



Açıköğretimde kullanılan sınavlardan Klasik Test Kuramına ve Madde Tepki Kuramına göre elde edilen yetenek ölçülerinin karşılaştırılması

Yrd. Doç. Dr. Murat AKYILDIZ^a
Dr. Murat Doğan ŞAHİN^a

^a Anadolu Üniversitesi Açıköğretim Fakültesi

Özet

Öğrenci başarısının gerçeğe uygun kestirilmesi her zaman önemli bir sorun olmuştur. Klasik Test Kuramı ve Madde Tepki Kuramı öğrenci başarısını gerçeğe yakın kestirmek amacıyla geliştirilmiş yaklaşımlardır. İki kuram için de testin çok boyutlu olup olmadığı önemli bir problemdir. Bu çalışmada 150000 kişiden daha fazla kişinin katıldığı geniş katımlı bir sınavın çok boyutlu olup olmadığı DIMTEST ve NOHARM yöntemiyle incelenmiştir. Yapılan analizler sonucunda sınavın iki boyutlu olduğu görülmüştür. Verilerin hangi madde tepki kuramı modeline uyum gösterdiğini bulmak amacıyla yapılan analizler sonucunda verilerin 3 parametrelili lojistik modele daha iyi uyum gösterdiği görülmüştür. Klasik test kuramına ve madde tepki kuramına dayalı yetenek kestirimleri arasındaki ilişkilerin farklı kitapçıklarda 0,801 ile 0,894 arasında değiştiği saptanmıştır. İki kurama dayalı yetenek kestirimlerinin benzer olmakla birlikte birbirlerinin yerine kullanılamayacağı sonucuna ulaşılmıştır.

Anahtar Sözcükler: Klasik Test Kuramı, Madde Tepki Kuramı, Yetenek Kestirimi, Açıköğretimde Puanlama, Öğrenci Başarısı

Abstract

Estimating true scores of students has always been a problem. Classical Test Theory and Item Response Theory are two different theories developed for estimating true score of students. Multidimensionality is a problem for both theories. In this study multidimensionality is examined in a test more than 150000 student attended using DIMTEST and NOHARM. Results showed that the exam found two dimensional. It is found that data fits the 3 parameter logistic model best. The correlation between Classical Test Theory and Item Response Theory ability estimations are differ between 0,801 and 0,894 based on exam booklet. It is found that both ability estimates are not interchangeable.

Keywords: Classcial Test Theroy, Item Response Theory, Ability Estimates, Scoring in Open Education, Student Success

Kaynak Gösterme

Akyıldız, M. ve Şahin, M. D. (2017). Açıköğretimde kullanılan sınavlardan Klasik Test Kuramına ve Madde Tepki Kuramına göre elde edilen yetenek ölçülerinin karşılaştırılması. *AUAAd*, 3(4), 141-159.

Giriş

Öğrenci başarısının gerçek değerine daha yakın kestirilmesi eğitim işinin en baştan bu yana süregelen isteğidir. Bilindiği gibi bireylerin gerçek yetenekleri (gerçek puanları) birer gizil değişkendir ve doğrudan ölçülmesi mümkün değildir. Bu nedenle bireylerin gerçek yeteneklerinin göstergesi olduğuna inanılan simülatif durumlar (sınav soruları) oluşturulur. Bireylerin bu simülatif durumlardaki tepkilerinin gerçek yeteneklerinin göstergesi olduğu düşünülür. Bu simülatif durumlardan elde edilen bilgilerin gerçek yeteneği yansıtması için hatasız tamamlanması gereken süreçlerden birisi de simülatif durumu oluşturan sınav sorularından elde edilen puanların ölçülmesi amaçlanan yeteneği temsil etme derecesidir. Sınav sorularından elde edilen puanların amaçlanan beceriyi temsil edip etmediğini test etmek için çeşitli yollar bulunmaktadır. Bu yollardan en yaygın kullanılanı, sınav sorularının boyutluluğunun incelenmesidir. Sınavı oluşturan soruların boyut ya da boyutlara yerleşme durumları sınavın ölçülmek istenen özelliği ne derecede ölçebildiğinin (temsil ettiğinin) önemli bir göstergesidir.

Soruların eğer temsil ediyorsa gerçek yeteneğin ne kadarını ölçebildiği ise başka ve önemli bir diğer konudur. Örneğin akıl yürütme becerisini ölçen bir maddeden elde edilen puanların ne kadarının akıl yürütme becerisi ile ilgili olduğu cevaplanması gereken bir sorudur. Psikolojik özellikler söz konusu olduğunda sınav sorusu gibi uyarıcıların davranışın ne kadarından sorumlu olduğu sorusunun sorulması gerekir. Bir iktisat sorusuna verilen doğru yanıtın her yanıtlayıcı için aynı olasılıkta olup olmadığının, aynı derecede şanstın arınık verildiğinin, aynı derecede zihinsel emek gerektirip gerektirmediğinin cevabı her bir sorunun her bireyden alabildiği bilginin miktarında değişiklik olup olmadığını da belirlemektedir. Kısaca söylemek gerekirse bir soru her bireyden aynı miktarda bilgiyi mi ölçmektedir? İşte bu soru yukarıda da söylendiği gibi bir soruyla gerçek yeteneğin ne kadarını ölçüldüğü sorusuyla aynı anlama gelmektedir. Bu çalışmada Açıköğretim sistemi ile eğitim vermekte olan bir programın bir dersinin ara sınavında soruların boyutluluk durumu incelenmiş ve farklı yöntemlerle elde edilen yetenek kestirimlerinin birbirine benzerlik taşıyıp taşımadığı test edilmiştir. Böylece büyük Açıköğretim sistemlerinde farklı puanlama yöntemlerinin avantaj ve dezavantajları gösterilmeye çalışılmıştır.

