students online exam proctoring: A case study using 360 degree security cameras. In Proceedings of the 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Bangladesh, 1–5.

Turner, S. W., & Uludag, S. (2013). Student perceptions of cheating in online and traditional classes.

Proceedings of the Frontiers in Education Conference (FIE), 1131–1137. DOI: https://doi.org/10.1109/FIE.2013.6685007

[1] aperwitasari@etsglobal.org, ORCID: 0000-0003-3153-6388, ETS EMEA, Amsterdam, The Netherlands

Yüksek Beklentiler, Düşen Memnuniyetler

Tülin OTBİÇER ACAR[1]

LGS, YKS, KPSS gibi ulusal ve merkezi olarak yapılan seçme ve yerleştirme sınavlarının mahiyeti, bu sınavlara atfedilen önem dikkate değer olduğu kadar bu sınavlara hazırlık süreci de tartışmaya değerdir. Öğretmenler ve öğrenciler özelinde başarı kavramı, memnuniyet, doyum, mutluluk, motivasyon vb kavramlarla bir bakıma temsil edilmektedir. Nitekim araştırmacılar tarafından okulda, sınıfta veya eğitimde başarı ile memnuniyet arasındaki ilişki çoğu zaman çeşitli açılardan irdelenmiştir (Jennings& Greenberg, 2019; Hascher&Waber, 2021; Marquez et al, 2024). Sınavlara hazırlık sürecinde çözülen soru sayısıyla ilişkilendirilen her ne kadar başarı olsa da mutluluk ya da memnuniyet kavramları düşünmeye değerdir. Genel olarak öğretmenler, öğrencilerine daha fazla soru çözdürerek memnuniyetlerini artırırken çözdükleri soru sayısı arttıkça öğrencilerde memnuniyet kaybı yaşanmaktadır. Bu durum, eğitimde önemli bir etkiye sahiptir. Bu etkinin çözümlenmesinde arz, talep, marjinal fayda, azalan verimlilik başta olmak üzere iktisat biliminin temel kavramlarından yararlanılabilir (Thweatt, 1983; Brewer, 2002).

Bu araştırma, iktisattaki arz ve talep kanunlarını öğretmenlerin ve öğrencilerin memnuniyet düzeyleri ile çözülen soru sayısı arasındaki ilişkiye uygulamayı amaçlamaktadır. Dolayısıyla ölçme bilimi ile iktisat bilimi arasındaki bağıntı kurularak ölçme bilimine yeni bir yaklaşım sunulacaktır. Araştırmada arz ve talep kanunları, azalan marjinal fayda kanunu ve fırsat maliyeti kavramları kullanılarak düşünsel çözümlemeler yapılmıştır. Öğrencilerin çözdükleri test sorularının sayısı ve memnuniyet düzeyleri ile öğretmenlerin memnuniyet düzeyleri matematiksel fonksiyonlar ve grafiklerle gösterilmiştir. Arz ve talep fonksiyonları belirlenerek denge noktası irdelenmiştir. Tıpkı piyasadaki arz ve talep dengesi gibi eğitimde de bir dengenin sağlanması önemlidir. Bu dengeyi sağlamak ve korumak için politika yapıcıların uzmanlarla iletişim kurması ve esnek bir yaklaşım benimsemesi önemlidir. Eğitimde önce denge ve çeşitlilik sağlanarak ölçme biliminin çıktılarından yararlanılması önerilmektedir.

Anahtar Kelimeler: *soru sayısı, memnuniyet düzeyi, arz- talep kanunları, azalan, marjinal fayda*

Kaynakça

Brewer, P.J., Huang, M., Nelson, B. *et al.* (2002). On the Behavioral Foundations of the Law of Supply and Demand: Human Convergence and Robot Randomness. *Experimental Economics* 5, 179–208.

Hascher, T., and Waber, J. (2021). Teacher well-being: a systematic review of the research literature from the years 2000–2019. Educ. Res. Rev. 34, 100411.

Jennings, P. A., & Greenberg, M. T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research*, *79*(1), 491-525.

Marquez, J., Taylor, L., Boyle, L., Zhou, W., & De Neve, J.-E. (2024). Child and adolescent well-being: global trends, challenges and opportunities (J.-E. De Neve, Ed.; Vol. 2024, pp. 61–102). Wellbeing Research Centre, University of Oxford.

Thweatt, W.O. (1983). Origins of the terminology, supply and demand. *Journal of Political Economy* 30, November, 287-94.

[1] totbicer@gmail.com, ORCID: 0000-0001-7976-5521, Eğitimde Ölçme ve Değerlendirme Uzmanı, Ankara, Türkiye

High Expectations vs. Falling Satisfactions

Tülin OTBİÇER ACAR[1]

The nature of national and centralized selection and placement exams such as LGS, YKS, KPSS, and the importance attributed to these exams is noteworthy, and the preparation process for these exams is also worthy of discussion. For teachers and students, the concept of success is represented in a way by concepts such as satisfaction, fulfillment, happiness, motivation and so on. As a matter of fact, researchers have often examined the relationship between success and satisfaction in school, classroom or education from various angles (Jennings & Greenberg, 2019; Hascher & Waber, 2021; Marquez et al, 2024). Although success is associated with the number of questions solved in preparation for exams, the concepts of happiness or satisfaction are worth thinking about. In general, while teachers increase their satisfaction by making their students solve more questions, students lose satisfaction as the number of questions they solve increases. This has a significant impact on education. In analyzing this effect, the basic concepts of economics such as supply, demand, marginal utility and diminishing returns can be utilized (Thweatt, 1983; Brewer, 2002).

This study aims to apply the laws of supply and demand in economics to the relationship between teachers' and students' satisfaction levels and the number of questions solved. Therefore, a new approach to measurement science will be presented by establishing a link between measurement science and economics. In the research, intellectual analyses were made using the laws of supply and demand, the law of diminishing marginal utility and the concepts of opportunity cost. The number of test questions solved by the students and their satisfaction levels and the satisfaction levels of the teachers were shown with mathematical functions and graphs. Supply and demand functions were determined and the equilibrium point was analyzed. Just like the supply and demand balance in the market, it is important to achieve a balance in education. To achieve and maintain this balance, it is important for policy makers to communicate with experts and adopt a flexible approach. It is recommended that the outputs of measurement science should be utilized by ensuring balance and diversity in education first.

Keywords: *number of questions, satisfaction level, supply-demand laws, diminishing marginal utility*

Reference

Brewer, P.J., Huang, M., Nelson, B. *et al.* (2002). On the Behavioral Foundations of the Law of Supply and Demand: Human Convergence and Robot Randomness. *Experimental Economics* 5, 179–208.

Hascher, T., and Waber, J. (2021). Teacher well-being: a systematic review of the research literature from the years 2000–2019. Educ. Res. Rev. 34, 100411.

Jennings, P. A., & Greenberg, M. T. (2009). The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes. *Review of Educational Research*, *79*(1), 491-525.

Marquez, J., Taylor, L., Boyle, L., Zhou, W., & De Neve, J.-E. (2024). Child and adolescent well-being: global trends, challenges and opportunities (J.-E. De Neve, Ed.; Vol. 2024, pp. 61–102). Wellbeing Research Centre, University of Oxford.

Thweatt, W.O. (1983). Origins of the terminology, supply and demand. *Journal of Political Economy* 30, November, 287-94.

[1] totbicer@gmail.com, ORCID: 0000-0001-7976-5521, Eğitimde Ölçme ve Değerlendirme Uzmanı, Ankara, Türkiye

Otomatik Madde Üretiminde Görsellerin İşlevleri Ve Dinamik Olarak Oluşturulması

Erol KARAKIRIK[1]

Kavramların anlaşılması, öğrenilmesi ve problemlerin çözülmesine katkıları açısından görsellerin işlevleri üzerine çok geniş bir literatür bulunmaktadır. Ancak önceki araştırmalar, otomatik madde üretiminde (OMÜ) görsellerin dinamik olarak oluşturulması ve oluşturulan soru maddelerinin açıklığa kavuşturulmasındaki rollerinin ve işlevlerinin tanımlanması ve sınıflandırılmasıyla ilgili zorlukları büyük ölçüde göz ardı etmektedir. Ayrıca, mevcut literatürde tanımlanan görsellerin klasik işlevleri olan bilgilendirici, dekoratif, düzenleyici, temsil edici, yorumlayıcı ve dönüştürücü işlevler (Carney & Levin, 2002), bu sınıflandırmanın OMÜ sırasında görsellerin nasıl oluşturulacağı veya kullanılacağı konusunda kapsamlı bir çerçeve sunmaması ve bu işlevlerden bazılarının bu süreçte çakışabileceğinden dolayı OMÜ'de kolayca kullanılamayabilir. Görsellerin soru maddelerinin kök (radical) ve arızi (incidental) kısımlarına nasıl katkıda bulunduğunu anlamak, üretilen soru maddelerinin psikometrik özelliklerini tanımlayabilmek için zorunludur. Bu nedenle, bu çalışma öncelikle, görsellerin nasıl oluşturuldukları ve soru maddelerini netleştirmeye nasıl yardımcı oldukları açısından, görsellerin belirli tür ve özelliklerini, mesela sorunun tanımına ilişkin temel bilgileri içerip içermedikleri, sorun bağlamının gerçek temsilini yansıtıp yansıtmadıkları veya gerçek mi yoksa figüratif mi oldukları gibi, dikkate alarak OMÜ sırasındaki işlevlerini özetlemek için kapsamlı bir çerçeve sunmayı amaçlamaktadır. Daha sonra bu çerçeve kullanılarak ÖSYM'nin ülke çapında yaptığı ve soruları açıklanan, TYT ve AYT gibi, bazı sınavlarda görselleri nasıl kullandığını ve yıllar içinde ÖSYM tarafından kullanılan görsellerin işlevleri açısından bir değişiklik olup olmadığı incelenecektir. Ülke çapında ÖSYM tarafından yapılan sınavlardaki görsel kullanımının son yıllarda önemli ölçüde arttığı aşikardır. Buna ilave olarak, OMÜ sırasında geleneksel madde modelleri kullanıldığında dinamik olarak görsel oluşturmanın teknik yönleri önemli zorluklar ortaya çıkarmaktadır. Bu nedenle soru maddelerinin içeriklerini ve yapılarını tanımlamak için algoritmik bir yaklaşım kullanılan üretken soru modeli kavramı tanıtılacaktır. Online test uygulaması sırasında dinamik olarak oluşturulan görsellerin gösterilmesiyle ilgili bazı güvenlik sorunları da ele alınacaktır. Son olarak, OMÜ sistemimiz tarafından üretken bir soru modeli kullanılarak dinamik olarak oluşturulan görseller içeren bazı soru maddeleri örnekleri sergilenecektir.

Anahtar Kelimeler: *Otomatik madde* üretimi*,* OMÜ, *otomatik soru* üretimi*, görsellerin işlevleri, görsellerin dinamik oluşturulması.*

Kaynakça

Carney, R. N., Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. Educational psychology review, 14, 5-26. https://doi.org/10.1023/A:1013176309260

[1] karakirik_e@ibu.edu.tr, ORCID: 0000-0001-7783-2006,Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

Dynamically Creation of Visuals And Their Functions In Automatic Item Generation

Erol KARAKIRIK [1]

There exists a very extensive literature on functions of visuals with respect to their contributions on understanding and learning concepts and solving problems. However, previous research has largely overlooked the challenges associated with the dynamic creation of visuals in automatic item generation (AIG) and defining and classifying their roles and functions in clarifying the generated question items. Furthermore, the classical functions of visuals identified in extant literature, namely informative, decorative, organizational, representative, interpretational and transformational functions (Carney & Levin, 2002), may not be easily employed in AIG since not only this classification fails to offer a comprehensive framework about how to create or make use of visuals during AIG but also some of these functionalities may overlap in this process. Furthermore, understanding how visuals contribute to the radical and incidental parts of question items is imperative to identifying their psychometric properties. Hence, this paper first and foremost aims to provide a comprehensive framework for outlining functions of visuals during AIG by taking certain types and properties of the visuals into account in terms of how they are created and how they help to clarify the question items, for instance whether they include essential information about the description of the problem, whether they portray actual representation of the problem context or whether they are veridical or figural. This framework will then be used to investigate how OSYM employed the visuals in some nationwide examinations whose question items were published and whether there has been a change in terms of functions of the visuals employed by OSYM over the years. It is evident that the use of visual in nationwide examinations by OSYM have increased significantly in recent years. Moreover, technical aspects of creating visuals dynamically during AIG present significant challenges when traditional item models are used. The concept of a generative item model employing an algorithmic approach to define the content and structure of an item will be introduced to simplify the creation of visuals dynamically. Some security issues about the online delivery of dynamically generated visuals during test administration will also be addressed. Finally, we will showcase examples of items including dynamically created visuals created with a generative item model by our AIG system.

Keywords: *automatic item generation, AIG, automatic question generation, functions of visuals, dynamically creation of visuals.*

Reference

Carney, R. N., Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. Educational psychology review, 14, 5-26. https://doi.org/10.1023/A:1013176309260

[1] karakirik_e@ibu.edu.tr, ORCID: 0000-0001-7783-2006, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Üretken Yapay Zekâ Modellerinin Çoktan Seçmeli Soru Üretme Performanslarının Karşılaştırması

Gökhan Akçapınar[1], Erkan Er[2], Ali Emre Karagül[3], Dilara Bakan Kalaycıoğlu[4]

ÖSYM çeşitli sınavlarda kullanmak üzere her yıl 10.000'in üzerinde yeni soruya ihtiyaç duymaktadır. Elektronik sınav uygulamalarının artmasıyla özellikle İngilizce sınavlarda (e-YDS, YÖKDİL, YDS, e-YÖKDİL ve YDT) ihtiyaç duyulan soru sayısı artmıştır. Bu artan ihtiyacın karşılanabilmesi için çok sayıda insan gücü gerekmektedir. Diğer taraftan, son yıllarda hayatımıza giren yapay zekâ temelli büyük dil modelleri insanlar tarafından gerçekleştirilen ve doğal dil işlemeyi gerektiren birçok görevi başarı ile yapabilmektedir. Bu dil modelleri soru yazma sürecinde insanlara yardımcı olabilir. Ancak, internet üzerinden erişilen ChatGPT vb. uygulamaların özellikle güvenliğin ve veri gizliliğinin ön planda olduğu durumlarda kullanılması uygun değildir. Diğer taraftan benzer amaçlarla ve aynı görevleri gerçekleştirmek için geliştirilen açık kaynak dil modelleri ise gelişmiş donanımlara sahip olmayan bilgisayarlarda dahi sorunsuz bir şekilde çalışabilmektedir. Aynı zamanda bu modellerin kullanımı için internet bağlantısı gerekmediğinden kullanıcı verilerinin gizliliği açısından da daha güvenilirdir. Bu çalışmada çevrimiçi ve çevrimdışı çalışan üretken dil modellerinin yabancı dil sınavlarına yönelik çoktan seçmeli soru üretme performansının karşılaştırmalı olarak incelenmesi amaçlanmıştır. Bu amaçla OpenAI firması tarafından geliştirilen GPT-4 büyük dil modeli ile açık kaynak bir dil modeli karşılaştırılmıştır. Bu çalışma kapsamında soru üretme süreci ve değerlendirme sonucu elde edilen bulgular paylaşılmıştır.

Anahtar Kelimeler: Üretken *yapay zekâ, büyük dil modeli, yabancı dil sınavı, soru* üretme*, GPT-4*

[1] gokhana@hacettepe.edu.tr, ORCID: 0000-0002-0742-1612, Hacettepe Üniversitesi, Ankara, Türkiye

[2] erkane@metu.edu.tr, ORCID: 0000-0002-9624-4055, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[3] me@emrekaragul.com, ORCID: 0000-0002-5820-8643, TOBB ETU, Ankara, Türkiye

[4] dilarabakan@gazi.edu.tr, ORCID: 0000-0003-1447-6918, Gazi Üniversitesi, Ankara, Türkiye

Comparison of the Performance of Generative AI Models in Multiple Choice Question Generation

Gökhan Akçapınar[1], Erkan Er[2], Ali Emre Karagül[3], Dilara Bakan Kalaycıoğlu[4]

OSYM requires over 10,000 new questions each year for various exams. With the rise of electronic exam applications, the number of questions needed for English exams (e-YDS, YÖKDİL, YDS, e-YÖKDİL, and YDT) has increased. Meeting this growing demand necessitates a significant amount of manpower. On the other hand, AI-based large language models, which have become prevalent, can successfully perform many tasks traditionally done by humans and that require natural language processing. These language models can assist humans in the question-writing process. However, using applications like ChatGPT, which are accessed via the internet, is not suitable in situations where security and data privacy are paramount. In contrast, open-source language models developed for similar purposes and to perform the same tasks can operate seamlessly even on computers without advanced hardware. Additionally, since these models do not require an internet connection, they offer greater reliability in terms of user data privacy. This study aims to comparatively examine the multiple-choice question generation performance of online and offline generative language models for foreign language exams. For this purpose, the GPT-4 large language model developed by OpenAI and an open-source language model were compared. The study presents the question generation process and the findings obtained from evaluation results.

Keywords: *Generative AI, large language model, foreign language exam, question generation, GPT-4*

[1] gokhana@hacettepe.edu.tr, ORCID: 0000-0002-0742-1612, Hacettepe Üniversitesi, Ankara, Türkiye

[2] erkane@metu.edu.tr, ORCID: 0000-0002-9624-4055, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[3] me@emrekaragul.com, ORCID: 0000-0002-5820-8643, TOBB ETU, Ankara, Türkiye

[4] dilarabakan@gazi.edu.tr, ORCID: 0000-0003-1447-6918, Gazi Üniversitesi, Ankara, Türkiye

Geniş Dil Modeliyle Otomatik Madde Üretimi: 5. Sınıf Matematik Uygulaması

Semih AŞİRET [1], Esat AVCI [2]

Geniş dil modellerinin otomatik madde üretmede etkili ve çok yönlü olduğunu gösteren birçok çalışma vardır. Ancak otomatik madde üretimine yönelik çok sayıda geniş dil modelleri uygulamasının genellikle sağlam bir eğitim temelinden yoksun olduğu gözlenmektedir (Tan ve diğ., 2024). Bu çalışmanın amacı, Türkiye Yüzyılı Maarif Modeli çerçevesinde, 5. sınıf matematik öğretim programında yer alan "Kategorik veri ile çalışabilme ve veriye dayalı karar verebilme" ile "Başkaları tarafından oluşturulan kategorik veriye dayalı istatistiksel sonuç veya yorumları tartışabilme" öğrenme çıktılarının ölçülmesine yönelik, geniş dil modelleri (GPT-4o) kullanılarak oluşturulan test ve maddelerin psikometrik özelliklerinin incelenmesidir.

Bu çalışmada, GPT modeline dayalı otomatik madde üretimi GPT modelini eğitimi, madde üretimi ve son olarak maddelerin nihai hale getirilmesi olmak üzere üç aşamada gerçekleştirilmiştir. İlk aşamada, ölçülmek istenen öğrenme çıktıları ve süreç bileşenleri ayrıntılı olarak tanımlanmış; istatistiksel bağlam oluşturma ölçütleri ile madde formatı belirlenmiştir. Bu bağlamda, her bir öğrenme çıktısını ve süreç bileşenini ölçmeye yönelik olarak GPT modelinden her bağlam için 2-3 madde üretmesi istenmiştir. Üretilen maddeler, alanında uzman üç kişi tarafından içerik doğruluğu ve bir dil uzmanı tarafından dilbilgisi açısından incelenmiş; gerekli düzeltmeler yapılarak dört seçenekten oluşan 30 maddelik bir test geliştirilmiştir.

Test maddelerinin ölçme ve değerlendirme kriterlerine uygunluğunu, istatistiksel bağlamın geçerliliğini ve öğrenme çıktısını ölçme yeterliliğini değerlendirmek amacıyla bir kontrol listesi oluşturulmuştur. Bu liste, 5 alan uzmanı ve 3 ölçme-değerlendirme uzmanı olmak üzere toplam 8 uzman tarafından kullanılarak her madde detaylı bir şekilde değerlendirilmiştir. Uzman değerlendirmeleri sonucunda, 20 maddelik nihai test oluşturulmuştur.

Geliştirilen testin, Eylül ayının ikinci haftasında 500 altıncı sınıf öğrencisine uygulanması planlanmaktadır. Uygulama sonrasında, testin madde ve test düzeyindeki psikometrik özellikleri klasik test kuramı ve Rasch modele göre analiz edilerek raporlanacaktır. Çalışmanın sonucunda, GPT-4 tabanlı olarak geliştirilen testin yüksek geçerlik ve güvenirlik düzeyine sahip olması ve adil bir değerlendirme sunması beklenmektedir.

Anahtar Kelimeler: *Geniş dil modelleri, otomatik madde* üretme*, istatistik eğitimi, eğitsel* ölçme*.*

Kaynakça

Tan, B., Armoush, N., mazzullo, E., Bulut, O., & Gierl, M. J. (2024, April 8). A Review of Automatic Item Generation Techniques Leveraging Large Language Models. https://doi.org/10.35542/osf.io/6d8tj

[1] semihasiret@gmail.com:, ORCID ID: 0000-0002-0577-2603, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

[2] esatuavci@gmail.com:, ORCID: 0000-0003-2366-4515, Millî Eğitim Bakanlığı, Mersin, Türkiye

Automatic Item Generation with the Large Language Model: Grade 5 Mathematics Application

Semih AŞİRET [1], Esat AVCI [2]

There are many studies showing that large language models are effective and versatile in automatic item generation (Tan et al., 2024). However, it is observed that many applications of large language models for automatic item generation often lack a solid educational foundation. The purpose of this study is to examine the psychometric properties of the tests and items created by using the GPT-4o to measure the learning outcomes of "Working with categorical data and making data-based decisions" and "Discussing statistical results or interpretations based on categorical data generated by others" in the 5th grade mathematics curriculum within the framework of the Turkish Century Education Model.

In this study, automatic item generation based on the GPT-4o model was carried out in three stages: training the GPT model, item generation and finally finalization of the items. In the first stage, the learning outcomes and components of them to be measured were defined in detail; statistical contextualization criteria and item format were determined. In this context, the GPT model was asked to generate 2-3 items for each context to measure each learning outcomes and components of them. The items were examined for content accuracy by three field-experts and for grammar by a language expert; a 30-item test consisting of four options was developed by making the necessary corrections.

A checklist was developed to assess the appropriateness of the test items to the measurement and evaluation criteria, the validity of the statistical context and the adequacy of measuring the learning outcome. This checklist was used by a total of 8 experts, 5 field experts and 3 measurement and evaluation experts, and each item was evaluated in detail. As a result of the expert evaluations, a final test of 20 items was created.

The developed test is planned to be administered to 500 sixth grade students in the second week of September. After the implementation, the psychometric properties of the test at item and test level will be analyzed and reported according to classical test theory and Rasch model. As a result of the study, the test developed based on GPT-4o is expected to have a high level of validity and reliability and to provide a fair assessment.

***Keywords**: Large language models, automatic item generation, statistics education, educational measurement.*

Reference

Tan, B., Armoush, N., mazzullo, E., Bulut, O., & Gierl, M. J. (2024, April 8). A Review of Automatic Item Generation Techniques Leveraging Large Language Models. https://doi.org/10.35542/osf.io/6d8tj

[1] semihasiret@gmail.com:, ORCID ID: 0000-0002-0577-2603, Mersin Üniversitesi Eğitim Fakültesi, Mersin, Türkiye

[2] esatuavci@gmail.com:, ORCID: 0000-0003-2366-4515, Ministry of National Education, Mersin, Turkiye

Klinik Akıl Yürütme Becerisinin Değerlendirmesinde Otomatik Soru Üretim Yöntemlerinin Karşılaştırılması

Betül Nalan Karahan [1], Emre Emekli [2]

Tıp fakültesi lisans eğitimde klinik akıl yürütme (KAY) becerisinin kazandırılması gerekmekte ve klinik öncesi dönemde KAY değerlendirmesi çoktan seçmeli sorular (ÇSS) ile yapılmaktadır (Daniel ve ark.,2019). ÇSS, yapay zeka (YZ) ile üretilen şablon tabanlı otomatik soru üretimi (ŞT-OSÜ) ve yazılımlarla üretilen şablon tabanlı olmayan OSÜ (ŞTO-OSÜ) olmak üzere iki yöntemle otomatik üretilebilir (Gierl ve ark.,2013). Bu çalışmanın amacı ŞT-OSÜ, ŞTO-OSÜ ile üretilen ve halihazırda kullanılan sorular hakkında uzman görüşü alarak, soruların KAY değerlendirilmesinde kullanılabilirliğini araştırmaktır. Abdominal aciller hakkında soru üretilmesi kararlaştırıldı. ŞT-OSÜ (içeriğin belirlenmesi, soru şablonu oluşturulması, yazılım ile soru üretilmesi) ile yaş, cinsiyet, semptom, öykü, laboratuvar, fizik muayene başlıkları belirlendi ve olası değişkenler yazıldı. Soru şablonu oluşturularak 7260 soru üretildi. ŞTO-OSÜ için ChatGPT-4o ile 17 soru üretildi (Kıyak&Emekli,2024). Rastgele altışar soru (18) seçilerek form ile değerlendirildi (a-Soru metni anlaşılırdır, b-Soru, klinik açıdan uygundur, c-Sorunun tek bir doğru cevabı vardır, d-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, e-Çeldiriciler mantıklıdır, f-Soru uzmanların cevaplaması için bile zordu, g-Soru tıp öğrencileri için uygun zorluktaydı). Formu 14 kişi (radyoloji hekimi) cevapladı. Katılımcılar soruların KAY ölçtüğünü ifade etmekteydi (ŞT- OSÜ 66/84 (%78,57), güncel sorular 69/84 (%82.14), ŞTO-OSÜ (%84,52)). Evet cevabı ortalamaları ŞT- OSÜ, güncel sorular, ŞTO-OSÜ için sırasıyla “a” 11,83-11,67-13; “b” 8,33-9,17-11,5; “c” 7,17-7,67-9; “d” 8-9,33-10,17; “e” 6,67-8,67-10,67; “f” 2,83-1-0,5; “g” 6,17-6,5-8,17 şeklindeydi. Üç soru üretme yöntemi ile de KAY ölçebilecek sorular üretilebilmektedir. Soru metninin anlaşılırlığı, klinik açıdan uygunluğu ve doğru cevabı bulmak için verilen bilgileri ölçen başlıklar üç soru üretme tekniği için de yüksektir. ŞTO-OSÜ ile üretilen sorulara birçok parametrede daha çok evet cevabı verilmiştir. YZ ile ölçmede kullanılabilecek soru üretilebilir.

Anahtar Kelimeler: *Otomatik Soru* Üretimi*, Klinik Akıl Yürütme, Yapay Zeka*

Kaynakça

Daniel, M., Rencic, J., & Durning, S. J., et al. (2019). Clinical reasoning assessment methods: A scoping review and practical guidance. *Academic Medicine*, 94(6), 902-912.

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, 46(8), 757-765.

Kıyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, qgae065

[1] nalanksgl@gmail.com, ORCID: 0000-0003-4958-330X, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

[2] emreemekli90@gmail.com, ORCID: 0000-0001-5989-1897, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

Comparison of Automatic Question Generation Methods in Assessing Clinical Reasoning Skills

Betül Nalan Karahan [1], Emre Emekli [2]

Clinical reasoning (CR) skills are essential in undergraduate medical education, and in the preclinical period, CR assessment is conducted using multiple-choice questions (MCQs) (Daniel et al., 2019). MCQs can be automatically generated using two methods: template-based automatic item generation (TB- AIG) with artificial intelligence (AI) and non-template-based AIG (NTB-AIG) produced with software (Gierl et al., 2013). This study aims to investigate the usability of questions generated by TB-AIG, NTB- AIG, and currently used questions in assessing CR, based on expert opinions. For TB-AIG, headings such as age, gender, symptoms, history, laboratory, and physical examination were defined, and potential variables were written. A question template was created, and 7260 questions were generated. For NTB-AIG, 17 questions were generated using ChatGPT-4 (Kıyak & Emekli, 2024). A random selection of six questions (18 in total) was evaluated using a form, assessing the understandability, clinical appropriateness, single correct answer, sufficiency of information, logical distractors, difficulty for experts, and appropriate difficulty for students. Fourteen radiology physicians participated. They indicated that the questions assessed CR effectively: TB-AIG 66/84 (78.57%), current questions 69/84 (82.14%), NTB-AIG (84.52%). The average "Yes" responses for TB-AQG, current questions, and NTB- AQG were as follows: "a" 11.83-11.67-13; "b" 8.33-9.17-11.5; "c" 7.17-7.67-9; "d" 8-9.33-10.17; "e" 6.67-8.67-10.67; "f" 2.83-1-0.5; "g" 6.17-6.5-8.17. Questions capable of measuring CR can be produced using all three question generation methods. The headings measuring the understandability of the question text, clinical appropriateness, and the sufficiency of the information provided to find the correct answer are high for all three question generation techniques. Questions generated by NTB- AQG received more "Yes" responses in many parameters. Questions that can be used in assessments with AI can be produced.

Keywords: *Automatic Question Generation, Clinical Reasoning, Artificial Intelligence*

Reference

Daniel, M., Rencic, J., & Durning, S. J., et al. (2019). Clinical reasoning assessment methods: A scoping review and practical guidance. *Academic Medicine*, 94(6), 902-912.

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, 46(8), 757-765.

Kıyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, qgae065

[1] nalanksgl@gmail.com, ORCID: 0000-0003-4958-330X, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

[2] emreemekli90@gmail.com, ORCID: 0000-0001-5989-1897, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

Öğrenci ve Okul Özellikleri, Öğretme ve Öğrenme Süreci ile İlgili Değişkenler ve Bilişsel Olmayan Öğrenme Çıktılarının PISA 2015 Fen Okuryazarlığı Üzerindeki Etkilerinin İncelenmesi*

Burcu ORHAN [1], Ömer GEBAN [2], Eren CEYLAN [3]

Nitelikli bireylerin yetiştirilmesinde ülkelerin benimsediği eğitim politikaları önemli bir yere sahiptir. Bu bağlamda ülkelerin öğrencilerinin başarılarını artırmak, eğitim sistemlerini iyileştirmek ve sosyo- ekonomik ve kültürel statüleri farklı olan öğrencilere fırsat eşitliği sağlamak için eğitim çıktılarını ulusal veya uluslararası düzeyde değerlendirmeleri gerekmektedir. Uluslararası büyük ölçekli değerlendirmelerden biri olan Uluslararası Öğrenci Değerlendirme Programı (PISA), zorunlu eğitim gören 15 yaşındaki öğrencilerin modern toplumda yerlerini almak için ihtiyaç duydukları temel bilgi ve becerilere ne ölçüde sahip olduklarını belirlemeyi amaçlayan bir araştırma projesidir. Küreselleşen dünyada bilim ve teknolojideki hızlı gelişme ve ilerlemeye ayak uydurabilen fen okuryazarı bireyler yetiştirmek birçok fen eğitimi programının temel amacıdır. PISA 2015 değerlendirme çerçevesinde fen okuryazarlığının boyutlarına odaklanılmıştır (OECD, 2017). Bu çalışmada, PISA 2015'e ait veriler kullanılarak fen ilişkili öğrenci ve okul özelliklerinin Türkiye'deki 15 yaşındaki öğrencilerin fen okuryazarlığına etkisinin araştırılması amaçlanmaktadır. Veri analizi hiyerarşik lineer modelleme (HLM) kullanılarak yapılmıştır. Okul dışında fen çalışmaya ayrılan zaman, fen aktiviteleri ve çevresel optimizmin fen okuryazarlığını negatif yönde etkilediği; fen öğrenme zamanı, epistemolojik inançlar, çevresel farkındalık ve öğrencilerin sahip olmayı umdukları mesleki statünün ise fen okuryazarlığını pozitif yönde etkilediği bulunmuştur. Fen öğretmenlerinin sayısının tüm öğretmenlerin sayısına oranı ve fen bilimlerine özgü kaynakların fen okuryazarlığını pozitif yönde etkilediği, eğitim materyali eksikliği ve okuldaki toplam fen öğretmeni sayısının ise fen okuryazarlığı üzerinde negatif etkisi olduğu tespit edilmiştir.

Anahtar Kelimeler: *Uluslararası* Öğrenci *Değerlendirme Programı (PISA 2015), Fen Okuryazarlığı*

Kaynakça:

Organisation for Economic Co-operation and Development (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (Revised Edition). Paris: PISA, OECD Publishing.

[1] burcu.orhan@osym.gov.tr, ORCID: 0000-0001-8555-5643, Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı, Ankara, Türkiye

[2] geban@metu.edu.tr, ORCID: 0000-0002-9433-0056, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

[3] eceylan@ankara.edu.tr, ORCID: 0000-0001-8244-8260, Ankara Üniversitesi, Ankara, Türkiye

*Bu bildiri Prof. Dr. Ömer GEBAN danışmanlığında yürütülen "Investigation of the effect of student and school background variables, teaching and learning variables and non-cognitive outcomes on the components of scientific literacy in Programme for International Student Assessment (PISA 2015)" adlı doktora tezinden üretilmiştir.

Investigation of the Effect of Student and School Background Variables, Teaching and Learning Variables and Non-Cognitive Outcomes on the Components of Scientific Literacy in PISA 2015*

Burcu ORHAN [1], Ömer GEBAN [2], Eren CEYLAN [3]

Education policies adopted by countries play a crucial role in fostering qualified individuals. In this context, it is essential for countries to assess educational outcomes at national or international levels to enhance student achievements, improve education systems, and ensure equal opportunities for students with diverse socio-economic and cultural backgrounds. One of the major international assessments is the Programme for International Student Assessment (PISA), which aims to determine the extent to which 15-year-old students, who are undergoing compulsory education, possess the fundamental knowledge and skills needed to participate effectively in modern society. Training individuals who are adept at navigating the rapid developments in science and technology in a globalized world is a primary goal of many science education programs. The PISA 2015 assessment framework focused on the dimensions of scientific literacy (OECD, 2017). This study focuses on investigating the impact of student and school-related factors related to science on the science literacy of 15-year-old students in Turkey using data from PISA 2015. Data analysis was conducted using hierarchical linear modeling (HLM). It was found that out-of-school study time for science, science activities and environmental optimism had negative effect on scientific literacy, while science learning time, epistemological beliefs, environmental awareness and students' expected occupational status had positive effect on scientific literacy. Proportion of science teachers by all teachers and science specific resources were found to be effecting scientific literacy positively, while shortage of educational material and total number of science teachers at school had negative effect on scientific literacy.

Keywords: *Programme for International Student Assessment (PISA 2015), Scientific Literacy*

Reference:

Organisation for Economic Co-operation and Development (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (Revised Edition). Paris: PISA, OECD Publishing.

---

[1] burcu.orhan@osym.gov.tr, ORCID: 0000-0001-8555-5643, The Measuring, Selection and Placement Center, Ankara, Türkiye

[2] geban@metu.edu.tr, ORCID: 0000-0002-9433-0056, Middle East Technical University, Ankara, Türkiye

[3] eceylan@ankara.edu.tr, ORCID: 0000-0001-8244-8260, Ankara University, Ankara, Türkiye

*This report is derived from the doctoral thesis titled "Investigation of the effect of student and school background variables, teaching and learning variables, and non-cognitive outcomes on the components of scientific literacy in the Programme for International Student Assessment (PISA 2015)", conducted under the supervision of Prof. Dr. Ömer GEBAN.

Test Eşitlemede Ortak Test Desenine Karşı Ortak Değişken Deseni Yaklaşımları

Feyzi Güneş [1], Burcu Atar [2]

Bireyler hakkında alınacak kararlarda testler önemli görevler üstlenmektedir. Eğitimde ve psikolojide ölçülmesi hedeflenen özelliklerin bireylerdeki var olan düzeyinin testler aracılığı ile ortaya konuyor olması testleri vazgeçilmez kılmaktadır. Özellikle de yüksek risk içeren sınavlarda testi alanlar için çok küçük puanlar dahi sıralamaya etki edebilmektedir. Hal böyle olunca test uygulayıcıları elde edilen puanların güvenirliğinin maksimum düzeyde sağlamak adına farklı form uygulama yoluna gidebilmektedir. Farklı form uygulamaları testlerin birbirine paralel olduğu sayıltısı üzerine kuruludur. Ancak test maddelerinin eşit güçlükte olmaması, kapsam ve psikometrik özellikler yönüyle ayrışması testlerin paralelliği üzerinde bir engeldir. Paralellik teorik olarak mümkün olsa da uygulamada formlar arasında güçlük farklılıkları kaçınılmaz olduğunu bu gerekçeyle bir testin farklı formlarından elde edilen puanların karşılaştırılabilir olması için eşitleme gibi bir sürecin gerekliliği doğmaktadır. Eşitleme işleminden sonra farklı formlardan elde edilen puanlar karşılaştırılabilir olduğu kabul edilmektedir (Angoff, 1984). Ülkemizde Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından Akademik Personel ve Lisansüstü Eğitim Giriş Sınavı (ALES), Yabancı Dil Sınavı (YDS) ve Yükseköğretim Kurumları Yabancı Dil Sınavlarından (YÖKDİL) elde edilen puanlar 5 yıl geçerli kabul edilmektedir (ÖSYM, 2020). Herhangi bir istatistiksel dönüşüm uygulamadan farklı formlardan elde edilen puanların eşdeğer kabul edilmesi adillik anlamında soru işaretleri oluşturmaya müsaittir. Ayrıca eşitleme gibi bir süreç yürütülmek istense formlar arasında ortak maddeler bulunmamaktadır. Böylesi durumlar için ölçülen özellikle ilişkili kovaryant değişkenler aracılığı ile bilgi toplanması ve bu bilgilerin ışığında eşitleme işleminin yürütülmesi yararlı bir yol olarak kabul edilebilir. Bu araştırmada ortak madde içeren test tasarımlarında eğer ortak madde yerine kovaryant değişken kullanılması durumunda ortak değişkenlere bağlı yapılacak eşitleme sonuçlarının ortak madde içeren eşitleme durumlarına göre ne derecede etkili olacağının araştırılması planlanmaktadır. Ortak madde sayısının, ortak maddelerden elde edilen puanların ve kovaryant değişkenlerin toplam test puanı ile göstereceği benzerlik ölçüsünde ortak değişkenlere bağlı eşitleme yöntemlerinin etkililiği değerlendirilecektir.

Anahtar Kelimeler: *Test eşitleme, ankor, kovaryant.*

Kaynakça

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service. ÖSYM. (2020). *2020 Akademik personel ve lisansüstü eğitimi giriş sınavı (ALES) kılavuzu*.

https://dokuman.osym.gov.tr/pdfdokuman/2020/ALES-1/aleskilavuz15052020.pdf

[1] feyzigunes@gmail.com:, ORCID: 0000-0003-2386-5114, TTKB, Ankara, Türkiye

[2] burcua@hacettepe.edu.tr:, ORCID: 0000-0003-3527-686X, Hacettepe Üniversitesi, Ankara, Türkiye

Anchor Test Design Versus Covariates Design Approaches in Test Equating

Feyzi Güneş [1], Burcu Atar [2]

Tests play an important role in decisions to be taken about individuals. The fact that the existing level of the characteristics aimed to be measured in education and psychology is revealed through tests makes tests indispensable. Especially in high-stakes exams, even very small scores can affect the ranking for those taking the test. In this case, test practitioners may choose to apply different forms in order to ensure maximum reliability of the scores obtained. Different form applications are based on the assumption that the tests are parallel to each other. However, the fact that the test items are not of equal difficulty and that they are separated in terms of scope and psychometric properties is an obstacle to the parallelism of the tests. Although parallelism is theoretically possible, in practice, difficulty differences between the forms are inevitable, and for this reason, a process such as equating arises in order for the scores obtained from different forms of a test to be comparable. After the equating process, the scores obtained from different forms are accepted to be comparable (Angoff, 1984). In our country, the scores obtained from the Academic Personnel and Postgraduate Education Entrance Exam (ALES), Foreign Language Exam (YDS) and Higher Education Institutions Foreign Language Exams (YÖKDİL) by the Measurement, Selection and Placement Center (ÖSYM) are considered valid for 5 years (ÖSYM, 2020). Accepting the scores obtained from different forms as equivalent without applying any statistical transformation is likely to create question marks in terms of fairness. In addition, if a process such as equating is desired to be carried out, there are no common items between the forms. For such cases, collecting information through covariant variables related to the measured feature and carrying out the equating process in the light of this information can be considered as a useful way. In this study, it is planned to investigate to what extent the equating results based on common variables will be effective compared to equating situations that include common items if a covariant variable is used instead of a common item in test designs containing common items. The effectiveness of equating methods based on common variables will be evaluated in terms of the similarity between the number of common items, the scores obtained from common items and the total test score of the covariant variables.