Araştırma Sorunsalı

Ölçme ve değerlendirme alanında gerçek yeteneğin (gerçek puanın) kestirilmesi için iki kuram bulunmaktadır. Bu kuramlardan birisi klasik test kuramı diğeri ise madde tepki

kuramıdır. Bu çalışmada Açıköğretim sistemiyle eğitim vermekte olan 150000 öğrenciden daha fazla kişinin kayıtlı olduğu bir programın bir dersinin 2014-2015 öğretim yılında uygulanmış olan ara ve final sınavından elde edilen puanların boyutluluk durumunun nasıl olduğu belirlenmeye çalışılacaktır. Ayrıca sınavlardan elde edilen puanların madde tepki kuramı modellerinden hangisine daha çok uyum gösterdiği ve sınavdan elde edilen puanların Klasik Test Kuramına (KTK) ve Madde Test Kuramına (MTK) göre kestirilmesinin anlamlı bir farklılık gösterip göstermediği ortaya çıkarılacaktır. Bu çalışma ile çok kaba bir değerlendirme ile her sorunun aynı puana sahip olduğu bir ölçme yaklaşımı ile her sorunun her birey için farklı puana sahip olduğu bir yaklaşımın farklı sıralama durumları oluşturup oluşturmayacağı belirlenmiş olacaktır.

Yukarıdaki genel amaçlar doğrultusunda aşağıdaki sorulara cevap aranmıştır:

1. Sınavdan elde edilen veriler kaç boyutludur?
2. Madde tepki kuramına göre kestirilecek yetenek ölçüleri hangi madde tepki kuramı modeline daha iyi uyum göstermektedir?
3. Klasik test kuramına göre kestirilen yetenek ölçüleri ile madde tepki kuramına göre kestirilen yetenek ölçüleri arasında yüksek ve anlamlı bir ilişki var mıdır?
4. Klasik test kuramına ve madde tepki kuramına göre elde edilen yetenek ölçüleri kitapçık türüne göre anlamlı farklılık göstermekte midir?

İlgili Alanyazın

Çalışmanın Kuramsal Temelleri

Klasik Test Kuramı:

Klasik test kuramı bireylerin gerçek puanlarının aşağıdaki eşitlikle bulunabileceğini kabul eder

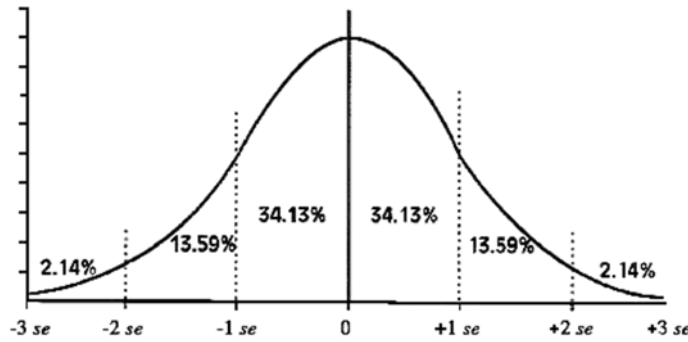
$$X = T + E \quad (1)$$

Eşitlikte X gözlenen puanı, T gerçek puanı, E ise tesadüfi hata miktarını göstermektedir. Bu eşitliğe dayanarak klasik test kuramı bazı varsayımları da içinde barındırmaktadır. Bu varsayımlar aşağıdaki gibi sıralanabilir.

1. Tesadüfi hata, ortalaması sıfır standart sapması her test uygulamasında değişen bir normal dağılım göstermektedir.
2. Tesadüfi hatanın gerçek puanla korelasyonu sıfırdır.
3. İki farklı özelliğin ölçme sonuçlarının hata puanları arasındaki korelasyon sıfıra eşittir.

Bu üç varsayımın da temelinde hata kavramı bulunmaktadır. Klasik Test Kuramı, hatanın iyi tanımlanmasının ve tespitine yönelik çalışma yapılmasının ölçme yapanı gerçek puana yaklaştıracakını kabul eder. Hata, klasik test kuramında yukarıdaki eşitlikten de görüleceği gibi gerçek puan ile gözlenen puan arasındaki fark olarak tanımlanabilir. Yönü ve miktarı belirlenebilen sabit ve sistematik hatalardan ziyade KTK, kuramın öngördüğü üçüncü hata türü olan tesadüfi hatalara odaklanır. Tesadüfi hatanın varlığı klasik test kuramında ölçümlerin güvenilirliği kavramının ortaya çıkmasına yol açmıştır. Bir ölçmenin güvenilirliği o ölçümün tesadüfi hatalardan arınlık derecesi olarak tanımlanabilir.

Ölçme süreçlerinde puanlara karışan tesadüfi hatanın aşağıdaki dağılıma uygun şekilde karıştığı kabul edilmektedir.



Şekil 1. Tesadüfi Hataların Dağılımı

KTK’da biri madde güçlüğü, diğeri ise madde ayırt ediciliği olarak anılan iki madde istatistiği yer alır. Madde güçlüğü (p_j), bir maddenin uygulandığı grupta doğru cevaplanma yüzdesidir. Madde ayırtediciliği (r_{jx}) ise maddenin ölçmeye çalışılan beceriye sahip olanlarla olmayanları ayırt etme gücüdür. Madde güçlüğü, maddeyi doğru yanıtlayanların tüm yanıtlayıcılara oranıyken, madde ayırtediciliği madde puanları ile testin bütününden elde edilen puanlar arasındaki korelasyon hesaplanarak elde edilir. Çoktan seçmeli testlerde madde ayırt ediciliği için, maddenin test sonuçlarına göre başarılı olan gruptan başarısız olan gruptaki yanıtlanma oranının çıkarılmasıyla elde edilen “alt-üst grup metodu”yla da elde edilebilir. Maddeye dair bu iki temel istatistiksel özellik, testin uygulandığı gruptan etkilenir. Öyle ki bir madde setinin bir gruptan elde edilmesiyle elde edilen güçlük ve ayırt edicilik değerleri, söz konusu madde setinin başka bir gruba uygulanmasıyla elde edilen güçlük ve ayırt edicilik değerlerinden farklı olabilmektedir. Bu durum da uygulanmış maddelerin parametrelerinden yararlanılarak sonraki uygulamalarda elde edilecek değerleri tahmin etmeyi mümkün kılmamaktadır. Bir başka ifadeyle KTK’da madde istatistikleri gruba bağımlıdır.