Keywords: *Test equating, Anchor, Covariate*

Reference

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

ÖSYM. (2020). *2020 Akademik personel ve lisansüstü eğitimi giriş sınavı (ALES) kılavuzu*. https://dokuman.osym.gov.tr/pdfdokuman/2020/ALES-1/aleskilavuz15052020.pdf

[1] feyzigunes@gmail.com:, ORCID: 0000-0003-2386-5114, TTKB, Ankara, Türkiye

[2] burcua@hacettepe.edu.tr:, ORCID: 0000-0003-3527-686X, Hacettepe Üniversitesi, Ankara, Türkiye

Standart Belirlemede Yapay Zekâ Kullanımı

Mahmut Sami Koyuncu[1], Mehmet Şata[2]

Yapay zekâ araçları modern hayatın birçok alanında giderek daha yaygın olarak kullanılmaya başlanmıştır. Belirgin olarak kullanıldığı alanlardan biriside eğitimdir. Eğitimde ölçme ve değerlendirme süreçlerinde kullanılan testler, ölçekler ve diğer ölçme araçlarından elde edilen puanlar, bireylerin başarı düzeylerini belirlemek ve yeterliliklerini değerlendirmek için kritik bir rol oynamaktadır. Bu puanların doğru ve adil bir şekilde yorumlanabilmesi için kesme puanlarının doğru belirlenmesi gerekmektedir. Geleneksel yöntemlerle kesme puanı belirleme süreci, genellikle uzman görüşlerine dayanan sübjektif kararlarla yapılmaktadır. Ancak bu süreç hem zaman alıcı olabilmekte hem de değerlendirmede tutarsızlıklara yol açabilmektedir. Yapay zekanın eğitimde yaygın olarak kullanılmasıyla birlikte, ölçme ve değerlendirme alanında kullanılan değerlendirme yöntemlerinde, otomatik puanlama, içerik analizi, kişiselleştirilmiş geri bildirim verme vb. konularda devrim niteliğinde ilerlemeler sağlamıştır (Bulut vd., 2024). Bu çalışmada, yapay zekâ da dayalı araçların kesme puanı belirleme sürecine nasıl entegre edilebileceği ve daha nesnel, hızlı ve tutarlı sonuçların nasıl elde edilebileceği ele alınmıştır. Çalışma kapsamında ayrıca yapay zekâ yöntemleri ile kesme puanı belirlemenin avantajları, olası zorlukları ve uygulanabilirliği tartışılmıştır. Çalışma da ChatGPT'ye yer alan Data Analyst, Data Analysis & Report Al vb. eklentileri kullanılarak çeşitli veri setleri üzerinden farklı standart belirleme yöntemlerine (ROC Eğrisi, Kümeleme Analizi vb.) göre kesme puanları belirlenmiştir. Yapay zekaya dayalı elde edilen kesme puanları istatistik paket programlarında elde edilen sonuçlarla karşılaştırılarak incelenmiştir. Elde edilen sonuçlar incelendiğinde yapay zekâ ile belirlenen kesme puanlarının, grafiklerin, anlamlılık test sonuçlarının istatistik paket programları kullanılarak elde edilen kesme puanları ile örtüştüğü görülmüştür. Sonuç olarak, yapay zekâ destekli kesme puanı belirleme yöntemlerinin, eğitimde ölçme ve değerlendirme alanında devrim niteliğinde yenilikler getirebileceği görülmektedir. Özellikle istatistik paket program kullanımı konusunda yeterli bilgi ve beceriye sahip olmayan araştırmacılara fayda sağlayacağı aşikardır.

Anahtar Kelimeler: *Standart Belirleme, Yapay Zekâ, Kesme Puanı, Geçerlik, ChatGPT*

Kaynakça

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D.,... & Morilova, P. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *arXiv preprint arXiv:2406.18900*.

[1] mskoyuncu@aku.edu.tr, ORCID: 0000-0002-6651-4851, Afyon Kocatepe Üniversitesi, Afyonkarahisar, Türkiye

[2] mehmetsata@yyu.edu.tr, ORCID: 0000-0003-2683-4997, Van Yüzüncü Yıl Üniversitesi, Van, Türkiye

Using Artificial Intelligence in Standard Setting

Mahmut Sami Koyuncu [1], Mehmet Şata [2]

Artificial intelligence tools have become increasingly widely used in many areas of modern life. One of the areas where they are used prominently is education. Scores obtained from tests, scales, and other measurement tools used in measurement and evaluation processes in education play a critical role in determining the success levels of individuals and evaluating their competencies. For these scores to be interpreted accurately and fairly, cut-off scores need to be determined correctly. Determining cut scores with traditional methods is usually done with subjective decisions based on expert opinions. However, this process can be time-consuming and lead to inconsistencies in the evaluation. The widespread use of artificial intelligence in education has revolutionized the assessment methods used in assessment and evaluation, such as automatic scoring, content analysis, personalized feedback, etc. (Bulut et al., 2024). This study discusses how artificial intelligence-based tools can be integrated into determining the cut score and how more objective, fast, and consistent results can be obtained. Within the scope of the study, the advantages, possible difficulties, and applicability of determining cut-off scores with artificial intelligence methods were also discussed. The study determined cut-off scores according to different standard setting methods (ROC Curve, Cluster Analysis, etc.) over various data sets using Data Analyst, Data Analysis & Report AI, etc. plug-ins included in ChatGPT. The cut-off scores obtained based on artificial intelligence were compared with the results obtained in statistical package programs. When the results obtained were examined, it was seen that the cut-off scores, graphs, and significance test results determined by artificial intelligence overlapped with the cut-off scores obtained using statistical package programs. As a result, it is seen that artificial intelligence-supported cut score determination methods can bring revolutionary innovations in the measurement and evaluation of education. It will significantly benefit researchers who need more knowledge and skills in the use of statistical package programs.

Keywords: *Standard Setting, Artificial Intelligence, Cut Score, Validity, ChatGPT*

Reference

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D.,... & Morilova, P. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *arXiv preprint arXiv:2406.18900*.

[1] mskoyuncu@aku.edu.tr, ORCID: 0000-0002-6651-4851, Afyon Kocatepe University, Afyonkarahisar, Türkiye

[2] mehmetsata@yyu.edu.tr, ORCID: 0000-0003-2683-4997, Van Yüzüncü Yıl Üniversitesi, Van, Türkiye

KTK, MTK ve D-Puanlama Yöntemlerinin Karşılaştırılması ve Standart Belirleme Süreçleri Üzerine Bir Uygulama

Esra Oyar[1], Ayşenur Erdemir[2], Serpil Çelikten-Demirel[3]

Öğrenci yanıtlarının puanlanmasında Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) sıklıkla kullanılmaktadır. KTK'da yanıtlayıcıların puanı toplam doğru yanıt sayısı ile oluşturulmakta iken MTK'da madde ve ilgili yetenek arasındaki ilişki dikkate alınmaktadır. Bu yöntemlerin birtakım sınırlılıkları bulunmaktadır. Bu sınırlılıklar göz önünde bulundurularak Dimitrov (2017) tarafından yeni bir yöntem olan D-puanlama yöntemi önerilmiştir. Burada kişinin d-puanı kişinin yanıt örüntüsünün madde güçlükleri ile ağırlıklandırılmasına dayanmaktadır. Madde ve yetenek ilişkisini dikkate alan ve MTK varsayımlarına karşı güçlü olan d-puanlama yöntemi ile elde edilen puanlar d-ölçeğinde [0,1] sınırlandırılmıştır. Elde edilen puanların kullanım amaçlarından biri de kişilerin sınıflandırılmasıdır. Bu sınıflandırmalar için oluşturulan performans standartları, bir sınava katılan kişilerin belirli bir performans kategorisinde sınıflandırılması için ilgili sınavda hangi düzeyde performans göstermesi gerektiğini belirtir. Bu düzeylerin belirlenmesi için standart belirleme çalışmaları yapılmaktadır. Bu çalışmanın ilk aşamasında KTK, MTK ve d-puanlamaya göre madde ve yetenek parametreleri elde edilip karşılıklı değerlendirilecektir. Bir sonraki aşamada hem kişi merkezli hem de veri merkezli yaklaşımlar ile standart belirleme yapılacak ve belirlenen kesme puanları ve bu kesme puanlarından yola çıkılarak yapılan sınıflama indekslerine dayalı olarak kesme puanlarının performansları karşılaştırılacaktır. Kişi merkezli yaklaşımlar kapsamında düzenlenmiş-Angoff; veri merkezli yaklaşımlardan ise kümeleme analizi, gizil sınıf analizi ve d-puanlama yanıt örüntüsü yöntemi (Response Vector Mastery [RVM]; Dimitrov, 2021) ile kesme puanı belirleme ele alınacak olup ele alınan bu yöntemlerin performansı karşılaştırılacaktır. RVM yönteminde maddeler uzmanlara sunularak her alt boyut için minimum yeterlik seviyesindeki kişinin hangi maddeleri doğru yanıtlamasının yeterli olacağı sorulmaktadır. Buna dayalı olarak yapılan 0-1 kodlama ile yeterlik için yanıt örüntüsü oluşturulmaktadır. Bu örüntüye göre her bir alt boyut ve ölçme aracının tamamı için d-puanı hesaplanarak kesme puanı belirlenmektedir. Çalışmada 25 maddeden oluşan ve 340 kişiye uygulanan ölçme ve değerlendirme başarı testi kullanılmıştır. Düzenlenmiş-Angoff yöntemi için yaklaşık 12-15; RVM yöntemi için yaklaşık 5-6 ölçme ve değerlendirme alanında doktora derecesine sahip uzmanlara ulaşılması hedeflenmektedir. Sonuç olarak üç yöntemden elde edilen madde ve yetenek parametreleri karşılaştırılacaktır. Her bir yönteme ilişkin kesme puanları raporlanacaktır. Sınıflama doğruluğu ve tutarlılık indeksleri tüm yöntemler için hesaplanacak ve karşılaştırmalı bir şekilde literatürle tartışmalı olarak değerlendirilecektir.

Anahtar Kelimeler: *standart belirleme, d-puanlama, sınıflama doğruluğu.*

[1] esratas@gazi.edu.tr, ORCID: 0000-0002-4337-7815, Gazi Üniversitesi, Ankara, Türkiye

[2] erdemiraysenur@gmail.com, ORCID: 0000-0001-9656-0878, Polis Akademisi, Ankara, Türkiye

[3] sserpilcelikten@gmail.com, ORCID: 0000-0003-3868-3807, Dicle Üniversitesi, Diyarbakır, Türkiye

Comparison of KTK, MTK and D-Scoring Methods and an Application on Standard Setting Processes

Esra Oyar[1], Ayşenur Erdemir[2], Serpil Çelikten-Demirel[3]

Classical Test Theory (CTT) and Item Response Theory (IRT) are two of the most commonly employed approaches to scoring student responses. In CTT, the score of the respondents is formed by the total number of correct answers, while in IRT, the relationship between the item and the related ability is taken into account. However, there are some limitations about these methods. In light of these constraints, a new method, D-scoring method, was proposed by Dimitrov (2017) in this study. In this method, the d-score of the individual is based on weighting the response pattern of the individual with the item difficulties. One of the key applications of the obtained scores is the classification of individuals. The performance standards that have been established for these classifications specify the level of performance that a test-taker must demonstrate in the relevant test in order to be classified in a particular performance category. Standard setting studies are being carried out to determine these levels. In the initial phase of this investigation, item and ability parameters will be obtained and evaluated in conjunction with the CTT, IRT and d-scoring. In the subsequent phase, both person- centred and data-centred methodologies will be employed for the establishment of standards and the evaluation of the performance of the cut-off scores will be conducted in comparison with the aforementioned cut-off scores and the classification indices derived from these cut-off scores. Within the scope of person-centered approaches, Angoff; and from data-centered approaches, cluster analysis, latent class analysis and d-scoring response pattern method (Response Vector Mastery [RVM]; Dimitrov, 2021) will be discussed to determine the cut-off score, and the performance of these methods will be compared on classification indices. In the RVM method, the items are presented to the experts. Based on this information, a response pattern for competence is created using a 0-1 coding system. By this pattern, the cut-off score is calculated by determining the D-score for each sub- dimension and the entire measurement tool in the study, a measurement and evaluation achievement test, comprising 25 items, was administered to a sample of 340 individuals. For the Modified-Angoff method, approximately 12-15 experts with a doctorate degree in the field of measurement and evaluation, and for the RVM method, 5-6 experts will be consulted. Subsequently, the item and ability parameters obtained from the three methods will be subjected to a comparative analysis. The results will be presented in the form of cut-off scores for each method. The classification accuracy and consistency indices will be calculated for all methods and evaluated in comparison with the existing literature.

Keywords: *standard setting, d-scoring, classification indices.*

[1] esratas@gazi.edu.tr, ORCID: 0000-0002-4337-7815, Gazi University, Ankara, Türkiye

[2] erdemiraysenur@gmail.com, ORCID: 0000-0001-9656-0878, Police Academy, Ankara, Türkiye

[3] sserpilcelikten@gmail.com, ORCID: 0000-0003-3868-3807, Dicle University, Diyarbakır, Türkiye

İngilizce Yeterlik Sınavı Dinleme Alt Testinin Geçerlik Kanıtlarının İncelenmesi

Sebahat GÖREN [1], Selahattin GELBAL [2]

Ortak bir uyarana bağlı maddelerden oluşan madde takımlarının kullanımı; yabancı dil testlerine zaman ve çaba açısından ekonomiklik sağlama, içerik/bağlam etkisini azaltma, üst düzey becerilerin ölçülebilmesi gibi önemli avantajlar sunar (Eckes, 2014; Wainer & Kiely, 1987). Madde takımlarının bu avantajlarından yararlanabilmek için öncelikle en uygun Madde Tepki Kuramı (MTK) modelinin seçilmesi ve Bilgisayarda Bireyleştirilmiş Test (BBT) uygulamalarına entegre etmede gerekli stratejilerin dikkatli bir şekilde uygulanması gerekir (Min & He, 2014; Wainer vd., 2000). Bu çalışma, bir BBT uygulamasının madde havuzunun geçerliğiyle ilgili kanıtlar toplamayı amaçlamaktadır. Bu amaçla madde havuzu kalibrasyonu için uygun MTK modellerinin belirlenmesi ve ölçekleme işleminin gerçekleştirilmesi adımları uygulanacaktır. Bu çalışmanın veri setini, özel bir üniversitedeki 915 hazırlık sınıfı öğrencilerine dokuz farklı form şeklinde uygulanan İngilizce yeterlik sınavının dinleme alt testinden alınan cevaplar oluşturmaktadır. MTK ile elde edilen sonuçlar Rasch modelinin diğer MTK modellerinden daha iyi veri uyumu sağladığını göstermiştir. Yapı geçerliğine ilişkin MTK'nın varsayımları da olan faktör analizi (tek boyutluluk) ve Yen'in Q3 istatistiği (yerel bağımsızlık) ile araştırılmıştır. Ortak maddeye dayalı eş zamanlı ölçekleme işlemi yapılarak her biri üçer madde içeren 47 madde takımı olmak üzere toplam 141 maddeden oluşan madde havuzu oluşturulmuştur. Uyum geçerliği için aynı sınavın okuma bölümünden alınan puanlar ile dinleme bölümünden alınan puanlar arasındaki korelasyon incelenecektir. Ek olarak DMF analizleri ve maddelerin güçlük parametreleri ile cevaplayıcıların yetenek parametrelerinin aynı ölçek üzerinde yer aldığı birey-madde haritası (Wright- Map) hesaplanacaktır. Sonuç olarak, İngilizce dinleme becerilerini ölçmek için madde havuzunda farklı yetenek seviyelerine uygun yeterli sayıda madde olduğu sonucunun elde edilmesi beklenmektedir.

Anahtar Kelimeler: *Madde takımı, bilgisayarda bireyselleştirilmiş test, geçerlik kanıtları, madde havuzu*

Kaynakça

Eckes, T. (2014). Examining testlet effects in the TestDaF listening section: A testlet response theory modeling approach. *Language Testing*, *31*(1), 39-61.

Min, S. & He, L. (2014). Applying unidimensional and multidimensional item response theory models

in testlet based reading assessment. *Language Testing*, 31(4), 453-477.

Wainer, H., Bradlow, E. T., & Du. Z. (2000). *Testlet response theory. An analog for the 3PL useful in testlet-based adaptive testing*. In W. J. van der Linden & G. A. Glas (Ed.). Computerized adaptive testing: Theory and practice (pp. 245-269). Springer, Dordrecht.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets.

Journal of Educational Measurement, 24, 185–201.

[1] sebahatgoren@gmail.com, ORCID: 0000-0002-6453-3258, Hacettepe Üniversitesi, Ankara, Türkiye

[2] sgelbal@gmail.com, ORCID: 0000-0001-5181-7262, Hacettepe Üniversitesi, Ankara, Türkiye

The Validation of English Proficiency Exam- Listening Subtest

Sebahat GÖREN [1], Selahattin GELBAL[2]

The use of testlets related to a common stimulus offers important advantages such as economy of time and effort in language tests, reducing content effects and measuring higher-level skills (Eckes, 2014; Wainer & Kiely, 1987). In order to benefit from these advantages of testlets, first of all, the most appropriate Item Response Theory (IRT) model should be selected and the necessary strategies for integrating into Computerized Adaptive Testing (CAT) applications should be carefully implemented (Min & He, 2014; Wainer et al., 2000). This study aims to collect evidence on the validity of the item pool of a CAT application. For this purpose, the steps of determining appropriate is study consists of the responses taken from the listening subtest of the English proficiency exam administered in nine different forms to 915 prep class students at a private university in Türkiye. The results obtained withIRT showed that the Rasch model provided better data fit than other IRT models. The assumptions of IRT regarding construct validity were investigated with factor analysis (unidimensionality) and Yen's Q3 statistic (local independence). A total of 141 items were created by performing concurrently scalingbased on the anchor item. For convergent validity, the correlation between the scores obtained from the reading section and the listening section of the same exam will be examined. In addition, DIF analyses will be conducted to examine item bias, and a person-item map (Wright-Map) will be generated, where the difficulty parameters of the items and the ability parameters of the respondents are on the same scale. As a result, it is expected that there will be a sufficient number of items suitablefor different ability levels in the item pool to measure English listening skills.

Keywords: *Testlet, computerized adaptive testing, item pool, validation*

Reference

Eckes, T. (2014). Examining testlet effects in the TestDaF listening section: A testlet response theory modeling approach. *Language Testing*, *31*(1), 39-61.

Min, S. & He, L. (2014). Applying unidimensional and multidimensional item response theory models in testlet based reading assessment. *Language Testing*, *31*(4), 453-477.

Wainer, H., Bradlow, E. T., & Du. Z. (2000). *Testlet response theory. An analog for the 3PL useful in testlet-based adaptive testing.* In W. J. van der Linden & G. A. Glas (Ed.). Computerized adaptive testing: Theory and practice (pp. 245-269). Springer, Dordrecht.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185–201.

[1] sebahatgoren@gmail.com, ORCID: 0000-0002-6453-3258, Hacettepe University, Ankara, Türkiye

[2] sgelbal@gmail.com, ORCID: 0000-0001-5181-7262, Hacettepe University, Ankara, Türkiye

Yapay Zekâ ile C1 Seviyesi Türkçe Okuma Sınavı Oluşturma ve Uygulama

Şeyda YEŞİLYURT [1]

Eğitim-öğretimde, ölçme-değerlendirme süreçleri dersin anlatılması kadar önem arz etmektedir. Türkçenin yabancı dil olarak öğretiminde seviye sınavlarının hazırlanması ve değerlendirilmesi alan uzmanlığı gerektiren bir hususdur. Özellikle C1 seviyesinin ölçülmesi, öğrencilerin dilde yetkinlik düzeyini belirlediği için bu seviye için yazılacak soruların titizlikle hazırlanması gerekmektedir. Bu doğrultuda yapay zekânın yabancı dil olarak Türkçe öğreticilerinin işini kolaylaştırıp kolaylaştırmayacağı bir merak unsurudur. Bu çalışmanın amacı, C1 okuma-anlama sorularını ChatGPT ile hazırlama aşamasında yaşanan olumlu ve olumsuz noktaları tespit etmek, hazırlanan sorularla ilgili alan uzmanlarının görüşlerini belirlemek, ChatGPT ile hazırlanan sorulardan öğrencilerin aldığı puanlarla, uzmanlar tarafından hazırlanan sorulardan aldıkları puanlar arasında anlamlı bir farklılık olup olmadığını belirlemektir. Keşfedici karma yöntem kullanılan çalışmada önce nitel veriler toplanıp daha sonra elde edilen nicel verilerle bir sonuca ulaşılmaya çalışılmıştır. Nitel veriler, uzman görüşü formu, araştırmacı günlüğü kullanılarak nicel veriler ise testlerden alınan puanlarla ilgili aritmetik ortalama, frekans ve t-testi değerleri hesaplanarak elde edilmiştir. Araştırmanın bulgularından hareketle ChatGPT'nin verilen konularla ilgili metinleri çok kısa sürede hazırlanması ve bu metinlere ait alt soruları hızlı bir şekilde (on soru-bir dakika gibi bir sürede) tamamlanması gibi olumlu yönlerinin yanında hazırlanan sorularda seviyeye uygunluk problemi olması, şıklar arasında uzunluk-kısalık, zorluk-kolaylık gibi uyumsuzluklarla karşılaşılması nedeniyle yapay zekânın insan zekâsına ihtiyaç duyduğu ve uzman eşliğinde eğitilerek doğru komutlar verilerek soru hazırlamada faydalı olabileceği sonucuna ulaşılmıştır.

Anahtar Kelimeler: *Yabancı dil olarak Türkçe, yapay zekâ, C1 okuma sınavı*

Kaynakça

Kic-Drgas, J. S., & Kılıçkaya, F. (2024). Using Artificial Intelligence (Al) to create language exam questions: A case study. XLinguae: European Scientific Language Review, 17.

Topsakal, O., & Topsakal, E. (2022). Framework for a foreign language teaching software for children utilizing AR, voicebots and ChatGPT (large language models). The Journal of Cognitive Systems, *7*(2), 33-38.

Zileli, E. N. (2023). Yabancı dil olarak Türkçe öğreniminde ChatGPT örneği. Uluslararası Karamanoğlu Mehmetbey Eğitim Araştırmaları Dergisi, *5*(1), 42-51.

[1] syesilyurt@gazi.edu.tr: ORCID: 0000-0001-7721-5323, Gazi Üniversitesi, Ankara, Türkiye

C1 Level Turkish Reading Test with Artificial Intelligence Creation and Implementation

Şeyda YEŞİLYURT [1]

In education and training, assessment and evaluation processes are as important as the teaching of the course. In teaching Turkish as a foreign language, the preparation and evaluation of level exams is a matter that requires expertise in the field. Especially since the measurement of C1 level determines the level of competence of the students in the language, the questions to be written for this level should be prepared meticulously. In this respect, it is a matter of curiosity whether artificial intelligence will facilitate the work of Turkish as foreign language teachers. The aim of this study is to determine the positive and negative points experienced during the preparation of C1 reading- comprehension questions with ChatGPT, to determine the opinions of field experts about the prepared questions, and to determine whether there is a significant difference between the scores obtained by the students from the questions prepared with ChatGPT and the scores obtained from the questions prepared by the experts. In the study using exploratory mixed method, qualitative data were first collected and then a conclusion was tried to be reached with the quantitative data obtained. Qualitative data were obtained by using expert opinion form, researcher diary and quantitative data were obtained by calculating arithmetic mean, frequency and t-test values related to the scores obtained from the tests. Based on the findings of the study, in addition to the positive aspects of ChatGPT, such as preparing the texts related to the given topics in a very short time and completing the sub-questions of these texts quickly (in a period of ten questions-one minute), it was concluded that artificial intelligence needs human intelligence and can be useful in preparing questions by training with an expert and giving the correct commands.

Keywords: *Turkish as a foreign language, artificial intelligence, C1 reading exam*

Reference

Kic-Drgas, J. S., & Kılıçkaya, F. (2024). Using Artificial Intelligence (AI) to create language exam questions: A case study. XLinguae: European Scientific Language Review, 17.

Topsakal, O., & Topsakal, E. (2022). Framework for a foreign language teaching software for children utilizing AR, voicebots and ChatGPT (large language models). The Journal of Cognitive Systems, *7*(2), 33-38.

Zileli, E. N. (2023). ChatGPT example in learning Turkish as a foreign language. International Journal of Karamanoğlu Mehmetbey Educational Research, *5*(1), 42-51.

[1] syesilyurt@gazi.edu.tr, ORCID: 0000-0001-7721-5323, Gazi University, Ankara, Türkiye

Okuduğunu Anlama Seviyesinin Ölçümünde Boşluk Tamamlama Testleri

M. Rahman Kalyoncu [1], Muhammet Memiş [2]

Nicel araştırma yöntemlerinden korelasyonel araştırma ile yapılandırılmış bu çalışma; okuyucunun okuduğunu anlama düzeyini, genel dil becerisini ve yazılı materyallerin okunabilirliğini ölçmek gibi birçok amaç için kullanılan (Mariotti ve Homan, 2001; Oller, Bowen, Dien ve Mason, 1972; Taylor, 1953) ancak Türkçe için geçerliliği yeterince araştırılmamış olan, ayrıca hazırlanmasında, uygulanmasında ve puanlanmasında belirsizlikler bulunan boşluk tamamlama testlerinin okuduğunu anlama düzeyinin belirlenmesinde Türkçe için geçerli sonuçlar verip vermediğini sorgulamak ve bu testlerin hazırlanmasına, uygulanmasına ve değerlendirilmesine yönelik net bir prosedür ortaya koymak amacıyla yürütülmüştür. Bu doğrultuda, okuduğunu anlama düzeyinin ölçümünde en sık tercih edilen, uygulama ve nesnel değerlendirme kolaylığı sebebiyle ulusal sınavlarda başat ölçme aracı olarak kullanılan çoktan seçmeli testler ile farklı prosedürler izlenerek oluşturulmuş ve puanlanmış boşluk tamamlama testleri arasındaki tutarlılık incelenmiştir. İnceleme sürecinde 4 farklı metin esas alınmış ve bu metinlere yönelik hazırlanmış 4 ayrı çoktan seçmeli okuduğunu anlama testi ile her birinde farklı sıradaki bir sözcüğün düzenli bir şekilde silindiği 4 ayrı boşluk tamamlama testi, ölçme aracı olarak 7. sınıf düzeyindeki öğrencilerden oluşan 90 kişilik çalışma grubuna 2023-2024 eğitim/öğretim döneminde yüz yüze uygulanmıştır. Boşluk tamamlama testlerinin puanlanmasında, birincisi yalnızca orijinal sözcüklerin doğru kabul edildiği, ikincisi bağlamı yani cümlenin anlamını koruyan sözcüklerin doğru kabul edildiği iki farklı puanlama yöntemi kullanılmıştır. Daha sonra çoktan seçmeli testler ile farklı şekillerde oluşturulan ve puanlanan boşluk tamamlama testleri arasındaki korelasyonlar, Pearson korelasyon analizi yardımıyla hesaplanmıştır. Gerçekleştirilen analizlere göre; genel olarak boşluk tamamlama testleri ile çoktan seçmeli okuduğunu anlama testleri arasında *orta düzeyde* pozitif bir ilişki olduğu, yalnızca orijinal kelimelerin doğru kabul edildiği boşluk tamamlama testlerinin çoktan seçmeli okuduğunu anlama testleri ile daha tutarlı sonuçlar verdiği, çoktan seçmeli testler ile boşluk tamamlama testleri arasında en yüksek düzeyde korelasyonun düzenli olarak 6. sıradaki sözcüğün silindiği durumlarda elde edildiği sonuçlarına ulaşılmıştır. Elde edilen sonuçlardan hareketle ise okuduğunu anlama düzeyinin nesnel olarak ölçülmesinde boşluk tamamlama testlerinden de yararlanılabileceği bağlamında araştırmacı ve eğiticilere bazı önerilerde bulunulmuştur.

Anahtar kelimeler: *Boşluk tamamlama testi, çoktan seçmeli test, okuduğunu anlama, Türkçe.*

Kaynakça

Mariotti, A. S., & Homan, S. P. (2001). *Linking reading assesment to* ıntruction*: An application worktext for elementary classroom teachers*. Lawrence Erlbaum Associates.

Oller, J. W., Bowen, D. J., Dien, T. T. ve Mason, V. W. (1972). Cloze tests in English, Thai, and Vietnamise: Native and non-native performance. *Language Learning*, *22*(1), 1-13.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*(1), 415-4

[1] mrkalyoncu@sinop.edu.tr, ORCID: 0009-0000-4644-1088, Sinop Üniversitesi, Sinop, Türkiye

[2] muhammet.memis@omu.edu.tr, ORCID: 0000-0001-6037-3958, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye

Cloze Tests in Measuring Reading Comprehension Levels

M. Rahman Kalyoncu[1], Muhammet Memiş[2]

This study, structured as a correlational study using quantitative research methods, was conducted to investigate whether cloze tests, which are used for purposes such as measuring reading comprehension, general language skills, and the readability of written materials (Mariotti & Homan, 2001; Oller, Bowen, Dien & Mason, 1972; Taylor, 1953), provide valid results for Turkish. The validity of these tests for Turkish has not been sufficiently researched, and their preparation, administration, and scoring procedures remain unclear. The primary goal was to determine whether cloze tests yield valid results for assessing reading comprehension in Turkish and to establish a clear procedure for their preparation, administration, and evaluation. In this context, the consistency between multiple-choice tests widely used for measuring reading comprehension due to their ease of administration and objective assessment, particularly in national exams and cloze tests, which were created and scored using different procedures, was examined. The study was conducted with 90 seventh-grade students during the 2023-2024 academic year. Four different texts were selected, and four corresponding multiple-choice reading comprehension tests, along with four cloze tests in which a word was systematically deleted in varying positions, were administered face-to-face to the study group. Two different scoring methods were employed for the cloze tests: in the first, only the original word was accepted as correct, while in the second, any word that preserved the context and meaning of the sentence was accepted as correct. Pearson correlation analysis was conducted to calculate the correlations between the multiple-choice tests and the various cloze tests. The results showed a moderate positive correlation between the cloze tests and multiple-choice reading comprehension tests overall. Cloze tests that accepted only the original words as correct demonstrated higher consistency with multiple-choice reading comprehension tests. Additionally, the highest correlation between the two test types was observed when the word in the sixth position was systematically deleted in the cloze tests. Based on these findings, several suggestions were made for researchers and educators regarding the potential use of cloze tests for the objective measurement of reading comprehension.

Keywords: *Cloze test, multiple-choice test, reading comprehension, Turkish*

Reference

Mariotti, A. S., & Homan, S. P. (2001). *Linking reading assesment to* ıntruction*: An application worktext for elementary classroom teachers*. Lawrence Erlbaum Associates.

Oller, J. W., Bowen, D. J., Dien, T. T., & Mason, V. W. (1972). Cloze tests in English, Thai, and Vietnamise: Native and non-native performance. *Language Learning*, *22*(1), 1-13.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*(1), 415-433.

[1] mrkalyoncu@sinop.edu.tr, ORCID: 0009-0000-4644-1088, Sinop Üniversitesi, Sinop, Türkiye

[2] muhammet.memis@omu.edu.tr, ORCID: 0000-0001-6037-3958, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye

Aynı Özelliği Ölçen Ortak Köklü ve Bağımsız Maddelerden Oluşan Yabancı Dil Olarak Türkçe Dinleme Testlerinin Psikometrik Özelliklerinin Karşılaştırılması

Ceren TUNABOYLU DEMİR [1], Havva Gökçe ÇAVDAR PAKSOY [2], Duygu ANIL[3]

İnsan sosyal bir varlık olarak doğası gereği çevresiyle etkileşim kurmak için dil öğrenmektedir. Dil öğrenimde iletişimin gerçekleşebilmesi dinleme, konuşma, okuma ve yazma becerilenin geliştirilmesine bağlıdır. Dil öğrenme sürecinde bu becerilerin dengeli bir şekilde geliştirilmesine, etkili ve doğru ölçme-değerlendirme yapılmasına ihtiyaç duyulmaktadır. Dinleme becerisinin ölçülmesinde, dinleme materyallerinin doğru geliştirilmesi ve kullanılması ön plana çıkmaktadır. Dinleme metinlerinin dil bilgisi yapısına uygun olarak hazırlanmasından seslendirme kalitesine gelinceye kadar her aşama ayrıntılı planlanmalı ve sunulmalıdır. Dinleme becerisinin ölçülmesinde, uyarıcının doğru iletilebilmesi, bireyin uyarana karşı doğru tepkide bulunabilmesinin temel gerekliliğidir. Bu araştırmada, aynı özelliği ölçen ortak köklü ve bağımsız maddelerden oluşan Türkçe dinleme testlerinin psikometrik özelliklerinin karşılaştırılması ve testlerin kullanışlılığına yönelik öğrenci görüşlerinin alınması amaçlanmaktadır. Bu amaç doğrultusunda yabancı dil olarak Türkçe alanında B1 düzeyinde geliştirilen dinleme testleri eş zamanlı olarak uygulanarak madde güçlük indeksleri, madde ayırt edicilik indeksleri, test ortalamaları, test güvenirlikleri ve yanıtlama süreleri arasında anlamlı fark olup olmadığını incelenecek, öğrenci görüşleri doğrultusunda testlerin kullanışlılığı değerlendirilecektir. Bu araştırmada karma desen yöntemlerinden açımlayıcı sıralı karma yöntem kullanılmış olup betimsel bir araştırmadır. Test geliştirme sürecinde, dinleme testleri Milli Eğitim Bakanlığı Türkçenin Yabancı Dil Olarak Öğretimi programı esas alınarak hazırlanmıştır. Katılımcılar, Türkçe Öğrenim Merkezlerinde öğrenim gören yabancı uyruklu öğrenciler arasından gönüllülük esasına dayalı olarak seçilecek olup yaklaşık 150-200 öğrenciye ulaşılması hedeflenmektedir. Alan yazın incelendiğinde testte bağımsız maddeler kullanıldığında, her sorunun bağımsız bir bilgi parçasını ölçmesi nedeniyle, öğrenci performansını madde madde değerlendirme açısından avantaj sağladığı; ortak köklü maddeler kullanıldığında ise, test süresini verimli kullanma ve okuma-anlama becerilerini entegre bir şekilde ölçme avantajı sağladığı görülmüştür. Bu araştırma kapsamında, dinleme becerisinin ölçülmesinde ortak köklü madde kullanımı ile bağımsız madde kullanımının öğrenci performansı üzerine etkilerinin ortaya konulmasının dil öğretiminde ölçme araçlarının geliştirilmesine ve uygulamalara katkı sağlayacağı değerlendirilmektedir.

Anahtar Kelimeler: *dinleme becerisinin* ölçülmesi*,* çoktan *seçmeli madde, madde ve test* özellikleri*.*

1 cerentunaboylu@gmail.com, ORCID: 0000-0001-8090-8913, Milli Eğitim Bakanlığı, Ankara, Türkiye

2 gkccavdar@gmail.com, ORCID: 0000-0003-1813-2725, Abant İzzet Baysal Üniversitesi, Türkçe Eğitimi, Bolu, Türkiye

3 aduygu@hacettepe.edu.tr, ORCID: 0000-0002-1745-4071, Hacettepe Üniversitesi, Ankara, Türkiye

A Comparison of the Psychometric Properties of Common Stem and Independent Items in Turkish as a Foreign Language Listening Tests

Ceren TUNABOYLU DEMİR[1], Havva Gökçe ÇAVDAR PAKSOY[2], Duygu ANIL[3]

As a social being, humans inherently learn language to interact with their environment. For communication to occur in language learning, it is essential to develop listening, speaking, reading, and writing skills. In the language learning process, there is a need for the balanced development of these skills and for effective and accurate measurement and evaluation. In assessing listening skills, the proper development and use of listening materials come to the forefront. Every stage, from preparing listening texts in accordance with grammatical structures to ensuring the quality of narration, must be meticulously planned and presented. In measuring listening skills, the fundamental requirement is the accurate transmission of stimuli and the individual's correct response to these stimuli. This study aims to compare the psychometric properties of Turkish listening tests composed of common stem and independent items that measure the same feature and to gather student feedback on the usability of these tests. For this purpose, listening tests developed at the B1 level in the field of Turkish as a foreign language will be administered simultaneously, and it will be examined whether there are significant differences in item difficulty indices, item discrimination indices, test means, test reliability, and response times. The usability of the tests will be evaluated based on student feedback. This study employs an explanatory sequential mixed-methods design and is descriptive in nature. During the test development process, the listening tests were prepared based on the Ministry of National Education's Turkish as a Foreign Language Teaching program. Participants will be selected on a voluntary basis from among foreign students studying at Turkish Learning Centers, with the aim of reaching approximately 150-200 students. The literature review reveals that using independent items provides an advantage in evaluating student performance item by item, as each question measures an independent piece of information. On the other hand, using common stem items offers the advantage of efficiently utilizing test time and measuring reading comprehension skills in an integrated manner. Within the scope of this study, it is considered that revealing the effects of using common stem items versus independent items on student performance in measuring listening skills will contribute to the development and application of assessment tools in language teaching.

Keywords: *assessment of listening skills, multiple choice items, item and test characteristics.*

1 cerentunaboylu@gmail.com, ORCID: 0000-0001-8090-8913, MEB, Ankara, Turkey

2 gkccavdar@gmail.com, ORCID: 0000-0003-1813-2725, Abant İzzet Baysal University, Bolu, Turkey

3 aduygu@hacettepe.edu.tr, ORCID: 0000-0002-1745-4071, Hacettepe University, Ankara, Turkey

Yabancı Dil Olarak Türkçe Öğretiminde Kriter Bazlı Soru Yazımı Eğitimi Tasarımı ve Değerlendirmesi

Uğur Okur [1], Kevser Yılmaz[2], Aslı Çelik[3]

Dil öğretimi ölçme değerlendirme alanında sınav sorularının madde yazma kriterlerine uygun şekilde soruların hazırlanması önem arz etmektedir. Ancak yabancı dil olarak Türkçe öğretimi alanında ölçme ve değerlendirme çalışmalarının sınırlı kalması ve öğreticilerin alan ve mesleki bilgisinin ölçme aracı geliştirmeye yönelik ihtiyacı tam olarak karşılayamaması nedeniyle konuyla ilgili öğreticilere eğitim verilmesinin uygun olacağı düşünülmektedir (Durmuş, 2013; Işıkoğlu, 2015; Yahşi Cevher&Güngör, 2015; Karataş Demirtaş &Karataş Acer, 2016; Boylu, 2019). Bu çalışma, yabancı dil olarak Türkçe öğretiminde alımlayıcı becerileri ölçen çoktan seçmeli sorularının hazırlanmasında ampirik çalışma sonrasında belirlenen ihtiyaçlara göre bir eğitim sağlanması hedeflenmektedir. Araştırmanın çalışma grubunu bir vakıf üniversitesinin Yabancı Diller Yüksekokulu Türkçe Hazırlık Programında öğretici olarak görev yapan 10 kişi oluşturmaktadır. Çalışma kapsamında, nitel araştırma yöntemlerinden doküman analizi yöntemi kullanılmıştır. Araştırma kapsamında ölçme değerlendirme alanyazınında yer alan çoktan seçmeli sınav sorusu yazma kriterleri incelenmiş ve çalışma grubunu oluşturan vakıf üniversitesinin Türkçe hazırlık programında hazırlanan sınavlar bu kriterlere uygunluk bakımından değerlendirilmiştir. Sınav sorularına verilen geri bildirimler, Haladyna ve Rodriguez (2013) ile Selçuk ve Uçar (2023) temel alınarak belirlenen temaların altına tümdengelimsel nitel analiz ile oluşturulmuştur. Eğitim tamamen ampirik veri üzerine planlanarak analiz bulguları doğrultusunda ADDIE modeli kullanılarak (REF) 10 saatlik bir sınav sorusu yazma eğitimi tasarlanarak, çalıştay formatında 5 gün içerisinde tamamlanmıştır. Çalıştay sonrası katılımcıların algıları anket ile değerlendirilmiştir. Araştırma sonucunda, yapılmış olan çalıştayın yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirmenin önemli bir unsuru olan sınav sorularının yazımında soru kalitesinin artırılmasına katkıda bulunacağı beklenmektedir.

Anahtar Kelimeler: *yabancı dil olarak Türkçe* öğretimi*,* ölçme *ve değerlendirme, kriter bazlı sınav hazırlama,* çoktan *seçmeli madde yazımı.*

Kaynakça

Boylu, E. (2019). *Yabancılara Türkçe* öğretiminde ölçme *değerlendirme uygulamaları ve standart oluşturma.* (Yayımlanmamış doktora Tezi). Çanakkale Onsekiz Mart Üniversitesi, Çanakkale.

Durmuş, M. (2013). Türkçenin yabancılara öğretimi: sorunlar, çözüm önerileri ve yabancılara Türkçe öğretiminin geleceğiyle ilgili görüşler. *Adıyaman* Üniversitesi *Sosyal Bilimler Enstitüsü Dergisi Türkçenin Eğitimi* Öğretimi *Özel Sayısı, 6*(11), 207-228.

Haladyna, T.M., & Rodriguez, M.C. (2013). Developing and Validating Test Items (1st ed.). Routledge. Özdemir, A. (Ed.). (2023). İkinci */ Yabancı Dil* Öğretiminde Ölçme *ve Değerlendirm*e. Altınbaş Üniversitesi Yayınları.