Klasik Test Kuramının Sınırlılıkları:

Klasik test kuramı, ilkeleri yaklaşık 100 yıl önce ortaya atılmış bir kuramdır. Gruba bağımlı oluşu nedeniyle içinde bazı sınırlılıklar taşımaktadır. Bu sınırlılıklar şöyle sıralanabilir (Embretson & Reise, 2000).

1. Ölçmenin standart hatası örneklemdaki tüm bireyler için aynıdır.
2. Az sayıdaki maddeyle yüksek bir güvenilirlik değeri elde etmek zordur.
3. Bir testten elde edilen puanları başka bir testten elde edilen puanlarla karşılaştırmak ancak testler eşdeğerse mümkündür.
4. Madde parametrelerinin doğru tahmin edilmesi ancak örneklemin çok iyi seçilmiş olmasına bağlıdır.
5. Testten elde edilen puanların ne anlama geldiğini sadece puandan çıkarsamak mümkün değildir. Örneğin testten 80 puan almış bir öğrencinin başarılı olup olmadığı norm grupla karşılaştırıldıktan sonra kararlaştırılabilmektedir.
6. Test puanları sıralama ölçeği düzeyindedir. Puanların eşit aralıklı olabilmesi için özel işlemlere (standart puanlara döndürme gibi) tabi tutulması gerekmektedir.
7. Bir test içinde farklı madde formatlarını kullanmak testten elde edilecek toplam puanlarda dengesizliklere yol açmaktadır. Örneğin bir klasik, bir çoktan seçmeli sorudan oluşan iki soruluk durumda hangi maddenin kaç puan alması gerektiğini belirlemek güçtür.
8. Testten ve maddelerden edilecek bilgiler (ortalama, standart sapma, madde güçlüğü, madde ayırt ediciliği) sadece o grup için geçerlidir. Tüm parametreler gruba bağımlıdır. Aynı test başka bir gruba uygulandığında elde edilen madde ve test istatistikleri değişkenlik gösterebilir.
9. Her maddenin puan değeri aynıdır. Farklı olsa bile bunun matematiksel bir kuralı yoktur.

Klasik Test Kuramı'nın bu sınırlılıklarına çözüm getirdiği iddiasıyla Madde Tepki Kuramı ortaya çıkmıştır.

Madde Tepki Kuramı

Madde Tepki Kuramı (MTK), özellikle 1970'lerden sonra test geliştirme çalışmalarında, KTK'nın yöntemlerine alternatif olarak gelişmiş ve giderek daha fazla uygulama alanı bulmuştur. Her ne kadar Madde Tepki Kuramına ilişkin tarihçe Binet ve Simon'a kadar geri götürülse de bu yoldaki ilk çalışmalar 1952 yılında Lord'la başlamıştır (Hambleton, 1988; Linden & Hambleton, 1997). Lord, bu ilk çalışmada iki parametrelili normal ogive modelini

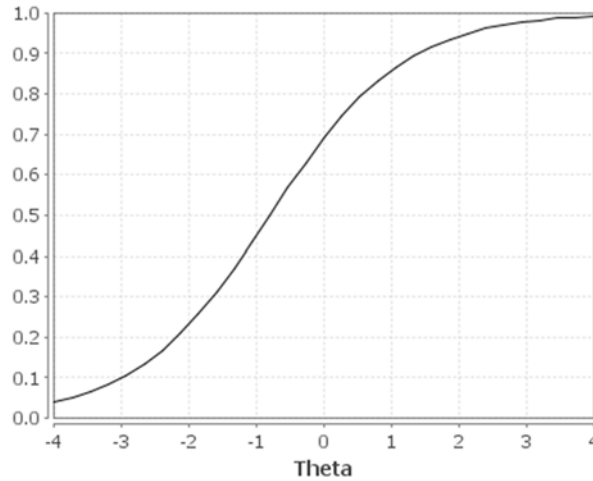
geliştirmiştir. Daha sonra Birnbaum'un normal ogive model için daha kolay kullanımlı lojistik modeller geliştirme çalışmaları olmuşsa da Rasch, 1960 yılında ileride kendi adıyla anılacak olan ilk Madde Tepki Kuramına dayalı lojistik modellerden birisini geliştirmiştir (Hambleton, 1988).

Bu gelişmelerden sonra MTK, hem testlerin farklı yetenek gruplarına uygulanırken o yetenek gruplarına uyarlanabilmesi kolaylığı getirdiği için hem de bireylerin yetenek ölçülerini teste bağımlı kalmadan, maddelerden yola çıkarak kestirebildiği için giderek daha sık kullanılmaya başlamıştır. 1977 yılında Educational and Psychological Measurement dergisi, Madde Tepki Kuramını konu alan bir özel sayı çıkarmıştır.