[1] ugur.okur@altinbas.edu.tr:, ORCID: 0000-0002-8670-1443, Altınbaş Üniversitesi, İstanbul, Türkiye

[2] kevser.yilmaz@altinbas.edu.tr:, ORCID: 0000-0001-6129-4303, Altınbaş Üniversitesi, İstanbul, Türkiye

[3] asli.celik@altinbas.edu.tr:, ORCID: 0000-0003-3590-5283, Altınbaş Üniversitesi, İstanbul, Türkiye

Design and Evaluation of Criteria-Based Question Writing Training in Teaching Turkish as a Foreign Language

Uğur Okur[1], Kevser Yılmaz [2], Aslı Çelik [3]

In the field of language teaching assessment, it is crucial to prepare exam questions in accordance with item-writing criteria. However, the limited scope of assessment studies within the domain of teaching Turkish as a foreign language, coupled with the instructors` lack of the requisite expertise in developing robust assessment tools, it would be appropriate to provide training to instructors on this subject (Durmuş, 2013; Işıkoğlu, 2015; Yahşi Cevher & Güngör, 2015; Karataş Demirtaş & Karataş Acer, 2016; Boylu, 2019). To address this gap, the study aims at providing training based on the needs identified through empirical research for the preparation of multiple-choice questions in order to assess receptive skills in teaching Turkish as a foreign language. The participant pool was comprised of 10 instructors affiliated with the Turkish Preparatory Program at the School of Foreign Languages of a foundation university. To examine the research question in the study, the document analysis method, one of the qualitative research methods, was employed. The multiple-choice question-writing criteria outlined in the assessment literature were examined, and subsequently the exams prepared in the Turkish preparatory program of the foundation university were evaluated for their alignment with these established criteria. Feedback on the exam questions was thematically categorized in accordance with the framework of Haladyna and Rodriguez (2013) and Selçuk and Uçar (2023) through deductive qualitative analysis. The training was entirely planned based on empirical data, and a 10-hour exam question-writing training was developed using the ADDIE model (REF) and completed in a workshop framework spanning five days. Upon the completion of the workshop, participants' perceptions of the training were assessed via a survey. As a result of the study, it is expected that the workshop will significantly contribute to improving the exam question quality, which is an essential facet of assessment in teaching Turkish as a foreign language.

Keywords: *Teaching Turkish as a foreign language, assessment and evaluation, criteria-based exam preparation, multiple-choice item writing.*

Reference

Boylu, E. (2019). *Yabancılara Türkçe* öğretiminde ölçme *değerlendirme uygulamaları ve standart oluşturma.* (Yayımlanmamış doktora Tezi). Çanakkale On sekiz Mart Üniversitesi, Çanakkale.

Durmuş, M. (2013). Türkçenin yabancılara öğretimi: sorunlar, çözüm önerileri ve yabancılara Türkçe öğretiminin geleceğiyle ilgili görüşler. *Adıyaman* Üniversitesi *Sosyal Bilimler Enstitüsü Dergisi Türkçenin Eğitimi* Öğretimi *Özel Sayısı, 6*(11), 207-228.

Haladyna, T.M., & Rodriguez, M.C. (2013). Developing and Validating Test Items (1st ed.). Routledge. Özdemir, A. (Ed.). (2023). İkinci*/Yabancı Dil* Öğretiminde Ölçme *ve Değerlendirm*e. Altınbaş Üniversitesi Yayınları.

[1] ugur.okur@altinbas.edu.tr:, ORCID: 0000-0002-8670-1443, Altınbaş Üniversitesi, İstanbul, Türkiye

[2] kevser.yilmaz@altinbas.edu.tr:, ORCID: 0000-0001-6129-4303, Altınbaş Üniversitesi, İstanbul, Türkiye

[3] asli.celik@altinbas.edu.tr:, ORCID: 0000-0003-3590-5283, Altınbaş Üniversitesi, İstanbul, Türkiye

Türkçenin Yabancı/İkinci Dil Olarak Öğretiminde Yenilenmiş Bloom Taksonomisi'ne Göre Hazırlanmış Soru ve Etkinliklerin Öğretici Farkındalıkları Açısından Değerlendirilmesi

Mustafa Durmuş[1], Eser Kocaman Gürata[2], Mine Akdemir[3]

Türkçe öğretimi sürecinde öğrenicilerin en çok karşılaştıkları materyalin ders kitapları olduğu düşünüldüğünde ders kitaplarının nasıl ve ne şekilde hazırlandığı önemli hale gelmektedir. Özellikle ilgili kitaplarda yer alan soru ve etkinliklerin değerlendirilmesi öğretim sürecinin pozitif çıktılarını görebilmek için önemlidir. Yabancı/ikinci dil öğretiminde ezberlediklerini söyleyen öğreniciler yerine öğrendiklerini sorgulayan, eleştiren ve yaratıcı düşünen; değerlendirme, yaratma gibi üst düzey düşünme süreçlerini de aktif kullanan öğreniciler hedeflenmektedir. Bu hedef doğrultusunda ders kitaplarında yer alan soru ve etkinliklerin üst düzey düşünme süreçlerine katkı sağlayacak şekilde hazırlanması gerekmektedir. Öğreticilerin ise soru ve etkinliklerin nasıl ve ne şekilde hazırlandığı, kazanımların üst bilişsel becerileri karşılayıp karşılamadığı ve hangi kazanımların seçildiği gibi birtakım durumlara karşı farkındalıklarının olması beklenmektedir. Bu kapsamda araştırmada, Türkçenin yabancı/ikinci dil olarak öğretiminde Yenilenmiş Bloom Takosnomisi'ne göre hazırlanmış soru ve etkinliklerin öğretici farkındalıkları açısından değerlendirilmesi amaçlanmıştır. Araştırmanın çalışma grubunu Türkçe Öğretim Merkezlerinde (TÖMER) görev alan 23-60 yaş aralığındaki 25 öğretici oluşturmaktadır. Araştırma, betimsel tarama modeline göre desenlenmiştir. Araştırma hem nicel hem nitel özelliklere sahiptir. Çalışmanın verileri öğreticilere uygulanacak anket formu ve yarı yapılandırılmış görüşme formu ile elde edilecektir. Anketin açık uçlu sorularından ve yarı yapılandırılmış görüşme formundan elde edilen nitel verilerin analizinde içerik analizinden yararlanılacaktır. Anketin kapalı uçlu sorularından elde edilen veriler SPSS programında analiz edilerek yüzde ve frekanslar tablolar halinde sunulacaktır. Bunun sonucunda öğreticilerin bilişsel düzeylere ait fiilleri, soru ve etkinlikleri hangi bilişsel düzeyi esas alarak eşleştirdikleri, hangi düzeylere ağırlık verdikleri ve hangi düzeyi göz ardı ettikleri belirlenecektir. İlgili çalışma, öğreticilere, öğretim programı geliştiricilerine, sınav hazırlayan komisyon üyelerine, ölçme ve değerlendirme uzmanlarına, alanda çalışmak isteyen aday öğreticilere bakış açısı sunması bakımından önemlidir.

Anahtar Kelimeler: Yabancı/ikinci dil olarak Türkçe, Bloom Taksonomisi, öğretici farkındalığı, üst düzey düşünme süreçleri, bilişsel beceriler

[1] mustafadurmus@hacettepe.edu.tr, ORCID: 0000-0002-5337-3113, Hacettepe Üniversitesi, Ankara, Türkiye

[2] eserkocaman@gmail.com, ORCID:0000-0002-0774-3208, Hacettepe Üniversitesi, Ankara, Türkiye

[3] mineakdemir@hacettepe.edu.tr, ORCID: 0000-0002-5047-3020, Hacettepe Üniversitesi, Ankara, Türkiye

Evaluation of Questions and Activities Prepared According to the Revised Bloom's Taxonomy in Teaching Turkish as a Foreign/Second Language in Terms of Teachers' Awareness

Mustafa Durmuş[1], Eser Kocaman Gürata[2], Mine Akdemir[3]

When it is considered that the material learners encounter the most during the Turkish teaching process is textbooks, how textbooks are prepared becomes important. Especially, evaluating the questions and activities in the textbooks is important to see the positive outcomes of the teaching process. Learners who question, criticize, and think creatively about what they have learned instead of memorizing in foreign/second language teaching; are targeted to actively use higher-level thinking processes such as evaluating and creating. In that regard, questions and activities in textbooks should be prepared in a way that contribute to higher-order thinking processes. Teachers are expected to be aware of several situations such as how and in what way questions and activities are prepared, whether the learning outcomes meet higher cognitive skills and which learning outcomes are selected. In this context, the aim of the research is to evaluate questions and activities prepared according to the Revised Bloom Taxonomy in Turkish as a foreign/second language instruction in terms of teachers' awareness. The study group of the research consists of 25 instructors aged between 23-60 working at Turkish Teaching Centers (TÖMER). The research is designed according to the descriptive survey model. The research has both quantitative and qualitative features. The data of the study will be obtained through a questionnaire and a semi-structured interview form to be applied to the teachers. Content analysis will be used in the analysis of qualitative data obtained from open-ended questions in the questionnaire and semi-structured interview form. The data obtained from the closed-ended questions in the questionnaire will be analyzed in the SPSS program and presented in tables as percentages and frequencies. As a result, it will be determined on the basis of which cognitive level the teachers match the verbs, questions and activities belonging to the cognitive levels, which levels they emphasize and which level they ignore. The relevant study is important in terms of providing a perspective to teachers, curriculum developers, language test developers, assessment and evaluation experts, and prospective teachers who want to work in the field.

Key words: Turkish as a foreign/second language, Bloom Taxonomy, teachers' awareness, higher-level thinking processes, cognitive skills

[1] mustafadurmus@hacettepe.edu.tr, ORCID: 0000-0002-5337-3113, Hacettepe Üniversitesi, Ankara, Türkiye

[2] eserkocaman@gmail.com, ORCID:0000-0002-0774-3208, Hacettepe Üniversitesi, Ankara, Türkiye

[3] mineakdemir@hacettepe.edu.tr, ORCID: 0000-0002-5047-3020, Hacettepe Üniversitesi, Ankara, Türkiye

Üniversite Eğitimlerini İngilizce Alan Öğrencilerin Yabancı Dil Yazma Becerilerinin Yapay Zekâ Tabanlı Puanlama ve İnsan Tarafından Değerlendirilmesinin Karşılaştırmalı Bir Analizi

Volkan İnceçay [1], Nafiye Çiğdem Aktekin [2], Cihan Kocabaş[3]

Yapay zekâdaki son gelişmeler, yazılı metinlerin verimli, güvenilir ve nesnel bir şekilde değerlendirilmesini vadeden otomatik kompozisyon puanlama sistemlerinin ortaya çıkmasına neden olmuştur. Ancak, yapay zekâ tabanlı bu sistemlerinin etkinliği ve güvenilirliği daha fazla kanıt gerektirmektedir. Bu nedenle, bu çalışmada, üniversite eğitimlerini İngilizce alan öğrenciler tarafından yazılan kompozisyonları değerlendirirken yapay zekâ tabanlı bir dil modelinin (ChatGPT-4.0) güvenilirliğini insan değerlendiricilere kıyasla araştırma ve karşılaştırma amaçlanmıştır. Bu amaçla, öğrenci kompozisyonlardan oluşan bir veri seti toplanmıştır (n=41). ChatGPT-4.0 ve iki deneyimli İngilizce eğitmeni, yerleşik bir değerlendirme rubriği olan TOEFL Yazma Değerlendirme Ölçütünü (Bağımsız versiyon) kullanarak toplanan kompozisyonları olarak puanlamıştır. İnsan değerlendiriciler için güvenilir ve tutarlı puanlama sağlamak amacıyla çeşitli kalibrasyon oturumları gerçekleştirilmiştir. Değerlendirmeler arasındaki anlaşma ve tutarsızlık alanlarını belirlemek için istatistiksel ölçümler olarak korelasyon analizleri yapılmıştır. Bulgular, yapay zekanın yabancı dil yazımına entegre edilmesinin, eğitimcilerin iş yükünü hafifleterek ve tutarlı değerlendirmeler sağlayarak önemli faydalar sağlayabileceğine işaret etmektedir.

Anahtar Kelimeler: *Yapay zekâ tabanlı puanlama, güvenilirlik analizi, yabancı dilde yazma değerlendirmesi.*

Kaynakça

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT-4.0. *Innovations in Education and Teaching International*, 1-12. https://doi.org/10.1080/14703297.2023.2190148

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT3).

Ethics in Science and Environmental Politics, *2*, 17-23. https://doi.org/10.3354/esep00195 Educational Testing Service. (n.d.). *TOEFL iBT writing rubrics*. https://www.ets.org/pdfs/toefl/toefl-ibt-writing-rubrics.pdf

Yamashita, T. (2024). An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, *3*(3), 100133. https://doi.org/10.1016/j.rmal.2024.100133

[1] volkan.incecay@medipol.edu.tr, ORCID: 0000-0001-9156-7624, İstanbul Medipol Üniversitesi, İstanbul, Türkiye

[2] nafiye.aktekin@ acibadem.edu.tr, ORCID: 0000-0003-0829-3744, Acıbadem Üniversitesi, İstanbul, Türkiye

[3] cihankocabas@gmail.com, ORCID: 0000-0001-6511-9094, Türkiye Maarif Vakfı, İstanbul, Türkiye

A Comparative Analysis of Artificial Intelligence-Based Scoring and Human Rater Evaluation in Assessing Foreign Language Writing Skills of English-Medium Instruction Students

Volkan İnceçay [1], Nafiye Çiğdem Aktekin[2], Cihan Kocabaş[3]

The latest advancements in artificial intelligence (AI) have led to the emergence of automated essay scoring systems that promise efficient, reliable, and objective evaluation of written texts. However, the effectiveness and reliability of these AI-based systems still require further evidence. Therefore, this study aims to investigate and compare the reliability of an AI language model (ChatGPT-4.0) against human raters when evaluating essays written by university students receiving their education in English. For this purpose, a dataset of student essays was collected (n=41). ChatGPT-4.0 and two experienced English instructors rated the collected essays using the TOEFL Writing Assessment Rubric (Independent version), a well-established scoring rubric. Several calibration sessions were conducted to ensure reliable and consistent scoring for the human raters. Correlation analyses were performed as statistical measures to identify areas of agreement and inconsistency between the evaluations. The findings suggest that the integration of AI into foreign language writing could provide significant benefits by reducing educators' workload and offering consistent evaluations.

Keywords: *AI-based scoring, reliability analysis, foreign language writing assessment.*

References

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT-4.0. *Innovations in Education and Teaching International*, 1- 12. https://doi.org/10.1080/14703297.2023.2190148

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT3).

Ethics in Science and Environmental Politics, *2*, 17-23. https://doi.org/10.3354/esep00195 Educational Testing Service. (n.d.). *TOEFL iBT writing rubrics*. https://www.ets.org/pdfs/toefl/toefl-ibt-writing-rubrics.pdf

Yamashita, T. (2024). An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, *3*(3), 100133. https://doi.org/10.1016/j.rmal.2024.100133

1 volkan.incecay@medipol.edu.tr, ORCID: 0000-0001-9156-7624, Istanbul Medipol University, İstanbul, Türkiye

2 nafiye.aktekin@ acibadem.edu.tr, ORCID: 0000-0003-0829-3744, Acıbadem University, İstanbul, Türkiye

3 cihankocabas@gmail.com, ORCID: 0000-0001-6511-9094, Turkish Maarif Foundation, İstanbul, Türkiye

Transforming Speaking Assessments with AI-based Virtual Characters in the Metaverse

Selin SOYDAN [1], Emre YILDIRIM [2]

This systematic review examines the impact of artificial intelligence (AI)-based and virtual reality (VR)- based technologies, such as chatbots and metaverse environments, on enhancing and evaluating EFL learners' speaking abilities, based on previous studies. The primary findings from the literature indicate that AI-based tools improve speaking fluency, coherence, pronunciation, and willingness to communicate (WTC) by offering personalized, instant feedback that reduces public speaking anxiety and boosts learner confidence. The review highlights the importance of future research to improve language assessment methods and create and integrate new AI-based tools that could increase the reliability, accuracy, and effectiveness of language skills assessments. It suggests that combining AI- based tools like ChatGPT with virtual human characters within the setting of a metaverse environment, enabling students to engage in virtual environments, scenarios, and conversations similar to those with native speakers, can be an efficient method for assessing and evaluating EFL learners' speaking skills. Through the advantages of artificial intelligence and the metaverse, this interactive experience could offer measuring learners' speaking skills in a real-like environment with real-like people, immediate data collection on speaking skills, providing an in-depth evaluation of learners' speaking performance, and detailed reports for assessment, which would add valuable insight both to the student and the teacher.

Keywords: *Metaverse environments, Speaking abilities, AI-based language assessment*

[1] selinsoydan022@gmail.com, ORCID: 0009-0008-7486-2670, National Defense University, Istanbul, Türkiye

[2] emre.yildirim@bogazici.edu.tr, ORCID: 0009-0005-3385-017X, Bogazici University, Istanbul, Türkiye

The practical effectiveness of ChatGPT as an automated scoring tool for writing assessment

Rabia Damla KARAÇEPER[1]

Testing writing proficiency without the use of technology has become a burden for teachers and language testers due to its complex nature, which requires excessive amount of time and effort to evaluate the multidimensional aspects such as grammar, vocabulary, content and organization. With the global increase in the number of the test takers, the automated scoring systems have been internationally utilized by major testing agencies such as Educational Testing Service (ETS) in order to obtain accurate, consistent and rapid test results. This study investigates the effectiveness of ChatGPT as an automated scoring tool for EFL learners' written outputs. It particularly aims to discover to what extent ChatGPT can produce reliable and valid scores in writing assessment and whether it can be used as an alternative scoring tool to the traditional human scoring. A number of 240 argumentative essays, pre-scored by two different human raters, are gathered to obtain scores from ChatGPT. 40 of them are planned to be used in the training process of ChatGPT based on the rubric criteria and scoring patterns of human raters. Statistical analysis is achieved through the inter-rater reliability Cohen's Kappa with a focus on the correlation and agreement among human and AI raters. Preliminary results suggest that ChatGPT will strongly correlate with humans, offering an effective, scalable and quick scoring tool for larger amount of student writings in second language assessment.

Keywords: *automated scoring tool, ChatGPT, writing assessment*

[1] k.rabiadamla@ogr.iuc.edu.tr, ORCID: 0000-0001-6155-9973, Milli Savunma Üniversitesi, İstanbul, Türkiye

Topluluk Önünde Konuşma Becerisi Değerlendirmesinin Öğrenen Özerkliği Süreci (ÖÖS) ile Geliştirilmesi

Begüm Ceylan [1]

Yabancı dilde konuşma becerisi, sözel üretim ve sözel etkileşim olmak üzere iki başlık altında toplanmaktadır (CEFR, 2018). Plansız olan ve akışta şekillenen sözel etkileşim, bireyin bir başkası ile konuşmasını ifade ederken; planlı olan sözel üretim ise bireyin duygu, düşünce, görüş ve deneyimlerini tek başına ya da topluluk önünde konuşmasını ifade etmektedir. Bu çalışmaya konu olan İngilizcede topluluk önünde konuşma becerisi, öğrencilere akademik ve iş hayatlarında büyük kolaylıklar ve fırsatlar sağlayacak olan bir beceridir. Ancak, topluluk önünde konuşma becerisi öğrenciler üzerinde kaygı ve gerginlik yaratabilmektedir. Bunun nedenleri arasında, teknik bilgi eksikliği, hata yapma duygusu, özgüven eksikliği ve negatif özyeterlik algısı gibi psikolojik etmenler ile içeriğe yönelik bilgi eksikliği sıralanabilir. Bu sebeplerle, topluluk önünde konuşma becerisinin öğretiminin ve değerlendirilmesinin etkili ve titiz bir şekilde gerçekleştirilmesi önemlidir. Ancak, Türkiye'de bu konuda eksiklik olduğu gözlemlenmektedir. Dolayısıyla, konuşma becerisinin değerlendirilmesi için önerilen performansa dayalı değerlendirme ile otantik değerlendirme yöntemlerinden öğrenen özerkliği süreci konusunda farkındalığın artırılması gerekmektedir.

Bu araştırmanın amacı, yabancı dil olarak İngilizce öğretiminde topluluk önünde konuşma becerisinin değerlendirilmesinin öğrenen özerkliği süreci ile geliştirilmesidir. Bu kapsamda, araştırma karma araştırma yöntemlerinden iç içe geçmiş karma desen kullanılarak yürütülmüştür. Araştırmanın çalışma grubunu, İstanbul'daki bir devlet üniversitesinin Meslek Yüksekokulu Sivil Havacılık Programında 2023- 2024 akademik yılı bahar döneminde ikinci sınıfta eğitim görmekte olan 24 öğrenci oluşturmaktadır. Öğrenen özerkliği sürecine uygun tasarlanan İngilizce Konuşma II dersinin başında ve sonunda öğrencilere, Topluluk Önünde Konuşma Öğrenen Özerkliği Ölçeği uygulanmış, eğitimin sonunda öğrencilerden sözel bir sunum yapmaları beklenmiş, son olarak öğrencilerden öz-değerlendirme formları toplanmıştır. Ön ve son testlerden elde edilen nicel veriler SPSS programında incelenmiş, nitel veriler ise içerik analizi yöntemi ile incelenmiştir. Araştırma sonunda, öğrencilerin ölçek kapsamında eğitim öncesi ve sonrası aldıkları puanlar arasında farklılıklar saptanmıştır. Ayrıca, öğrencilerin yazmış oldukları öz-değerlendirme raporlarının, öğrencilerin kendi performansları konusundaki farkındalıklarını arttırdığı gözlemlenmiştir. Dolayısıyla, öğrenen özerkliği sürecinin yabancı dilde topluluk önünde konuşma becerisinin değerlendirilmesinde etkili olduğu sonucuna ulaşılmıştır.

Anahtar Kelimeler: *Konuşma, konuşma becerisinin değerlendirilmesi, topluluk* önünde *konuşma*

Kaynakça

CEFR (2018). *Common european framework of reference for languages: learning, teaching, assessment companion volume with new descriptors*, Council of Europe.

[1] begum.ceylan@iuc.edu.tr, ORCID: 0000-0002-6549-5951, İstanbul Üniversitesi-Cerrahpaşa, İstanbul, Türkiye

Enhancing Public Speaking Skill Assessment with Autonomous Learning Process (ALP)

Begüm Ceylan [2]

Speaking skill in a foreign language is categorized under two main headings: oral production and interaction (CEFR, 2018). Unplanned oral interaction, which evolves in the flow, refers to an individual's conversation with another person, while planned oral production refers to an individual expressing their feelings, thoughts, opinions, and experiences alone or in front of a community. Public speaking skill in English, which is the subject of this study, is a skill that will provide students with great ease and opportunities in their academic and professional lives. However, speaking to an audience can create anxiety and tension among students. Among the reasons for this are psychological factors such as lack of technical knowledge, fear of making mistakes, lack of self-confidence, and negative self-efficacy perception, as well as a lack of content knowledge. Therefore, it is important that the teaching and assessment of public speaking skills be carried out effectively and meticulously. However, it is observed that there is a deficiency in this regard in Turkey. Therefore, it is necessary to increase awareness about the performance-based assessment and the autonomous learning process among the authentic assessment methods for the evaluation of public speaking skills.

The aim of this research is to develop the evaluation of public speaking skills in English through the autonomous learning process. In this context, the research was conducted using the embedded mixed design. The study group of the research consisted of 24 students studying in the second year of the Civil Aviation Program at a Vocational School of a public university in Istanbul during the spring semester of the 2023-2024 academic year. At the beginning and end of the English-Speaking II course, which was designed in accordance with the autonomous learning process, the Public Speaking Learner Autonomy Scale was applied to the students. At the end of the training, students were expected to make an oral presentation, and finally, self-assessment forms were collected from the students. Quantitative data obtained from pre and post-tests were analyzed using the SPSS program, and qualitative data were analyzed using the content analysis. At the end of the research, differences were found between the scores students received before and after the training within the scope of the scale. In addition, it was observed that the self-assessment reports written by the students increased their awareness of their own performance. Therefore, it was concluded that the autonomous learning process is effective in evaluating public speaking skills in English.

Keywords: *Speaking, Speaking Assessment, Public speaking*

Reference

CEFR (2018). *Common european framework of reference for languages: learning, teaching, assessment companion volume with new descriptors*, Council of Europe.

[1] begum.ceylan@iuc.edu.tr, ORCID: 0000-0002-6549-5951, İstanbul University-Cerrahpaşa, İstanbul, Turkiye

Yazılı Sınavlar için Yapay Zekâ Destekli Web Tabanlı Ön Değerlendirme Sistemi Tasarımı: Rusça Dil Sınavları Örneği

Cihan KARA [1]

Yazılı sınavların değerlendirilmesi, öğretim elemanları için zaman alıcı ve yoğun emek gerektiren bir süreç olabilmektedir (Jauhiainen & Guerra, 2024). Özellikle yabancı dil eğitiminde, Rusça gibi karmaşık dil yapılarına sahip dillerde, bu zorluk daha da belirgindir. Bu çalışmanın amacı, yazılı sınavların değerlendirilmesinde öğretim elemanlarına yardımcı olacak, yapay zekâ destekli web tabanlı bir ön değerlendirme sistemi tasarlamaktır.

Sistemin teknik altyapısı şu şekildedir: Önerilen sistem, PHP ve MySQL tabanlı bir web sunucusu üzerine kurulmuştur. Kullanıcı arayüzü HTML, CSS ve JavaScript kullanılarak geliştirilmiştir. Uzun metinler üzerinde ücretsiz olarak yapay zekâ işlemleri gerçekleştirmek için Gemini API (Google, 2024) tercih edilmiştir. Sistemde, JavaScript Fetch API kullanılarak asenkron HTTP istekleri yapılmaktadır. Bu istekler, PHP'ye yönlendirilerek veritabanı işlemleri ve Gemini API çağrıları sunucu tarafında gerçekleştirilmektedir.

Sistemin çalışma prensibinin özeti şu şekildedir: Öğretim elemanı tarafından sınav soruları, yanıtları ve değerlendirme kriterleri sisteme tanımlanır. Ardından öğrenci sınav kâğıdı taranarak görsel formatında sisteme yüklenir. Yüklenen görseller Gemini API ile metne dönüştürülerek öğrenci sınav kâğıdı dijital olarak oluşturulur. Yapay zekâ, öğrenci yanıtlarını analiz ederek her soru için gerekçeleriyle birlikte bir puanlama yapar. Bu şekilde sistem, öğretim elemanına detaylı bir ön değerlendirme raporu sunar ve nihai puanlamaya olanak tanır.

Bu sistemin uygulanmasıyla, yazılı sınavların değerlendirilme sürecinin önemli ölçüde hızlanması beklenmektedir. Rusça dil sınavları özelinde, dilin karmaşık yapılarının daha tutarlı ve hızlı bir şekilde değerlendirilmesi mümkün olabilir. Öğretim elemanlarının değerlendirme için harcadıkları zamanın azalması, daha detaylı geri bildirim verebilmelerine olanak sağlayabilir.

Anahtar Kelimeler: *Yapay zekâ, yazılı sınavlar,* ölçme *ve değerlendirme, Rusça dil sınavları*

Kaynakça

Jauhiainen, J. S. & Guerra, A. G. (2024). Evaluating students' open-ended written responses with LLMs: Using the RAG framework for GPT-3.5, GPT-4, Claude-3, and Mistral-Large. arXiv. https://doi.org/10.48550/arXiv.2405.05444

Google. (2024). Gemini API. Google for Developers. https://ai.google.dev/gemini- api/docs/models/gemini?hl=tr (Erişim Tarihi: 25 Temmuz 2024)

[1] cihan_kara@anadolu.edu.tr, ORCID: 0000-0002-2638-7237, Anadolu Üniversitesi, Eskişehir, Türkiye

Ai-Supported Web-Based Pre-Assessment System Design for Written Exams: Example of Russian Language Exams

Cihan Kara [1]

The evaluation of written exams can be a time-consuming and labor-intensive process for instructors (Jauhiainen & Guerra, 2024). This challenge is particularly pronounced in foreign language education, especially in languages with complex structures like Russian. The aim of this study is to design an AI- supported web-based pre-assessment system to assist instructors in evaluating written exams.

The technical infrastructure of the system is as follows: The proposed system is built on a PHP and MySQL-based web server. The user interface is developed using HTML, CSS, and JavaScript. Google's Gemini API (Google, 2024) is chosen to perform AI operations on long texts free of charge. The system uses JavaScript Fetch API to make asynchronous HTTP requests. These requests are directed to PHP, where database operations and Gemini API calls are performed server-side.

The summary of the system's working principle is as follows: The instructor defines exam questions, answers, and evaluation criteria in the system. Then, student's exam paper is scanned and uploaded to the system in image format. The uploaded images are converted to text using the Gemini API, and the student exam paper is created digitally. AI analyzes student responses and scores each question with justifications. In this way, the system presents a detailed pre-assessment report to the instructor and allows for final scoring.

With the implementation of this system, it is expected that the evaluation process of written exams will be significantly accelerated. Specifically for Russian language exams, it may be possible to evaluate the complex structures of the language more consistently and quickly. The reduction in time spent by instructors on assessment may allow them to provide more detailed feedback.

Keywords: *Artificial intelligence, written exams, assessment and evaluation, Russian language exams*

Reference

Jauhiainen, J. S. & Guerra, A. G. (2024). Evaluating students' open-ended written responses with LLMs: Using the RAG framework for GPT-3.5, GPT-4, Claude-3, and Mistral-Large. arXiv. https://doi.org/10.48550/arXiv.2405.05444

Google. (2024). Gemini API. Google for Developers. https://ai.google.dev/gemini- api/docs/models/gemini?hl=en (Retrieved July 25, 2024)

[1] cihan_kara@anadolu.edu.tr, ORCID: 0000-0002-2638-7237, Anadolu University, Eskişehir, Türkiye

Öğrencilerin Test Çözme Davranışlarının eTIMSS 2019 Süreç Verileri Kullanılarak İncelenmesi: Türkiye Verisi için Bulgular

Derya Akbaş [1]

Teknolojik gelişmeler, bilgisayar tabanlı testlerin uygulanmasının yanı sıra kâğıt-kalem testlerinde mevcut olmayan ek verilerin toplanmasına da olanak sağlamıştır. Bilgisayar log dosyaları ve göz izleme gibi yöntemlerle toplanabilen veriler arasında yanıt süresi, göz hareketleri ve fare tıklamaları gibi veriler yer almaktadır. Bu süreç verileri, öğrencilerin maddeleri nasıl okuduklarını, hangi adımları izlediklerini ve yanıt verirken hangi yetkinlikleri ve bilgileri kullandıklarını ve maddelerle gerçekten ilgilenip ilgilenmediklerini ortaya çıkarmaya yardımcı olmaktadır (Ercikan ve Pellegrino, 2017). Yanıt süreçleri hakkında ayrıntılı bilgi sunan bu tür veriler, test tasarımı, alan testi, kalite güvencesi, performans ve etkileşim ile test ve madde geçerliliği dâhil olmak üzere değerlendirme döngüsü içinde çeşitli amaçlar için kullanılabilir (Maddox, 2023). Bu çalışma, eTIMSS 2019 matematik testinde dördüncü ve sekizinci sınıf öğrencilerinin test çözme davranışlarını incelemek için süreç verilerini analiz etmektedir. Çalışmada, madde ve genel başarı verileri, her bir madde ekranında harcanan toplam süre ve madde ekranı başına ziyaret sayısı dahil olmak üzere Türkiye örnekleminden elde edilen veriler kullanılmıştır (Fishbein vd., 2021). Bulgular, farklı performans düzeylerine (örneğin, orta seviye uluslararası ölçütlerin altında, düzeyinde veya üstünde) sahip öğrencilerin test çözme davranışlarını ve madde güçlüğünün bu davranışlar üzerindeki potansiyel etkisini belirleme açısından incelenecektir. Bulgular, test çözme davranışları ve bunların öğrenci performansıyla ilişkisi hakkında detaylı bilgiler sağlayacaktır.

Anahtar Kelimeler: *yanıt süreci, test*-çözme *davranışı, eTIMSS*

Kaynakça

Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. K. Ercikan ve J.W. Pellegrino (Ed.), Validation of score meaning for the next generation of assessments (s. 1-8) içinde. Routledge. https://doi.org/10.4324/9781315708591

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://www.iqb.hu-berlin.de/fdz/studies/TIMSS_2019/TIMSS2019UserGui.pdf

Maddox, B. (2023). The uses of process data in large-scale educational assessments. OECD Education Working Papers, No. 286. OECD Publishing. https://doi.org/10.1787/5d9009ff-en

[1] derya.akbas@adu.edu.tr, ORCID: 0000-0001-9852-4782, Aydın Adnan Menderes Üniversitesi, Aydın, Türkiye

Exploring Students' Test-Taking Behaviors Using Process Data in the eTIMSS 2019: Findings for Türkiye Data

Derya Akbaş [1]

Technological advancements have enabled the administration of computer-based tests as well as the collection of additional data that was not available with paper-and-pencil tests. Data, that can be collected through, such as, computer log files and eye-tracking, includes response time, eye movement, and mouse clicks. These process data help to reveal how students read items, what steps they follow and which competencies and knowledge they use when responding, and whether they truly engage with the items (Ercikan & Pellegrino, 2017). Such data, which presents detailed information about response processes, can be used for a variety of purposes within the assessment cycle, including test design, field testing, quality assurance, performance and engagement, and test and item validation (Maddox, 2023). This study analyzes process data to investigate the test-taking behaviors of fourth and eighth grade students in eTIMSS 2019 mathematics test. The study uses data from Türkiye sample including item and overall achievement data, total time spent on each item screen, and the number of visits per item screen (Fishbein et al., 2021). The findings will be examined in terms of identifying test-taking behaviors of students with varying performance levels (e.g., below, at, or above the intermediate international benchmark) and potential impact of item difficulty on these behaviors. The findings will provide detailed information into test-taking behaviors and their relationship to student performance.

Keywords: *Response process, Test-taking behavior, eTIMSS*

Reference

Ercikan, K., & Pellegrino, J. W. (2017). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J.W. Pellegrino (Eds.), Validation of score meaning for the next generation of assessments (pp. 1-8). Routledge. https://doi.org/10.4324/9781315708591

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). https://www.iqb.hu-berlin.de/fdz/studies/TIMSS_2019/TIMSS2019UserGui.pdf

Maddox, B. (2023). The uses of process data in large-scale educational assessments. OECD Education Working Papers, No. 286. OECD Publishing. https://doi.org/10.1787/5d9009ff-en

[1] derya.akbas@adu.edu.tr, ORCID: 0000-0001-9852-4782, Aydın Adnan Menderes University, Aydın, Türkiye

Tepki Sürelerinin Karma Modellemeye Dahil Edilmesi, Matematik Okuryazarlığı Örtük Grupların Belirlenmesini Nasıl Etkiler?

Halime Yıldırım Hoş [1], Menekşe Uysal Saraç[2]

Birçok araştırmacı, sınav katılımcılarının zaman baskısı altında kullandıkları "tahmin stratejilerini" yakalamak için çeşitli modelleme yaklaşımları önermiştir (Chang, Tsai, & Hsu, 2014). Ayrıca, düşük ölçekli değerlendirmelerde başarılı olmak için motive olmayan bireylerin de tahmin stratejilerine başvurabildiği gözlemlenmiştir (Wise ve DeMars, 2006). Bu çalışmalarda tahmin genellikle yetenekten bağımsız bir madde çözme stratejisi olarak tanımlanmaktadır.

Jeon ve De Boeck'in (2019), Van der Linden (2007) tarafından geliştirilen hız ve doğruluk modellemesinin hiyerarşik çerçevesine dayanan karma MTK modeli, tepki sürelerini ve stratejik davranışları dikkate alarak test alma süreçlerini daha kapsamlı bir şekilde analiz etmektedir. Bu yaklaşım, bireylerin yanıt süreçlerini hem yetenek düzeylerine hem de yanıt sürelerine göre değerlendirerek model uyumunu iyileştirmektedir. Bu çalışmanın amacı, PISA 2022 matematik okuryazarlığı verisinde yetenek ve yeteneksizlik örtük sınıflarını belirlemek ve yanıt sürelerinin bu sınıflar üzerindeki etkisini değerlendirmektir. Burada "yeteneksiz sınıf" terimi, matematiksel yetenekleri ölçtüğü varsayılan "normal" yetenek sınıfının dışında kalan yanıt sürelerine dayalı olarak belirlenen alt grupları ifade etmektedir.

Sonuç olarak, tepki sürelerinin bir kişinin yeteneğini tahmin etmek için önemli bir bilgi kaynağı olduğu sıklıkla vurgulanmaktadır, bireylerin yeteneklerinin daha doğru tahmin edilmesine ve hız-doğruluk dengesinin daha etkili bir şekilde değerlendirilmesine olanak tanıyarak daha kesin değerlendirmelere yol açmaktadır (Van der Linden, 2007). Bu çalışma, yanıt sürelerinin PISA 2022 matematik okuryazarlığı verilerinde nasıl kullanılabileceğini ve bu kullanımların ikincil örtük sınıfları nasıl etkilediğini belirleyecektir.

Anahtar Kelimeler: *Tepki süresi, Karma Madde Tepki Kuramı, Matematik Okuryazarlığı*

Kaynakça

Chang, Y. C., Tsai, C. C., & Hsu, H. C. (2014). The impact of guessing strategies on item response theory model parameters**.** *Educational and Psychological Measurement*, *74*(1), 69-85.

Jeon, M. H., & De Boeck, P. (2019). A hierarchical model for response times and accuracy in educational testing**.** *Applied Psychological Measurement*, *43*(1), 3-21.

Van der Linden, W. J. (2007). The impact of response times on test scoring and test design. *Psychometrika*, *72*(1), 1-17.

Wise, S. L., & DeMars, C. E. (2006). Using response time data to improve the quality of test scoring. *Educational Assessment*, *11*(2), 77-96.

[1] halime.yldrm@gmail.com: ORCID: 0000-0001-6479-4469, İstanbul Medeniyet University, İstanbul, Türkiye

[2] menekseysl@gmail.com: ORCID: 0000-0002-9847-124, Çankırı Karatekin University, Çankırı, Türkiye

How Integrating Response Times into Mixture Modeling Effect the Identification of Latent Groups in Mathematical Literacy?

Halime Yıldırım Hoş [1], Menekşe Uysal Saraç[2]

Many researchers have proposed a variety of modeling approaches to capture the "guessing strategies" employed by test-takers under time pressure (Chang, Tsai, & Hsu, 2014). Furthermore, it has been observed that individuals who are not motivated to succeed in low-stakes assessments may also engage in guessing strategies (Wise & DeMars, 2006). In these studies, guessing is typically defined as an item-solving strategy that is independent of ability.

Jeon and De Boeck's (2019) mixture IRT model, based on the hierarchical framework of speed and accuracy modeling developed by Van der Linden (2007), analyzes test-taking processes more comprehensively by considering response times and strategic behaviors. This approach aims to improve model fit by evaluating individuals' response processes based on both their abilities and response times. This study aims to identify latent classes of ability and no-ability in PISA 2022 mathematical literacy data and to assess the effect of response times on these classes. Here, the term "no-ability class" refers to subgroups identified based on response times that fall outside the "normal" ability class assumed to measure mathematical abilities.

As a result, it is often emphasized that response times are an important source of information for estimating a person's ability, allowing for more accurate estimation of individuals' abilities and more effective assessment of the speed-accuracy, leading to more precise assessments (Van der Linden, 2007). This study will determine how response times can be used in PISA 2022 mathematical literacy data and how these uses affect secondary latent classes.

Keywords: *Response time, Mixture Item Response Theory, Mathematical Literacy*

Reference

Chang, Y. C., Tsai, C. C., & Hsu, H. C. (2014). The impact of guessing strategies on item response theory model parameters**.** *Educational and Psychological Measurement*, *74*(1), 69-85.

Jeon, M. H., & De Boeck, P. (2019). A hierarchical model for response times and accuracy in educational testing**.** *Applied Psychological Measurement*, *43*(1), 3-21.

Van der Linden, W. J. (2007). The impact of response times on test scoring and test design. *Psychometrika*, *72*(1), 1-17.

Wise, S. L., & DeMars, C. E. (2006). Using response time data to improve the quality of test scoring. *Educational Assessment*, *11*(2), 77-96.