MTK, iki temel önermeye dayanır:

- Bir kişinin bir test maddesindeki performansı, örtük özellik ya da yetenek denilebilecek bir faktör aracılığıyla tahmin edilebilir.
- Bir kişinin bir maddedeki performansı ile o maddeye cevap vermesini sağlayan (performansını belirleyen) özellik arasındaki ilişki, madde karakteristik eğrisi (item characteristic curve) adı verilen bir eğri ile gösterilebilir (Hambleton, Swaminathan, & Rogers, 1991)

Bir kişinin yeteneğinin o kişinin bir maddedeki başarısını açıklamak için “madde karakteristik eğrisi” adı verilen matematiksel bir açıklama kullanılır. Bu eğri, maddeyle ölçülmeye çalışılan yeteneğin o maddeye doğru yanıt verme olasılığı üzerindeki regresyonunu ifade eder. Bu regresyon eğrisinin alacağı şekil, regresyon eğrisini tanımlamak için kullanılacak parametre sayısına bağlı olarak değişmektedir. Madde karakteristik eğrisini belirlemek üzere madde güçlüğü (b), madde ayırt ediciliği (a), ve şans başarısı (c) olmak üzere üç parametre kullanılır ve bu parametrelerin hangilerinin kullanıldığına bağlı olarak madde tepki kuramında bir, iki veya üç parametrelilik olmak üzere farklı modeller belirlenmiştir (Crocker & Algina, 1986; Linden & Hambleton, 1997). Aşağıda örnek bir madde karakteristik eğrisi gösterilmektedir.



Şekil 2. Madde Karakteristik Eğrisi

MTK'nın getirdiği en önemli yenilikler, madde parametrelerinin belirli bir gruptan ya da örneklemden bağımsız olarak kestirilebilmesi kolaylığı (Baykul, 1979; Crocker & Algina, 1986; Linden & Hambleton, 1997) yetenek ölçülerinin uygulanan madde örneklemden bağımsız olarak kestirilmesi (Ronald K. Hambleton ve ark., 1991), her bir maddenin testle ölçülmek istenen beceriler ile eşleştirilerek henüz kazanılmamış olan ve hâlihazırda edinilmiş bulunan becerilerin birey bazında ortaya çıkarılması (Embretson & Reise, 2000) ve testlerin bilgisayar ortamına aktarılarak her bireye kendi yetenek seviyesine uygun madde sunulması (Embretson & Hershberger, 1999; Embretson & Reise, 2000) olarak sıralanabilir.

Ayrıca madde tepki kuramının pratikte getirdiği başka bazı yenilikler de şöyle sıralanabilir:

- Her soruya zorluk, ayırteçicilik gibi özelliklerini dikkate alarak farklı puan değeri tanımlayabilmek
- Her sorunun hangi yetenek düzeyinde doğru cevaplanacağını tahmin edebilmek
- Her soru için, şansla doğru cevap verme ihtimalini soru için planlanan puandan çıkartabilmek
- Öğrencilerin puanlarını eşit aralıklı ölçek düzeyine daha yakın olarak bildirebilmek
- Öğrenci puanlarını doğru cevapların toplamı olarak alan klasik test kuramına oranla gerçeğe daha yakın kesinlikte hesaplamak
- Hangi yetenek grubundan hesaplanmış olursa olsun madde parametrelerinin grup değişiminden etkilenmemesi

- Bilgisayar ortamında bireye uyarlanmış test uygulamalarına imkan tanımak (An & Yung, 2014; Hambleton & Jones, 1993).

Çalışmayla Doğrudan İlişkili Araştırmalar

Progar ve Sočan (2008) yaptıkları çalışmada matematik ve fen başarısını ölçen ve çok sayıda ülkede uygulanan TIMMS sınavına katılmış 8000 öğrencinin puanlarını klasik test kuramına ve madde tepki kuramına göre kestirerek karşılaştırmışlardır. Araştırma sonucunda verilerinin tek boyutlu ve 2 parametrelili modele daha iyi uyum gösterdiğini bulmuşlar ve iki kurama göre kestirilen yetenek ölçüleri arasında yüksek ve manidar bir ilişki tespit etmişlerdir. İki kuramın yüksek ilişkili yetenek ölçüsü bildirmesine rağmen uyum yüksek olduğunda madde tepki kuramı parametrelerinin gruptan bağımsızlık konusunda daha iyi olduklarını bildirmişlerdir.

Courville (2004) 11. Ve 12. Sınıf öğrencilerine uygulanan ACT Assessment Test sonuçlarını iki kurama göre karşılaştırdığı çalışmada iki kuramdan elde edilen yetenek ölçüleri arasında 0,90 ile 1,00 arasında değişen korelasyon katsayıları elde etmiştir.

Wiberg (2004) ehliyet sınavından elde edilen puanları klasik test kuramına ve madde tepki kuramına göre kestirdiği çalışmada iki kuramdan elde edilen bilgilerin ilişki gösteriyor olsalar bile birbirinden farklı olduğunu ve madde tepki kuramının 3. Bir parametre barındırması nedeniyle madde parametrelerinin gerçeğe daha yakın kestirime izin verdiğini söylemiştir.

Fan (1998) tarafından yapılan bir araştırmada Madde Tepki Kuramı'nın ve Klasik Test Kuramının karakteristik ölçüleri arasında, bir kuramı diğerine tercih ettirecek kadar anlamlı bir fark bulunmamıştır.

MacDonald ve Paunonen (2002) yaptıkları monte carlo çalışmada madde tepki kuramına ve klasik test kuramına göre ürettikleri birey parametreleri karşılaştırmışlar ve 1.00'e çok yakın değerler bulmuşlardır. Fakat madde tepki kuramına göre yapılan hesaplamalarda madde parametrelerinin büyük bir uyumluluk gösterdiğini tespit etmişlerdir. Bu nedenle madde seçimi söz konusu olduğunda klasik test kuramı yerine madde test kuramını önermişlerdir.

Young (1991), Madde Tepki Kuramını ve klasik yığılımlı (cumulative) puan hesaplama (Klasik Test Kuramı) yönteminin kız ve erkek öğrencilerin yeteneklerini tahmin etmesi bağlamında karşılaştırdığı çalışmada, iki cinsiyet grubu için de Madde Tepki Kuramına dayanan puanlama sisteminin klasik yığılımlı puan hesaplama yöntemine göre daha yordayıcı olduğunu bulmuştur.

Çelen ve Aybek (2013) çalışmalarında, sınıf içi ders başarısını ölçmeyi amaçlayan bir ölçme aracından, KTK ve MTK'ya dayalı olarak elde edilen öğrenci başarıları arasındaki

ilişkiyi karşılaştırmışlardır. Elde edilen sonuçlara göre, her iki kurama göre hesaplanan başarı puanları arasında pozitif, yüksek ve manidar bir ilişki bulunmuştur. Elde edilen puanlarla öğrencilerin dersten geçme-kalma durumu karşılaştırıldığında ise öğrencilerin %87.4'ünde iki kurama göre de aynı karar verildiği görülmüştür.

Gelbal (1994) çalışmasında, ilkököl 5. sınıf Türkçe ve matematik dersleri için geliştirilen başarı testlerinin uygulaması sonucu, madde tepki kuramının Rasch modeliyle ve klasik test kuramı ile kestirilen yetenek parametrelerini karşılaştırmıştır. Elde edilen sonuçlara göre her iki kuramla elde edilen yetenek kestirimleri arasında pozitif ve yüksek bir ilişki bulunmuş, öğrencilerin yetenek düzeyi arttıkça bu ilişkinin de artış gösterdiği sonucuna ulaşılmıştır.

Özkan (2012) çalışmasında, MEB tarafından düzenlenen Öğrencilerin Başarılarının Belirlenmesi Sınavının (ÖBBS) Türkçe ve matematik alt testlerinden Klasik test kuramı (KTK), tek boyutlu ve çok boyutlu madde tepki kuramı (MTK) ile kestirilen başarı puanlarını karşılaştırmıştır. Araştırmanın genel sonucu olarak çok boyutlu MTK modelleri kapsamında elde edilen yetenek parametrelerinin, KTK ve tek boyutlu MTK'ye göre daha az hata içerdiği ve daha duyarlı ölçümler sağladığı sonucuna varılmıştır.

Görüleceği gibi madde tepki kuramı ile klasik test kuramına göre hesaplanan yetenek ölçüleri arasındaki korelasyonlar için kesin bir netlikten bahsedilmesi mümkün görünmemektedir. Kimi çalışmalar ilişkiyi yüksek hesaplarırken kimi çalışmalar düşük bir ilişki olduğunu göstermişlerdir. Fakat hemen hemen tüm araştırmacılar madde seçimi ve öğrenciler arası farkın gösterilmesi söz konusu olduğunda madde tepki kuramını önermektedir.

Yapılan bu çalışma ile Anadolu Üniversitesi Açıköğretim Fakültesi sınavları için madde tepki kuramı ile yetenek ve parametre kestirimi ilk kez yapılmıştır. Böylece Üniversitenin uygulamış olduğu testler üzerindeki hakimiyeti artmış ve yeni test geliştirme becerisini bünyesinde barındırır hale gelmiştir. Ayrıca bu çalışma geniş ölçekli sınavlarda yetenek kestiriminde çok boyutluluğu dikkate alması bakımından ülkemizde yapılmış ilk çalışmadır.

Yöntem

Araştırma Modeli

Araştırma nicel araştırma yöntemlerinden betimsel model kullanacak şekilde tasarlanmıştır. Nicel araştırmalar “bir fenomeni, istatistik gibi matematiksel yöntemler kullanarak analiz edilmiş miktarsal (sayısal) veri toplayarak açıklamak” (Aliaga & Gunderson;

2002) olarak tanımlanmaktadır. Betimsel çalışmalar varolan durumu olduğu haliyle ortaya koyan çalışmalardır.

Araştırma Grubu

Bu çalışmada bir örnekleme yöntemi kullanılmamıştır. Araştırmada kullanılacak veriler 2014-2015 öğretim yılı güz döneminde öğrenci sayısı 150000 üzerinde olan bir dersin ara ve final sınav sonuçlarıdır.

Veri Toplama Araçları

Veriler sınav sonuçlarının tutulduğu veritabanından İstatistik Bilgi Edinme ve Değerlendirme biriminden alınan izinle çekilmiştir.

Veri Toplama Süreci

Veriler veritabanından çekildikten sonra işlenebilir hale getirmek amacıyla SPSS 23 programına aktarılmıştır. Veriler ham haliyle alınmış, kitapçık kodlarına uygun cevap anahtarları ile 1-0 matrisi olacak şekilde yeniden kodlanmıştır. Verilerin doğruluğu veritabanında bulunan öğrenci toplam puanları ile eşlenerek sağlanmıştır.

Veri Analizi

Veriler için öncelikle boyutluluk analizleri yapılmıştır. Sınavın kaç boyutlu olduğu iki yöntemle belirlenmiştir. Önce Stout'un 1987 yılında geliştirdiği nonparametrik bir yaklaşım olan DIMTEST yöntemiyle boyutluluk durumu test edilmiştir. Daha sonra eğrisel ilişkileri dikkate alan bir boyutluluk yaklaşımı olan NOHARM testiyle boyutluluk analizi yapılmıştır.

İkinci soruya cevap bulabilmek amacıyla dersin boyutluluğu da göz önünde bulundurularak IRTPRO yazılımı kullanılarak uygun madde tepki kuramı modeli için analize sokulmuştur. IRTPRO kullanılarak derse ilişkin verinin kaç parametrelili madde tepki kuramı modeline uyum gösterdiği tespit edilmiştir.