[1] halime.yldrm@gmail.com: ORCID: 0000-0001-6479-4469, İstanbul Medeniyet University, İstanbul, Türkiye

[2] menekseysl@gmail.com: ORCID: 0000-0002-9847-124, Çankırı Karatekin University, Çankırı, Türkiye

Yapay Zeka Tarafından Oluşturulan Olgu Temelli Çoktan Seçmeli Maddelerin Değerlendirilmesi: Pediatri Örneği

Gülşen Taşdelen Teker [1], Dilara Bakan Kalaycıoğlu[2], Melek Gülşah Şahin[3], Saliha Esenboğa[4]

Çoktan seçmeli maddeler, tıp eğitiminin her düzeyinde sıklıkla kullanılan bir madde türüdür. Mezuniyet öncesi tıp eğitiminde özellikle klinik öncesi dönem olmak üzere klinik dönemde de sıklıkla kullanılmaktadır (Farley, 1989; Royal vd., 2018). Klinik öncesi dönemde daha çok salt bilginin yoklandığı geleneksel çoktan seçmeli maddeler yaygın olarak kullanılmaktadır. Klinik dönemde ise klinik akıl yürütme becerilerinin işe koşularak cevaplanabilen olgu temelli çoktan seçmeli maddeler tercih edilmektedir. Yaygın kullanılmakla birlikte nitelikli çoktan seçmeli madde yazmak görece zorlu bir süreçtir. Özellikle olgu temelli çoktan seçmeli maddelerin yazılması sürecinde hem olgunun gerektiği kadar bilgiyi içerek şekilde iyi yapılandırılması gerekmektedir hem de seçeneklerin her birinin dikkatli seçilmesi gerekmektedir. Tıp eğitimi için test geliştirmedeki en önemli zorluklardan biri, yüksek kaliteli, klinik açıdan uygun olgu temelli çoktan seçmeli sorular oluşturmaktır. Bu araştırmada, yapay zeka kullanılarak üretilen olgu temelli çoktan seçmeli maddelerin bilimsel doğruluğu, kapsama uygunluğu ile güçlük ve ayırt edicilik düzeyleri açısından hedef gruba uygulanabilirliği açısından değerlendirilmesi amaçlanmıştır. Bu amaçla 2023-2024 eğitim öğretim yılında Hacettepe Üniversitesi Tıp Fakültesi Dönem 4 öğrencilerinden pediatri stajını alan öğrencilere uygulamak üzere ChatGPT4 kullanılarak pediatrik immünoloji alanında soru üretilmiştir. Dönem 4 pediatri stajında pediatrik immünoloji kapsamında ele alınan sekiz konudan ikişer soru üretilmiş ve bu sorular pediatrik immünoloji alanından üç uzmanının görüşüne sunulmuştur. Alan uzmanları görüşü sonrası uygun olduğu düşünülen her konudan birer soru olmak üzere toplam sekiz soru 34 Dönem 4 öğrencisine uygulanmıştır. Elde edilen veriler ışığında madde güçlüğü ortalamasının 0.67 (minimum 0.32, maksimum 0.97), madde ayırt edicilik ortalamasının ise 0.32 (minimum 0.13, maksimum 0.47) çıktığı görülmüştür. Seçeneklerin işaretlenme oranları incelendiğinde ise iki madde için sadece ikişer seçeneğin işaretlendiği diğer altı maddede ise tüm seçeneklerin işaretlendiği görüşmüştür. Sonuç olarak sınırlı bir öğrenci grubuna uygulanmakla birlikte madde parametreleri açısından kabul edilebilir değerlerin elde edildiği görülmüştür. Alan uzmanlarının görüşleri doğrultusunda ise çok küçük düzeltmelerle nitelikle olgu yazmanın ve seçeneklerin gözden geçirilmesiyle amaca hizmet edecek nitelikte olgu temelli çoktan seçmeli madde yazmanın mümkün olduğu belirtilmiştir. Özetle yapay zeka desteğiyle olgu temelli çoktan seçmeli madde yazımı sürecinin kolaylaştığı görülmüştür.

Anahtar Kelimeler: *yapay zeka, olgu temelli* çoktan *seçmeli madde, pediatri*

Kaynakça

Farley, J.K. (1989). The multiple choice test: writing the questions. Nurse Educator, 14(6), 10–39. https://doi.org/10.1097/00006223-198911000-00003

Royal, K. D., Hedgpeth, M., Jeon, T. & Colford, C. M. (2018). Automated item generation: the future of medical education assessment? European Medical Journal, 2(1), 88-93.

[1] gulsentasdelen@hacettepe.edu.tr, ORCID: 0000-0003-3434-4373, Hacettepe Üniversitesi, Ankara, Türkiye

[2] dilarabakan@gazi.edu.tr, ORCID: 0000-0003-1447-6918, Gazi Üniversitesi, Ankara, Türkiye

[3] mgulsahsahin@gazi.edu.tr, ORCID: 0000-0001-5139-9777, Gazi Üniversitesi, Ankara, Türkiye

[4] saliha.esenboga@hacettepe.edu.tr, ORCID: 0000-0003-0562-9863, Hacettepe Üniversitesi, Ankara, Türkiye

Evaluation of Case-Based Multiple-Choice Items Generated by Artificial Intelligence: A Pediatrics Example

Gülşen Taşdelen Teker [1], Dilara Bakan Kalaycıoğlu[2], Melek Gülşah Şahin[3], Saliha Esenboğa[4]

Multiple-choice items are a frequently used type of question at all levels of medical education. They are commonly employed in undergraduate medical education, especially during the preclinical period, and also during the clinical period (Farley, 1989; Royal et al., 2018). In the preclinical period, traditional multiple-choice items that primarily test factual knowledge are widely used. During the clinical period, case-based multiple-choice items that can be answered through clinical reasoning skills are preferred. Although commonly used, writing high-quality multiple-choice items is a relatively challenging process. In particular, when writing case-based multiple-choice items, it is necessary to structure the case to include adequate information and carefully select each option. One of the most significant challenges in test development for medical education is creating high-quality, clinically relevant case-based multiple-choice questions. This study aims to evaluate the scientific accuracy, content relevance, and applicability to the target group of case-based multiple-choice items generated using artificial intelligence in terms of difficulty and discrimination levels. For this purpose, questions were generated in the field of pediatric immunology using ChatGPT-4 for fourth-year medical students during their pediatrics internship at Hacettepe University Faculty of Medicine in the 2023-2024 academic year. Two questions were generated for each of the eight topics covered in pediatric immunology during the fourth-year pediatrics internship, and these questions were presented to three experts in the field of pediatric immunology for their opinions. Following expert feedback, one question deemed appropriate from each topic was administered to a total of 34 fourth-year students. The data obtained revealed that the average item difficulty was 0.67 (minimum 0.32, maximum 0.97), and the average item discrimination was 0.32 (minimum 0.13, maximum 0.47). When examining the response rates for the options, it was found that only two options were selected for two items, while all options were selected for the other six items. In conclusion, although applied to a limited group of students, acceptable values were obtained in terms of item parameters. According to expert opinions, it was suggested that with minor adjustments, it is possible to write high-quality case-based multiple-choice items that serve the intended purpose by reviewing the options. In summary, it was observed that the process of writing case-based multiple-choice items was facilitated with the support of artificial intelligence.

Keywords: *artificial intelligence, case based multiple choice questions, pediatric immunology*

References

Farley, J.K. (1989). The multiple choice test: writing the questions. Nurse Educator, 14(6), 10–39. https://doi.org/10.1097/00006223-198911000-00003

Royal, K. D., Hedgpeth, M., Jeon, T. & Colford, C. M. (2018). Automated item generation: the future of medical education assessment? European Medical Journal, 2(1), 88-93.

[1] gulsentasdelen@hacettepe.edu.tr, ORCID: 0000-0003-3434-4373, Hacettepe Üniversitesi, Ankara, Türkiye

[2] dilarabakan@gazi.edu.tr, ORCID: 0000-0003-1447-6918, Gazi Üniversitesi, Ankara, Türkiye

[3] mgulsahsahin@gazi.edu.tr, ORCID: 0000-0001-5139-9777, Gazi Üniversitesi, Ankara, Türkiye

[4] saliha.esenboga@hacettepe.edu.tr, ORCID: 0000-0003-0562-9863, Hacettepe Üniversitesi, Ankara, Türkiye

Geri Bildirimin Başarıya Etkisi: Bibliyometrik Bir Analiz

Burcu DEMİRÖZ [1], Nuri DOĞAN[2]

Geri bildirim, kişinin performansının belirli yönleri hakkında sağlanan bilgidir (Hattie & Timperley, 2007). Geri bildirimin başarı üzerindeki etkisinin incelendiği birçok araştırma bulunmaktadır. Yapılan araştırmalarda geri bildirimin başarı üzerinde etkisi olduğu belirlenmiştir. Bu araştırmada geri bildirimin başarı üzerindeki etkisi konusunda yapılan araştırmaların genel durumunu, dönemin en çok atıf alan makaleleri ve yazarları, en çok çalışma yapan ülkeleri ve yayın evlerini belirlemek amaçlanmıştır.

Araştırmanın amacı doğrultusunda bibliyometrik analiz yöntemi kullanılmıştır. Farklı yaklaşımları temsil eden birçok veri tabanı vardır. Web of Science (WoS), Scopus, Google Scholar ve PubMed bunlardan bazılarıdır. Araştırma kapsamında WoS ve Scopus veri tabanları incelenmiştir. Veri tabanlarında yer alan kaynakların başlık, özet ve anahtar kelimelerinde "feedback" and "student performance or achievement or individual performance or participant performance" and "impact" or "effect" and "student" or "individual" or "participant" and "education" bağlantısı kurularak tarama yapılmıştır. Tarama sonucunda WoS veri tabanında 627, Scopus veri tabanında 1334 araştırmaya ulaşılmıştır. Ulaşılan araştırmalar, 1992 ile 2024 yılları arasında WoS veri tabanında eğitim ve Scopus veri tabanın da sosyal bilimler alanında İngilizce olarak yayınlanan makaleler ile sınırlandırılmıştır. Bu sınırlandırmalar sonucunda WoS veri tabanında 639, Scopus veri tabında 1366 makaleden oluşan bibliyometrik veri seti elde edilmiştir. Veri setinde konu ile ilişkisi olmayan araştırmalarla karşılaşıldığı için tüm araştırmalar incelenmiştir. Yapılan inceleme sonucunda nihai veri seti oluşturulmuştur. Nihai veri seti Wos veri tabanından 172; Scopus veri tabanından 176 makaleyi içermektedir. Veri analizleri RStudio (R Core Team, 2024) yazılımı "bibliometrix" paketi (Aria & Cuccurullo, 2017) kullanılarak gerçekleştirilmiştir.

Araştırmaların yapıldığı ülkeler incelendiğinde en fazla araştırmanın Amerika'da yapıldığı belirlenmiştir. En çok atıf 2016 yılında yapılmıştır. En çok atıfı Ma, W. (2017) ve "Computers and Education" almıştır.

Anahtar Kelimeler: *Geri bildirim, başarı, bibliyometrik analiz.*

Kaynakça

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis.

Journal of Informetrics, 11(4), 959-975.

Hattie, J., & Zierer, K. (2019). *Visible learning insights*. Routledge.

R Development Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[1] burcu199514.bdz@gmail.com, ORCID: 0000-0002-8326-4671, MEB, Gaziantep, Türkiye

[2] nurid@hacettepe.edu.tr:, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

The Effect of Feedback on Achievement: A Bibliometric Analysis

Burcu DEMİRÖZ [1], Nuri DOĞAN[2]

Feedback is information provided regarding specific aspects of an individual's performance (Hattie & Timperley, 2007). Numerous studies have examined the effect of feedback on achievement, and it has been determined that feedback influences achievement. This study aims to determine the general state of research on the effect of feedback on achievement, identify the most cited articles and authors of the period, the countries conducting the most research, and the publishers.

In line with the aim of the study, the bibliometric analysis method was employed. There are many databases representing different approaches, including Web of Science (WoS), Scopus, Google Scholar, and PubMed. In the scope of the research, the WoS and Scopus databases were examined. A search was conducted in the databases by establishing a connection between the terms "feedback" and "student performance or achievement or individual performance or participant performance" and "impact" or "effect" and "student" or "individual" or "participant" and "education" in the titles, abstracts, and keywords of the sources. As a result of the search, 627 studies were found in the WoS database and 1,334 in the Scopus database. The studies found were limited to articles published in English between 1992 and 2024 in the field of education in the WoS database and in the field of social sciences in the Scopus database. After these limitations, a bibliometric dataset consisting of 639 articles from the WoS database and 1,366 articles from the Scopus database was obtained. Since some studies unrelated to the topic were encountered in the dataset, all studies were reviewed. As a result of the review, the final dataset was created. The final dataset includes 172 articles from the WoS database and 176 articles from the Scopus database. Data analyses were conducted using the "bibliometrix" package (Aria & Cuccurullo, 2017) in RStudio (R Core Team, 2024).

The most research was carried out in the USA. The most cited author was Ma, W. (2017), and the journal "Computers and Education" received the most citations.

Keywords: *Feedback, achievement, bibliometric analysis.*

Reference

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959-975.

Hattie, J., & Zierer, K. (2019). *Visible learning insights*. Routledge.

R Development Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[1] burcu199514.bdz@gmail.com, ORCID: 0000-0002-8326-4671, MEB, Gaziantep, Türkiye

[2] nurid@hacettepe.edu.tr:, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Baskın Maddeler Faktör Sayısı Belirleme Yöntemlerinin Performansını Nasıl Etkiler?

Abdullah Faruk Kılıç [1], Tugay KAÇAK[2], Meltem ACAR GÜVENDIR[3], Gül GÜLER[4]

Bu çalışmada, Zwick & Velicer'in (1986) bir faktörde en az 2 baskın madde olacağına dair önerisi, yani yüksek yüklemeler olan maddeler, analiz ve simüle edildi. Çok sayıda Monte Carlo simülasyon çalışmasında, faktör yükleri dar bir aralıkta belirlenir. Örneğin, ortalama faktör yüklemesi 0.40 olan koşullarda, faktör yüklemeleri 0.35-0.45 arasında eşit olarak dağıtılarak konjenik ölçümleri temsil eder. Ancak, ampirik çalışmalarda genellikle faktör yüklemeleri bu şekilde gözlenmemektedir. Faktörün doymamış olması da bir sorun olarak görülmektedir. Bu doğrultuda, iki baskın madde, beş baskın madde, dar bir faktör yük aralığına sahip maddeler ve eşit faktör yükleri olan (paralel) maddeleri modelledik. Örneklem büyüklüğü (50, 200, 500), çarpıklık (-2.5, 0, +2.5) ve ortalama faktör yükü (0.440 ve 0.575) diğer koşullardı ve test uzunluğu 10, kategori sayısı 5 ve tek boyutlu yapılar sabit tutuldu. Toplamda, 72 koşul ve 1000 replikasyon, göreli yanlılık ve doğru tahmin yüzdesi kullanılarak değerlendirildi.

Faktör sayısını belirleme yöntemlerine dair iki temel sonuç çıkarılabilir: i)EGA-TMFG ve EGA-Glasso, gerçek faktör sayısını olduğundan daha yüksek belirlemiştir ii) MAP/MAP4, örneklem boyutları küçük olsa bile kabul edilebilir bir yanlılık ile performans göstermiştir. Zwick & Velicer'in önerisi, tüm maddelerin faktör yükleri arttığında küçük ve orta örneklem büyüklükleri için daha uygulanabilir olarak yorumlanabilir. Baskın madde sayısındaki artış yöntemlerin performansı üzerinde olumlu bir etkiye sahiptir. Bununla birlikte, her iki EGA yöntemi ve Hull küçük örneklemlerde zıt yanlılık sergilemiştir.

Anahtar Kelimeler: *faktör sayısı belirleme, baskın madde, faktör analizi.*

Kaynakça

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99(3), 432–442. https://doi.org/10.1037/0033- 2909.99.3.432

[1] afarukkilic@trakya.edu.tr:, ORCID: 0000-0003-3129-1763, Trakya Üniversitesi, Edirne, Türkiye

[2] tugaykacak@trakya.edu.tr:, ORCID: 0000-0002-5319-7148, Trakya Üniversitesi, Edirne, Türkiye

[3] meltemacar@trakya.edu.tr:, ORCID: 0000-0002-3847-0724, Trakya Üniversitesi, Edirne, Türkiye

[4] gulguler@trakya.edu.tr:, ORCID: 0000-0001-8626-4901, Trakya Üniversitesi, Edirne, Türkiye

To What Extent Dominant Items Effect on Factor Retention Methods' Performance?

Abdullah Faruk KILIÇ[1],Tugay KAÇAK[2], Meltem ACAR GÜVENDIR[3], Gül GÜLER[4]

In this study, the suggestion by Zwick & Velicer (1986) that there should be at least 2 dominant items per factor, i.e., items with high loadings, was analyzed and simulated. In many Monte Carlo simulation studies, factor loadings are determined within a narrow range. For instance, under conditions where the average factor loading is 0.40, the factor loadings are evenly distributed between 0.35 and 0.45 to represent congeneric measurements. However, in empirical studies, factor loadings are often not observed in this way. The issue of a factor being under-saturated is also seen as a problem. In this context, we modelled two dominant items, five dominant items, items with a narrow range of factor loadings, and items with equal factor loadings (parallel). Other conditions included sample size (50, 200, 500), skewness (-2.5, 0, +2.5), and average factor loadings (0.440 and 0.575), while the test length was fixed at 10, the number of categories at 5, and the structures were unidimensional. In total, 72 conditions and 1000 replications were evaluated using relative bias and the percentage of correct estimates.

Two main results can be drawn regarding methods for determining the number of factors: i) EGA-TMFG and EGA-Glasso tended to overestimate the true number of factors ii) MAP/MAP4 showed acceptable performance with acceptable bias even with small sample sizes. Zwick & Velicer's suggestion can be interpreted as more applicable for small and medium sample sizes when all item factor loadings are increased. The increase in the number of dominant items positively affects the performance of methods. However, EGAs and Hull exhibited contrast biased estimates in small samples.

Keywords: *factor retention, dominant items, factor analysis*

Reference

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99(3), 432–442. https://doi.org/10.1037/0033- 2909.99.3.432

[1] afarukkilic@trakya.edu.tr:, ORCID: 0000-0003-3129-1763, Trakya Üniversitesi, Edirne, Türkiye

[2] tugaykacak@trakya.edu.tr:, ORCID: 0000-0002-5319-7148, Trakya Üniversitesi, Edirne, Türkiye

[3] meltemacar@trakya.edu.tr:, ORCID: 0000-0002-3847-0724, Trakya Üniversitesi, Edirne, Türkiye

[4] gulguler@trakya.edu.tr:, ORCID: 0000-0001-8626-4901, Trakya Üniversitesi, Edirne, Türkiye

Madde Parametrelerinin Tahmininde Ortak Madde Bağlantısı Yönteminin Geçerliliği

İbrahim Hakkı TEZCİ [1]

Ortak madde bağlantısı (Common Item Linking), farklı test formlarında yer alan ortak maddeler aracılığıyla, bu test formlarının sonuçlarını karşılaştırma ve birleştirme sürecidir. Bu yöntem, test formlarının her birinde yer alan belirli sayıda ortak madde kullanarak, farklı testler arasında bir bağlantı kurulmasını sağlar. Bu sayede, tüm madde havuzunun her bir katılımcıya uygulanması gerekmeksizin, madde parametreleri doğru bir şekilde tahmin edilebilir. Ortak madde bağlantısı yöntemi, madde parametrelerinin doğru tahmin edilmesinde etkili bir yöntem olarak kabul edilse de, bu yöntemin parametreleri ne kadar doğru ifade ettiği konusunda daha fazla araştırma yapılması gerekmektedir. Bu çalışmanın amacı, ortak madde bağlantısı yönteminin madde parametrelerini ne derece doğru tahmin ettiğini araştırmaktır. Bu araştırma, yöntemin geçerliliğini ve güvenilirliğini test ederek, eğitim değerlendirmelerinde ve psikometrik analizlerde kullanılabilirliğini belirlemeyi hedeflemektedir.

Bu araştırma kapsamında, belirli sayıda öğrenciye 6 farklı test formu uygulanmıştır. Her bir öğrenci, toplamda 155 soruya cevap vermiştir. Ancak, bu öğrenciler 6 gruba bölünerek, her bir grup 6 farklı formda 5 ortak soru cevaplandırmıştır. Araştırmanın dizaynı, ortak maddeler aracılığıyla farklı test formlarından elde edilen verilerin birleştirilmesini ve madde parametrelerinin tahmin edilmesini içermektedir. Araştırmada, ortak madde bağlantısı yöntemi ile tahmin edilen madde parametrelerinin doğruluğu ve geçerliliği test edilmektedir.

Araştırma kapsamında, ortak madde bağlantısı yöntemi kullanılarak elde edilen madde parametreleri, farklı analiz ve karşılaştırmalarla değerlendirilecektir. İlk olarak, simülasyon çalışmaları ile tahmin edilen parametrelerin gerçek parametrelerle karşılaştırılması yapılacaktır. Daha sonra, gerçek veri çalışmaları kapsamında farklı test formlarından elde edilen parametrelerin tutarlılığı ve geçerliliği incelenecektir. İstatistiksel analizler, ortalama mutlak hata (MAE), kök ortalama kare hata (RMSE) ve korelasyon analizleri gibi ölçütler kullanılarak yapılacaktır. Ayrıca, ANOVA testleri ile farklı test formlarındaki madde parametreleri arasında anlamlı fark olup olmadığı değerlendirilecektir.

Araştırmanın sonucunda, ortak madde bağlantısı yönteminin madde parametrelerini doğru bir şekilde tahmin ettiğini ve eğitim değerlendirmelerinde güvenilir bir şekilde kullanılabileceğini göstermeyi amaçlamaktadır. Beklenen sonuçlar, yöntemin simülasyon ve gerçek veri çalışmaları ile desteklendiğini ve parametre tahminlerinde yüksek doğruluk ve tutarlılık sağladığını ortaya koyabilir. Ayrıca, farklı test formları arasındaki tutarlılık ve geçerliliğin yüksek olması, yöntemin eğitim değerlendirmelerinde pratik bir araç olarak kullanılabileceğini gösterecektir.

Anahtar Kelimeler: *Madde tepki kuramı, madde parametre tahmini, ortak madde bağlama*

[1] ihtezci@akdeniz.edu.tr, ORCID: 0000-0003-0273-8853, Akdeniz Üniversitesi, Antalya, Türkiye

The Validity of Common Item Linking Method in Estimating Item Parameters

İbrahim Hakkı TEZCİ [1]

Common Item Linking is a process that uses common items present in different test forms to compare and combine the results of these test forms. By employing a specific number of common items in each test form, this method establishes a connection between different tests. Consequently, it allows for accurate estimation of item parameters without necessitating the administration of the entire item pool to each participant. Although Common Item Linking is recognized as an effective method for accurately estimating item parameters, further research is needed to determine the extent to which this method accurately reflects the parameters. The purpose of this study is to investigate how accurately the Common Item Linking method estimates item parameters. This research aims to determine the validity and reliability of the method and its applicability in educational assessments and psychometric analyses.

In this study, six different test forms were administered to a specific number of students. Each student answered a total of 155 questions. However, the students were divided into six groups, and each group answered five common questions in the six different forms. The design of the study involves combining the data obtained from different test forms through common items and estimating the item parameters. The accuracy and validity of the item parameters estimated using the Common Item Linking method are being tested in this study.

The item parameters obtained using the Common Item Linking method will be evaluated through various analyses and comparisons. First, simulation studies will be conducted to compare the estimated parameters with the actual parameters. Subsequently, the consistency and validity of the parameters obtained from different test forms in real data studies will be examined. Statistical analyses will be performed using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and correlation analyses. Additionally, ANOVA tests will be used to assess whether there are significant differences in item parameters across different test forms.

The study aims to demonstrate that the Common Item Linking method accurately estimates item parameters and can be reliably used in educational assessments. Expected results may show that the method is supported by both simulation and real data studies, providing high accuracy and consistency in parameter estimation. Furthermore, high consistency and validity across different test forms would indicate that the method can be practically used in educational evaluations.

Keywords: *Item Response Theory, Item Parameter Estimation, Common Item Linking*

[1] ihtezci@akdeniz.edu.tr:, ORCID: 0000-0003-0273-8853, Akdeniz University, Antalya, Türkiye

Katılımcıların Test Sürecine Dahil Olma Durumlarının İncelenmesi: Bir Yöntem Önerisi ve Farklı Yöntemlerin Karşılaştırılması

Başak ERDEM-KARA [1], Murat Doğan ŞAHİN [2]

Başarı testleri, öğrenci performansını değerlendirmek için eğitimde yaygın olarak kullanılmaktadır. İdeal olarak, başarı testlerinde öğrencilerin test maddelerinde ellerinden gelen en iyi performansı gösterdikleri ve böylece test puanlarının ölçülen yapıyı gerçekten yansıttığı varsayılır. Ancak bazen bu varsayım ihlal edilir çünkü öğrenciler bir maddeyi doğru bir şekilde yanıtlamak için gereken azami çabayı göstermeyebilir ve gerçek yetenek düzeyleriyle tutarlı olmayan bir yanıt verebilirler. Bu yanıt verme davranışı 'kopuk yanıt verme – disengaged responding' olarak adlandırılır. Kopuk yanıt verme olgusu, test puanlarının geçerliğini tehdit eden, yapıyla ilgisiz bir varyans faktörü yaratır. Bu yanıtların neden olabileceği potansiyel sorunlar göz önüne alındığında, bu konu araştırmacılar için önemli ve ilginç bir araştırma konusu haline gelmiş ve kopuk yanıtların etkilerini tespit etmek ve düzeltmek için çeşitli yöntemler geliştirilmiştir.

Bu araştırma kapsamında, kopuk-bağlantısız yanıtları tespit etmek için kullanılabilecek sıralı bir argüman önerilmesi amaçlanmaktadır. Daha önceki araştırmalarda, bir tespit yöntemi olarak test eşit şekilde bölünerek öğrencilerin testlerin birinci ve ikinci yarısındaki performanslarındaki değişim incelenmiştir. Bu araştırmada yeni bir argüman olarak, testi eşit olarak değil, sırayla iki parçaya bölen sıralı bir yöntem önerilmeye çalışılacaktır. Bu doğrultuda araştırmacılar tarafından 30 maddelik bir başarı testi geliştirilmiş ve yaklaşık 350 ortaokul öğrencisinden veri toplanmıştır. 30 maddelik testte maddeler 15-15 maddelik iki bölüme değil, 11-19, 12-18,.... 18-12, 19-11 vb olacak şekilde iki bölüme ayrılacaktır. Amacımız, kopmanın gerçekleşmeye başladığı en uygun noktayı belirlemektir. Daha sonra, belirlediğimiz noktaya göre öğrencilerin testin iki bölümündeki yeteneklerini kestirerek (maksimum olabilirlik kestirimi) süreçten ayrışma davranışı gösteren bireyler belirlenecektir. MTK'ya dayalı kişi- uyum indeksleri gibi diğer yöntemler de süreçten ayrışan öğrencileri (disengaged responders) belirlemek için kullanılacaktır. Bu yöntemlerin bizim önerdiğimiz yöntemle tutarlılığı incelenecektir.

Anahtar Kelimeler: *kopuk yanıtlama davranışı, başarı testi, kişi-uyum indeksleri*

[1] basakerdem@anadolu.edu.tr, ORCID: 0000-0003-3066-2892, Anadolu Üniversitesi, Eskişehir, Türkiye

[2] mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu Üniversitesi, Eskişehir, Türkiye

Are Participants Fully Engaged During a Test Process? A Sequential Argument and Comparison with Different Methods

Başak ERDEM-KARA [1], Murat Doğan ŞAHİN[2]

Achievement tests are widely used in education to assess student performance. Ideally, in achievement testing it is assumed that students perform to the best of their ability on test items so that test scores truly reflect the construct being measured. Sometimes, however, this assumption is violated because students may not use the maximum effort necessary to process an item accurately and provide an answer that is not consistent with their true ability level. This response behavior is referred to as 'disengaged responding'. The phenomenon of disengaged responding (dr) creates a construct-irrelevant variance factor that threatens the validity of test scores. Given the potential problems that 'dr' can cause, it has become an important and interesting research topic for researchers, and several methods have been developed to detect and correct the effects of dr.

Recent research aims to propose a sequential argument that can be used to identify disengaged responders. In previous studies, as a detection method, the test was equally divided into half and the change in students' performance in the first and second half of the tests was analysed. As a new argument, we will try to propose a sequential method dividing the test into two parts not equally but sequentially. In this direction, a 30-item achievement test that was developed by researchers and data was collected from approximately 350 secondary school students. For the 30-item test, we will not divide the items into two parts of 15-15 items, but 11-19, 12-18,.... 18-12, 19-11 etc. Our aim is to specify an optimal point at which disengagement begins to occur. We will then identify disengaged responders by estimating the students' ability (maximum likelihood estimation) on the two parts of the test based on our specified point. Other methods such as IRT person fit analyses will also be conducted to identify disengaged students. The consistency of these methods with our proposed method will be examined.

Keywords: *disengaged responding, achievement tests, person-fit indices*

[1] basakerdem@anadolu.edu.tr, ORCID: 0000-0003-3066-2892, Anadolu Üniversitesi, Eskişehir, Türkiye

[2] mdsahin@anadolu.edu.tr, ORCID: 0000-0002-2174-8443, Anadolu Üniversitesi, Eskişehir, Türkiye

Matlab Programında Destek Vektör Makinesi İle Sınıflandırma Doğrulunun İncelenmesi

Fazilet TAŞDEMİR [1]

Verilerin sınıflanması amacıyla kullanılan istatistiksel programlar ve veri analizleri bulunmaktadır. MATLAB (Matrix laboratory) yazılım programı da eleman, birimler arasındaki uzaklıkları veren sınıflama amacı ile de kullanılabilen ve hesaplama kolaylığı sunan bir programdır. MATLAB, daha çok sayısal lineer cebir hesaplamaları, bulanık sistemler, data analizi ve görselleştirme, kodlama, mühendislik, makine öğrenmesi, derin öğrenme, simülasyon ve yazılım alanlarında araştırmacılar tarafından tercih edilmektedir.

Bu araştırmada, MATLAB programında Destek Vektör Makinesi (Support Vector Machine) ile sınıflandırma doğrulunun incelenmesi amaçlanmıştır. Bu amaç doğrultusunda belirlenen sorulara yanıt aranmıştır: 1. Destek Vektör Makinesi ile belirlenen modelin hata oranı nedir? 2.Kernel Function ve Gaussian parametrelerine göre oluşan modelde belirlenen hata oranı nedir? Araştırma sorularının yanıtlanması ve MATLAB programının kullanılabilmesi için belirli sayısal değerlere ihtiyaç duyulmaktadır. Bu sebeple araştırmanın çalışma grubu için simülatif veri üretilmiştir. Simülasyon, kullanıcının geniş kontrole sahip olmasını sağlar.

Çalışma grubunda online olarak kamuya açık olan 2023 Yüksek Öğretim Kurulu Tercih listesinde yer alan devlet üniversitelerinin tıp fakültelerinin kontenjan, başarı sıralamaları, başarı puanları ve yerleşen adayları ölçüt alınarak oluşturulmuştur. Bu verilerden 99'u tıp fakültesi tercih listesinden aynen alınmış, 101 veri ise bu veriler doğrultusunda simülatif üretilmiştir. Veri analiz 200 adaylık bir simülatif veri üzerinden yürütülmüştür. Veri analizinde kullanılan*>>pt=cvpartition(veri.yerl esme,"HoldOut",0.2);>>hdTrain=veri(training(pt),:);>>hdTest=veri(test(pt),:);>>svmModel=fitcsvm(hd Train,"yerlesme"); >> tahmin=predict(svmModel,hdTest) ve > hata = loss(svmModel,hdTest); >> disp("Test Hatası:" +hat*a) komutları sonrasında "Test Hatası:0.40" olarak bulunmuştur.

Verinin normalliği dikkate alınarak Kernel ve Gaussian fonksiyonu model için işletildiğinde de*>>svmModelGauss=fitcsvm(hdTrain,"yerlesme","KernelFunction","gaussian")>>hataGauss=loss(svmM odelGauss,hdTest); >> disp("Test Hatası: " + hataGauss)* "Test Hatası: 0.40 " olarak bulunmuştur. MATLAB Support Vector Machine komutunun yerleştirme puan ve sıralaması için genel anlamda (0.60) oranında doğru sonuçlar ürettiği görülmektedir. MATLAB programının makine öğrenmesinde iyi sonuçlar verdiği görülmekte olup diğer eğitim çıktıları için de kullanımı önerilmektedir.

Anahtar Kelimeler: *MATLAB programı, destek vektör makinesi, sınıflama, tahmin.*

[1] fazilet.tasdemir@erdogan.edu.tr:, ORCID: 000-0002-0430-9094, Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye

Investigation of Classification Accuracy with Support Vector Machine in MATLAB Program

Fazilet TAŞDEMİR[1]

Statistical programs and data analyses are used to classify data. MATLAB (Matrix laboratory) software program is a program that can also be used for classification purposes that gives distances between elements and units and offers ease of calculation. MATLAB is mostly preferred by researchers in the fields of numerical linear algebra calculations, fuzzy systems, data analysis and visualization, coding, engineering, machine learning, deep learning, simulation and software.

In this research, it is aimed to examine the classification accuracy with Support Vector Machine (SVM) in MATLAB program. In line with this purpose, answers to the following questions were sought:

1. What is the error rate of the model determined by Support Vector Machine? 2. What is the error rate of the model based on Kernel Function and Gaussian parameters? In order to answer the research questions and to use the MATLAB program, certain numerical values are needed. For this reason, simulative data was generated for the study group of the research. Simulation allows the user to have extensive control.

In the study group, the quotas, success rankings, success scores and placed candidates of the medical faculties of the state universities included in the 2023 Higher Education Council Preference list, which is publicly available online, were formed by taking into consideration. Of these data, 99 were taken exactly from the medical faculty preference list, and 101 data were generated simulatively in line with these data. Data analysis was conducted on a simulative data of 200 candidates. After the*>>pt=cvpartition(data.settlement,"HoldOut",0.2);>>hdTrain=datatraining(pt),:);>>hdTest=data(tes t(pt),:);>>svmModel=fitcsvm(hdTrain," settlement "); >> tahmin=predict(svmModel,hdTest) and > error= loss(svmModel,hdTest); >> disp("Test Error:" +error*) commands used in data analysis, "Test Error: 0.40" was found.

Considering the normality of the data, *>> svmModelGauss = fitcsvm (hdTrain, "settlement","KernelFunction","gaussian")>>hataGauss=loss(svmModelGauss,hdTest); >>disp("Test Error:"+errorGa uss)* the "Test Error: 0.40" was found when the Kernel and Gaussian function was run for the model. It is seen that the MATLAB Support Vector Machine command produces (0.60) correct results for placement scores and rankings in general. The MATLAB program seems to give good results in machine learning and its use is recommended for other training outputs.

Keywords: *MATLAB program, support vector machine, classification, prediction.*

[1] fazilet.tasdemir@erdogan.edu.tr:, ORCID: 000-0002-0430-9094, Recep Tayyip Erdogan University, Rize, Turkey

Uluslararası Okuma Becerilerinde Gelişim Araştırması (PIRLS) 2021 Türkiye Verisinin Bilişsel Tanı Modellemesi İle Analizi

Nihal Burçin BİRCAN [1], Sedat ŞEN [2]

Bu araştırma, 2021 yılında yapılmış olan Uluslararası Okuma Becerilerinde Gelişim Araştırması'nda (PIRLS) Türkiye'nin verilerini bilişsel tanı modellemesi ile analiz etmeyi amaçlamaktadır. Türkiye PIRLS sınavına 2021 yılından önce en son 2001 yılında girmiştir. Yirmi yılın ardından uluslararası yapılan bu sınava Türkiye'nin de girmesi büyük bir önem taşımaktadır. Sınavdan elde edilen verilerin bilişsel tanı modelleri (BTM) ile analiz edilmesi sayesinde, sınava girenlerin okuma becerisi ile ilgili güçlü ve zayıf yönleri görülebilir ve daha detaylı geri bildirim alınabilir. Ayrıca BTM ile eğitim-öğretim sürecinde öğrencilerin okuma becerisini daha iyi geliştirmek amacıyla izlenmesi gereken yol hakkında bilgi alınabilir. Bu çalışmada analiz etmek için "The Empty Pot" metni seçilmiştir ve Türkiye'den bu metni çözen 667 öğrencinin verileri analize alınmıştır. Öncelikle alan uzmanlarının görüşü de alınarak PIRLS nitelikleri arasında yer alan "Fikirleri/bilgileri yorumlama ve bütünleştirme" niteliği "Yorumlama" ve "Kişisel bilgi ve deneyimleri kullanma" olmak üzere birbirinden ayrılmıştır. Nihai Q-matriste "Çıkarım yapma", "Yorumlama", "Kişisel bilgi ve deneyimleri kullanma" ve "İnceleme ve Değerlendirme" nitelikleri bulunmaktadır. Çalışmada BTM'ler arasından G-DINA, DINA, DINO, A-CDM, LLM ve R-RUM modelleri karşılaştırılmış ve G-DINA modelinin veri ile daha uyumlu olduğu belirlenmiştir. Aynı zamanda Q-matriste oluşabilecek yanlış belirlemeleri düzeltmek amacıyla Q-matris doğrulama adımı uygulanmıştır. Elde edilen bulgular 39 öğrencinin (%5,85) tüm niteliklerde uzman (1111) olarak sınıflandırıldığını göstermektedir. Bir başka dikkat çeken sonuç ise, sınava giren 69 öğrenci hiçbir nitelikte (0000) uzmanlaşamamıştır ve bu da öğrencilerin yaklaşık %10'unun ölçülen hiçbir niteliğe sahip olmadığını göstermektedir. Tüm analiz sonuçlarına göre, dört nitelik arasında çıkarım yapma uzmanlaşması en kolay nitelik iken, kişisel bilgi ve deneyimleri kullanma ise en zor nitelik olarak görülmektedir.

Anahtar Kelimeler: *Okuma Becerisi, Uluslararası Sınavlar, PIRLS, Türkiye, Bilişsel Tanı Modellemesi.*

[1] nhlbrcn@gmail.com, Milli Eğitim Bakanlığı, Şanlıurfa, Türkiye

[2] sedatsen06@gmail.com, ORCID: 0000-0001-6962-4960, Harran Üniversitesi, Şanlıurfa, Türkiye

Analysis Of Progress in International Reading Literacy Study (PIRLS) 2021 Türkiye Data Using Cognitive Diagnostic Modeling

Nihal Burçin BİRCAN [1], Sedat ŞEN[2]

This research aims to analyze Türkiye's data from the International Readıng Literacy Study (PIRLS) conducted in 2021 with cognitive diagnostic modeling. Turkiye last took the PIRLS exam in 2001 before 2021. It is of great importance for Turkiye to enter this international exam after twenty years. By analyzing the data obtained from the exam with cognitive diagnostic models (CDM), the strengths and weaknesses of the test takers regarding reading skills can be seen and more detailed feedback can be received. In addition, with CDM, information can be obtained about the path to be followed in order to better improve students' reading skills during the education process. In this study, the text "The Empty Pot" was chosen to analyze and the data of 667 students from Türkiye who solved this text were analyzed. First of all, after taking the opinions of field experts, the quality of "Interpreting and integrating ideas/information" among the PIRLS qualifications was divided into "Interpreting" and "Using personal knowledge and experiences". The final Q-matrix includes the attributes of "Inferencing", "Interpreting", "Using personal knowledge and experiences" and "Examining and Evaluation". In the study, G-DINA, DINA, DINO, A-CDM, LLM and R-RUM models were compared among the CDMs and it was determined that the G-DINA model was more compatible with the data. At the same time, a Q-matrix verification step was applied to correct any misdeterminations that may occur in the Q-matrix. The findings show that 39 students (5,85%) were classified as experts in all attributes (1111). Another striking result is that 69 students who took the exam could not master any attribute (0000), indicating that approximately 10% of the students did not have any of the measured attributes. According to all analysis results, among the four attributes, inferencing is the easiest attribute to master, while using personal information and experiences is seen as the most difficult attribute.

Keywords: *Reading Skill, International Exams, PIRLS, Turkiye, Cognitive Diagnostic Modeling*

[1] nhlbrcn@gmail.com, Milli Eğitim Bakanlığı, Şanlıurfa, Türkiye

[2] sedatsen06@gmail.com, ORCID: 0000-0001-6962-4960, Harran Üniversitesi, Şanlıurfa, Türkiye

Türkiye ve Amerika'da Uygulanan Merkezî Sınavların Dil Becerileri Bakımından Karşılaştırılması

Tahir TAĞA [1]

Kademeler arası geçiş, mesleğe kabul gibi çeşitli amaçlarla başvurulan merkezî sınavların Türkiye'de uzun bir geçmişi vardır (TEDMEM, 2023). Zaman içinde etkisi ve kapsamı genişleyen bu sınavlar, eğitim sisteminin baskın bir unsuru hâline gelmiştir. Bazen güncellenen eğitim sistemi ve ders programları bazen de olumsuz etkileri göz önünde bulundurularak bu sınavların içerik ve uygulamaları birçok kez değiştirilmiştir. Bu değişikliklerde başka ülkelerin ve son dönemde artan etkisine bağlı olarak uluslararası sınavların içerik ve uygulamalarının etkisi de gözlemlenmektedir. Son olarak Millî Eğitim Bakanlığının 2023 Eğitim Vizyonu'nda sistemdeki bütün sınavların amacı, içeriği, soru tiplerine bağlı yapısı ve sağlayacağı yarar bakımından yeniden düzenleneceği ifade edilmiş (MEB, 2018) ve süreç içinde merkezî sınavların içerik ve uygulamalarında köklü değişikliklere gidilmiştir. Bu değişiklikler her konu alanında olduğu gibi sınavın dil becerilerini ölçen bölümlerinde de meydana gelmiştir. Öte yandan öğrencilerin dil becerilerine ilişkin yeterliklerinin tespit edilmesi amacıyla Millî Eğitim Bakanlığınca dört temel becerinin ölçülmesine yönelik izleme sınavı da yapılmaya başlanmıştır. Son beş yılda merkezî sınavların dil becerilerini temele alan bölümlerinde meydana gelen değişikliklerin sebepleri, sınav içerikleri, uygulamaları ve bunların etkileri üzerine derinlemesine incelemeler yeterince yapılamamıştır. Bu bakımdan başka ülkelerde dil becerilerinin merkezî sınavlarda nasıl ele alındığını ortaya koymak mevcut durumun anlaşılmasına ve bundan sonraki değişikliklere katkıda bulunacaktır. Bu düşünceyle, bu çalışmada, Türkiye ve Amerika'daki merkezî sınavların dil becerileri bakımından karşılaştırılması amaçlanmıştır. Temel nitel araştırma (Merriam & Tisdell, 2016) olarak tasarlanan bu çalışmanın incelemeye konu olacak verileri dökümün incelemesi yöntemine göre toplanacaktır (Yıldırım & Şimşek, 2008). Bu çerçevede liseye, üniversiteye ve lisansüstü eğitime giriş sınavları incelenecektir. Bu sınavlar, ilgili literatür ve yayımlanmış soru örnekleri temele alınarak amacı, içeriği, soru tipleri vb. yönlerden betimlenecek ve karşılaştırmalı olarak ele alınacaktır. Bulgulardan hareketle de dil becerilerinin merkezî sınavlarda ölçülmesi ile ilgili hususlara dair teklifler sunulacaktır.

Anahtar Kelimeler: *Merkezî sınavlar, dil becerileri, karşılaştırmalı eğitim.*

Kaynakça

MEB. (2018). *2023 eğitim vizyonu*. Millî Eğitim Bakanlığı.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation.* Jossey-Bass.

TEDMEM. (2023). *Sınavların gölgesinde eğitim sistemi*. Türk Eğitim Derneği.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin.

[1] tahirtaga@sdu.edu.tr, ORCID: 0000-0002-6797-6556, Süleyman Demirel Üniversitesi, Isparta, Türkiye

Comparison of High-Stakes Test in Turkey and the United States in terms of Language Skills

Tahir TAĞA [1]

High-stakes tests, which are used for various purposes such as transition between levels and admission to professions, have a long history in Turkey (TEDMEM, 2023). These tests, which have expanded in scope and impact over time, have become a dominant element of the education system. The content and implementation of these tests have been changed many times, sometimes due to the updated education system and curricula and sometimes due to their negative effects. In these changes, the influence of the content and implementations of other countries and international exams that their increasing influence in the recent period are also observed. Finally, in the 2023 Education Vision of the Ministry of National Education, it was stated that all tests in the system will be restructured in terms of their purpose, content, and structure depending on question types and the benefits they will provide (MoNE, 2018). In the process, radical changes were made in the content and implementation of high- stakes tests. These changes occurred in every subject area as well as in the domain of the test that measures language skills. On the other hand, the Ministry of National Education started to conduct a monitoring test to measure four basic skills to determine students' proficiency in language skills. Over the past five years, in-depth analyses of the reasons for the changes in the language skills-based domain of high-stakes tests, test contents, implementations, and their effects have not been conducted sufficiently. In this respect, identifying how language skills are handled in high-stakes tests in other countries will contribute to the understanding of the current situation and future changes. With this in mind, this study aims to compare the high-stakes tests in Turkey and the USA in terms of language skills. The data to be analyzed in this study, which is designed as basic qualitative research (Merriam & Tisdell, 2016), will be collected according to the document analysis method (Yıldırım & Şimşek, 2008). In this framework, entrance tests for high school, university, and postgraduate education will be examined. These tests will be described and analyzed comparatively in terms of their purpose, content, question types, etc. based on the relevant literature and released items. Based on the findings, proposals regarding the measurement of language skills in high-stakes tests will be presented.

Keywords: *High-stakes tests, language skills, comparative education*

Reference

MEB. (2018). *2023 eğitim vizyonu*. Millî Eğitim Bakanlığı.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation.* Jossey-Bass.

TEDMEM. (2023). *Sınavların gölgesinde eğitim sistemi*. Türk Eğitim Derneği.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin.

[1] tahirtaga@sdu.edu.tr, ORCID: 0000-0002-6797-6556, Suleyman Demirel University, Isparta, Türkiye

PISA2018 Okuma Okuryazarlığı Temel Testinin Ölçme Değişmezliğinin İncelenmesi

Tuba Yazıkan [1], Kübra Atalay Kabasakal [2]

Uluslararası eğitim araştırmalarından biri olan PISA'nın sonuçları farklı alt gruplarda karşılaştırmalar yapmak üzere kullanılmaktadır. Yapılan karşılaştırmaların anlamlı olabilmesi için uygulanan ölçme aracının ilgili alt gruplarda aynı yapıyı ölçtüğünün kanıtlanması gerekir. Ölçme aracının farklı gruplarda aynı yapıyı ölçtüğü çoklu grup doğrulayıcı faktör analizi (ÇGDFA) ile incelenmektedir (Tabacknick & Fidel, 2013). Bu araştırmanın amacı, PISA 2018 uygulamasında okuma becerilerini yordayan ve PISA için önemli olan güvenilirliği değerlendirme, özetleme, *sosyoekonomik statü* değişkenlerine göre oluşturulan alt gruplarda ölçme değişmezliğinin incelenmesidir. Çalışmanın örneklemini PISA 2018 uygulamasına Türkiye'den katılan 6890 öğrencinin öğrenci anketine ve okuma becerileri temel testindeki beş üniteye verdikleri yanıtlar oluşturmaktadır. Önce ÇGDFA yapılmış ve ölçme değişmezliği aşama aşama WLS kestirim yöntemi ile test edilmiştir. Muthen ve Muthen'e (1998-2019) göre 1-0 şeklindeki ikili kategorik verilerde parametre kestirimi için WLS yöntemi kullanıldığından iki model arasındaki farkı değerlendirmek için doğrudan ki-kare farkı değil Satorra-Bentler ölçeklenmiş ki-kare farkı hesaplanmalıdır. Bu çalışmada her bir aşamada elde edilen ölçeklenmiş uyum katsayılarının değerlendirilmesinin yanı sıra aşamaların karşılaştırılması ölçeklenmiş CFI ve ölçeklenmiş TLI uyum indeksleri farkı kullanılarak yapılmıştır. Araştırmanın beş ünite ve üç değişken ile elde edilen 15 analiz sonucundan dört tanesinin yapısal değişmezlik, 10 tanesinin metrik değişmezlik aşamasında kaldığı, bir ünitede model yakınsamadığı için uyum katsayılarının elde edilemediği sonucuna ulaşılmıştır. Bu araştırma sonuçları hiç bir ünitede hiç bir değişkene göre katı değişmezliğin sağlanamadığını ortaya koymaktadır. Bu durum alt gruplardaki öğrencilerin okuma okuryazarlıklarının karşılaştırılmasının uygun olmadığı anlamına gelmektedir.

Anahtar Kelimeler: ölçme *değişmezliği, sosyoekonomik statü,* çok *gruplu doğrulayıcı faktör analizi*

Kaynakça

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

[1] tubayazikan@gmail.com, ORCID: 0009-0000-2397-9440, MEB, Kütayha, Türkiye

[2] katalay@hacettepe.edu.tr, ORCID: 0000-0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

Examining the Measurement Invariance of the PISA2018 Reading Literacy Core Test

Tuba Yazıkan [1], Kübra Atalay Kabasakal[2],

The results of PISA, one of the international education researches, are used to make comparisons in different subgroups. In order for the comparisons to be meaningful, it must be proven that the applied measurement tool measures the same structure in the relevant subgroups. (MGCFA) examines whether the measurement tool measures the same construct in different groups. The purpose of this research is to examine the measurement invariance in the subgroups created for three variables (reliability assessment, summarization, economic-social-cultural situation) that explain the success of reading skills and are important for PISA in the PISA 2018 cycle. The sample of the study consists of the responses of 6890 students from Turkey who participated in the PISA 2018 cycle to the student survey and 5 units in the reading skills basic test. MGCFA was performed and measurement invariance was tested step by step using the WLS estimation method. According to Muthen and Muthen (1998-2019), since the WLS method is used for parameter estimation in 1-0 binary categorical data, the Satorra- Bentler scaled chi-square difference should be calculated to evaluate the difference between the two models, not the direct chi-square difference. In addition to evaluating the scaled fit coefficients obtained at each stage, the comparison of the stages was made using the difference in scaled CFI and scaled TLI fit indices. It was concluded that out of 15 analysis results obtained with 5 units and 3 variables, 4 of them remained at the structural invariance stage, 10 of them remained at the metric invariance stage, and the fit coefficients could not be obtained because the model did not converge in 1 unit. The results of this study reveal that strict invariance could not be achieved in any unit according to any variable. This means that it is not appropriate to compare the reading literacy of students in subgroups.

Keywords: *measurement invariance, socioeconomic status, multigroup confirmatory factor analysis*

Reference

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

[1] tubayazikan@gmail.com, ORCID: 0009-0000-2397-9440, MEB, Kütayha, Türkiye

[2] katalay@hacettepe.edu.tr, ORCID: 0000-0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

Sürdürülebilir Kapsayıcı Başarı: Öğretmenlerle İlgili Faktörlerin Rolü*

Emine Yavuz[1], Elizabeth A. Sanders[2]

Son yıllarda gerçekleşen göçlerle birlikte Türkiye'deki sınıflarda farklı etnik kökenden gelen öğrencilerin birlikte eğitim alma ihtimalleri artmış, öğrenme ortamları hızla çok kültürlü bir yapıya bürünmüştür. Çok kültürlü sınıf ortamlarında eşit ve kapsayıcı bir eğitim için öğretmenlerin rolü tartışılmayacak kadar büyüktür. Bu nedenle öğretmen-öğrenci ilişkisi, öğretmen tutumu, öğretmen beklentisi, öğretmen inancı gibi öğretmen faktörlerinin öğrenci başarısı üzerindeki etkileri sıklıkla araştırmalara konu olmaktadır ve uygulayıcılara gerekli öneriler sunulmaktadır. Fakat öğretmen faktörleri ne kadar nitelikli hale getirilse de eğer öğrencilerin, öğretmenlerinin davranışlarına ilişkin algıları değişmiyorsa bu iyileştirmenin etkisi tam olarak gözlemlenemeyecektir. İlgili alanyazın incelendiğinde öğretmen davranışları hakkındaki öğrenci algılarının, öğrencilerin başarılarına olan etkileri hakkında herhangi bir çalışmaya ulaşılamamıştır. Bu düşünceden hareketle bu araştırmanın amacı, öğrenci ve okul yöneticilerinin görüşlerine göre öğretmenlerle ilgili faktörlerinin öğrencilerin PISA 2018 matematik başarısına etkilerini 2 düzeyli model (MLM) ile belirlemektir. Korelasyonel çalışma olarak tasarlanan bu çalışmanın evreni, Türkiye'de PISA 2018 matematik uygulamasına katılan tüm öğrenciler ve okul yöneticilerinden elde edilen verilerdir. Araştırma örneklemini ise 5968 öğrenci ve 169 okul yöneticisinden elde edilen veriler oluşturmaktadır. Tüm olası değerlerin bağımlı değişken olarak ele alınıp öğrenci düzeyinde bir, okul düzeyinde 19 bağımsız değişkenin temel ve etkileşim etkileri örneklem ağırlığı dikkate alınarak R programının lme4 paketi ile test edilmiştir. Analiz sonucunda öğrenci düzeyinde, öğretmenlerin kültürel farklılıklara karşı davranışları; okul düzeyinde, öğretmenlerin kültürel farklılıklara karşı davranışları, öğretmenlerin başka ülkelerden gelen öğretmenlerle görüşmeleri, öğretmen koçluğu, okuldaki öğretim elemanının eksikliği, öğretmenlerin teknoloji kullanım kalitesi ana etkileri istatistiksel olarak anlamlı bulunmuştur. Ek olarak, öğrenci ve okul düzeyindeki öğretmenlerin kültürel farklılıklara karşı davranışları değişkenlerinin etkileşimi istatistiksel olarak anlamlıdır. Araştırma sonuçları öğretmenlerin bireysel ve profesyonel gelişimlerinin öğrenci başarısına etkisini inceleyen literatürün, öğrencilerin bu faktörleri algılayışlarına da odaklanması gerekliliğini ortaya koymaktadır. Ayrıca, okullarda öğrenci başarısının genel olarak artması için öğretmenlerin sınıf ortamlarında eşitlikçi davranışlara sahip olması ve okul yöneticilerinin okul genelinde öğretmenlerin birbirleriyle ve farklı kültürden öğretmenlerle etkileşim halinde olmaları için onlara olanak tanımaları önerilmektedir.

Anahtar Kelimeler: *Multilevel Model, PISA,* öğretmen *faktörleri*

---

[1] emineyavuz@erciyes.edu.tr, ORCID: 0000-0002-1991-1416, Erciyes Üniversitesi, Kayseri, Türkiye

[2] lizz@uw.edu, ORCID: 0000-0002-1008-8435, Universtiy of Washington, Seattle, USA

*Bu çalışma TÜBİTAK 2219A Projesi kapsamında kullanılan verilerden üretilmiştir.

Sustainable Inclusive Success: The Role of Teacher-Related Factors*

Emine Yavuz[1], Elizabeth A. Sanders[2]

With the migrations that have taken place in recent years, the possibility of students from different ethnic backgrounds to study together in classrooms in Turkey has increased, and learning environments have rapidly taken on a multicultural structure. The role of teachers in providing equal and inclusive education in multicultural classroom environments is indisputably great. For this reason, the effects of teacher factors such as teacher-student relationship, teacher attitude, teacher expectation, and teacher belief on student success are frequently the subject of research, and necessary suggestions are presented to practitioners. However, no matter how qualified the teacher factors are, if students' perceptions of their teachers' behaviors do not change, the effect of this improvement will not be fully observed. No study was found on the effects of student perceptions of teacher behaviors on students' achievement. Based on this idea, this research aims to determine the effects of teacher-related factors on students' PISA 2018 mathematics success according to the views of students and school administrators using a two-level model (MLM). Designed as a correlational study, the universe of this study is the data obtained from all students and school administrators who participated in the PISA 2018 mathematics application in Turkey. The research sample consists of data obtained from 5968 students and 169 school administrators. All possible values were considered as dependent variables and the main and interaction effects of one independent variable at the student level and 19 independent variables at the school level were tested with the lme4 package of the R program by taking into account the sample weight. As a result of the analysis, the main effects of teachers' behaviors towards cultural differences at the student level; teachers' behaviors towards cultural differences, teachers' meetings with teachers from other countries, teacher mentoring, teaching staff absenteeism at the school, and teachers' technology use quality were found to be statistically significant at the school level. In addition, the interaction of the variables of teachers' behaviors towards cultural differences at the student and school levels was statistically significant. The research results reveal that the literature examining the effects of teachers' individual and professional development on student achievement should also focus on students' perceptions of these factors. In addition, it is recommended that teachers should have egalitarian behaviors in classroom environments and that school administrators provide opportunities for teachers to interact with each other and with teachers from different cultures throughout the school in order to increase student achievement in schools in general.

Keywords: *Multilevel Model, PISA, Teacher factors.*

[1] emineyavuz@erciyes.edu.tr, ORCID: 0000-0002-1991-1416, Erciyes Üniversitesi, Kayseri, Türkiye

[2] lizz@uw.edu, ORCID: 0000-0002-1008-8435, Universtiy of Washington, Seattle, USA

*Bu çalışma TÜBİTAK 2219A Projesi kapsamında kullanılan verilerden üretilmiştir.

Özel Amaçlı İngilizce Öğretmenlerinin Mesleki Bilgilerinin Ölçülmesi

Mert TOPKARAOĞLU[1], Hossein FARHADY[2]

Öğretmenlerin mesleki bilgisi, başarılı bir öğretim için temel bir ön koşuldur. Bir öğretmenin mesleki bilgisinin yetersiz olması durumunda, etkili yönlendirme yapabilme becerileri de kaçınılmaz bir şekilde etkilenir ki bu da öğretmeni yetersiz kılarak farklı alternatif yollara yönlendirir. Mesleki bilgiye sahip olmayan öğretmenler genellikle sınıf ortamlarında da kendilerine güvenmezler. Bu türden öğretmenler genel olarak tutarlı bir pedagoji geliştirmekte, öğrencilerin ilerleme ve başarıları hakkında bilinçli kararlar vermekte güçlük yaşarlar ve ders kitabının içeriğine takılıp kalırlar. Dolayısıyla, öğretim sırasında meydana gelen aksiliklerle başa çıkmakta sorun yaşarlar. Alandaki son gelişmeler ve akademik toplulukların araştırmaları, öğretmenlerin mesleki bilgilerinin daha ayrıntılı ve kapsamlı bir modele kavuşturulmasına yol açmıştır. Ancak, birçok çerçeve çoğu mesleki bilginin bileşenlerini bir şekilde ortaya koysa da farklı öğretim ortamlarında kullanılabilecek çok boyutlu bir model sunmamaktadırlar. Buna ek olarak, ortaya koydukları kavramlar benzerdir ve öğrencilerin genişleyen ihtiyaçlarını içermemektedir. Bununla birlikte, yalnızca birkaç çalışma, farklı konu alanları için gerçekten tanımlanabilir bir mesleki bilgi bütünü olup olmadığını test etmiştir. Bu bağlamda, genel olarak Özel Amaçlı İngilizce ve özellikle de Denizcilik İngilizcesi eksik bir paradigmadır. Bu sorunu ele almak için amacıyla çalışma, Yabancı Dil Olarak İngilizce ortamlarında özel amaçlar için İngilizce öğretimine özgü özel amaçlı mesleki bilginin gösterilmesi ve değerlendirilmesi teorisi üzerine inşa edilmiştir. Çalışmanın bir diğer amacı, İngilizce öğretmenlerinin özel amaçlı mesleki bilgilerini ölçmek için geçerli, güvenilir ve veriye dayalı bir mesleki bilgi testi geliştirmektir. Bu amaçla doğrultusunda araştırmacı ilk olarak, Denizcilik Fakültelerindeki İngilizce öğretmenleri, Denizcilik İngilizcesi için eğitim almış İngilizce öğretmenleri, konu uzmanları, profesyoneller, öğrenciler, koordinatörler ve yabancı dil programlarının yöneticilerinden oluşan 423 katılımcıya ulaşarak, katılımcıların Denizcilik İngilizcesi öğretimine özgü belirli bir bilgi türüne olan ihtiyaçlarını ortaya koymak kesitsel bir anket çalışması yapmıştır. İkinci olarak, uygun içerik kapsamını sağlamak amacıyla belirlenen ihtiyaçları çapraz kontrol etmek için takip gözlem kontrol listeleri doldurulmuş, saha notları kaydedilmiş ve yarı yapılandırılmış görüşmeler gerçekleştirilmiştir. Üçüncü olarak, öğretmenlerin ihtiyaç değerlendirmesinden, gözlem kontrol listelerinden, saha notlarından ve görüşmelerden elde edilen veriler, özel amaçlı bilgi testinin özellikler tablosunu hazırlamak için kullanılmıştır. Ayrıca, testin pilot uygulaması yapılmış, uzmanlar tarafından gözden geçirilmiş ve gerekli düzeltmeler yapılmıştır. Son hali verilen test uygulanmış ve psikometrik özellikleri belirlenmiştir. Genel olarak, sınırlılıklarına rağmen, çalışmanın sonuçları Yabancı Dil Olarak İngilizcenin öğretildiği bağlamlarda Özel Amaçlı İngilizce öğretimine özgü belirli bir bilgi türü olabileceğini göstermektedir ve yeni geliştirilen Denizcilik İngilizcesi Mesleki Bilgi Testi aracılığıyla Özel Amaçlı İngilizce öğretmenlerinin sınıf ortamlarında belirli amaçlar için bu gerekli bilgiyi kapsamlı bir şekilde kullandıkları dikkatle gözlemlenmiştir.

Anahtar Kelimeler: *Mesleki bilgi,* özel *amaçlı* İngilizce*, denizcilik* İngilizcesi

[1] mtopkaraoglu@hotmail.com:, ORCID: 0000-0002-2134-0464, Milli Savunma Üniversitesi, İstanbul, Türkiye

[2] farhadyh@gmail.com:, ORCID: 0000-0002-2019-6157, Yeditepe Üniversitesi, İstanbul, Türkiye

Assessing Professional Knowledge of English For Specific Purposes (ESP) Teachers

Mert TOPKARAOĞLU [1], Hossein FARHADY[2]

Professional knowledge of teachers is an essential precondition for successful teaching. It is fundamental to teaching in that when a teacher has insufficient professional knowledge, his/her ability to give effective instructions inevitably gets affected, which leaves him/her incapable and leads to unproductive alternative routes. Teachers, who do not have professional knowledge, generally do not have confidence in their classroom settings either. They, by and large, are not able to develop a coherent pedagogy, cannot make informed decisions about the progress and achievement of students, get stuck with the contents of the coursebook, and thus are unable to cope with the mishaps during teaching. The latest developments and research of the academic associations have led to a more detailed and comprehensive model of teachers' professional knowledge at length. However, though most of the frameworks demonstrate the components of professional knowledge in one way or another, they do not seem to propose a multidimensional model that could be utilized in different settings. In addition, the concepts they demonstrate are similar and fail to include the expanding needs of learners. Nonetheless, only a few studies have tested whether there is indeed a distinct body of identifiable professional knowledge for different subject areas. In this respect, English for Specific Purposes (ESP) in general, and Maritime English, in particular, is a missing paradigm. To address this issue, this study builds on the theory of the demonstration and the assessment of specific professional knowledge (SPK) sui generis to the teaching of English for Specific Purposes (ESP) in English as a Foreign Language (EFL) settings. Another aim of the study is to develop a valid, reliable, and data-driven professional knowledge test to measure EFL teachers' SPK. For this purpose, the researcher first investigated 423 stakeholders including English language teachers, English-trained maritime English (ESP) teachers, subject specialists, professionals, students, coordinators, and administrators of ESP programs at Maritime Faculties in terms of their needs for a specific type of knowledge unique to the teaching of Maritime English through a cross-sectional survey. Secondly, follow- up observation checklists were filled out, field notes were recorded, and semi-structured interviews were carried out to cross-check the identified needs to ensure appropriate content coverage. Thirdly, the data obtained from the needs assessment of teachers, observation checklists, field notes, and interviews were used to prepare the table of specifications of the test of SPK. Furthermore, the test was piloted, and reviewed by experts, and the necessary revisions were made. The finalized test was administered, and the psychometric properties were determined. Overall, despite its limitations, the results of the study indicate that there might be a specific type of knowledge unique to the teaching of ESP in EFL contexts and it is observed carefully through the newly developed novel Professional Knowledge of Maritime English Test that ESP teachers are making extensive use of that necessary knowledge for specific purposes in their ESP settings.

Keywords: *Professional knowledge, English for specific purposes, maritime English*

[1] mtopkaraoglu@hotmail.com:, ORCID: 0000-0002-2134-0464, Milli Savunma Üniversitesi, İstanbul, Türkiye

[2] farhadyh@gmail.com:, ORCID: 0000-0002-2019-6157, Yeditepe Üniversitesi, İstanbul, Türkiye

Dil Becerilerinin Ölçülmesi Alanında Yapılan Çalışmaların Konu Modellemesi ile İncelenmesi

Kübra Atalay Kabasakal [1], Rabia Akcan[2], Duygu Koçak[3]

Bu araştırmada dil becerilerinin ölçülmesinde son yıllarda yapılan çalışmaların hangi konulara odaklandığını belirlemek amacıyla birinci, ikinci ve yabancı dil değerlendirmesinde teori, araştırma ve uygulamanın ilerlemesine odaklanan "Language Assessment Quarterly" dergisinin 2020-2024 yılları arasında yayımladığı makaleler konu modellemesi yöntemiyle incelenmiştir. Konu modelleme, metin belgelerinde altta yatan gizil konu veya temaları belirlemeye yönelik denetimsiz makine öğrenme algoritması tekniğidir. Konu modellemesinin çeşitli varyantları mevcut olsa da, orijinal konu modellerinden biri olan Gizil Dirichlet Tahsisi (GDT, Latent Dirichlet Allocation – LDA) ilk olarak Blei ve arkadaşları tarafından 2003 yılında literatüre tanıtılmıştır (Blei vd., 2003). Bu çalışmada, metin belgesi olarak 93 çalışmanın özeti kullanılmıştır. Konu modellemesinin ilk aşaması metin verisinde veri ön işleme yapmaktır. Bu aşamadaki adımlar sadece konu modellemesi yapmadan önce temiz ve uygun şekilde biçimlendirilmiş bir veri kümesi geliştirmek için değil, aynı zamanda konu modellemesi çıktısının yorumlanabilirliğini kolaylaştırmak için de gereklidir. Ön işleme sürecinde, modelleme analizinin kalitesini etkileyebilecek tüm noktalama işaretleri ve sayıların belgelerden çıkarılması, tüm kelimelerin küçük harfe dönüştürülmesi ve araştırma bağlamında konuların belirlenmesine değer katmayacak kadar sık kullanılan etkisiz kelimelerin (stopwords) değerlendirilmesi gerekir (Gencoglu vd., 2023). Bu bağlamda ön işleme keşifsel, yinelemeli bir aşamadır ve analizin ilerleyen aşamalarında bile bu aşamaya tekrar geri dönmek gerekebilir. Bu analizler sonucunda belgelerde en yüksek konu ağırlıklarının genç/azınlık dil öğrencilerinde dil değerlendirmesi ve okuryazarlığı düşük dil öğrenenlerde dil değerlendirmesi olduğu görülmüştür. Bu iki başlığın yanı sıra dil değerlendirmelerinde geçerlik, teknoloji destekli dil değerlendirmeleri, dil testlerinde dinamik değerlendirme, Ortak Avrupa Dil Referans Çerçevesine dayalı çalışmalar, yapay zekâ ve insan değerlendiricilerle puanlama, etkileşimsel yeterliliğin değerlendirilmesinde sözlü diyalog sistemleri üzerine yapılan çalışmaların konu modellemesinde ayrı başlıklar altında toplandığı görülmektedir.

Anahtar Kelimeler: *Dil becerilerin* ölçülmesi*, Konu modellemesi, Gizil dirichlet tahsisi*

Kaynakça

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.

Gencoglu, B.,Helms-Lorenz, M., Maulana, R., Jansen, E. P., &Gencoglu, O. (2023). Machine and expert judgments of student perceptions of teaching behavior in secondary education: Added value of topic modeling with big data. *Computers&Education*, *193*, 104682. https://doi.org/10.1016/j.compedu.2022.104682.

[1] katalay@hacettepe.edu.tr, ORCID: 0000-0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

[2] rabiselt@gmail.com, 0000-0003-3025-774X, Milli Eğitim Bakanlığı, Afyonkarahisar, Ülke

[3] duygu.kocak@alanya.edu.tr, ORCID: 0000-0003-3211-0426, Alanya Alaaddin Keykubat Üniversitesi, Antalya, Türkiye

Investigation of Research on Measurement of Language Skills with Topic Modeling

Kübra Atalay Kabasakal [1], Rabia Akcan[2], Duygu Koçak[3]

In this research, articles published between 2020 and 2024 in "Language Assessment Quarterly" journal, which focuses on the advancement of theory, research and practice in first, second and foreign language assessment, have been examined using topic modeling to determine which topics the studies on language skill measurement have focused in recent years. Topic modeling is an unsupervised machine learning algorithm technique for identifying hidden topics or themes in text documents. Although several variants of topic modeling exist, one of the original topic models, Latent Dirichlet Allocation (LDA), was first introduced to the literature in 2003 by Blei et al. (2003). In this study, abstracts of 93 studies were used as the text corpus. The first stage of topic modeling is preprocessing of the text data. Steps in this stage are necessary not only to develop a clean and appropriately formatted dataset before conducting topic modeling, but also to facilitate the interpretability of the topic modeling output. Preprocessing involves removing all punctuation and numbers from the documents that may affect the quality of the modeling analysis, converting all words to lowercase, and assessing stop-words that are used too frequently to add value to the identification of issues in the research context (Gencoglu et al., 2023). In this context, preprocessing is an exploratory, iterative stage and it may be necessary to return to it even in later stages. As a result of the analyses, the highest topic weights were found in language assessment for young/young minority language learners and language assessment for low-literate language learners. In addition, studies on validity in language assessments, technology-enhanced language assessments, dynamic assessment in language testing, the Common European Framework of Reference for Languages, scoring with artificial intelligence and human raters, and spoken dialogue systems in the assessment of interactional competence are grouped under separate headings in the topic modeling.

Keywords: *Measurement of language skills, Topic modeling, Latent dirichlet allocation*

Reference

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.

Gencoglu, B.,Helms-Lorenz, M., Maulana, R., Jansen, E. P., &Gencoglu, O. (2023). Machine and expert judgments of student perceptions of teaching behavior in secondary education: Added value of topic modeling with big data. *Computers&Education*, *193*, 104682. https://doi.org/10.1016/j.compedu.2022.104682.

[1] katalay@hacettepe.edu.tr, ORCID: 0000-0002-3580-5568, Hacettepe Üniversitesi, Ankara, Türkiye

[2] rabiselt@gmail.com, 0000-0003-3025-774X, Milli Eğitim Bakanlığı, Afyonkarahisar, Ülke

[3] duygu.kocak@alanya.edu.tr, ORCID: 0000-0003-3211-0426, Alanya Alaaddin Keykubat Üniversitesi, Antalya, Türkiye

Akıllı Teknolojiler, Dijital Araçlar ve YZ Uygulamaları Üzerinden Konuşma Becerisinin Değerlendirilmesi

Dinçay KÖKSAL [1], Ayşe DAĞ PESTİL [2]

Bu çalışmada, dinleme-konuşma becerisi ölçme değerlendirmesinde kullanılabilecek akıllı teknolojiler, yapay zeka ve dijital araçlarla desteklenen yöntem ve teknikler üzerinde durulmuştur. Konuşma yeterliliğini değerlendirmek için dijital ve yüz yüze yöntemleri karşılaştıran çalışmalar, ücretsiz ve açık kaynaklı yazılımlara yönelik olumlu öğrenci ve eğitmen algılarını ortaya koymaktadır (Swanson, 2015). Bu araştırma dil becerileri öğretimi çerçevesinde, öğrencilerin hedef dildeki metni ve sözcükleri seslendirme, dildeki ritim, tonlama ve vurgu hatalarına çözüm bulmaya odaklanmıştır. Çünkü telaffuzu yapılamayan bir sözcük duyulduğunda da anlaşılamamaktadır. Araştırmanın yöntemi, öğrenci merkezli özdeğerlendirme ve geri bildirim özelliklerini temel alarak, dijital araçlar ve akıllı teknolojiler kullanılarak şekillendirilmiş bir uygulama sürecine dayanmaktadır. Araştırma, bir devlet üniversitesinde yabancı dil öğrenen yaklaşık 200 öğrenciye yönelik olarak gerçekleştirilmiştir. Öğrenciler, seslendirme, ritim, tonlama ve vurgu becerilerini geliştirmeleri için belirli ses dosyalarını yoğun bir şekilde dinleyerek kendi ses kayıtlarını oluşturmuşlar, bu ses kayıtlarını, Teams üzerinden tanımlanan ödev uygulaması aracılığıyla sisteme yüklemişler, ayrıca, bu sürece dair geri bildirimlerini, Teams'in etkileşimli reflection özelliğini kullanarak vermişlerdir. Elde edilen *reflection* verileri, sıfat-tanım ve renklerle kodlanmış emojiler yoluyla gruplandırılmış, kodlanmış ve başarı değerlendirmeleri kıyaslanarak analiz edilmiştir. Bu çalışmada akıllı telefonlar, bilgisayarlar, internet ve dijital platformlar etkin bir şekilde kullanılarak, öğrencilerin dinleme-konuşma becerisinde yaşadıkları zorlukların üstesinden gelmeleri amaçlanmış ve akıllı teknolojilerin, yapay zeka ve dijital platformların öğrencilerin konuşma ve telaffuz becerilerini geliştirmesinde önemli bir rol oynayabileceği sonucuna varılmıştır Bondarenko ve Gudkova, araştırmalarında görsel-işitsel teknoloji ve internet kaynaklarının entegrasyonunun, konuşma ve dinleme becerilerinin gelişiminde önemli katkı sağladığına dikkat çekmektedir (Bondarenko ve Gudkova, 2023). Bu çalışmanın bulguları, dil becerileri öğretimi ve değerlendirmesinde YZ, akıllı ve dijital teknolojilerinin etkili bir şekilde nasıl kullanılabileceğine dair önemli bilgiler sunmakta ve muhtemel öneriler ortaya koymaktadır.

Anahtar Kelimeler: Akıllı teknolojiler, konuşma becerisi ölçme değerlendirme, dijital platformlar.

Kaynakça

Swanson, P. B. (2015). Measuring language learners' speaking proficiency in a second language using economical digital tools. In Open Source Technology: Concepts, Methodologies, Tools, and Applications (pp. 1504-1523). IGI Global.

Bondarenko, A., & Gudkova, N. (2023). Enhancing foreign language teaching through digital tools and resources. Collection of scientific papers «ΛΟΓΟΣ», (June 23, 2023; Oxford, UK), 129-132.

[1] koksal.dincay@gmail.com:, ORCID: https://orcid.org/0000-0002-8681-4093, Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

[2] aysedag@comu.edu.tr:, ORCID: https://orcid.org/0000-0002-3460-3390, Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

Evaluation of Speaking Skill Through Smart Technologies, Digital Tools and AI Apps

Dinçay KÖKSAL [1], Ayşe DAĞ PESTİL[2]

This study focuses on the methods and techniques supported by smart technologies, artificial intelligence, and digital tools for assessing listening and speaking skills. Research comparing digital and face-to-face methods for evaluating speaking proficiency reveals positive perceptions among students and instructors toward free and open-source software (Swanson, 2015). The study addresses issues related to pronunciation, rhythm, intonation, and stress in students' spoken language, as mispronounced words are often misunderstood when heard. The methodology is based on a student-centered approach that incorporates self-assessment and feedback, using digital tools and smart technologies. Conducted at a state university with approximately 200 students learning a foreign language, the study involved students intensively listening to specific audio files to improve their pronunciation, rhythm, intonation, and stress. They then created their own recordings and uploaded them via a task application on Teams, providing feedback using the platform's interactive reflection feature. The reflection data, categorized by adjective-descriptions and color-coded emojis, were grouped, coded, and analyzed through performance comparisons. The study aimed to help students overcome difficulties in listening and speaking skills by effectively using smartphones, computers, the internet, and digital platforms. The findings indicate that smart technologies, AI, and digital platforms play a significant role in enhancing students' speaking and pronunciation abilities. Bondarenko and Gudkova (2023) highlight the significant contributions of integrating audiovisual technology and internet resources in developing speech and listening skills (Bondarenko & Gudkova, 2023). The results of this study provide valuable insights and potential recommendations for the effective use of AI, digital and smart technologies in language skills teaching and assessment.

Keywords: *Smart technologies, assessment and evaluation of speaking skills, digital platforms.*

Reference

Swanson, P. B. (2015). Measuring language learners' speaking proficiency in a second language using economical digital tools. In Open Source Technology: Concepts, Methodologies, Tools, and Applications (pp. 1504-1523). IGI Global.

Bondarenko, A., & Gudkova, N. (2023). Enhancing foreign language teaching through digital tools and resources. Collection of scientific papers «ΛΟΓΟΣ», (June 23, 2023; Oxford, UK), 129-132.

[1] koksal.dincay@gmail.com:, ORCID: https://orcid.org/0000-0002-8681-4093, Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

[2] aysedag@comu.edu.tr:, ORCID: https://orcid.org/0000-0002-3460-3390, Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

Geleneksel Dört Beceri Değerlendirmesi ile Alternatif Portföy Değerlendirmesinin B2 Seviye İngilizce Hazırlık Öğrencilerinde Karşılaştırılması

Gizem ARSLAN DEĞİRMENCİ [1]

Bu çalışma, B2 seviye İngilizce hazırlık öğrencilerinde geleneksel dört beceri değerlendirme yöntemlerinin alternatif portföy değerlendirmesi ile etkinliğini karşılaştırmayı amaçlamaktadır. Dil öğreniminde bütüncül ve kapsamlı değerlendirmeye yönelik artan vurgu ile bu araştırma, bu iki değerlendirme yaklaşımının öğrenci performansı ve katılımını nasıl etkilediğini araştırmaktadır.

Çalışma, öğrencilerin geleneksel dört beceri sınavlarından ve portföy değerlendirmelerinden elde ettikleri puanlara odaklanan nicel bir yaklaşımı kullanmaktadır. Geleneksel değerlendirme, dinleme, konuşma, okuma ve yazma becerilerini ölçen standart testleri içermektedir. Buna karşılık, portföy değerlendirmesi, öğrencilerin dönem boyunca topladıkları çalışmaları ve yansımaları, önceden belirlenmiş kriterlere göre değerlendirir.

Veri analizi, tanımlayıcı istatistikleri ve çift örneklem t-testlerini içermektedir. Tanımlayıcı istatistikler, portföy değerlendirmeleri için ortalama puanın 95.60 (SD = 2.74) olduğunu, dört beceri testleri için ise ortalama puanın 76.38 (SD = 7.45) olduğunu göstermiştir. Çift örneklem t-testi sonuçları, iki değerlendirme yöntemi arasında istatistiksel olarak anlamlı bir fark olduğunu göstermiştir ($t = 12.60$, $p < 0.001$).

Sonuçlar, geleneksel değerlendirmelerin belirli bir zamanda belirli dil becerilerinin güvenilir bir ölçüsünü sunduğunu, portföy değerlendirmelerinin ise öğrenci ilerlemesi ve iyileştirilmesi gereken alanlar hakkında daha derin bir içgörü sağladığını göstermektedir. Ayrıca, portföy değerlendirmelerinin öğrenci katılımını ve öğrenme sürecine olan sahipliği teşvik ettiği görülmektedir. Bu araştırma, dil eğitiminde daha etkili ve öğrenci odaklı değerlendirme stratejileri uygulamak isteyen eğitimciler ve politika yapıcılar için önemli bir tartışmaya katkıda bulunmaktadır.

Anahtar Kelimeler: *Dil Değerlendirmesi, Portföy Değerlendirmesi, Dört Beceri Değerlendirmesi,* İngilizce *Hazırlık* Öğrencileri*, Alternatif Değerlendirme Yöntemleri.*

Kaynakça

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology-TOJET*, *2*(3), 13-19.

Goctu, R. (2012). Comparison of traditional and portfolio assessment efficiency in English language teaching in high schools. *Journal of Education*, *1*(2), 43-52.

[1] gizemarslandegirmenci@gmail.com:, ORCID: 0000-0003-2204-4891, Yüksek İhtisas Üniversitesi, Ankara, Türkiye

Comparing Traditional Four-Skill Assessment and Alternative Portfolio Evaluation in B2 Level English Preparatory Students

Gizem ARSLAN DEĞİRMENCİ[1]

This study aims to compare the effectiveness of traditional four-skill assessment methods with alternative portfolio evaluation for B2 level English preparatory students. With the increasing emphasis on holistic and comprehensive evaluation in language learning, this research investigates how these two assessment approaches influence student performance and engagement.

The research utilizes a quantitative approach, focusing on the scores obtained by students from traditional four-skill exams and portfolio evaluations. The traditional assessment includes standardized tests that measure listening, speaking, reading, and writing skills. In contrast, the portfolio assessment consists of students' collected works and reflections throughout the semester, evaluated based on predefined criteria.

Data analysis involves descriptive statistics and paired t-tests to compare the performance outcomes of the two assessment methods. Descriptive statistics revealed that the mean score for portfolio evaluations was 95.60 (SD = 2.74), whereas the mean score for the four-skill tests was 76.38 (SD = 7.45). Paired t-test results indicated a statistically significant difference between the two assessment methods ($t = 12.60$, $p < 0.001$).

The results indicate that while traditional assessments provide a reliable measure of specific language skills at a given point in time, portfolio evaluations offer deeper insights into student progress and areas of improvement. Furthermore, portfolio assessments seem to foster greater student involvement and ownership of their learning process. The findings are particularly relevant for educators and policymakers seeking to implement more effective and learner-centered evaluation strategies in language education.

Keywords: *Language Assessment, Portfolio Evaluation, Four-Skill Assessment, English Preparatory Students, Alternative Assessment Methods.*

Reference

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology-TOJET*, *2*(3), 13-19.

Goctu, R. (2012). Comparison of traditional and portfolio assessment efficiency in English language teaching in high schools. *Journal of Education*, *1*(2), 43-52.

[1] gizemarslandegirmenci@gmail.com:, ORCID: 0000-0003-2204-4891, Yüksek İhtisas Üniversitesi, Ankara, Türkiye

Kısa Yanıtlı Soruları Özdevinimli Puanlanma Platformu: Cevaplab

Şevket Polan[1], Tarık Kışla [2], Elif K. Demir [3], Hakan Atılgan[4], Burak Aydın[5]

Testlerde yer alan madde türlerinden olan Kısa Yanıtlı Sorular (KYS), yanıtının cevaplayıcı tarafından üretilmesi gereken maddelerdir. Eğitimin tüm basamaklarında sıklıkla kullanılan ve bilişsel alanın hemen her düzeyine uygun olarak yazılabilecek olan bu madde türünün; bilginin yeniden çağırılarak ifade edilmesi, yazma becerisine olumlu katkısı, şans başarısının olmaması gibi avantajları bulunmaktadır. Ancak KYS gibi yanıtın bireyler tarafından üretildiği tüm soru türlerinde puanlayıcılar tarafından yapılan puanlamanın zaman alıcı olması ve puanlama maliyeti önemli bir problemdir. Bu nedenle, özellikle birey sayısının fazla olduğu geniş ölçekli testlerde KYS'nin kullanımı sınırlanmaktadır. Bu probleme çözüm olarak KYS'nin özdevinimli, bir diğer ifadeyle otomatik puanlanması (OP) önemli bir ölçme ve değerlendirme konusudur. Burrows, Gurevych ve Stein (2015), OP yapabilen, biri Çince, üçü Almanca, dördü İspanyolca ve geri kalanı İngilizce için olmak üzere, 37 farklı yazılım tespit etmiştir. Geniş ölçekli uluslararası sınavlar dahil olmak üzere KYS-OP sistemleri geliştirilmeye devam edilmektedir (Tyack, Khorramdel & von Davier, 2024). Bu bildiri ile KYS için Türkçe diline uygun olarak, sınav hazırlama, uygulama ve değerlendirme süreçlerini desteklemek amacıyla 2024 yılı itibariyle geliştirilmeye başlanan Cevaplab adlı web tabanlı platformun tasarım ve uygulama süreçleri tanıtılacaktır. Cevaplab.com, kullanıcıların sınav sorularını sisteme girmesine, bu sorularla sınavlar hazırlamasına, sınavları gerçekleştirmesine ve sonuçlarını OP dahil olmak üzere değerlendirmesine olanak tanıyan kapsamlı bir sistem sunmaktadır. Agile yazılım geliştirme metodolojisi ile inşa edilen platform; sürekli geri bildirime, kullanıcı ile yakın iş birliğine, esnek planlamaya ve değişen gereksinimlere adaptasyona uygundur. Platform, bir veri tabanı üzerinde çalışmakta olup ön uç, kullanıcıların sorunsuz ve etkileşimli bir deneyim yaşaması için HTML, CSS ve JavaScript gibi modern web teknolojileri kullanılarak geliştirilmiştir. Arka uç tarafı ise *php* ile hazırlanmıştır ve veri tabanı yönetiminden sorumludur. Veri tabanı yönetim sistemi olarak, performans ve ölçeklenebilirlik sağlayan MySQL ilişkisel veri tabanı tercih edilmiştir. Sonuç olarak, Cevaplab, eğitim ve sınav süreçlerinde verimliliği artırmak, kullanıcıların iş yükünü hafifletmek ve değerlendirme süreçlerinde tutarlılığı artırmak amacıyla geliştirilmiş bir platformdur.