Üçüncü soruya cevap bulmak amacıyla SPSS programı aracılığı ile öğrencilerin ara sınav ve final sınavı cevaplarından yararlanılarak ham puanları belirlenmiştir. Daha sonra IRTPRO yazılımı kullanılarak uygunluğun daha yüksek olduğu madde tepki kuramı modeli kullanılarak yetenek kestirimleri yapılmıştır. Bu yetenek kestirimlerinin anlamlı farklılık

gösterip göstermediği iki farklı yetenek kestirimi arasındaki korelasyon katsayıları hesaplanarak gösterilmiştir.

Son soruya cevap bulmak amacıyla hem Klasik Test Kuramına hem de Madde Tepki Kuramına göre elde edilmiş olan yetenek madde parametre kestirimleri arasındaki ilişkiler kitapçık türü bağlamında korelasyon katsayıları aracılığı ile incelenmiştir. Böylece kitapçık türünün madde parametreleri üzerinde anlamlı bir etkisinin bulunup bulunmadığı ortaya çıkarılmıştır. Madde tepki kuramına dayalı ayırtedicilik (a), zorluk (b) ve soruyu şansla cevaplama ihtimali (c) madde parametreleri üçüncü soruya cevap bulurken elde edilmiştir.

Bulgular ve Yorumlar

Araştırmanın birinci sorusu olan verilerin kaç boyutlu olduğu sorusunu yanıtlamak amacıyla verilere ilişkin DIMTEST ve NOHARM sonuçları aşağıdaki gibidir.

Dimtest, bir testi oluşturan maddeler iki gruba ayrıldığında bu iki grup maddenin birbirlerinden aralarındaki korelasyonlar bakımından anlamlı bir fark gösterip göstermediğini test eder. Birinci grup maddeler aralarındaki korelasyonu diğerlerine göre yüksek olan maddelerdir. İkinci grup maddeler ise geriye kalan maddeler olurlar. Dimtest bu iki grup maddeler arasında anlamlı farklılık olup olmadığını test eder. Dimtest sonucunda bir T değeri ve bu T değerinin anlamlılığı elde edilir. T değerinin anlamlı olması durumunda iki grup madde arasında anlamlı bir farklılık olduğu dolayısı ile ölçme aracının birden çok boyutunun olduğu kararı verilir.

Bu çalışmada uygulanan Dimtest analizi sonucunda elde edilen değerler aşağıda Tablo 1’de gösterilmiştir.

Tablo 1		
<i>DIMTEST Analizi Sonuçları</i>		
DIMTEST	T	p
Ara Sınav	0,4591	0,3231
Final	5,5807	0,0000

Dimtest analizi sonuçları ara sınav sonuçlarının tek boyutlu fakat final sınavının çok boyutlu olduğuna dair sonuçlar edilmesine yol açmıştır.

Dimtest her ne kadar geleneksel faktör analizi gibi boyutluluk belirleme yaklaşımlarından daha güçlü bir boyutluluk belirleme yöntemi olsa da temelde maddeler arası ve öğrenci yetenek arası ilişkilerin doğrusal kabul edildiği bir yaklaşımla analiz yapmaktadır. Bu nedenle Madde tepki Kuramının doğasına daha uygun olan yetenek ve sorulara verilen tepkiler arasındaki ilişkilerin eğrisel olabilmesini dikkate alarak analiz yapan NOHARM yaklaşımıyla boyutluluk analizleri tekrarlanmıştır. Elde edilen sonuçlar aşağıda Tablo 2’de gösterilmiştir.

NOHARM analizi sonucunda 1 boyutlu, 2 boyutlu ve varsa daha çok boyutlu olma durumları için sırasıyla analiz tekrar edilir. Her bir koşul için Tanaka’nın GFI değeri ve RMSR (Root Mean Score of Residuals) değeri elde edilir. GFI değerinin 1’e yakın olması beklenir. Farklı boyutluluk durumları için elde edilen Tanaka GFI indeksleri gibi dağılımı olmayan indeksler arasında, literatürde bildirilmiş bir ölçüt olmamakla birlikte, Cheung ve Rensvold’un (2002) yaptıkları çalışmalar sonrasında uyum indeksleri için genel olarak 0,01 ve daha fazla fark olması durumunda farkın anlamlı bir fark gibi değerlendirilmesi eğilimi doğmuş görünmektedir. NOHARM analizleri 2 faktöre kadar tekrar edilebilmiş, 3 ve sonraki faktörler için 2 faktörlü olma durumundan farklı sonuçlar elde edilmediğinden analizler 2 faktörlü sonuçlardan sonra devam ettirilmemiştir. Aşağıda Tablo 2’de NOHARM sonucunda elde edilen uyum indisleri verilmiştir.

Tablo 2					
<i>Tek ve İki Boyutlu Olma Durumları için NOHARM Boyutluluk Analizi Uyum İndeksleri</i>					
		Tahmin Düzeltmesi olmadan		Tahmin Düzeltmesi Kullanarak	
Ara Sınav	1 Faktör	RMSR=0,009	TANAKA GFI= 0,970	RMSR=0,009	TANAKA GFI= 0,970
	2 Faktör	RMSR=0,004	TANAKA GFI= 0,990	RMSR=0,004	TANAKA GFI= 0,990
Final	1 Faktör	RMSR=0,009	TANAKA GFI= 0,968	RMSR=0,009	TANAKA GFI= 0,969
	2 Faktör	RMSR=0,006	TANAKA GFI= 0,984	RMSR=0,006	TANAKA GFI= 0,984

Tablo 2’den görülebileceği gibi dersten elde edilen veriler çok boyutlu bir modele daha çok uyum göstermektedir. Bu nedenle dersin verileri ileride yapılacak analizlerde çok boyutlu olarak değerlendirilmiştir.