Anahtar Kelimeler: *Kısa yanıtlı sorular, otomatik puanlama, sınav platformu*

Kaynakça

Burrows, S., Gurevyc, I., & Stei, B. (2015). The Eras and Trends of Automatic Short Answer Grading. *International Journal of Artificial Intelligence in Education*, 25-60.

Tyack, L., Khorramdel, L., & von Davier, M. (2024). Using Convolutional Neural Networks to Automatically Score Eight TIMSS 2019 Graphical Response Items. *Computers and Education: Artificial Intelligence*, 100249.

[1] sevket.polan@ege.edu.tr: ORCID: 0000-0002-1096-1548, Ege Üniversitesi, İzmir, Türkiye

[2] tarik.kisla@ege.edu.tr: ORCID: 0000-0001-9007-7455, Ege Üniversitesi, İzmir, Türkiye

[3] elif.kubra.demir@ege.edu.tr: ORCID: 0000-0002-3219-1644, Ege Üniversitesi, İzmir, Türkiye

[4] hakan.atilgan@ege.edu.tr: ORCID: 0000-0002-5562-3446Ege Üniversitesi, İzmir, Türkiye

[5] burak.aydin@ege.edu.tr: ORCID: 0000-0003-4462-1784, Ege Üniversitesi/ Leuphana Üniversitesi, İzmir/Hamburg, Türkiye/Almanya

An Automatic Short Answer Grading Platform: Cevaplab

Şevket Polan[1], Tarık Kışla [2], Elif K. Demir [3], Hakan Atılgan[4], Burak Aydın[5]

Short answer (SA) questions are classified as constructed response items. SAs are frequently used at different levels of education to measure different levels of cognitive demands. SAs have advantages such as their potential to assess complex skills, to address guessing and they can promote students' writing skills. However, it is an important issue that scoring is time-consuming in all question types where the answer is produced by individuals. For this reason, the use of SA is limited especially in large- scale exams with a large number of individuals. As a promising solution, automatic scoring (AS) of SA has become an important research topic in the field of assessment. Burrows, Gurevych, and Stein (2015) identified 37 different software programs that can perform AS, including one for Chinese, three for German, four for Spanish, and the rest for English. SA-AS systems continue to be developed, including large-scale international exams (Tyack, Khorramdel & von Davier, 2024).

In this presentation, the design and implementation processes of the web-based platform named Cevaplab will be introduced. We started to develop Cevaplab in 2024 to support exam preparation, administration and evaluation processes for the SA item in Turkish. Cevaplab offers a comprehensive system that allows users to enter exam questions into the system, prepare exams with these questions, conduct exams and evaluate the results, including AS. Built with Agile software development methodology, the platform is suitable for continuous feedback, close collaboration with the user, flexible planning and adaptation to changing requirements. The platform runs on a database and the front-end is developed using modern web technologies such as HTML, CSS and JavaScript to provide users with a seamless and interactive experience. The backend is built in php and is responsible for database management. MySQL relational database, which provides performance and scalability, was preferred as the database management system. Overall, Cevaplab is a platform developed to increase efficiency in education and exam processes, lighten the workload of users and increase consistency in assessment processes.

Keywords: *Short Answer, Constructed response, Automated Scoring, Assessment Platform*

Reference

Burrows, S., Gurevyc, I., & Stei, B. (2015). The Eras and Trends of Automatic Short Answer Grading.

International Journal of Artificial Intelligence in Education, 25-60.

Tyack, L., Khorramdel, L., & von Davier, M. (2024). Using Convolutional Neural Networks to Automatically Score Eight TIMSS 2019 Graphical Response Items. *Computers and Education: Artificial Intelligence*, 100249.

[1] sevket.polan@ege.edu.tr: ORCID: 0000-0002-1096-1548, Ege Üniversitesi, İzmir, Türkiye

[2] tarik.kisla@ege.edu.tr: ORCID: 0000-0001-9007-7455, Ege Üniversitesi, İzmir, Türkiye

[3] elif.kubra.demir@ege.edu.tr: ORCID: 0000-0002-3219-1644, Ege Üniversitesi, İzmir, Türkiye

[4] hakan.atilgan@ege.edu.tr: ORCID: 0000-0002-5562-3446Ege Üniversitesi, İzmir, Türkiye

[5] burak.aydin@ege.edu.tr: ORCID: 0000-0003-4462-1784, Ege Üniversitesi/ Leuphana Üniversitesi, İzmir/Hamburg, Türkiye/Almanya

Açık Uçlu Maddelerin Puanlanmasında Yapay Zekâ Performansının İncelenmesi

Dr. Mahmut Sami Yiğiter [1], Dr. Erdem Boduroğlu [2]

Yapay Zeka (YZ), açık uçlu soruların puanlanması da dahil olmak üzere çeşitli eğitim araçlarına giderek daha fazla entegre edilmektedir (Chen vd., 2020). Bu çalışmanın amacı, öğrencilerin açık uçlu maddelere el yazısıyla verdiği yanıtların değerlendirilmesinde YZ'nin puanlama performansını incelemektir. Araştırmada bir devlet üniversitesinde Eğitimde Ölçme ve Değerlendirme dersi kapsamında 4 açık uçlu sorudan oluşan bir başarı testi geliştirilmiştir. Çoktan seçmeli maddeler 0-1 şeklinde puanlanırken, açık uçlu maddeler 0-1-2 şeklinde puanlanmıştır. 84 katılımcının yer aldığı çalışmada yer alan açık uçlu sorular uzman grubu ve bir YZ aracı (ChatGPT-4o) tarafından puanlanmıştır. Araştırmada YZ aracına iki farklı senaryoda öğrencilerin el yazıları ile yazdıkları görsel yanıtlar puanlatılmıştır. Birinci senaryoda YZ'ye herhangi bir puanlama ölçütü verilmeden YZ aracının puanlama yapması istenirken, ikinci senaryoda standart değerlendirme ölçütlerine göre YZ'den puanlama yapması istenmiştir. Araştırmanın sonuçları, uzman puanları ile ölçütsüz YZ puanları arasındaki Kappa uyum istatistikleri sırasıyla 0.04, 0.07, 0.03 ve 0.18 iken, uzman puanları ile standart değerlendirme ölçütleriyle YZ puanları arasındaki Kappa uyum istatistikleri 0.55, 0.85, 0.73, 0.78'dir. Puanlamalar arasındaki farklılığı test etmek için yapılan t test sonuçlarına göre uzman puanları ile ölçütsüz YZ puanları arasında anlamlı farklılık var iken, uzman puanları ile standart değerlendirme ölçütleriyle YZ puanları arasında anlamlı fark bulunmamaktadır. Sonuçlar, YZ'nin standart değerlendirme ölçütleriyle el yazısıyla yanıtlanmış açık uçlu maddeleri iyi düzeyde puanlayabildiğini göstermektedir. Gelecekte YZ'nin gelişim ve dönüşümüyle birlikte tutarlılık açısından insan puanlayıcılarla karşılaştırılabilir puanlama doğruluğuna ulaşabileceği düşünülmektedir.

Anahtar Kelimeler: *açık uçlu madde puanlama, yapay zeka, chatGPT, puanlama.*

Kaynakça

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264-75278.

[1] mahmutsami.yigiter@asbu.edu.tr:, ORCID: 0000-0002-2896-0201, Ankara Sosyal Bilimler Üniversitesi, Ankara, Türkiye

[2] erdemboduroglu@gmail.com.:, ORCID: 0000-0001-8318-4914, Milli Eğitim Bakanlığı, Niğde, Türkiye

Investigation of Artificial Intelligence Performance in Scoring Open-Ended Items

Dr. Mahmut Sami Yiğiter [1], Dr. Erdem Boduroğlu [2]

Artificial Intelligence (AI) is increasingly being integrated into various educational tools, including the scoring of open-ended questions (Chen et al., 2020). The aim of this study is to examine the scoring performance of AI in the evaluation of students' handwritten responses to open-ended items. In the study, an achievement test consisting of 4 open-ended questions was developed within the scope of Measurement and Evaluation in Education course at a state university. While multiple-choice items were scored as 0-1, open-ended items were scored as 0-1-2. 84 participants took part in the study and the open-ended questions were scored by an expert group and an AI tool (ChatGPT-4o).

In the study, the AI tool scored the students' handwritten visual responses in two different scenarios. In the first scenario, the AI was given any scoring criteria.

In the second scenario, the AI was asked to score according to the standard evaluation criteria. The results of the study show that the Kappa agreement statistics between expert scores and AI scores without criteria are 0.04, 0.07, 0.03 and 0.18 respectively, while the Kappa agreement statistics between expert scores and AI scores with standard evaluation criteria are 0.55, 0.85, 0.73, 0.78. According to the t-test results conducted to test the difference between the ratings, there is a significant difference between expert scores and AI scores without criteria, while there is no significant difference between expert scores and AI scores with standard evaluation criteria. The results show that AI can score open-ended items answered in handwriting with standard assessment criteria at a good level. In the future, with the development and transformation of AI, it is thought that it can reach scoring accuracy comparable to human raters in terms of consistency.

Keywords: *Scoring open-ended items, artificial intelligence, chatGPT.*

Reference

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264-75278

[1] mahmutsami.yigiter@asbu.edu.tr:, ORCID: 0000-0002-2896-0201, Ankara Sosyal Bilimler Üniversitesi, Ankara, Türkiye

[2] erdemboduroglu@gmail.com.:, ORCID: 0000-0001-8318-4914, Milli Eğitim Bakanlığı, Niğde, Türkiye

Boyutlar Arasındaki İlişki Düzeyinin Çok Boyutlu Test Eşitlemeye Etkisi *

Burcu DEMİRÖZ [1], Nuri DOĞAN[2]

Çok boyutlu testler için çok boyutlu test eşitleme yöntemleri geliştirilmiştir. ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (Brossman, 2010) ve bi-faktör ÇB-MTK gözlenen puan (Lee ve Lee, 2016) yöntemleri bunlardan bazılarıdır. Yöntemlerin performansları çeşitli koşullar altında incelenmeye devam edilmektedir. Boyutlar arasındaki ilişki düzeyi arttıkça testler tek boyutluluğa yakınsamaktadır. Bu nedenle boyutlar arasındaki ilişki düzeyi yöntemlerin performansını etkileyen koşullardan biri olduğu düşünülmektedir. Bu araştırmada, boyutlar arasındaki ilişki düzeyinin ÇB-MTK gözlenen puan eşitleme tek boyutlu yaklaşım (AOSE) bi-faktör ÇB-MTK gözlenen puan eşitleme (BF-MIRT) yöntemlerinden denk olmayan gruplarda ortak madde deseni altında elde edilen eşitlenmiş puanlara ait hata değerlerine etkisini belirlemek amaçlanmıştır.

Monte Carlo simülasyon çalışması yürütülmüştür. Araştırmada, bir genel ve iki spesifik faktörden oluşan 3 boyutlu bi-faktör model kullanılmıştır. Örneklem büyüklüğü 1500, ortak madde oranı %20 ve boyutlar arasındaki ilişki düzeyi ρ=0, 0,3 ve 0,6 olarak belirlenmiştir. Test 60 maddeden oluşmaktadır. Eş zamanlı kalibrasyon yöntemi kullanılmıştır. Test eşitleme yöntemlerini incelemek amacıyla hata değerleri hesaplanmıştır.

BF-MIRT yöntemi için boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin arttığı gözlenmiştir. AOSE yönteminde ise boyutlar arasındaki ilişki düzeyi arttıkça hata değerlerinin azaldığı gözlenmiştir. Boyutlar arasındaki ilişki düzeyi BF-MIRT yöntemi kullanıldığında ρ=0'a yakın; AOSE yöntemi kullanıldığında ρ=0,60 ve üstü olmalıdır.

Anahtar Kelimeler: *Test eşitleme, eşitleme hatası,* çok *boyutluluk.*

Kaynakça

Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory* (Doctoral Dissertation). University of Iowa, USA.

Lee, G., & Lee, W. (2016). Bi-factor MIRT observed-score equating for mixed-format tests. *Applied Measurement in Education, 29*, 224-241.

* Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı doktora tezinden üretilmiştir.

[1] burcu199514.bdz@gmail.com:, ORCID: 0000-0002-8326-4671, MEB, Gaziantep, Türkiye

[2] nurid@hacettepe.edu.tr:, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Effect of the Level of Relationship Between Dimensions on Multidimensional Test Equalization *

Burcu DEMİRÖZ [1], Nuri DOĞAN[2]

Multidimensional test equating methods have been developed for multidimensional tests. The Unidimensional Approximation of MIRT Observed Score Equating (Brossman, 2010) and Bi-Factor MIRT Observed Score Equating (Lee and Lee, 2016) methods are some of them. The performances of the methods continue to be examined under various conditions. As the level of relationship between the dimensions increases, the tests converge to unidimensionality. Therefore, the level of relationship between the dimensions is thought to be one of the conditions affecting the performance of the methods. In this study, it was aimed to determine the effect of the level of relationship between the dimensions on the error values of the equated scores obtained under the common item design in non- equivalent groups from the Unidimensional Approximation of MIRT Observed Score Equating (AOSE) and Bi-Factor MIRT Observed Score Equating (BF-MIRT) methods.

A Monte Carlo simulation study was conducted. In the study, a 3-dimensional bi-factor model consisting of one general and two specific factors was used. The sample size was 1500, the common item rate was 20%, and the relationship level between the dimensions was determined as ρ=0, 0.3, and 0.6. The test consisted of 60 items. The simultaneous calibration method was used. In order to examine the test equating methods, errors values were calculated.

For the BF-MIRT method, it was observed that as the level of correlation between dimensions increased, the error values increased. In the AOSE method, it was observed that as the level of correlation between dimensions increased, the error values decreased. When the BF-MIRT method is used, the level of correlation between dimensions should be close to ρ=0; when the AOSE method is used, ρ=0.60 and above.

Keywords: *Test equating, equating error, multidimensionality.*

Reference

Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory* (Doctoral Dissertation). University of Iowa, USA.

Lee, G., & Lee, W. (2016). Bi-factor MIRT observed-score equating for mixed-format tests. *Applied Measurement in Education, 29*, 224-241.

* Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı doktora tezinden üretilmiştir.

[1] burcu199514.bdz@gmail.com:, ORCID: 0000-0002-8326-4671, MEB, Gaziantep, Türkiye

[2] nurid@hacettepe.edu.tr:, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Açık Uçlu Sınavların Otomatik Değerlendirilmesi: Görüntü İşleme ve Yapay Zekâ Destekli iOS Tabanlı Mobil Uygulama Geliştirme Örneği

Nazmi Ekin Vural[1]

Bu çalışma kapsamında, açık uçlu sınavlarda el yazısıyla yazılmış cevap kağıtlarının tanınması ve değerlendirilmesi amacıyla bir iOS uygulaması geliştirilmiş, sınavların otomatik olarak puanlanması sağlanmış ve bu uygulama, kullanıcı deneyimleri doğrultusunda optimize edilmiştir. Çalışmanın amacı, sınav kağıtlarının otomatik olarak değerlendirilmesini sağlayan algoritmaların etkinliğini incelemek ve bu algoritmaların ölçme ve değerlendirme süreçlerindeki potansiyel etkisini ve uygulanabilirliğini değerlendirmektir.

El yazısıyla yapılan açık uçlu sınavları değerlendiren ve iOS uygulamaları geliştirmek için Apple firması tarafından oluşturulmuş bir programlama dili olan Swift ile kodlanan, *Sınav Okuma Uygulaması* adlı bu uygulama, iki temel bileşen üzerine kurulmuştur. Birincisi, el yazısının tanınması için Google Cloud Vision API (Uygulama Programlama Arayüzü) kullanılmasıdır. İkincisi ise, tanınan bu verilerin anlamlandırılması ve değerlendirme kriterlerine uygun bir biçimde puanlanması için kullanılan OpenAI API'dir. Çalışmanın teknik yönteminde, öğrencilerin sınav kağıtları OCR (Optik Karakter Tanıma) teknolojisi ile taranmış, ardından dil bilgisi hatalarının düzeltilmesi ve anlatımın düzenlenmesi için yapay zekâ teknolojisi kullanılmıştır. Son olarak, düzeltilen ve önceden belirlenen değerlendirme kriterlerine göre puanlanan sınav sonuç ve diğer detayları içeren bilgiler PDF belgesi olarak kullanıcıya sunulmuştur.

Ayrıca, uygulamanın potansiyel etkisini ve uygulanabilirliğini değerlendirmek amacıyla bir araştırma yapılmış ve uygulama 8 akademisyenin kullanımına sunulmuştur. Araştırmaya katılan akademisyenlerden, kendilerine ait 10 adet açık uçlu sınav cevap kağıdını uygulama üzerinden puanlamaları istenmiştir. Daha sonra, sınav puanlaması için uygulamayı kullanan akademisyenlerle, uygulamanın doğruluk oranı, performansı ve deneyimlerine yönelik yarı yapılandırılmış görüşmeler gerçekleştirilmiştir. Bu görüşmeler sonucunda, *Sınav Okuma Uygulaması* akademisyenlerin değerlendirmeleri doğrultusunda iyileştirilmiştir. Uygulamanın, el yazısının tanınması ve değerlendirilmesi süreçlerinde %80'e varan yüksek doğruluk oranları sağladığı ve akademik sınavların değerlendirme süreçlerinde zaman tasarrufu ve objektiflik sağladığı ortaya konulmuştur.

Anahtar Kelimeler: *Görüntü* İşleme*, El Yazısı Tanıma, Yapay Zekâ, Cloud Vision API, OpenAI API, iOS*

[1] ekin.vural@bozok.edu.tr, ORCID: 0000-0003-4198-0407, Yozgat Bozok Üniversitesi, Yozgat, Türkiye

Automatic Evaluation of Open-Ended Exams: A Case Study of Developing an iOS-Based Mobile Application Supported by Image Processing and AI

Nazmi Ekin Vural[1]

Within the scope of this study, an iOS application was developed to recognize and evaluate handwritten answer sheets in open-ended exams, automatically score the exams, and this application was optimized in line with user experiences. The aim of the study is to examine the effectiveness of algorithms that provide automatic evaluation of exam papers and to evaluate the potential impact and applicability of these algorithms in measurement and evaluation processes.

This application, called Exam Reading App, which evaluates handwritten open-ended exams and is coded in Swift, a programming language created by Apple Inc. for developing iOS applications, is based on two main components. The first one is the Google Cloud Vision API (Application Programming Interface) for recognizing handwriting. The second one is the OpenAI API, which is used to make sense of this recognized data and score it in accordance with the evaluation criteria. In the technical method of the study, students' exam papers were scanned with OCR (Optical Character Recognition) technology, and then artificial intelligence technology was used to correct grammatical errors and organize the narration. Finally, the information containing the exam results and other details, corrected and scored according to predetermined evaluation criteria, were presented to the user as a PDF document.

In addition, a study was conducted to evaluate the potential impact and applicability of the application, and the application was made available to 8 academicians. The academicians participating in the study were asked to score 10 open-ended exam answer sheets of their own through the application. Then, semi-structured interviews were conducted with the academics who used the application for exam scoring about the accuracy rate, performance and experiences regarding the application. As a result of these interviews, the Test Reading Application was improved in line with the evaluations of the academics. It was revealed that the application provides high accuracy rates of up to 80% in handwriting recognition and evaluation processes and enables time saving and objectivity in the evaluation processes of academic exams.

Keywords: *Image Processing, Handwriting Recognition, AI, Cloud Vision API, OpenAI API, iOS*

[1] ekin.vural@bozok.edu.tr, ORCID: 0000-0003-4198-0407, Yozgat Bozok University, Yozgat, Türkiye

PISA 2022 Matematik Okuryazarlığı Performansı ve Okul Etkililiği ile İlişkili Faktörlerin Belirlenmesi: Kapsamlı Bir Değerlendirme

Şeyma Nur Özsoy [1], Büşra Sarıhan Koçak[2]

Eğitim ve öğretim sürecinde tüm öğrencilere kaliteli eğitim sağlanması hedeflense de okullar arasında belirgin kalite ve etki farklılıkları mevcuttur. Bu çalışma, PISA 2022 öğrenci ve okul anketi verileriyle, öğrenci ve okul düzeylerinde Türkiye'de matematik okuryazarlığı performanslarını ve okul etkililiği ile ilişkili faktörleri kapsamlı bir şekilde incelemeyi hedeflemektedir. İlk olarak öğrenci matematik performansı üzerindeki etkileri incelenmiş ve etkili ve etkisiz okullar iki düzeyli modelle belirlenmiştir. Daha sonra karar ağacı yöntemi kullanılarak, öğrenci ve okul düzeyindeki çok sayıda değişkenin etkili ve etkisiz okulların sınıflandırılmasındaki etkileri ortaya koyulmuştur. Bu süreç, okul etkililiği çalışması, CIPO (bağlam-girdi-süreç-çıktı) modeli temel alınarak yürütülmüştür (Chapman ve diğerleri, 2016). Türkiye'nin 2022 yılı örneklemi, 7528 öğrenci ve 196 okuldan oluşmaktadır. Ancak Hiyerarşik Lineer Modelleme (HLM) analizi öncesinde 20'den az katılımcı ile değerlendirme yapılan okullar analiz dışında bırakılmıştır. Ayrıca ESCS (Ekonomik, sosyal ve kültürel durum indeksi) değişkeninde bulunan 5 kayıp veri analizden çıkarılmıştır. Böylece kullanılan örneklem ise 182 okul ve 7042 öğrenciye düşürülmüştür. Etkili ve etkisiz okul ayrımını yapmak ve okul etkililiğiyle ilişkili faktörleri belirleyebilmek için R 4.4.1 programından yararlanılmıştır. Yapılan analizler sonucunda, 182 okul, girdi değişkenleri aracılığıyla etkisiz (95 okul) ve etkili (87 okul) olmak üzere iki gruba ayrılmıştır. Sonrasında süreç değişkenleri kullanılarak karar ağacı oluşturulmuştur. Değerlendirme kriteri olarak duyarlılık, özgüllük, pozitif tahmin, negatif tahmin, tespit oranı, tespit yaygınlığı, dengelenmiş doğruluk ve AUC (ROC eğrisi altında kalan alan) değerleri ele alınmıştır. Her iki sınıf (etkili okul ve etkisiz okul) için ROC eğrileri yüksek duyarlılık ve özgüllük seviyelerinde belirgin bir performans gösterdiği, her iki sınıf için de AUC değerinin 0.89 olması modelin oldukça iyi bir ayrıştırma kapasitesine sahip olduğunu göstermektedir. Bu nedenle eğitimde fırsat eşitliğini sağlamak ve matematik performansını artırmak amacıyla okul etkililiğini artırmaya yönelik çalışmalar yapılmak istendiğinde bu modelin dikkate alınması faydalı olabilir.

Anahtar Kelimeler: *Etkili okul, Karar ağaçları, PISA.*

Kaynakça

Chapman, C., Muijs, D., Raynolds, D., Sammons, P., & Teddlie, C. (Eds.). (2016). *The Routledge International Handbook of Educational Effectiveness and Improvement Research, policy, and practice*. New York: Routledge.

[1] 1seymaozsoy@gmail.com, ORCID: 0000-0002-8306-6114, Ankara Üniversitesi Doktora Öğrencisi, Ankara, Türkiye

[2] busrasarihan12@gmail.com, ORCID: 0009-0001-0964-8247, Ankara Üniversitesi Doktora Öğrencisi, Ankara, Türkiye

Determining the Factors Associated with PISA 2022 Mathematics Literacy Performance and School Effectiveness: A Comprehensive Evaluation

Şeyma Nur Özsoy [1], Büşra Sarıhan Koçak[2]

Despite the goal of providing high-quality education to all students during the education and teaching process, there are significant differences in quality and impact between schools. This study aims to comprehensively examine the factors related to mathematics literacy performance and school effectiveness at the student and school levels in Turkey using the PISA 2022 student and school survey data. Firstly, the effects on student mathematics performance were examined, and effective and ineffective schools were identified using a two-level model. Subsequently, using the decision tree method, the effects of numerous variables at the student and school levels on the classification of effective and ineffective schools were revealed. This process was conducted based on the CIPO (context-input-process-output) model for school effectiveness (Chapman et al., 2016). The 2022 sample from Turkey consists of 7528 students and 196 schools. However, schools evaluated with less than 20 participants were excluded from the Hierarchical Linear Modeling (HLM) analysis. Additionally, 5 missing data points in the ESCS (Index of economic, social and cultural status) variable were excluded from the analysis. Thus, the sample used consisted of 182 schools and 7042 students. To differentiate between effective and ineffective schools and determine factors related to school effectiveness, the R 4.4.1 program was utilized. As a result of the analyses, 182 schools were divided into two groups: ineffective (95 schools) and effective (87 schools) through input variables. Then, a decision tree was created using process variables. Evaluation criteria included sensitivity, specificity, positive prediction, negative prediction, detection rate, detection prevalence, balanced accuracy, and AUC (area under the ROC curve) values. The ROC curves for both classes (effective school and ineffective school) showed a notable performance at high sensitivity and specificity levels, with an AUC value of 0.89 for both classes, indicating that the model has a very good discriminative capacity. Therefore, when aiming to ensure equal opportunities in education and improve mathematics performance, considering this model in efforts to enhance school effectiveness could be beneficial.

Keywords: Effective school, Decision trees, *PISA*

Reference

Chapman, C., Muijs, D., Raynolds, D., Sammons, P., & Teddlie, C. (Eds.). (2016). *The Routledge International Handbook of Educational Effectiveness and Improvement Research, policy, and practice*. New York: Routledge.

[1] 1seymaozsoy@gmail.com, ORCID: 0000-0002-8306-6114, Ankara University PhD Student, Ankara, Türkiye

[2] busrasarihan12@gmail.com, ORCID: 0000-0002-8306-6114, Ankara University PhD Student, Ankara, Türkiye

Yapay Zekâ İngilizce Yazma Becerilerini Puanlamada Ne Kadar Başarılı?

Nagihan Boztunç Öztürk [1], Melek Gülşah Şahin[2]

Günümüzde her alanda karşımıza çıkan yapay zekâ uygulamaları her geçen gün merakla takip edilmektedir. Eğitimde ölçme ve değerlendirme alanının en önemli konulardan olan geçerlik ve güvenirlik bağlamında yapay zekâ uygulamalarında kullanılması da araştırma konularındandır. Yapay zekâ birey ve makine özelliklerini bir araya getirmesinden dolayı bireylerden elde edilen sonuçlar ile karşılaştırılması alana katkı sağlamaktadır. Literatür incelendiğinde, birey ve yapay zekanın Pisa okuma becerisi (Demir, 2023), İngilizce yazama becerisi (Han ve Sarı, 2024; Schmidt-Fajlik, 2023; Yancey, Laflair, Verardi, ve Burstein; 2023) gibi alanlarda puanlama açısından elde edilen puanların tutarlılığı bağlamında karşılaştırıldığı görülmektedir.

Bu çalışmada da "İnsan puanlayıcı ve Chat GPT puanlaması arasındaki ilişki nasıldır?, İnsan puanlayıcı ve Chat GPT puanlaması arasındaki ilişki zamana göre nasıl değişim göstermektedir? ve Puanlayıcılararası Güvenirlik Genellenebilirlik kuramına göre nasıldır?" sorularına cevap aranacaktır. Araştırmada bir üniversitenin Yabancı Diller Yüksekokulu'nun İngilizce seviye belirleme amacıyla öğrencilerine uyguladığı sınavların verisi kullanılacaktır. Verilerin analizinde puanlayıcılar arasındaki korelasyon değeri hesaplanacak ve elde edilen bulgular tablolaştırılacaktır. Puanlayıcılararası güvenirlik hesaplanırken genellenebilirlik kuramından faydalanılacaktır.

Anahtar Kelimeler: *ChatGPT, yapay zeka, ingilizce yazma becerisi.*

Kaynakça

Demir, S (2023). Investigation of ChatGPT and real raters in scoring open-ended items in terms of inter-rater reliability. International Journal of Turkish Educational Sciences, 11 (21), 1072-1099.

Kevin P. Yancey, Geoffrey Laflair, Anthony Verardi ve Jill Burstein. 2023. Rating Short L2 Essays on the CEFR Scale with GPT-4. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), syf: 576–584, Toronto, Kanada. Association for Computational Linguistics.

Schmidt-Fajlik, R. (2023). ChatGPT as a Grammar Checker for Japanese English Language Learners: A Comparison with Grammarly and ProWritingAid. AsiaCALL Online Journal, 14(1), 105–119.

Turgay Han ve Elif Sari (2024) An investigation on the use of automated feedback in Turkish EFL students' writing classes, Computer Assisted Language Learning, 37:4, 961-985, DOI: 10.1080/09588221.2022.2067179.

[1] nagihanboztunc@hacettepe.edu.tr, ORCID: 0000-0002-2777-5311, Hacettepe Üniversitesi, Ankara, Türkiye

[2] melekgulsah@gmail.com, ORCID: 0000-0001-5139-9777, Gazi Üniversitesi, Ankara, Türkiye

How Successful Is Artificial Intelligence in Scoring English Writing Skills?

Nagihan Boztunç Öztürk [1], Melek Gülşah Şahin[2]

Artificial intelligence, which we encounter in every field today, are followed with curiosity day by day. The use of artificial intelligence applications in the context of validity and reliability, which is one of the most important issues in the field of measurement and evaluation in education, is one of the research topics. Since artificial intelligence combines individual and machine features, comparing the results obtained from individuals contributes to the field. When the literature is examined, it is seen that human raters and artificial intelligence are compared in terms of the consistency of the scores obtained in terms of scoring in areas such as reading skills in PISA (Demir, 2023) or English writing skills (Han & Sarı, 2024; Schmidt-Fajlik, 2023; Yancey, Laflair, Verardi, & Burstein; 2023). In this context, answers to these questions "How is the relationship between human raters and Chat GPT scoring?", "How does the relationship between human raters and Chat GPT scoring change over time?" and "How is inter- rater reliability according to Generalizability theory?" will be sought. In the study, the data of the exams administered by the School of Foreign Languages of a university to its students for English placement tests will be used. In the analysis of the data, the correlation value between the raters will be calculated and the findings obtained will be tabulated. Generalizability theory will be used to calculate inter-rater reliability.

Keywords: *Chat GPT, Artificial intelligence, English writing skills*

Reference

Demir, S (2023). Investigation of ChatGPT and real raters in scoring open-ended items in terms of inter-rater reliability. International Journal of Turkish Educational Sciences, 11 (21), 1072-1099.

Kevin P. Yancey, Geoffrey Laflair, Anthony Verardi ve Jill Burstein. 2023. Rating Short L2 Essays on the CEFR Scale with GPT-4. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), syf: 576–584, Toronto, Kanada. Association for Computational Linguistics.

Schmidt-Fajlik, R. (2023). ChatGPT as a Grammar Checker for Japanese English Language Learners: A Comparison with Grammarly and ProWritingAid. AsiaCALL Online Journal, 14(1), 105–119.

Turgay Han ve Elif Sari (2024) An investigation on the use of automated feedback in Turkish EFL students' writing classes, Computer Assisted Language Learning, 37:4, 961-985, DOI: 10.1080/09588221.2022.2067179.

[1] nagihanboztunc@hacettepe.edu.tr, ORCID: 0000-0002-2777-5311, Hacettepe University, Ankara, Türkiye

[2] melekgulsah@gmail.com, ORCID: 0000-0001-5139-9777, Gazi University, Ankara, Türkiye

Yapay Zeka ile Sınav Kağıdının Otomatik Değerlendirilmesi

Mustafa Çağrı Gürbüz [1], Tuba Gürbüz[2]

Yazılı sınav, eğitim alanında öğrenci performansını değerlendirmek için evrensel bir araçtır. Son zamanlarda önerilen makine öğrenimi tabanlı cevap notlandırma sistemleri, birçok uygulama alanında bulunmayan büyük bir eğitim verisi gerektirir. Yeterli eğitim verisi oluşturmak maliyetli ve zaman alıcıdır. Sonuç olarak, yapay zeka ile açık uçlu cevapları notlandırmak oldukça zordur. Bu makalede, küçük bir sınıf senaryosunda değerlendirme yapabilen uzun veya açıklayıcı cevap notlandırmak için pratik bir sistem önerilmektedir. Sistem, alan uzmanları tarafından fikir birliğine varılarak yazılmış bir referans cevabı kullanır ve bir öğrenci cevabının onunla benzerliğini hesaplar. Benzerlik hesaplaması için Anlamsal İndeksleme, Dirichlet Analizi, TextRank özetleyici ve sinirsel cümle yerleştirme tabanlı InferSent dahil olmak üzere çeşitli kelime düzeyi ve cümle düzeyi benzerlik ölçümlerini kullanılmaktadır. Cevaplara birkaç ön işleme adımı uygulanır. En önemlisi karmaşık ve bileşik cümlelerin basit cümlelere dönüştürülmesidir. Sistem, cümlelerin birebir eşlenmesinin istendiği cümle benzerliğini hesaplar. Bileşik cümleler birden fazla olgu içerir ve bire çok veya çoktan bire eşleme üretir. Bu nedenle, Eş-referans çözünürlüğü cümleler üzerinde bir diğer ön işleme adımı olarak gerçekleştirilir (Das ve Majumder, 2017). Öğrenci cevabı, model cevapta bulunmayan belirli gerçekleri içerebilir. Sistem bu tür cümleleri belirler, bunların alaka düzeyini ve doğruluğunu inceler ve buna göre ekstra puanlar atamaktadır. Değerlendirme için 50 öğrenci cevap fasikülü toplanmıştır. Veri setinde sekiz tanımlayıcı soru seçildi. Literatürde cevapları puanlamak için çeşitli şemalar benimsenmiştir. Son aşamada, sistem kümelemeye dayalı bir güven analizi kullanmaktadır. Sistem, matematik okuryazarlık sorularına verilen cevapların değerlendirilmesi üzerinden test edilmektedir. Deneysel sonuçlar, sistemin cevap kitaplarını yüksek doğrulukla değerlendirdiğini göstermektedir, en iyi kök ortalama kare hata değeri 0-5 puanlama ölçeğinde 0,59'dur. Önerilen tekniğin genel olduğunu ve diğer ilgili eğitim değerlendirmelerinde uygulanabileceği varsayılmaktadır. Ancak deneysel doğrulama gerektirmektedir. Sistem, güvenilir yanıtlar için yakın bir puan üretiyor. Aktif öğrenme benzeri bir yaklaşımın, güvensiz yanıtlar için de benzer yakın puan üretmek üzere minimum insan çabasını kullanacak şekilde uyarlanıp uyarlanamayacağı araştırılabilir. Model yanıta dahil olmayan ek cümlelerin tanımlanması ve doğrulanması, sistemin önemli bir modülüdür.

Anahtar Kelimeler: *Biçimlendirici geri bildirim, matematik okuryazarlığı, yapay zekâ.*

Kaynakça

Das, B, & Majumder, M. (2017). Factual open cloze question generation for assessment of learner's knowledge. *International Journal of Educational Technology in Higher Education, 14*(1).

[1] mcagrigurbuz@harran.edu.tr, ORCID: 0000-0003-1851-2672, Harran Üniversitesi, Eğitim Fakültesi, Şanlıurfa, Türkiye

[2] tubadag25@gmail.com, ORCID: 0000-0003-4862-8342, Milli Eğitim Bakanlığı, Şanlıurfa, Türkiye

Automatic Evaluation of Exam Paper with Artificial Intelligence

Mustafa Çağrı Gürbüz [1], Tuba Gürbüz[2]

The written exam is a universal tool for assessing student performance in the field of education. Recently proposed machine learning based answer grading systems require a large amount of training data, which is not available in many application domains. Generating sufficient training data is costly and time consuming. As a result, grading open-ended answers with AI is quite challenging. In this paper, we propose a practical system for long or descriptive answer grading that can perform evaluation in a small class scenario. The system uses a reference answer written by consensus by domain experts and calculates the similarity of a student answer to it. Several word-level and sentence- level similarity measures are used for similarity calculation, including Semantic Indexing, Dirichlet Analysis, TextRank summarizer and neural sentence embedding based InferSent. Several pre- processing steps are applied to the answers. The most important one is the conversion of complex and compound sentences into simple sentences. The system calculates sentence similarity, where a one- to-one mapping of sentences is desired. Complex sentences contain more than one fact and produce one-to-many or many-to-one mapping. Therefore, Co-reference resolution is performed as another preprocessing step on sentences (Das & Majumder, 2017). The student answer may contain certain facts that are not present in the model answer. The system identifies such sentences, analyses their relevance and accuracy, and assigns extra points accordingly. For the evaluation, 50 student answer fascicles were collected. Eight descriptive questions were selected in the data set. Various schemes for scoring answers have been adopted in the literature. In the final stage, the system uses a confidence analysis based on clustering. The system is tested through the evaluation of the answers to the mathematical literacy questions. Experimental results show that the system evaluates answer books with high accuracy, the best root mean square error value is 0.59 on the 0-5 scoring scale. It is assumed that the proposed technique is general and can be applied to other related educational assessments. However, it requires experimental validation. The system produces a close score for reliable responses. It could be investigated whether an active learning-like approach could be adapted to use minimal human effort to produce a similar close score for unsafe responses.

Keywords: *Artificial intelligence, formative feedback, mathematical literacy*

Reference

Das, B, & Majumder, M. (2017). Factual open cloze question generation for assessment of learner's knowledge. *International Journal of Educational Technology in Higher Education, 14*(1).

[1] mcagrigurbuz@harran.edu.tr, ORCID: 0000-0003-1851-2672, Harran University, Faculty of Education, Şanlıurfa, Türkiye

[2] tubadag25@gmail.com, ORCID: 0000-0003-4862-8342, Ministry of Education, Şanlıurfa, Türkiye

Ölçme ve Değerlendirme Uzmanlarının Yeni Öğretim Programındaki Ölçme Değerlendirme Anlayışı Açısından Görüşleri

Leyla Burcu DİNÇSOY [1], Uğur ALTUN [2], Hülya KELECİOĞLU[3]

Öğretim programı, Türk Dil Kurumu yayını Türkeçe Sözlük'te "Bir okulu bitirmek veya bir alanda uzmanlaşmak için okunması gereken ders ve konuları kapsayan plan; ders programı, müfredat, müfredat programı." olarak tanımlanmıştır (TDK, 2024). Ülkemizdeki son öğretim programı değişikliği 2024 yılında Türkiye Yüzyılı Maarif Modeli olarak adlandırılan program ile gerçekleştirilmiştir. Türkiye Yüzyılı Maarif Modeli'nde insanın çok yönlü gelişimini destekleyen, akademik başarılarının artırılmasının yanında, öğrencilerin sosyal, duygusal ve kültürel gelişimlerine de katkıda bulunmayı amaçlayan bütüncül eğitim yaklaşımı esas alınmıştır. Bu doğrultuda öğrenmeleri iyileştirmek için sürekli ve biçimlendirici ölçme ve değerlendirme anlayışı benimsenmiştir. Yeni Model, çağdaş teknikler ve araçlar kullanılarak öğrencilerin hayata ve geleceğe en iyi şekilde hazırlanmalarının sağlanması yanında programın felsefesinin amacına ulaşıp ulaşmadığını ölçme değerlendirme çalışmaları ve bunların dönütleriyle sağlanması gibi amaçlara ulaşma hedefindedir (MEB, 2024). Yeni öğretim programlarında kullanılacak ölçme değerlendirme araçlarının çeşitliliği, uygulanabilirliği, ekonomikliği, güncelliği, öğretmenlerin bu araçlarla ilgili yeterince bilgiye sahibi olmaları, ölçme ve değerlendirme sonuçlarını kullanabilme gibi konuların irdelenmesinin önemi büyüktür. Bu çalışmada, ölçme değerlendirme uzmanlarının 2024 yılında yenilenen öğretim programlarının ölçme ve değerlendirme açısından görüşlerinin betimlenmesi amaçlanmıştır. Bu amaç doğrultusunda hazırlanan anket, 150 ölçme ve değerlendirme uzmanına uygulanacaktır. Araştırmada soruları cevaplayabilmek için bir anket hazırlanmış ve bu anket soruları ölçme değerlendirme ve program geliştirme uzmanlarından görüş alınarak uygulamaya hazır hâle getirilmiştir. Daha sonra ölçme değerlendirme uzmanlarının anketi doldurmaları sağlanacaktır. Çalışmanın olası sonuçları ve ölçme değerlendirme uzmanlarından öğretim programının ölçme ve değerlendirme boyutu hakkındaki görüşlerine ulaşılarak programın olumlu ve olumsuz görülen yönleri belirlenip bu sonuçların program geliştiricilere bilgi vermesi amaçlanmakta, programın ölçme ve değerlendirme boyutunun geliştirilmesi ve uygulanmasına katkı sağlaması beklenmektedir.

Anahtar Kelimeler: öğretim *programı,* ölçme *ve değerlendirme, maarif modeli.*

Kaynakça

Millî Eğitim Bakanlığı (2024). *Türkiye Yüzyılı Maarif Modeli* Öğretim *Programları Ortak Metni*.

Türk Dil Kurumu (2024). Güncel Türkçe sözlük. https://sozluk.gov.tr/

[1] leylaburcuadak@gmail.com: ORCID: 0000-0002-5633-3520, MEB, Ankara, Türkiye

[2] uguraltun@hotmail.com: ORCID: 0009-0008-6185-9384, MEB, Antalya, Türkiye

[3] hulyakelecioglu@gmail.com: ORCID: 0000-0002-0741-9934, Hacettepe Üniversitesi, Ankara, Türkiye

Measurement and Evaluation Experts' Views on the Measurement and Evaluation Approach in the New Curriculum

Leyla Burcu DİNÇSOY [1], Uğur ALTUN [2], Hülya KELECİOĞLU[3]

The curriculum is defined in the Turkish Language Association dictionary as "a plan that includes the courses and subjects to be studied in order to graduate from a school or to specialize in a field; course program, curriculum" (TDK, 2024). The most recent curriculum change in our country occurred in 2024 with the introduction of the Türkiye Yüzyılı Maarif Modeli. This model adopts a holistic educational approach that supports the multi-faceted development of individuals, aiming not only to enhance academic success but also to contribute to students' social, emotional, and cultural development. To this end, a continuous and formative assessment approach is adopted to improve learning. The new model aims to prepare students for life and the future in the best possible way using contemporary techniques and tools, and to ensure that the program's philosophy is achieved through feedback and assessment studies (MEB, 2024). The importance of examining the diversity, applicability, cost- effectiveness, and up-to-date nature of the assessment tools to be used in the new curriculum, as well as the teachers' sufficient knowledge about these tools and their ability to use the assessment results, is significant. This study aims to describe the views of measurement and evaluation experts on the measurement and evaluation aspects of the revised 2024 curriculum. A survey prepared for this purpose will be administered to 150 measurement and evaluation experts. The survey questions were developed and made ready for implementation by consulting experts in measurement, evaluation, and curriculum development. Subsequently, measurement and evaluation experts will be asked to complete the survey. The potential outcomes of the study aim to determine the positive and negative aspects of the curriculum based on the opinions of measurement and evaluation experts, and to provide this information to curriculum developers. It is expected that the results will contribute to the improvement and implementation of the measurement and evaluation aspects of the program.

Keywords: *curriculum, measurement and evaluation, educational model.*

Reference

MEB (2024). *Türkiye Yüzyılı Maarif Modeli Curriculum Common Text*

Turkish Language Association (2024). Current Turkish Dictionary. Retrieved from https://sozluk.gov.tr/

[1] leylaburcuadak@gmail.com: ORCID: 0000-0002-5633-3520, MEB, Ankara, Türkiye

[2] uguraltun@hotmail.com: ORCID: 0009-0008-6185-9384, MEB, Antalya, Türkiye

[3] hulyakelecioglu@gmail.com: ORCID: 0000-0002-0741-9934, Hacettepe Üniversitesi, Ankara, Türkiye

Görme Engelli YKS Adaylarının Sınav Uyarlamaları Hakkındaki Bilgi Düzeyi: Mevcut Durum ve Geliştirme Önerileri

Huzeyfe Aydın [1], Tuba Boduroğlu [2], Ayşenur Bahçivancıoğlu Yazıcı [3]

Kişilerin hayatlarını çeşitli yönlerden etkileyen yükseköğretime geçiş sürecinin engelliler için uygun şekilde yürütülmesi, engellilerin kişisel gelişimleri için önem arz etmektedir.

Bu nedenle diğer birçok sınavda görüldüğü üzere üniversiteye yerleşmek amacıyla girilen Yükseköğretim Kurumları Sınavı'nda (YKS) da görme engelli adaylar için çeşitli tedbirler alınmaktadır. Görme engelli adaylar ancak sınav uyarlamaları hakkında doğru ve yeterli bilgiye sahip olmaları halinde bu tedbirlerden verimli bir şekilde yararlanabilir.

Bu araştırma, görme engelli adayların YKS sürecinde adaylara sunulan sınav uyarlamaları hakkındaki bilgilerinin yeterli ve doğru olup olmadığının anlaşılması amacıyla, Parıltı Görmeyen Çocuklara Destek Derneği tarafından 2024 YKS'ye girecek adayları sınav uygulamaları hakkında bilgilendirmek üzere yürütülen proje kapsamında gerçekleştirilmiştir.

Araştırma kapsamında 2024 YKS'ye okuyucu görevli eşliğinde gireceğini belirten Türkiye genelinde farklı illerden görme engelli adaylarla çevrim içi şekilde yarı yapılandırılmış görüşmeler gerçekleştirilmiştir.

Görüşmelerde elde edilen veriler adayların; sınava başvuru süreci, okuyucu görevlilerle iletişim, muafiyet, sınav süresi gibi konularda yeterli bilgiye sahip olup olmadıkları açısından Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı (ÖSYM) tarafından paylaşılan bilgiler ve geçmiş yıllardaki sınav uygulamaları göz önünde bulundurularak değerlendirilmiştir.

Elde edilen bulgular adayların, görme engellilere yönelik sınav uygulamaları hakkındaki bilgilerinin yetersiz olduğunu göstermektedir.

Bilgi yetersizliğinin azaltılması amacıyla adayların; yükseköğretime geçiş sürecini yürütme yükümlülüğü kapsamında Yükseköğretim Kurulu ve ÖSYM; üst öğretime geçiş sürecindeki rehberlik yükümlülüğü kapsamında ise Milli Eğitim Bakanlığı tarafından sınavlarda sahip oldukları haklar ve sınav uyarlamaları hakkında daha açık, anlaşılabilir ve detaylı şekilde bilgilendirilmeleri gerektiği sonucuna ulaşılmıştır.

Anahtar Kelimeler: *Görme engelli, sınav uyarlamaları, bilgi düzeyi.*

[1] huzeyfeaydin@sakarya.edu.tr:, ORCID: 0000-0001-6318-7683, Sakarya Üniversitesi Hukuk Fakültesi, Sakarya, Türkiye

[2] tuba.boduroglu@parilti.org.tr:, ORCID: 0000-0003-4371-4532, Parıltı Görmeyen Çocuklara Destek Derneği, İstanbul, Türkiye

[3] aysenur.yazici@parilti.org.tr:, ORCID: 0009-0009-4364-2154, Parıltı Görmeyen Çocuklara Destek Derneği, İstanbul, Türkiye

Knowledge Level of Visually Impaired YKS Candidates About Exam Accommodations: Current Status and İmprovement Recommendations

Huzeyfe Aydın [1], Tuba Boduroğlu [2], Ayşenur Bahçivancıoğlu Yazıcı [3]

It is important for the personal development of people with disabilities that the transition to higher education, which affects people's lives in various ways, is carried out appropriately for people with disabilities.

For this reason, as seen in many other exams, various measures are taken for visually impaired candidates in the Higher Education Institutions Examination (YKS), which is taken to settle in a university.

Visually impaired candidates can benefit from these measures efficiently only if they have accurate and sufficient information about test accommodations.

This research was conducted within the scope of the project carried out by Parıltı Association for Supporting Blind Children to inform the candidates who will take the 2024 YKS about the exam practices during the YKS process in order to understand whether the visually impaired candidates' knowledge about the exam adaptations offered to the candidates is sufficient and accurate.

Within the scope of the research, semi-structured interviews were conducted online with visually impaired candidates from different provinces across Turkey who stated that they will take the 2024 YKS with reader proctors.

The data obtained from the interviews were evaluated by taking into consideration the information shared by the Presidency of Measurement, Selection and Placement Center (ÖSYM) and the exam practices in previous years in terms of whether the candidates have sufficient knowledge about the exam application process, communication with readers, exemption, and exam duration.

The findings obtained show that the candidates' knowledge about the exam practices for the visually impaired is insufficient.

In order to reduce the lack of information, it was concluded that candidates should be informed more clearly, comprehensibly and in detail about their rights in exams and exam adaptations by the Council of Higher Education and ÖSYM within the scope of the obligation to carry out the transition to higher education, and by the Ministry of National Education within the scope of the guidance obligation in the transition to higher education.

Keywords: *Visually impaired, Exam accommodations, Information level*

[1] huzeyfeaydin@sakarya.edu.tr:, ORCID: 0000-0001-6318-7683, Sakarya Üniversitesi Hukuk Fakültesi, Sakarya, Türkiye

[2] tuba.boduroglu@parilti.org.tr:, ORCID: 0000-0003-4371-4532, Parıltı Görmeyen Çocuklara Destek Derneği, İstanbul, Türkiye

[3] aysenur.yazici@parilti.org.tr:, ORCID: 0009-0009-4364-2154, Parıltı Görmeyen Çocuklara Destek Derneği, İstanbul, Türkiye

Uyumsuz Madde Puan Örüntülerinin Belirlenmesinde Kullanılan Tekniklerin Sistematik İncelemesi

Asiye ŞENGÜL AVŞAR [1]

Ölçme araçlarından elde edilen puanların geçerliği, önemli bir psikometrik çalışma konusudur. Geçerlik kanıtlama süreci, çeşitli tekniklerle yürütülür. Özellikle yapı geçerliğinin araştırılmasında ölçme aracının uygulandığı grup verileri üzerinden değerlendirmeler yapılır. Ancak bireysel test puanlarının geçerliği de araştırılması gereken ve geleneksel geçerlik kanıtlama çalışmalarını etkileyen bir konudur (Curtis, 2004; Liu vd., 2019; Sijtsma ve Molenaar, 2002, Şengül Avşar, 2023). Bireysel test puanları geçerli olmayan bireyler, uyumsuz madde puan örüntülerine sahiptir. Bu kişiler madde puan örüntüleri incelenerek belirlenir. Bunun için de kişi-uyum istatistikleri kullanılır. Çok sayıda kişi-uyum istatistikleri geliştirilmiştir. Geleneksel olarak ifade edilebilecek bu istatistiklerin yanı sıra makine öğrenme modelleriyle de uyumsuz madde puan örüntüleri belirlenebilmektedir. Bu araştırmanın problemini uyumsuz madde puan örüntülerinin belirlenmesinde kullanılan tekniklerin incelenmesi oluşturmaktadır. Araştırmanın genel amacı geleneksel ve makine öğrenme algoritmalarına dayalı tekniklerin karşılaştırılmasıdır. Bu amaç doğrultusunda tekniklerin; i) maddelerin puanlanma türüne göre sınıflandırılması, ii) uyumsuz madde puan örüntülerini belirlemede performanslarının değerlendirilmesi araştırmanın alt amaçlarını oluşturmaktadır.

Bu çalışma nitel araştırma yöntemlerinden doküman incelemesi modelinde tasarlanmıştır. Araştırma sonuçlarının hem ulusal çapta test düzenleyen kurumların hem profesyonel araştırmacıların uyumsuz madde puan örüntülerini belirlerken kullanacakları tekniklerin seçiminde karar vermelerine yardımcı olacak bir rehber niteliği taşıması beklenmektedir.

Anahtar Kelimeler: *Uyumsuz madde puanları, Kişi uyum istatistikleri, Makine* öğrenme *algoritmaları.*

Kaynakça

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications.

International Education Journal, *5*(2), 125-143. https://files.eric.ed.gov/fulltext/EJ903843.pdf

Liu, T., Sun, Y., Li, Z., & Xin, T. (2019). The impact of aberrant response on reliability and validity. Measurement: *Interdisciplinary Research and Perspectives*, *17*(3), 133-142.

https://doi.org/10.1080/15366367.2019.1584848

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Sage Publications.

Şengül Avşar, A. (2023) Aberrant individuals' effects on fit indices both of confirmatory factor analysis and polytomous IRT models. *Current Psychology*, *42*(3), 2157-2166. https://doi.org/10.1007/s12144- 021-01563-4

[1] asiye.sengul@erdogan.edu.tr, 0000-0001-5522-2514, Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye

A Systematic Review of Techniques Used to Identify Aberrant Item Score Patterns

Asiye ŞENGÜL AVŞAR [1]

The validity of the scores obtained from measurement tools is an important psychometric study. The validity proving process is carried out with various techniques. In particular, in the investigation of construct validity, evaluations are made on the group data to which the measurement tool is applied. However, the validity of individual test scores is also an issue that needs to be investigated and affects traditional validity proving studies (Curtis, 2004; Liu et al., 2019; Sijtsma and Molenaar, 2002, Şengül Avşar, 2023). Individuals whose individual test scores are not valid have aberrant item score patterns. These individuals are identified by examining item score patterns. Person-fit statistics are used for this purpose. A large number of person-fit statistics have been developed. In addition to these statistics that can be expressed traditionally, aberrant item score patterns can also be determined with machine learning models. The problem of this research is to examine the techniques used in determining aberrant item score patterns. The general aim of the research is to compare techniques based on traditional and machine learning algorithms. For this purpose, the sub-objectives of the research are; i) classification of the techniques according to the scoring type of the items, ii) evaluation of their performance in determining aberrant item score patterns.

This study was designed using the document review model, one of the qualitative research methods. The research results are expected to serve as a guide to help both national test-making institutions and professional researchers decide on the techniques they will use to determine aberrant item score patterns.

Keywords: *Aberrant item scores, Person fit statistics, Machine learning algorithms.*

Reference

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications.

International Education Journal, *5*(2), 125-143. https://files.eric.ed.gov/fulltext/EJ903843.pdf

Liu, T., Sun, Y., Li, Z., & Xin, T. (2019). The impact of aberrant response on reliability and validity. Measurement: *Interdisciplinary Research and Perspectives*, *17*(3), 133-142.

https://doi.org/10.1080/15366367.2019.1584848

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Sage Publications.

Şengül Avşar, A. (2023) Aberrant individuals' effects on fit indices both of confirmatory factor analysis and polytomous IRT models. *Current Psychology*, *42*(3), 2157-2166. https://doi.org/10.1007/s12144- 021-01563-4

[1] asiye.sengul@erdogan.edu.tr, 0000-0001-5522-2514, Recep Tayyip Erdoğan University, Rize, Türkiye

Yükseköğretim Kurumlarında Test Adaletinin İncelenmesine Yönelik Bir Kontrol Listesi Geliştirme Çalışması

Ezgi Mor [1],

Test adaleti, geçerli test sonuçlarına ulaşmada temel bir özellik olarak kabul edilmektedir, çünkü ancak bu sayede öğrenci yeteneğinin herhangi bir yanlılık olmaksızın, doğru bir şekilde belirlenmesi mümkün olur. Testlerde adalet sayesinde öğrencilerin geçmişleri ve kişisel özellikleri ne olursa olsun, iyi performans göstermeleri için tüm öğrenciler eşit fırsata sahip olur ve bu kavram eğitsel ve psikolojik ölçmelerde eşitliği teşvik eder. Eğitimsel ve Psikolojik Test Standartları (AERA, APA, NCME, 2014), testlerde adilliğin temel bir ilke olduğunu ve testlerin ırk, cinsiyet, etnik köken veya dil gibi kişisel özelliklerle ilgili haksız avantaj veya dezavantajlardan arınmış olmasını gerektiğini vurgulamaktadır. Bu durum özellikle farklı geçmişlere sahip öğrenci gruplarının yer aldığı geniş ölçekli sınavlar ve yükseköğretim seviyeleri gibi çok kültürlü eğitim ortamlarında yanlılıklardan arınmış test sonuçları elde etmek için çok daha önemli bir hale gelmektedir. Söz konusu niteliğin belirtilen kritik önemine rağmen, test uygulama ve geliştirme süreçlerinde, özellikle de kültürel çeşitliliğin en yüksek düzeyde olduğu yükseköğretim aşamasında kolaylıkla ihlal edilebileceği pek çok durum vardır. Her çeşit yanlılıktan kaçınmak ve test puanlarına dayalı geçerli ve doğru sonuçlar elde etmek için, test geliştiriciler ve uygulayıcılar tüm test süreci boyunca adalet kavramını göz önünde bulundurmalıdır. Test adaletinin ölçmelerdeki hayati rolüne istinaden, bu çalışma kapsamında araştırmacı, yükseköğretim düzeyindeki test süreçlerinde bu alanda dikkate alınması gereken temel kriterleri belirlemeyi amaçlamıştır. Eğitimsel ve Psikolojik Ölçme Standartları (AERA, APA, NCME, 2014) ile ITC kılavuzları (2013) temel alınarak bu konuda gözetilmesi gereken kriterlerin belirlenmesinin ardından, bu göstergelerin yükseköğretim ortamlarında uygulanabilirliği konusunda uzman görüşleri alınmıştır. Daha sonra bu göstergelere dayalı olarak bir kontrol listesi geliştirilmiş ve ilgili araştırmacıların, test kullanıcılarının, geliştiricilerin ve yükseköğretim düzeyindeki öğretim görevlilerinin kullanımına sunulmuştur. Oluşturulan kontrol listesi, test adaleti ile ilgili önemli noktaları belirten ve yükseköğretim düzeyindeki test süreçlerinde yanlılıkları önlemeye yönelik tedbirleri ortaya koyan bir kılavuz olarak kabul edilebilir. Yapılan ölçme ve değerlendirme etkinliklerinden daha geçerli ve doğru sonuçlar elde etmek için öğretim görevlileri, test geliştiricileri ve test yöneticileri tarafından bir kaynak olarak kullanılabilir.

Anahtar Kelimeler: *Test adaleti, Geçerlik, Yükseköğretim.*

Kaynakça

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing.*

International Test Commission. (2018). ITC guidelines for the large-scale assessment of linguistically and culturally diverse popula-tions (version 4.2). Retrieved from September 21, 2018, https://www.intestcom.org/files/guideline

[1] emor@kastamonu.edu.tr:, ORCID: 0000-0003-0250-327X, Kastamonu Üniversitesi, Kastamonu, Türkiye.

The Checklist Development Study to Investigate the Test Fairness in The Higher Education

Ezgi Mor [1]

The test fairness has been accepted is a central issue in achieving valid test results, because it ensures that assessments accurately measure student ability without any bias. Fairness in testing promotes equity by ensuring that all students, regardless of their backgrounds and characteristics, have an equal opportunity to perform well. The Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) emphasize that fairness is a foundational principle in testing, which requires that tests are free from unfair advantages or disadvantages related to personal characteristics such as race, gender, ethnicity, or language. That is crucial especially in multicultural educational settings such as largescale assessment and higher education levels, in which diverse student populations necessitate careful consideration to avoid bias. Despite the stated importance of this feature in the assessments, there are so many situations in which fairness may be violated in the test administration and development process especially in the higher education level, in which cultural variety is thought at the maximum level. In order to avoid bias and get valid and accurate results based on the test scores, the test developers and users should consider the fairness issues throughout all the testing process. Regarding the central role of fairness in testing, within the context of this study, the researcher aimed to determine the basic criteria that should be taken into consideration in testing processes at higher education level. After the determination of the criteria based on the Standards of (AERA, APA, NCME, 2014) and ITC guidelines (2018), the expert views were gathered about the applicability of these indicators in higher education settings. Then a checklist was developed and presented to the usage of related researchers and test users, developers and lecturers at higher education level. The composed checklist might be accepted as a guideline stating the critical issues related with test fairness and presenting the precautions to prevent biases in testing processes in higher education level. It may be used as a source by lecturers, test developers and test administrators so as to get more valid and accurate results.

Keywords: *Test fairness, Validity, Higher education*

Reference

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing.*

International Test Commission. (2018). ITC guidelines for the large-scale assessment of linguistically and culturally diverse popula-tions (version 4.2). Retrieved from September 21, 2018, https://www.intestcom.org/files/guideline

[1] emor@kastamonu.edu.tr:, ORCID: 0000-0003-0250-327X, Kastamonu University, Kastamonu, Türkiye

Testlerde Adillik ve Test Düzenlemeleri: Bibliyometrik Analiz ve Sistematik İnceleme Çalışması

Sevda ÇETİN [1], Zeynep Neveser KIZILÇİM [2]

Sınavlarda fırsat eşitliği ve adillik, sınava giren tüm alt grupların (cinsiyet, ırk, sosyoekonomik düzey, engel durumu vb.) öğrenme fırsatlarını göz önünde bulundurarak sağlanabilir. *Eğitimde ve psikolojide uygulanan test standartları* adilliği (fairness), tüm test katılımcılarına test süreci boyunca ilgili yapıdaki durumlarını göstermeleri için eşit fırsat sağlanması olarak tanımlamaktadır (AERA vd., 2014). Kapsayıcı değerlendirme, eğitim sisteminde yer alan tüm bireylerin değerlendirme sürecine olabildiğince fazla katılımını sağlamayı amaçlamaktadır (Gong & Marion, 2006). Kapsayıcı değerlendirme, engelli öğrencilerin sınava yalnızca erişimini değil, aynı zamanda sınav öncesinde ve sınav sürecinde bazı düzenlemeleri ve uyarlamaları (accomodation) da içermektedir. Pek çok ülkede bireylerin değerlendirilmesinde, engeli olan öğrencilerin ihtiyaçları göz önüne alınmış ve sınavı erişmelerinde engel olan durumların ortadan kaldırılması için gerekli birtakım önemler alınmıştır. Bu önlemlerden bazıları ekstra süre, kelime işlemci, yazma yardımı, okuma yardımı ve gözetimli dinlenme molaları şeklinde sırlanabilir (Ofqual, 2020).

Bu araştırmada “test fairness” and “test accommodations” or “access arrangements” and “disabilities” anahtar kelimelerinden yararlanılarak, 2000 yılı ve sonrasında İngilizce dilinde yapılmış olan çalışmalar, Wos’ta (Web Of Science) taranmış, PRISMA raporlama yöntemine uygulandıktan sonra ulaşılan 120 makale bibliyometrik ve sistematik analiz yöntemleriyle incelenmiştir.

Araştırma ile “Test uyarlamaları konusunda yapılan çalışmaların yıllara göre dağılımı nasıldır? Bu konuda yapılan çalışmaların yazarların bulundukları ülkelere göre dağılımı nasıldır? Yapılan çalışmalarda en güncel konular nelerdir?” sorularına cevap aranmıştır. Ayrıca sistematik analiz ile örnek çalışmalar üzerinden en etkili yöntemlerin neler olduğu sorusu üzerinde durulmuştur. Bu incelemeler ile eğitim paydaşlarına, alan uzmanlarına ve politika yapıcılara genel görünüm hakkında bilgi sağlamak amaçlamaktadır.

Anahtar Kelimeler: *Test adilliği, sınav düzenlemeleri ve uyarlamaları, engel grupları*

Kaynakça

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing.

Washington, DC: American Educational Research Association.

Gong, B. & Marion, S. (2006). Dealing with flexibility in assessments for students with significant cognitive disabilities (Synthesis Report 60). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Ofqual. (2021). Systematic divergence between teacher and test-based assessment: Literature review.Office of Qualifications and Examinations Regulation Publishers.

[1] tsevda@hacettepe.edu.tr, ORCID: 0000-0001-5483-595X, Hacettepe Üniversitesi, Ankara, Türkiye

[2] Zeynep.kizilcim@ahievran.edu.tr, ORCID: 0000-0002-0164-5682, Kırşehir Ahi Evran Üniversitesi, Kırşehir, Türkiye

Test Fairness and Test Accommodations: A Bibliometric Analysis and Systematic Literature Review

Sevda ÇETİN [1], Zeynep Neveser KIZILÇİM [2]

Equality of opportunity and fairness in testing can be achieved by considering the learning opportunities of all subgroups (gender, race, socioeconomic status, disability, etc.) taking the test. Testing standards in education and psychology define fairness as providing equal opportunity for all test takers to demonstrate their status in the relevant construct throughout the testing process (AERA et al, 2014). Inclusive assessment aims to ensure that all individuals in the education system participate in the assessment process as much as possible (Gong & Marion, 2006). Inclusive assessment includes not only access to testing for students with disabilities, but also accommodations and adaptations before and during the testing process. In many countries, the needs of students with disabilities have been taken into account in the assessment of individuals and some measures have been taken to eliminate the obstacles that prevent them from accessing the exam. Some of these measures include extra time, word processing, writing assistance, reading assistance and supervised rest breaks (Ofqual, 2020).

In this study, using the keywords “test fairness” and “test accommodations” or “access arrangements” and “disabilities”, studies conducted in English in 2000 and later were searched in Wos (Web Of Science), and 120 articles obtained after applying the PRISMA reporting method were examined by bibliometric and systematic analysis methods.

With the research, “How is the distribution of studies on test accommodations according to years? What is the distribution of the studies on this subject according to the countries of the authors? What are the most current topics of the studies?” questions were sought to be answered. In addition, systematic analysis will be used to answer the question of what the most effective methods are through case studies. With these reviews, it is aimed to provide information about the general outlook to education stakeholders, field experts and policy makers.

Keywords: *Test fairness,* test accommodations, *disability groups*

Reference

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing.

Washington, DC: American Educational Research Association.

Gong, B. & Marion, S. (2006). Dealing with flexibility in assessments for students with significant cognitive disabilities (Synthesis Report 60). Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Ofqual. (2021). Systematic divergence between teacher and test-based assessment: Literature review.Office of Qualifications and Examinations Regulation Publishers.

[1] tsevda@hacettepe.edu.tr, ORCID: 0000-0001-5483-595X, Hacettepe Üniversitesi, Ankara, Türkiye

[2] Zeynep.kizilcim@ahievran.edu.tr, ORCID: 0000-0002-0164-5682, Kırşehir Ahi Evran Üniversitesi, Kırşehir, Türkiye

Kayıp Veri Bağlamında Bayesyen Rasch ve Örtük Uzay Madde Tepki Modellerinin Karşılaştırılması[1]

Fatma Gül UZUNER [2], Tuncay ÖĞRETMEN [3]

Kayıp veriler, örtük özelliklerin açıklanmasına yönelik girişimlerde her zaman potansiyel bir problem kaynağı olarak ele alınmıştır. Nitekim geliştirilen örtük özellik modelleri ile kayıp verilere yönelik yapılan çalışmalar her zaman güncelliğini korumaktadır. Rasch model, örtük özelliklerin anlaşılması için geliştirilen Madde Tepki Kuramı (MTK) modellerinden biridir. Rasch modelin varsayımlarının karşılanması noktasında karşılaşılan güçlüklerden dolayı Örtük Uzay Madde Tepki Modeli (ÖUMTM) geliştirilmiştir. ÖUMTM, Rasch modelin genelleştirilmiş bir uzantısı olarak kavramsallaştırılabilir. Kayıp veriler her iki model için de önemli bir problem olarak ele alınabilir. Dolayısıyla iki modelin kayıp veriler bağlamında karşılaştırılması literatüre önemli bir katkı sağlayabilir.

Bu araştırmanın amacı kayıp veri bağlamında Bayesyen Rasch ve ÖUMTM'nin madde ve yetenek parametrelerinin karşılaştırılmasıdır. İlgili karşılaştırma için öncelikle 1-0 yanıt örüntüsüne sahip ve kayıp veri içermeyen simülatif Rasch model verileri üretilmiştir. Bu verilerin örneklem büyüklükleri 100 ve 500 olarak belirlenmiştir. Belirlenen simülatif verilerdeki kayıp veri oranları da %10 ve %20 şeklinde düzenlenmiştir. Bu süreçte yalnızca tamamen rassal olan kayıp veri mekanizması esas alınmıştır. Kayıp veriler noktasında hem tekil hem de çoklu atamaya imkân sağlayan ve esnek bir atama yöntemi olarak değerlendirilen Tahmine Dayalı Ortalama Eşleştirme (TDOE) yöntemi kullanılmıştır. Kayıp veri içermeyen ve kayıp yöntemlerinin kullanıldığı veri setlerinin tümü Rasch model ve ÖUMTM analizine tabi tutulmuştur. Bununla birlikte herhangi bir atama yöntemi kullanılmadan kayıp veri içeren veri setleri de aynı analiz sürecinden geçirilmiştir. Analizler sonucunda elde edilen parametre değerlerinin her iki model için birbirine oldukça yakın olduğuna dair kanıtlara ulaşılmıştır. Ancak elde edilen değerler arasındaki küçük farklar esas alındığında, küçük örneklem büyüklüğündeki kayıp verilerde tekil TDOE, büyük örneklem büyüklüğündeki kayıp verilerde ise çoklu TDOE'nin daha iyi performans gösterdiği görülmüştür.

Anahtar Kelimeler: *Örtük Uzay Madde Tepki Modeli, Rasch model, kayıp veri.*

[1] Bu çalışma Fatma Gül UZUNER'in Eğitimde Ölçme ve Değerlendirme alanındaki doktora tez çalışmasından üretilmiştir.

[2] fgu61@hotmail.com, ORCID: 0000-0001-8483-6021, Kafkas üniversitesi, Kars, Türkiye

[3] togretmen@gmail.com, ORCID: 0000-0001-7783-1409, Ege Üniversitesi, İzmir, Türkiye

Comparison of Bayesian Rasch and Latent Space Item Response Models in the Context of Missing Data[1]

Fatma Gül UZUNER [2], Tuncay ÖĞRETMEN[3]

Missing data has always been considered a potential source of problems in attempts to explain latent traits. As a matter of fact, studies on missing data with the latent trait models developed always remain up-to-date. The Rasch model is one of the Item Response Theory (IRT) models developed to understand latent traits. Due to the difficulties encountered in meeting the assumptions of the Rasch model, the Latent Space Item Response Model (LSIRM) was developed. LSIRM can be conceptualized as a generalized extension of the Rasch model. Missing data can be considered a significant problem for both models. Therefore, comparing the two models in the context of missing data can make an important contribution to the literature.

The aim of this research is to compare the item and ability parameters of Bayesian Rasch and LSIRM in the context of missing data. For the relevant comparison, firstly, simulative Rasch model data with a 1-0 response pattern and no missing data was produced. The sample sizes of these data were determined as 100 and 500. The missing data rates in the determined simulative data are arranged as 10% and 20%. In this process, only the missing data mechanism, which is completely random, was taken as basis. Predictive Mean Matching (PMM) method, which allows both single and multiple imputation and is considered a flexible imputation method, was used for missing data. All data sets that did not contain missing data and where loss methods were used were subjected to Rasch model and LSIRM analysis. In addition, data sets containing missing data were also subjected to the same analysis process without using any imputation method. As a result of the analysis, evidence was found that the parameter values obtained for both models were very close to each other. However, based on the small differences between the obtained values, it was observed that single PMM performed better in missing data with small sample sizes, and multiple PMM performed better in missing data with large sample sizes.

Key Words: *Latent Space Item Response Model, Rasch model, missing data.*

[1] This study was produced from Fatma Gül UZUNER's doctoral thesis in the field of Measurement and Evaluation in Education.

[2] fgu61@hotmail.com, ORCID: 0000-0001-8483-6021, Kafkas University, Kars, Türkiye

[3] togretmen@gmail.com, ORCID: 0000-0001-7783-1409, Ege University, Izmir, Türkiye

Birleştirilmiş Ağaç BBT ve Ağaç BBT performansının Ağacın Büyümesini Kontrol Eden Ana Parametrelere Göre İncelenmesi

Demet ALKAN [1], Nuri Doğan[2]

Birleştirilmiş Ağaç Tabanlı Bilgisayarda Bireyselleştirilmiş Test (Birleştirilmiş Ağaç BBT) yöntemi Karar Ağaçları ve Ağaç Tabanlı Bilgisayarda Bireyselleştirilmiş Test (Ağaç BBT) yöntemlerini önemli ölçüde geliştirerek herhangi bir kişisel bilgisayarda BBT'lerin hızlı bir şekilde oluşturulmasını sağlamaktadır. Bu nedenle önerilen bu yöntem, çok sayıda katılımcıya aynı anda uygulanması gereken testlerin oluşturulması için ideal bir araçtır. Bu araştırmada R programlama dili ile Birleştirilmiş Ağaç Tabanlı BBT ve Ağaç tabanlı BBT performansı karşılaştırılmıştır. Araştırmada RMSE, Yanlılık, Ortalama Test Uzunluğu, gerçek yeteneklerle kestirilen yetenekler arası korelasyon (r), her bir sınav katılımcısının bir sonraki maddeyi alması için geçen süre değerlerinin ortalaması hesaplanmıştır. Test etkililiğinin ve ölçme kesinliğinin ağacın büyümesini kontrol eden ana parametre olan seviye başına maksimum düğüm sayısı, birleştirilecek iki düğümün yoğunluk fonksiyonlarının minimum kesişimi (birleşme parametreleri), madde havuzundaki madde sayısı, maddenin maruz kalma oranına farklı bir ifade ile her bir madde ile uygulanan katılımcı sayısına göre belirlenen çeşitli durumlarda nasıl değiştiği araştırılmıştır. Araştırma gerçek veri ile yapılmıştır. Madde havuzunda 916 bireye uygulanmış 256 maddeye ait cevap örüntüsü bulunmaktadır. Tüm koşullarda gerçeğe en yakın sonuç elde edebilmek için 25 tekrarın ortalaması alınmıştır. Araştırma sonunda madde havuzundaki madde sayısı arttıkça yeteneğin kestirilmesinde Birleştirilmiş Ağaç BBT'lerde daha düşük RMSE değerleri elde edilmiştir. Maddenin maruz kalma oranı arttıkça, seviye başına maksimum düğüm sayısı ve birleştirme parametreleri arttıkça test etkililiği ve ölçme kesinliğinin Birleştirilmiş ağaç BBT de daha yüksek olduğu görülmektedir. Birleştirilmiş ağaç BBT için seviye başına maksimum dal sayısının belirleyici parametre olduğu anlaşılmıştır.

Anahtar Kelimeler: Birleştirilmiş ağaç BBT, Ağaç BBT, Maksimum dal sayısı

Kaynakça

Delgado-G´omez, D., Laria, J. C., and Ruiz-Hern´andez, D. (2019). Computerized adaptive test and decision trees: A unifying approach. Expert Systems with Applications, 117:358–366.

Yan, D., Lewis, C., and Stocking, M. (2004). Adaptive testing with regression trees in the presence of multidimensionality. Journal of Educational and Behavioral Statistics, 29(3):293–316

[1] alkandemet@hotmail.com, ORCID: 0000-0002-1478-9183, Milli Eğitim Bakanlığı, Ankara, Türkiye

[2] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Investigation of Merged Tree CAT and Tree CAT Performance According to the Main Parameters Controlling the Growth of the Tree

Demet ALKAN [1], Nuri Doğan[2]

The Merged Tree-Based Computer Adaptive Testing (Merged Tree-CAT) method significantly improves on the Decision Trees and Tree-Based Computer Adaptive Testing (Tree CAT) methods, enabling the rapid generation of CATs on any personal computer. Therefore, the proposed method is an ideal tool for creating tests to be administered to a large number of participants at the same time. The current study compared the performance of Merged Tree CAT and Tree CAT combined with R programming language. Also, the study calculated RMSE, Bias, Average Test Length, correlation (r) between actual and predicted abilities, and the average time for each test taker to take the next item. Furthermore, the study investigated the main parameters that control the growth of the tree, namely the maximum number of nodes per level, the minimum intersection of the density functions of two nodes to be merged (merging parameters), the number of items in the item pool, the exposure rate of the item, which is the number of participants administered with each item. The research employed real data. The item pool contained response patterns of 256 items administered to 916 individuals. An average of 25 repetitions was taken in all conditions to obtain the closest result to reality. As a result of the study, as the number of items in the item pool increased, lower RMSE values were obtained in the Merged Tree CATs in predicting ability. As the exposure rate of the item increased, the maximum number of nodes per level and the combination parameters increased. Also, the test efficiency and measurement accuracy were higher in the Merged Tree- CAT. Additionally, the maximum number of branches per level was the determining parameter for the Merged Tree CAT.

Keywords: Merged Tree-CAT, Tree CAT, maximum number of branches

Reference

Delgado-G´omez, D., Laria, J. C., and Ruiz-Hern´andez, D. (2019). Computerized adaptive test and decision trees: A unifying approach. Expert Systems with Applications, 117:358–366.

Yan, D., Lewis, C., and Stocking, M. (2004). Adaptive testing with regression trees in thepresence of multidimensionality. Journal of Educational and Behavioral Statistics, 29(3):293–316.

[1] alkandemet@hotmail.com, ORCID: 0000-0002-1478-9183, Milli Eğitim Bakanlığı, Ankara, Türkiye

[2] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Lisansüstü Programlara Başvuruda Kullanılan Sınavların Karşılaştırmalı İncelenmesi

Tansu Alan [1], Seher Yalçın [2]

Lisansüstü eğitim, lisans derecesini tamamlayan bireylerin daha ileri düzeyde uzmanlaşmak, akademik veya profesyonel meslek alanlarında derinlemesine bilgi edinmek amacıyla aldıkları eğitimdir. Türkiye'de lisansüstü programlara başvurularda yaygın olarak Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) kullanılırken, uluslararası düzeyde Graduate Record Examination -GRE ve Graduate Management Admission Test -GMAT sınavları tercih edilmektedir. ALES, GMAT ve GRE, akademik ve mesleki başarıyı ölçmek için kullanılan önemli sınavlar olmasından dolayı bu sınavların kullanım amacı, alt testleri, ölçeklenme şekli, psikometrik özellikleri, test güvenliği, test kuramı ve uygulanma şekli bakımından karşılaştırılmasını amaçlayan bu araştırma nitel araştırma yöntemlerinden olan durum çalışması ile yürütülmüştür. Bu kapsamda alan yazın incelemesi yapılmıştır. Elde edilen verilerin çözümlenmesinde karşılaştırma analizi kullanılmış, analiz sonuçları karşılaştırmalı tablolar halinde verilmiştir. Araştırma sonucunda ALES'in GRE ve GMAT ile benzer becerileri ölçen benzer alt testlerden oluştuğu ve benzer güvenirlik yöntemlerinin kullanıldığı sonucuna ulaşılmıştır. Bunun yanı sıra ALES'in uygulanma şekli ve temel aldığı test kuramının GRE ve GMAT'ten farklı olduğu sonucuna ulaşılmıştır. Ayrıca GRE ve GMAT'ten alınan puanların geçerliği için test eşitleme, ölçme değişmezliği ve yanlılık çalışmaları yapılırken ALES'te bu çalışmaların yapıldığına yönelik bir rapora ulaşılamamıştır. Sadece bazı lisansüstü tezlerde bu tür çalışmaların yapıldığı görülmektedir. Bundan dolayı ALES'in GRE ve GMAT gibi test standartlarını yakalayabilmesi için test geliştirme ve iyileştirme süreçleri içerisinde test eşitleme ve yanlılık çalışmalarının yürütülmesi, madde tepki kuramı temelinde bilgisayar ortamında bireye uyarlanmış test olarak uygulanması çalışmalarının denenmesi önerilmektedir. Bu durum, ALES'in daha geniş kabul gören uluslararası düzeyde rekabet edebilir bir sınav haline gelmesini destekleyebilir, adayların dana geçerli ve güvenilir değerlendirilmesine ve adayların akademik ve profesyonel kariyerlerinde başarılı olmalarına katkıda bulunabilir.

Anahtar Kelimeler: *Lisansüstü eğitim, ALES, Test Geçerliği*

[1] tansualan@adiyaman.edu.tr, ORCID: 0000-0001-5855-0302, Adıyaman Üniversitesi, ADIYAMAN, TÜRKİYE

[2] yalcins@ankara.edu.tr, ORCID: 0000-0003-0177-6727, Ankara Üniversitesi, ANKARA, TÜRKİYE

Comparative Examination of Exams Used for Graduate Program Applications

Tansu Alan [1], Seher Yalçın [2]

Graduate education refers to the training individuals receive after completing their undergraduate degree, aimed at achieving advanced specialization and in-depth knowledge in academic or professional fields. In Turkey, the Academic Staff and Graduate Education Entrance Exam (ALES) is commonly used for applications to graduate programs, while internationally, the Graduate Record Examination (GRE) and the Graduate Management Admission Test (GMAT) are preferred. This research aims to conduct a comparative analysis of ALES, GRE, and GMAT—important exams used to measure academic and professional success—regarding their purposes, sub-tests, scaling methods, psychometric properties, test security, test theory, and administration methods through a qualitative case study approach. A literature review was conducted as part of this scope. Comparative analysis was employed in the data analysis, and the results were presented in comparative tables. The findings indicate that ALES consists of similar sub-tests measuring comparable skills to those in GRE and GMAT, using similar reliability methods. Additionally, it was found that the application method of ALES and the test theory it is based on differ from those of GRE and GMAT. Furthermore, while studies on test equating, measurement invariance, and bias are conducted for GRE and GMAT scores, no reports were found indicating that such studies have been performed for ALES. Therefore, to enable ALES to meet the testing standards of GRE and GMAT, it is recommended to conduct test equating and bias studies within the test development and improvement processes, and to experiment with computer-based adaptive testing based on Item Response Theory. This could support ALES in becoming a more widely accepted and competitive examination at the international level, contributing to more valid and reliable assessments for candidates and aiding their success in academic and professional careers.