Araştırmanın ikinci sorusu ise “Madde tepki kuramına göre kestirilecek yetenek ölçülerinin hangi madde tepki modeline daha iyi uyum göstereceği” şeklindedir. Bu aşamada dersin ara ve final sınavı, uygun MTK modelinin belirlenmesi amacıyla sırasıyla tek parametrelili, iki parametrelili ve üç parametrelili modellere göre analiz edilmiş, her bir modelden elde edilen uyum değerleri (RMSEA) karşılaştırılarak en iyi uyum gösteren modeller belirlenmiştir. Yapılan analizler sonucu elde edilen uyum değerleri ve kullanılması uygun görülen model aşağıdaki tablo 3’te gösterilmiştir.

Tablo 3 <i>MTK’ya göre Ölçeklenen Veri-Model Uyum Katsayıları</i>			
Sınav Dönemi	1PL için RMSEA	2PL için RMSEA	3PL için Uyum Değeri
Ara	0,07	0,06	0,04*
Final	0,07	0,06	0,04*

*En iyi uyum.

Yukarıdaki tabloda ara ve final sınavları için elde edilen uyum katsayıları (RMSEA) görülmektedir. RMSEA değerine bakarak bir verinin modele “iyi uyum” gösterdiğini söyleyebilmek için bu değer en az 0,05 olmalıdır. Ayrıca RMSEA değeri, 0,00’a yaklaştıkça model-veri uyumu artar. Bu nedenle dersin ara ve final sınavlarından Madde Tepki Kuramına göre kestirilecek yetenek ölçüleri için en uygun modelin iki boyutlu 3PL model olduğu sonucuna ulaşılmıştır.

Araştırmanın üçüncü sorusu ise “Klasik test kuramına göre kestirilen yetenek ölçüleri ile madde tepki kuramına göre kestirilen yetenek ölçüleri arasında yüksek ve anlamlı bir ilişki” olup olmadığını test etmeye yöneliktir. Dersin ara ve final sınavı için boyutluluğun ve en uygun MTK modelinin belirlenmesinin ardından bir sonraki aşamada veriye en iyi uyum gösteren MTK modeline dayalı olarak öğrencilerin yetenek kestirimleri IRTPRO paket programı kullanılarak elde edilmiştir. Sonraki adımda ise öğrencilerin MTK kullanılarak elde edilen yetenek ölçüleriyle, Klasik Test Kuramından (KTK) elde edilen yetenek ölçüleri arasındaki Pearson momentler çarpımı korelasyon katsayısı SPSS 23 paket programı kullanılarak hesaplanmıştır. Dersin her bir boyutu için MTK ve KTK’den elde edilen yetenek ölçüleri arasındaki korelasyona bakılmış, dolayısıyla ara ve final sınavları için iki korelasyon değeri elde edilmiştir. Elde edilen sonuçlar aşağıda Tablo 4’te görülmektedir.

Tablo 4		
<i>MTK Ve KTK ile Elde Edilen Öğrenci Yetenek Ölçüleri Arasındaki Korelasyon Değerleri</i>		
Sınav Dönemi	Korelasyon Katsayısı	
	1. boyut	2. boyut
Ara	0.939*	0.924*
Final	0.859*	0.870*

*p<.01

Yukarıdaki tabloda sınavın KTK ve MTK'ya göre ölçeklenmesinden elde edilen öğrenci yetenek ölçüleri arasında son derece yüksek ve anlamlı bir ilişki olduğu görülmektedir. Sırasıyla ara sınavda 1. Boyut için MTK ve KTK'ya göre ölçeklenen öğrenci başarı puanlarının 0.939 ilişki gösterdiği, ikinci boyut için bu ilişkinin 0.924 olduğu görülmüştür. Final sınavı için ise 1. Boyut için MTK ve KTK'ya göre ölçeklenen öğrenci başarı puanlarının 0.859 ilişki gösterdiği, ikinci boyut için bu ilişkinin 0.870 olduğu görülmüştür Tablo 4'te bildirilen değerler KTK ve MTK'ya göre ölçeklenen yetenek ölçülerinin yüksek düzeyde benzerlik gösterdiğini bildirmektedir.

Araştırmanın dördüncü ve son sorusu "Klasik test kuramına ve madde tepki kuramına göre elde edilen yetenek ölçüleri ve madde parametreleri kitapçık türüne göre anlamlı farklılık gösterip göstermediğini" bulmaya dönüktür.

Bu adımda, sınavın KTK ve MTK'ya göre ölçeklenmesiyle elde edilen yetenek ölçüleri arasındaki korelasyonlar her bir kitapçık için ayrı ayrı elde edilmiştir. A, B, C ve D kitapçıklarında ara ve final sınavlarının KTK ve MTK'ya göre ölçeklenmesiyle elde edilen öğrenci yetenek ölçüleri arasındaki korelasyonlar aşağıda Tablo 5'te gösterilmiştir.

Tablo 5								
<i>A, B, C ve D Kitapçıklarında MTK Ve KTK ile Elde Edilen Öğrenci Yetenek Ölçüleri Arasındaki Korelasyon</i>								
Sınav Dönemi	Korelasyon Katsayısı							
	A kitapçığı		B kitapçığı		C kitapçığı		D kitapçığı	
	Birinci boyut	İkinci boyut	Birinci boyut	İkinci boyut	Birinci boyut	İkinci boyut	Birinci boyut	İkinci boyut
Ara	0.891*	0.881*	0.889*	0.879*	0.883*	0.877*	0.894*	0.890*
Final	0.871*	0.820*	0.870*	0.801*	0.844*	0.830*	0.827*	0.872*

p<.01

Yukarıdaki tablo incelendiğinde çok boyutlu olduğu belirlenen dersin ara sınavındaki kitapçıklarının her biri için KTK ve MTK'ya göre ölçeklenen yetenek ölçüleri arasında hesaplanan korelasyon katsayılarının 0.877 ile 0.890 arasında değiştiği ve bu korelasyon katsayılarının tümünün istatistiksel olarak anlamlı olduğu bulunmuştur. Buna göre kitapçığa bağlı olarak öğrencilerin dersin ara sınavındaki yetenek ölçülerinin yetenek kestirim yöntemine bağlı olarak küçük değişikliklerin ortaya çıktığı sonucuna ulaşılmıştır.