Keywords: *Graduate education, ALES, Test validity*

[1] tansualan@adiyaman.edu.tr, ORCID: 0000-0001-5855-0302, Adıyaman Üniversitesi, ADIYAMAN, TÜRKİYE

[2] yalcins@ankara.edu.tr, ORCID: 0000-0003-0177-6727, Ankara Üniversitesi, ANKARA, TÜRKİYE

Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) Başarısını Makine Öğrenmesi Algoritmalarıyla Yordama

Ayşegül Bozdağ Kasap[1], Dilara Bakan Kalaycıoğlu[2]

Akademik personel ve lisansüstü eğitime geçiş sürecinde bireylerin yeteneklerinin belgelemesinde kullanılan ALES, geniş ölçekte uygulandığı için büyük veri yapısına sahiptir. Bu sınavlardaki başarı durumlarının, veriden anlamlı sonuçlar elde etmek amacıyla makine öğrenmesi algoritmaları kullanılarak incelenmesi ve başarıyı yordamada değişkenlerin önem düzeylerinin belirlenmesi, araştırmanın amacını oluşturmaktadır. Bu doğrultuda çalışma, değişkenler arasındaki ilişkilerin belirlendiği korelasyonel araştırma desenindedir (Fraenkel, Wallen & Hyun, 2012). ALES'te Sayısal, Sözel ve Eşit Ağırlık olmak üzere üç puan türü bulunduğundan, analizler puan türüne göre ayrı ayrı yapılmıştır. Bu kapsamda, ALES başarısını yordamada rastgele orman, destek vektör makineleri, K-en yakın komşu ve yapay sinir ağları gibi makine öğrenmesi yöntemleri kullanılarak bireylerin akademik not ortalaması, ortaöğretim başarı puanı, lise, üniversite, fakülte, lisans programı, lisans alanı, Yükseköğretime Geçiş Sınavı (YGS), Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS) ve Kamu Personeli Seçme Sınavı (KPSS) gibi değişkenlerin önem düzeyleri incelenmiştir. Ayrıca, araştırmada kullanılan yöntemlerinin ortalama mutlak hata ve korelasyon katsayısı değerleri karşılaştırılarak, modellerin ALES başarısını yordama performansları değerlendirilmiştir. Birey ve toplum için sonuçları önemli olan ALES başarısını yordamada modeller arasında KPSS genel yetenek testinin önem düzeyi en yüksektir. YGS matematik testi ALES Sayısal başarısını, YGS Türkçe testi ALES Sözel başarısını yordamada önemli değişkenler olmakla birlikte ALES Eşit Ağırlık puan türünde her iki testin de önem düzeyi yüksektir. Araştırma sonuçları, KPSS ve YGS'nin ALES başarısını yordama performansının yüksek olduğunu göstermiştir. Benzer şekilde, Arıkan ve D'Costa (2016) çalışmasında üniversiteye giriş sınavı puanlarıyla lisansüstü eğitimi sınavı puanlarının yordanabildiği sonucuna ulaşmıştır. Ayrıca, Hsu ve Schombert (2010) GRE ile SAT puanları arasındaki korelasyonun yüksek olduğu ve genel bilişsel yeteneğin sabit olduğu görüşünü savunmaktadır. Makine öğrenmesi yöntemlerinin performansları incelendiğinde ise ALES başarısını yordamada en iyi performans gösteren modelin yapay sinir ağları olduğu görülmüştür.

Anahtar Kelimeler: *Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı, Makine* öğrenmesi

Kaynakça

Fraenkel, J.R., Wallen, N.E. & Hyun, H.H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Arıkan, S., & D'Costa, A. (2016). ÖSS ve LES ile Ölçülen Sayısal ve Sözel Beceriler Arasındaki İlişki. *Mersin* Üniversitesi *Eğitim Fakültesi Dergisi, 12*(1), 303-313. https://doi.org/10.17860/efd.71302

Hsu, S. D. H. & Schombert, J. (2010). Data Mining the University: College GPA Predictions from SAT Scores. *SSRN Electronic Journal,* 1-27. https://doi.org/10.48550/arXiv.1004.2731

[1] bozdagaysegul@yandex.com, ORCID: ID: 0000-0002-5333-8617, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, Türkiye

[2] dilarabakan@gmail.com, ORCID: 0000-0003-1447-6918, Gazi Üniversitesi, Ankara, Türkiye

Predicting Academic Personnel and Postgraduate Education Entrance Exam (ALES) Success with Machine Learning Algorithms

Ayşegül Bozdağ Kasap[1], Dilara Bakan Kalaycıoğlu[2]

The Academic Personnel and Postgraduate Education Entrance Exam (ALES), widely used to assess individuals' abilities during the transition to postgraduate education, features a large dataset. This study aims to analyze the success rates in these exams by applying machine learning algorithms to derive meaningful insights and identify the importance levels of variables in predicting success. The study follows a correlational research design that investigates the relationships between variables (Fraenkel, Wallen & Hyun, 2012). ALES includes three types of scores—quantitative, verbal, and equally weighted— so the analyses were conducted separately for each type. Machine learning methods such as random forest, support vector machines, k-nearest neighbors, and artificial neural networks were employed to evaluate the importance of variables including academic grade point average, secondary education success score, high school, university, faculty, undergraduate program, field of study, university entrance exam (YGS), foreign language exam (YDS), and public personnel selection examination (KPSS) in predicting ALES success. The performance of these models was compared using mean absolute error and correlation coefficient metrics. Among the variables, the KPSS general ability test had the highest importance in predicting overall ALES success. Specifically, the YGS mathematics test was significant in predicting ALES quantitative scores, while the YGS Turkish test was crucial for predicting ALES verbal scores. For the equally weighted score type, both tests were highly important. These findings demonstrate that KPSS and YGS exams have high predictive performance for ALES. Similarly, Arıkan and D'Costa (2016) found that university entrance exam scores could predict graduate education exam performance. Hsu and Schombert (2010) also reported a high correlation between GRE and SAT scores, suggesting that general cognitive ability remains stable over time. When evaluating the performance of machine learning methods, artificial neural networks emerged as the most effective model for predicting ALES success.

Keywords: *Academic Personnel and Postgraduate Education Entrance Exam, Machine Learning*

Reference

Fraenkel, J.R., Wallen, N.E. & Hyun, H.H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Arıkan, S., & D'Costa, A. (2016). ÖSS ve LES ile Ölçülen Sayısal ve Sözel Beceriler Arasındaki İlişki. *Mersin* Üniversitesi *Eğitim Fakültesi Dergisi, 12*(1), 303-313. https://doi.org/10.17860/efd.71302

Hsu, S. D. H. & Schombert, J. (2010). Data Mining the University: College GPA Predictions from SAT Scores. *SSRN Electronic Journal,* 1-27. https://doi.org/10.48550/arXiv.1004.2731

[1] bozdagaysegul@yandex.com, ORCID: ID: 0000-0002-5333-8617, Measurement, Selection and Placement Center, Ankara, Türkiye

[2] dilarabakan@gmail.com, ORCID: 0000-0003-1447-6918, Gazi University, Ankara, Türkiye

Afet Dönemlerinde e-Sınav Uygulaması için AUZEF Örneği

Zeynep KIROĞLU[1], Selma TOSUN [2], Metin ÜNVER[3]

Yıllar boyunca bilgi iletişim teknolojilerindeki gelişmeler hayatımızın her alanını etkilediği gibi eğitim öğretim sisteminde de büyük değişimleri beraberinde getirmiştir. Özellikle son yıllarda zamandan ve mekandan bağımsız olarak yürütülen online eğitim öğretim platformlarının yaygınlaşması ile bireylerin farklı branşlarda kişisel gelişimlerini destekleyen bilgi ve öğretici eğitim almaları kolaylaşmıştır. Bununla birlikte özellikle çalışanlar ve diğer dezavantajlı gruplar için yükseköğretimde açık ve uzaktan öğrenmeye talep de artmıştır. Pandemi ve doğal afet gibi beklenmedik olağanüstü durumların yaşandığı dönemlerde ise yüz yüze gerçekleştirilemeyen eğitim süreçlerinin, bilgi iletişim teknolojileri kullanılarak dijital platformlarda çevrimiçi olarak uygulanması ve sınavların bu platformlarda gerçekleştirilmesi kaçınılmaz olmuştur. 6 Şubat 2023 tarihinde gerçekleşen deprem nedeniyle Türkiye'de eğitim öğretim süreçlerinin yürütülmesine yönelik kararlar alınmış ve mümkün olduğunca süreçler dijital platformlara taşınmıştır. Ders materyallerinin (kitap, video vb.) öğrencilere bu platformlardan sunulması sağlanmıştır. Ancak dijital eğitimin sonucunda süreçlerin değerlendirilmesi ve sınavların yapılmasında klasik yöntemlerin uygulanması bir problem olarak ortaya çıkmaktadır. Şubat 2023 tarihinde Yükseköğretim Kurulu tarafından alınan bir karar ile yükseköğretimde sınavların dijital uygulanmasının önü açılmış; hatta birçok kurumda dijital sınav uygulamaları zorunlu hale gelmiştir. Sonuç olarak pek çok üniversitede sınıflarda yapılan gözetimli kâğıt kalem sınavlarına alternatif olarak farklı sınav uygulamaları için çalışmalar yürütülmüştür. Bu araştırma kapsamında İstanbul Üniversitesi açık ve uzaktan öğrenme programlarındaki elektronik sınav uygulaması incelenmiştir. Araştırma, nitel araştırma desenlerinden vaka analizi türünde olup; bu bağlamda araştırmanın kapsamını İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi (AUZEF) 2022-2023 bahar dönemi bitirme sınavı uygulaması oluşturmaktadır. İlgili döneme ait 50 lisans ve önlisans programında kayıtlı yaklaşık 400.000 bin öğrenci için hazırlanmış e-sınav süreçleri incelenmiştir. Araştırma sonucunda büyük öğrenci kitlelerine yönelik gerçekleştirilen çevrimiçi sınavların örneklem bağlamında soru havuzu ve test nitelikleri ile sınav uygulama kuralları ele alınmıştır. Bunun yanında çevrimiçi sınav uygulamalarında geçerlik ve güvenirlik tartışmaları ve buna yönelik alınan tedbirler ortaya çıkartılmıştır.

Anahtar Kelimeler: *yükseköğretimde e-sınav, açık ve uzaktan* öğrenme, çevrimiçi *sınav*

[1] zeynepacarer@istanbul.edu.tr, 0009-0006-4253-4280, İstanbul Üniversitesi, İstanbul, Türkiye

[2] selma.tosun@istanbul.edu.tr, 0000-0001-7444-7903, İstanbul Üniversitesi, İstanbul, Türkiye

[3] munver@istanbul.edu.tr, ORCID: 0000-0001-7417-3404, İstanbul Üniversitesi, İstanbul, Türkiye

AUZEF Example for e-Exam Application in Disaster Periods

Zeynep KIROĞLU[1], Selma TOSUN [2], Metin ÜNVER[3]

Over the years, developments in information and communication technologies have brought great changes in the education and training system as well as affecting every aspect of our lives. Especially in recent years, with the widespread use of online education and training platforms that are independent of time and space, it has become easier for individuals to receive training that supports their personal development in different branches. However, the demand for open and distance learning in higher education has also increased, especially for employees and other disadvantaged groups. In times of unexpected extraordinary situations such as pandemics and natural disasters, it has become inevitable that education processes that cannot be carried out face-to-face are implemented online on digital platforms using information communication technologies and exams are held on these platforms. On 6 February 2023, due to the earthquake, decisions were taken to carry out education and training processes in Turkey and the processes were moved to digital platforms as much as possible. Course materials (books, videos, etc.) were made available to students on these platforms. However, as a result of digital education, the application of classical methods in evaluating the processes and conducting exams emerges as a problem. With a decision taken by the Council of Higher Education in February 2023, the digital application of exams in higher education has been paved the way; in fact, digital exam applications have become compulsory in many institutions. As a result, studies have been carried out in many universities for different exam applications as an alternative to supervised paper and pencil exams in classrooms. Within the scope of this research, the electronic exam application in Istanbul University open and distance learning programs was examined. The research is a case study type of qualitative research design; in this context, the scope of the research consists of Istanbul University Open and Distance Education Faculty (AUZEF) 2022-2023 spring term graduation exam application. The e-exam processes prepared for approximately 400,000 thousand students enrolled in 50 pre-graduate and undergraduate degree programs for the relevant period were examined. As a result of the research, question pool and test qualifications and exam application rules were discussed in the context of the sample of online exams for large student masses. In addition, validity and reliability discussions in online exam applications and the measures taken for this purpose have been revealed.

Keywords: *e-exam in higher education, open and distance learning, online exam*

[1] zeynepacarer@istanbul.edu.tr, 0009-0006-4253-4280, İstanbul Üniversitesi, İstanbul, Türkiye

[2] selma.tosun@istanbul.edu.tr, 0000-0001-7444-7903, İstanbul Üniversitesi, İstanbul, Türkiye

[3] munver@istanbul.edu.tr, ORCID: 0000-0001-7417-3404, İstanbul Üniversitesi, İstanbul, Türkiye

Aday Gözünden Çevrimiçi Sınavların Kullanılabilirliği: Göz Hareketleri Izleme Çalışması

Serkan Alkan[1]

Bu çalışmanın amacı standardize testlere hazırlanan adayların hazırlık yapmaları için Internet'te serbest olarak erişim olanağı sunulan örnek test sayfalarının kullanılabilirliğinin karşılaştırılmasıdır. Günümüzde birçok örneği bulunmasına rağmen, uygulanabilirlik açısından çalışma, e-YDS ve TOEFL-IBT sınavlarının örnek oturumlarının gerçek kullanıcılar ile test edilmesi hedefi ile kısıtlı tutulmuştur. Çalışmada, daha önceden örnek test sayfalarını tecrübe etmemiş katılımcıların web sitesini kullanım davranışları göz hareketlerini izleme cihazı ile kaydedilecek, hemen ardından web sitesini kullanım davranışları sesli düşünme tekniği ile katılımcıların kendileri tarafından değerlendirilecektir. Davranışsal boyutu tanımlayan *fixation* ve *saccade* gibi tipik göz hareketi metriklerinin yanında, sonuçlar kullanılabilirlik testleri ile birlikte yaygın olarak kullanılan kullanılabilirlik tutum ölçekleri ile birlikte desteklenerek sunulacaktır. Çalışma ile amaçlanan hedef, hem sınavlara katılan toplam kişi sayısı dikkate alındığında hem de adayların daha sonraki yaşamları için önemli bir ölçüt teşkil eden bu sınavların kullanılabilirliği ve erişilebilirliği açısından yapılabilecek en ufak iyileştirmenin etkisi, değişikliğin kendisinden çok yüksek olacaktır.

Anahtar Kelimeler: *kullanılabilirlik, göz hareketleri, erişilebilirlik,* çevrimiçi *sınavlar*

Kaynakça

Erdinc, Oguzhan. (2015). Kullanılabilirlik Çalışmaları için Türkçe Computer System Usability Questionnaire (CSUQ) Anketi. 10.13140/RG.2.1.4035.3129.

ETS (n.d) Sample the TOEFL iBT Test and Unlock Your English Proficiency. Retrieved July 17, 2024, fromhttps://www.ets.org/forms/toefl-sampler-forms/test-taker.html

ÖSYM (n.d) YDS: Yabancı Dil Bilgisi Seviye Tespit Sınavı. Retrieved July 17, 2024, from https://www.osym.gov.tr/TR,10359/e-yds-ornek-sinav.html

[1] serkana@metu.edu.tr:, ORCID: 0000-0003-1243-4570, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

Usability of Online Exams from The Candidate's Eye: An Eye Tracking Study

Serkan Alkan[1]

The aim of this study is to compare the usability of free online sample tests that are accessible on the Internet for the candidates who are preparing for standardized tests. Although there are many examples today, current study was limited to test the sample sessions of e-YDS and TOEFL-IBT exams with real users. During the study, the website usage behavior of the participants who have not previously experienced the sample test pages will be recorded with an eye tracking device. Later, the behaviors on the website will be evaluated by the participants themselves using the think-aloud technique. In addition to typical eye movement metrics such as fixation and saccade that define the behavioral dimension, the results will be supported by usability questionnaires that are commonly used to cover attitudes of participants. The aim of the study is that, -considering both the total number of people who participate to the exams and to be an important criterion for them-, any slightest improvement at the usability or accessibility on these exams will constitute much higher impact than the change itself.

Keywords: *usability, eye tracking, accessibility, online exams*

Reference

Erdinc, Oguzhan. (2015). Kullanılabilirlik Çalışmaları için Türkçe Computer System Usability Questionnaire (CSUQ) Anketi. 10.13140/RG.2.1.4035.3129.

ETS (n.d) Sample the TOEFL iBT Test and Unlock Your English Proficiency. Retrieved July 17, 2024, fromhttps://www.ets.org/forms/toefl-sampler-forms/test-taker.html

ÖSYM (n.d) YDS: Yabancı Dil Bilgisi Seviye Tespit Sınavı. Retrieved July 17, 2024, from https://www.osym.gov.tr/TR,10359/e-yds-ornek-sinav.html

[1] serkana@metu.edu.tr:, ORCID: 0000-0003-1243-4570, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

Performans Değerlendirmede Puanlayıcılar Arası Uyum İle Göz İzleme Sonuçları Arasındaki İlişkinin İncelenmesi

Mine DEMİRBAŞ [1], Sümeyra SOYSAL[2], Nuri DOĞAN[3]

Bu araştırmada performans değerlendirme sürecinde puanlayıcılar arası uyum, uyum katsayıları, genellenebilirlik ve çok yüzeyli Rasch ile araştırılmıştır. Bu amaç doğrultusunda öğrencilerin sunum yapma performansını değerlendirmek üzere aynı puanlayıcıların sınıf içi performans esnasında verdikleri puanlar ve göz izleme tekniği ile ekran üzerinden verdikleri puanlar arasındaki ilişkiler incelenmiştir. Sınıf içi değerlendirme ve göz izleme tekniği ile ekran üzerinden yapılan değerlendirme sonucu elde edilen puanlar, klasik test kuramında yer alan puanlayıcılar arası uyum istatistikleri, genellenebilirlik kuramı ve çok yüzeyli Rasch modeline göre analiz edilmiştir. Araştırma sonucunda, idealizm ve realizm sunumları için puanlayıcılar arası uyumun iyi olduğu bulunmuştur. Genellenebilirlik analizi sonucuna göre en büyük varyans kaynağı sunum olarak belirtilmiştir. Çok yüzeyli Rasch analizi sonucu puanlamaların birbiriyle tutarlı olduğu, puanlamaların sunu ve maddelere göre anlamlı bir farklılık gösterdiği ancak uygulamaya göre anlamlı bir farklılık göstermediği sonucuna ulaşılmıştır. Göz izleme metriklerinden ortalama sabitlenme süreleri çok yüzeyli Rasch analiziyle incelenmiş ve puanlayıcıların ortalama sabitlenme sürelerine göre anlamlı bir farklılık göstermediği bulunmuştur Sonuç olarak, genellenebilirlik ve çok yüzeyli Rasch sonuçlarının birbiriyle uyumlu olduğu gözlemlenmiştir. Bu doğrultuda göz izleme metrislerinin, puanlayıcılar arası uyumu araştırmak için alternatif bir yöntem olarak kullanılabileceği söylenebilir.

Anahtar Kelimeler: *Göz izleme tekniği, Puanlayıcılar arası uyum, Genellenebilirlik kuramı, Çok Yüzeyli Rasch Modeli, Puanlayıcılar arası güvenirlik, Performans değerlendirme*

[1] zorlumine@gmail.com, ORCID: 0000-0003-2568-9038, Hakkari Üniversitesi, Hakkari, Türkiye

[2] sumeyrasoysal@hotmail.com, ORCID: 0000-0002-7304-1722, Necmettin Erbakan Üniversitesi, Konya, Türkiye

[3] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Investigation Of Interrater Agreement And Eye Tracking in Performance Evaluation

Mine DEMİRBAŞ[1], Sümeyra SOYSAL[2], Nuri DOĞAN[3]

In this study, the fit between raters, coefficients of fit, generalizability and many facet Rasch were investigated in the performance evaluation process. For this purpose, to evaluate the presentation performance of the students, the relationships between the scores given by the same raters in the classroom during the performance and the scores given on the screen by eye tracking technique were examined. The scores obtained as a result of the in-class evaluation and the evaluation on the screen with the eye tracking technique were analyzed according to the inter-rater fit statistics in the classical test theory, the generalizability theory and the many facet Rasch model. It was found that there was a good agreement between the raters in the presentations of idealism and realism. According to the results of the generalizability analysis, the biggest source of variance was the presentation. As a result of the many facet Rasch model, it was found that the scores were consistent with each other and the scores differed significantly according to the presentation and the items, but did not differ significantly according to the application. The mean fixation times of eye tracking metrics were examined by many facet Rasch model analysis and it was found that the raters did not show a significant difference according to the mean fixation times. As a result, it was found that generalizability and multi-surface Rasch model results were compatible with each other. The study indicated that it could be used as an alternative method to investigate the compatibility between eye tracking metrics and raters.

Keywords: *Eye tracking, Inter-Rater agreement, Generalizability theory, Many facet rasch model, Inter-Rater reliability, Performance assessment*

[1] zorlumine@gmail.com, ORCID: 0000-0003-2568-9038, Hakkari Üniversitesi, Hakkari, Türkiye

[2] sumeyrasoysal@hotmail.com, ORCID: 0000-0002-7304-1722, Necmettin Erbakan Üniversitesi, Konya, Türkiye

[3] nuridogan2004@gmail.com, ORCID: 0000-0001-6274-2016, Hacettepe Üniversitesi, Ankara, Türkiye

Dikkat Dağınıklığı Olan İlkokul Öğrencisinin Okuma Esnasındaki Göz Hareketlerinin Yapay Zeka Destekli Uygulamayla Değerlendirilmesi

Kemal Faruk BAKIR [1] Emrullah AKCAN [2]

Teknolojinin eğitim üzerindeki etkisi düşünüldüğünde ciddi yeniliklerin yapıldığı görülmektedir (Kafadar, 2022). Bu yeniliklerin başında yapay zekâ gelmektedir. Nitekim yapay zekâ, eğitimde dönüştürücü bir güç olarak ortaya çıkmıştır. Eğitimciler, yapay zekânın gücünden yararlanarak çok miktarda öğrenci verisini analiz edebilmekte, bireysel öğrenme ihtiyaçlarına ilişkin öngörüler sağlayabilmektedir (Çavuş, 2024). Bu durum özellikle ölçme ve değerlendirmenin önemine ve dolayısıyla teknolojik gelişmelerle birlikte yapay zekânın ölçme ve değerlendirmede kullanımının öğretme ve öğrenmedeki kritik rolüne işaret etmektedir. Araştırmalar, yapay zekâ tabanlı ölçme değerlendirmeye odaklanan yeni nesil ölçme-değerlendirme araçlarına/yöntemlerine ihtiyaç duyduklarını göstermektedir (Mede & Atay, 2017; Tsagari, 2011). Bu çalışmada da, dikkat dağınıklığı olan ilkokul öğrencisinin okuma esnasındaki göz hareketlerinin yapay zeka destekli uygulamayla değerlendirilmesi amaçlanmıştır. Bu amaç doğrultusunda çalışmada nitel araştırma yöntemlerinden biri olan eylem araştırması kullanılmıştır. Araştırmacılar tarafından uygulamalar titizlikle yapılmış, uygulama sonrası öğrenciyle görüşme gerçekleştirilmiş ve elde edilen veriler yorumlanmıştır. Araştırmadan beklenen sonucun ise dikkat eksikliği olan ilkokul öğrencisinin okuma esnasındaki göz hareketlerinin yapay zeka destekli uygulamayla değerlendirilmesinin işlevsel ve çok boyutlu veriler sunacağı yönündedir.

Anahtar Kelimeler: *Okumanın Değerlendirilmesi, Göz Hareketleri, Yapay Zeka.*

Kaynakça

Çavuş, M. N. (2024). Eğitimde Yapay Zekâ Tabanlı Ölçme ve Değerlendirme Üzerine Bir Derleme.

Uluslararası Özel Amaçlar için İngilizce *Dergisi, 2*(1), 39-54.

Kafadar, T. (2022). Oyunlaştırmanın eğitimdeki yeri. In T. Kafadar, ve C. A. Asena (Eds.), *Eğitimde oyunlaştırma,* (1-16). Nobel Akademik Yayıncılık.

Mede, E., ve Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi, 168*(1), 43-60.

Tsagari, D. (2011). Investigating the 'assessment literacy' of EFL state school teachers in Greece. In D. Tsagari, and I. Csépes (eds.), Classroom-based language assessment, (169-190).

Yıldırım, A. ve Şimşek, H. (2005). *Sosyal bilimlerde nitel araştırma yöntemleri.* Birinci baskı içinden. Seçkin Yayınları.

[1] kemalfarukbakir@gmail.com:, ORCID: https://orcid.org/0000-0002-4441-7209, Milli Eğitim Bakanlığı, Gaziantep, Türkiye

[2] emrullahakcan@gmail.com:, ORCID: https://orcid.org/0000-0002-5492-4159,Gaziantep Üniversitesi, Gaziantep, Türkiye

Evaluation of Eye Movements of Primary School Students with Attention Disorders During Reading with Artificial Intelligence Supported Application

Kemal Faruk BAKIR [1], Emrullah AKCAN[2]

Considering the impact of technology on education, it is seen that serious innovations have been made (Kafadar, 2022). Artificial intelligence comes first among these innovations. As a matter of fact, artificial intelligence has emerged as a transformative force in education. By taking advantage of the power of artificial intelligence, educators can analyze a large amount of student data and provide insights into individual learning needs (Çavuş, 2024). This situation especially points to the importance of measurement and evaluation and therefore the critical role of the use of artificial intelligence in measurement and evaluation in teaching and learning along with technological developments. Studies show that they need new generation measurement-evaluation tools/methods that focus on artificial intelligence-based measurement and evaluation (Mede & Atay, 2017; Tsagari, 2011). In this study, it was aimed to evaluate the eye movements of a primary school student with attention deficit during reading with an artificial intelligence-supported application. For this purpose, action research, one of the qualitative research methods, was used in the study. The applications were meticulously carried out by the researchers, an interview was conducted with the student after the application and the obtained data were interpreted. The expected result of the research is that the evaluation of eye movements of primary school students with attention deficit while reading with an artificial intelligence-supported application will provide functional and multi-dimensional data.

Keywords: *Reading Assessment, Eye Movements, Artificial Intelligence.*

Reference

Çavuş, M. N. (2024). Eğitimde Yapay Zekâ Tabanlı Ölçme ve Değerlendirme Üzerine Bir Derleme.

Uluslararası Özel Amaçlar için İngilizce *Dergisi, 2*(1), 39-54.

Kafadar, T. (2022). Oyunlaştırmanın eğitimdeki yeri. In T. Kafadar, ve C. A. Asena (Eds.), *Eğitimde oyunlaştırma,* (1-16). Nobel Akademik Yayıncılık.

Mede, E., ve Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi, 168*(1), 43-60.

Tsagari, D. (2011). Investigating the 'assessment literacy' of EFL state school teachers in Greece. In D. Tsagari, and I. Csépes (eds.), Classroom-based language assessment, (169-190).

Yıldırım, A. ve Şimşek, H. (2005). *Sosyal bilimlerde nitel araştırma yöntemleri.* Birinci baskı içinden. Seçkin Yayınları.

[1] kemalfarukbakir@gmail.com:, ORCID: https://orcid.org/0000-0002-4441-7209, Milli Eğitim Bakanlığı, Gaziantep, Türkiye

[2] emrullahakcan@gmail.com:, ORCID: https://orcid.org/0000-0002-5492-4159,Gaziantep Üniversitesi, Gaziantep, Türkiye

Test Geliştirmede Paradigma Değişimi: Yapay Zekâ Destekli Çeldiricilerin Etkililiği

Mehriban Ceylan [1], Semirhan Gökçe [2]

Bu çalışmanın amacı çoktan seçmeli testlerde yapay zeka (YZ) destekli çeldiricilerin etkililiğini inceleyerek bu yeniliğin test tasarımında nasıl bir paradigma değişikliğine yol açabileceğini araştırmaktır. Geleneksel test tasarımı, çeldiricileri oluşturmak için insan uzmanlığına dayanır, ancak bu süreç önyargılara açık olup değerlendirmelerin güvenilirliği ve geçerliliğini tehlikeye atabilir. Geleneksel madde yazımı genellikle madde kalitesini sağlamak için kapsamlı revizyonlar gerektirmektedir (Downing, 2005). YZ odaklı yaklaşımlar ise farklı ve bağlama uygun çeldiriciler üreterek bu sorunların üstesinden gelmeye çalışmaktadır (Gierl, Lai ve Turner, 2012). YZ'nın yalnızca ölçülen bilişsel süreçlerle de uyumlu çeldiriciler üretebileceğini göstermektedir (Gierl vd., 2017). YZ destekli çeldiricilere geçiş, test geliştirme verimliliğini artırsa da şeffaflık ve hesap verebilirlik gibi zorlukları beraberinde getirebilir (Williamson, 2019). Ancak, bu çeldiriciler verimlilik, psikometrik kalite ve uyarlanabilirlik açısından test geliştirmede dönüştürücü bir değişim potansiyeline sahiptir. Araştırma kapsamında açıklanan Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) soruları kullanılarak insan uzmanlığına dayalı seçenekleri ile YZ tarafından oluşturulan seçeneklerinin etkililiği incelenecektir. Çeldiricilerin etkililiği TAP programı kullanılarak yürütülecek madde ve seçenek analizi ile karşılaştırılacaktır. Araştırmanın ön bulguları, YZ'nın eğitimdeki potansiyelini vurgularken, insan uzmanlığına dayalı kontrol sürecinin önemini de ortaya koymaktadır. Bulgular, ölçme ve değerlendirme uzmanlarının test geliştirme sürecinde genel değerlendirme çerçevesini iyileştirmeye odaklanması gerektiğini göstermektedir.

Anahtar Kelimeler: Çeldirici, *test geliştirme, yapay zekâ.*

Kaynakça

Downing, S. M. (2005). The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education, 10*, 133-143. https://doi.org/10.1007/s10459-004-4019-5

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research, 87*(6), 1082-1116. https://doi.org/10.3102/0034654317726529

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education, 46*(8), 757-765. https://doi.org/10.1111/j.1365- 2923.2012.04289.xWilliamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology, 50*(6), 2974-2987. https://doi.org/10.1111/bjet.12849

[1] mehriban.ceylan@osym.gov.tr, ORCID: 0009-0008-6201-8675, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, Türkiye

[2] semirhan.gokce@ohu.edu.tr, ORCID: 0000-0002-4752-5598, Niğde Ömer Halisdemir Üniversitesi, Niğde, Türkiye

Paradigm Shift in Test Development: The Effectiveness of AI-Supported Distractors

Mehriban Ceylan[1], Semirhan Gökçe[2]

The aim of this study is to investigate the effectiveness of artificial intelligence (AI)-supported distractors in multiple-choice tests and explore how this innovation may lead to a paradigm shift in test design. Traditional test design relies on human expertise to create distractors; however, this process is vulnerable to biases, which can compromise the reliability and validity of assessments. Traditional item writing often requires extensive revisions to ensure item quality (Downing, 2005). AI- driven approaches, on the other hand, aim to overcome these issues by generating diverse and context-appropriate distractors (Gierl, Lai, & Turner, 2012), showing that AI can produce distractors that align with the cognitive processes being measured (Gierl et al., 2017). While the adoption of AI- supported distractors increases test development efficiency, it also introduces challenges related to transparency and accountability (Williamson, 2019). However, these distractors hold the potential for transformative changes in test development, improving efficiency, psychometric quality, and adaptability. In this research, items from the Academic Personnel and Graduate Education Entrance Exam (ALES) will be used to compare the effectiveness of AI-generated options with those created by human experts. The effectiveness of distractors will be evaluated using item and option analysis through the TAP program. Preliminary findings highlight the potential of AI in education while emphasizing the indispensable role of human expertise in maintaining quality control. The results indicate the need for measurement and evaluation experts to focus on improving the overall assessment framework in the test development process.

Keywords: Distractor, test development, artificial intelligence.

Reference

Downing, S. M. (2005). The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education, 10*, 133-143. https://doi.org/10.1007/s10459-004-4019-5

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research, 87*(6), 1082-1116. https://doi.org/10.3102/0034654317726529

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education, 46*(8), 757-765. https://doi.org/10.1111/j.1365- 2923.2012.04289.x

[1] mehriban.ceylan@osym.gov.tr, ORCID: 0009-0008-6201-8675, Ölçme, Seçme ve Yerleştirme Merkezi, Ankara, Türkiye

[2] semirhan.gokce@ohu.edu.tr, ORCID: 0000-0002-4752-5598, Niğde Ömer Halisdemir Üniversitesi, Niğde, Türkiye

Çoktan Seçmeli Testlerde Çeldiriciler için Puan Düzeltme Uygulamaları

Cahit POLAT [1]

Çoktan seçmeli sınavlarda katılımcıların test puanları doğru cevap sayısından yanlış cevap sayısının çeldirici sayısına bölünmesi ile elde edilen değerden (düzeltme puanı) çıkarılması ile hesaplanır. Örneğin, üniversite yerleştirme sınavlarında her bir soru için 1 doğru cevap ve 4 çeldiriciden oluşan toplamda 5 seçenek yer alır ve test puanı doğru sayısından yanlış sayısının dörtte biri çıkarılarak hesaplanır. Test puanı hesaplanırken doğru sayısından düzeltme puanı çıkarmanın önemli çıktılarından biri öğrenci puanındaki şans başarısını minimize etmektir (Holzinger, 1924; Thurstone, 1919). Diğer taraftan, böyle bir uygulamada bütün çeldiriciler aynı düzeyde değerlendirilmiş olur. Ancak, her bir madde için çeldiriciler faklı özellikler gösterebilmektedir (Terzi ve Yakar, 2018). Bu durum, çeldiricilerde doğru cevaba yakınlık ve şans başarısı açısından her çeldiriciye aynı düzeyde düzeltme uygulamanın puanlama adaletini zedelediğini göstermektedir. Bu bakımdan, puan düzeltmeleri yapılırken çeldiriciler arasındaki eşitsizliğin göz önünde bulundurularak değerlendirmelerin yapılması puanlama adaletini sağlamak açısından önem arz etmektedir. Bu çalışmada her bir seçenek için çeldiricilik düzeyi (ÇD) hesaplanarak düzeltme puanının uygulanabileceği öneriler verilmektedir.

Seçeneklerin çeldiricilik düzeylerine uzman görüşü veya sınavda seçeneklerin cevaplanma sıklıklarına bakılarak karar verilebilir. Cevaplanma sıklığı açısından, her bir seçeneğin çeldiricilik düzeyi seçeneği işaretleyenlerin sayısının toplam yanlış cevap verenlerin sayısına oranı hesaplanarak belirlenebilir. Bu durumda, düzeltme puanları seçeneklerin çeldiricilik düzeylerine ters orantılı olacak şekilde hesaplanır ve öğrencinin her bir soru için puanı ve soru puanlarının toplamı olarak test puanı hesaplanır.

Bu çalışmada faklı çeldiricilik düzeylerine sahip çeldiricilerin olduğu ve öğrencilerin farklı çeldiricileri seçtikleri bir senaryo ile üretilmiş veride, aynı doğru ve yanlış sayısına sahip öğrenciler içerisinde çeldiricilik düzeyi yüksek olan seçenekleri cevaplama eğilimi olan öğrencilerin daha yüksek puanlar aldıkları görülmüştür. Bu uygulama ile faklı çeldiricilik düzeyleri göz önünde bulundurularak puanlama adaletinin ve öğrenci puanlarında çeşitliliğinin artırıldığı görülmüştür.

Anahtar Kelimeler: Çeldiricilik *düzeyi, düzetme puanı,* çoktan *seçmeli testler.*

Kaynakça

Holzinger, K. J. (1924). On scoring multiple response tests. *Journal of Educational Psychology*, *15*(7), 445-447.

Terzi, R., & Yakar, L. (2018). Differential item and differential distractor functioning analyses on Turkish high school entrance exam. *Journal of Measurement and Evaluation in Education and Psychology*, *9*(2), 136-149.

Thurstone, L. L. (1919). A method for scoring tests. *Psychological Bulletin*, *16*, 235-240

[1] cahitpolat@harran.edu.tr, ORCID: 0000-0002-1423-5084, Harran Üniversitesi, Şanlıurfa, Türkiye

Score Correction Applications for Distractors in Multiple Choice Tests

Cahit Polat [1]

In multiple-choice exams, test takers' test scores are calculated by subtracting the value obtained by dividing the number of incorrect answers by the number of distractors (correction score) from the number of correct answers. For example, in the university placement exams, for each question there are five choices in total, consisting of one correct answer and four distractors, and the test score is calculated by subtracting one-fourth of the number of incorrect answers from the number of correct answers. One of the essential outcomes of subtracting the correction score from the number of correct answers when calculating the test score is to minimize the chance success in the student score (Holzinger, 1924; Thurstone, 1919). On the other hand, in such an application, all distractors are evaluated at the same level. However, some distractors in the questions may have different characteristics (Terzi & Yakar, 2018). This situation shows that applying the same correction level to each distractor regarding proximity to the correct answer damages the scoring justice. In this respect, it is critical to consider the inequality between the distractors when making score corrections to ensure scoring fairness. This study calculates the distractor level (DL) for each distractor, and suggestions are given for applying the correction score.

The distraction levels of the choices can be decided by expert opinion or by looking at the frequencies of the choices answered in the exam. In terms of frequencies, the distraction level of each option can be determined by calculating the ratio of the examinees who marked the certain distractor to the total number of wrong answers. In this case, the correction scores are computed inversely proportional to the distractor levels of the choices and scores for each question are determined. Finally, test scores are calculated as the sum of the question scores.

In this study, data was generated using a scenario where students selected distractors with varying levels of distractions. It was observed that students who tended to choose options with high levels of distraction achieved higher scores despite having an equal number of correct and incorrect answers. This application increased scoring fairness and plurality among test takers by accounting for varying distractions.

Keywords: *Distractor level, Score correction, Multiple choice tests*

References

Holzinger, K. J. (1924). On scoring multiple response tests. *Journal of Educational Psychology*, *15*(7), 445- 447.

Terzi, R., & Yakar, L. (2018). Differential item and differential distractor functioning analyses on Turkish high school entrance exam. *Journal of Measurement and Evaluation in Education and Psychology*, *9*(2), 136-149.

Thurstone, L. L. (1919). A method for scoring tests. *Psychological Bulletin*, *16*, 235-240.

[1] cahitpolat@harran.edu.tr, ORCID: 0000-0002-1423-5084, Harran University, Şanlıurfa, Türkiye

The multilingual education and its assessment-evaluation systems in schools for immigrants

Sima Ahdab [1], Onur Özcan[2]

As our world becomes increasingly interconnected, multilingual education is gaining traction. Especially, with the increasing number of immigrant students in several countries, including Türkiye, it became necessary to create an educational system that relies on multilingual education and its structuring. As well as curriculum, the assessment system in multilingual education requires careful design to ensure comprehensive and fair evaluation for students from different linguistic backgrounds and levels. Creating evaluation systems involves taking into account linguistic, cultural, and sociological factors. This system should be carried out through different types including several tools such as written and oral exams in addition to practical exams and projects. This aims to adequately assess listening, speaking, reading and writing skills. Besides, the evaluation system must also include clear and objective standards and criteria for evaluating students' performance across multiple languages to ensure fairness and comprehensiveness.

This study discusses a vision for multilingual education and the design of assessment-evaluation systems able to effectively target multilingual learners. In Türkiye, there are several international schools for immigrant students located in different cities like Istanbul, Ankara, Antalya and Mersin which offer multilingual education system, teaching many languages as: Arabic, Turkish, French and English, therefore they have different systems of evaluating each language, all based on the four skills: reading writing speaking and listening. As a case study we would like to focus on the students of an Iraqi school located in Ankara which offers the education for immigrant students who learn several languages (Arabic, Turkish, French and English) and the assessment-evaluation system in addition to the problems that these students may encounter.

We will carry out a descriptive analysis to adequately demonstrate the different forms of exams for the languages taught at school. This will serve as a mean to evaluate the effectiveness of the various assessment methods. Consequently, we will be able to analyze the exams of immigrant students in foreign language courses at school and conduct comparative content analyses. The discussion will focus on the foreign language exam procedures at the school in question in relation to multilingualism and immigrant concerns based on the data collected. Therefore, it is believed that this study will add value to the research on language education for foreign students and the assessment-evaluation systems for immigrant students living outside their home country by choice or necessity.

Keywords: *Multilingual education, migration, assessment-evaluation, foreign language, Türkiye.*

[1] simaahdab88@gmail.com, Ankara, Türkiye

[2] onurozcanfrancais@gmail.com, Ankara, Türkiye


TÜRKİYE CUMHURİYETİ MİLLÎ EĞİTİM BAKANLIĞI


ÖSYM
50 yıl