Final sınavında ise birinci boyut için en yüksek korelasyonun A kitapçığında, ikinci boyut içinse en yüksek korelasyonun ise D kitapçığında olduğu sonucuna ulaşılmıştır. Görüleceği gibi korelasyon katsayıları yüksek birbirine çok yakın değerler almıştır.

Bu sonuçlardan yola çıkarak, öğrencilerin yetenek ölçülerinin KTK'ya ve MTK'ya göre elde edilen kestirimleri arasındaki korelasyonların kitapçıklara göre küçük farklılıklar gösterdiği, kitapçıklar arasında gözlenen farklılığın ayrıca incelenmesi gerektiği sonucuna da ulaşılmıştır.

Sonuçlar

Bulgular bir bütün olarak incelendiğinde 150000 kişiden fazla öğrencinin katıldığı bir dersin çok boyutlu olduğu sonucuna ulaşılmıştır. Klasik Test Kuramına göre her soruya 1 puan verilerek hesaplanan ara ve final sınav toplam puanları ile Madde Tepki Kuramının her soruyu birey düzeyinde o bireyin yeteneğini, soru zorluğunu, sorunun ayırtediciliğini ve sorunun şansa doğru cevaplanmış olma ihtimalini dikkate alarak farklı değerlerde puanlayan yaklaşımıyla kestirilen yetenek ölçüleri arasında yüksek sayılabilecek korelasyon katsayıları bulunmuştur. Araştırma bulguları klasik test kuramına göre kestirilen yetenek ölçüleri ile Madde Tepki Kuramına göre kestirilen yetenek ölçülerinin 0,801 ile 0,894 arasında korelasyon ilişkisi gösterdiğini ortaya koymuştur. Bu ilişkiler yüksek olmakla birlikte bir kurama göre kestirilen yetenek ölçülerinin diğerine göre kestirilen ölçülerin yerini alabileceği anlamına gelmemektedir.

Öneriler

Yukarıdaki sonuçlar elde edilirken Klasik Test Kuramından elde edilen verilere şans başarısından arındırma işlemi uygulanmamıştır. Madde Tepki Kuramının 3 parametrelili modeli doğal olarak bireylerin puanlarını şanstan arındırarak hesaplamaktadır. İlerideki çalışmalarda Klasik Test Kuramına dayalı şans başarısından arındırma yöntemleri kullanılarak yetenek kestirimleri arasındaki ilişkiler incelenmelidir. Böylece en yaygın 4 yanlışın 1 doğruyu iptal

etmesi olarak bilinen Klasik Test Kuramına dayalı şans başarısından arındırma işlemi gibi işlemler ile Madde Tepki kuramına dayalı 3 parametrelili kestirimlerin şansın düzeltilmesi amacına ne kadar yardımcı olduğu karşılaştırılabilir.

Kaynakça

- Aliaga, M., & Gunderson, B. (2002). *Interactive statistics*. New Jersey: Prentice Hall.
- An, X., & Yung, Y.-F. (2014). Item Response Theory: what it is and how you can use the IRT procedure to apply It. Paper presented at the *SAS Global Forum*, Washington.
- Baykul, Y. (1979). Örtük Özellikler ve Klasik Test Kuramları Üzerine Bir Karşılaştırma. Doktora Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Cheung, G., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 233-255.
- Courville, T. G. (2004). An Empirical Comparison Of Item Response Theory and Classical Test Theory Item/Person Statistics (Doctor Of Philosophy), Texas A&M University.
- Crocker, L. M., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York, NY: Holt, Rinehart, and Winston.
- Çelen, Ü., ve Aybek, E.C. (2013). Öğrenci başarısının öğretmen yapımı bir testle Klasik Test Kuramı ve Madde Tepki Kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. 4(2). 64-75.
- Embretson, S. E., & Hershberger, S. L. (1999). *The new rules of measurement: what every psychologist and educator should know*. Mahwah, N.J.: L. Erlbaum Associates.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates.
- Fan, X. T. (1998). Item Response Theory and Classical Test Theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- Gelbal, S. (1994). p madde güçlük indeksi ile Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 10, 85-94.
- Hambleton, R. K. (1988). Principles and selected applications of Item Response Theory. In R. L. Linn (Ed.), *Educational Measurement* (pp. 147-200). New York, NY: Macmillan.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Karadağ, N. (2014). *Mega Üniversitelerde Ölçme ve Değerlendirme*. Yayınlanmamış Doktora Tezi. Anadolu Üniversitesi. Eskişehir.

- Linden, W. J. v. d., & Hambleton, R. K. (1997). *Handbook of Modern İtem Response Theory*. New York, NY: Springer.
- MacDonald, P., & Paunonen, S. V. (2002). A Monte Carlo comparison of item and person statistics based on İtem Response Theory versus Classical Test Theory. *Educational and Psychological Measurement*, 62(6), 921-943.
- Özer Özkan, Y. (2014). Öğrenci başarılarının belirlenmesi sınavından klasik test kuramı, tek ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması. *International Journal of Human Sciences*, 11(1), 20-44.
- Progar, Š., & Sočan, G. (2008). An empirical comparison of Item Response Theory and Classical Test Theory. *Horizons of Psychology*, 17(3), 5-24.
- Wiberg, M. (2004). *Classical Test Theory vs. Item Response Theory*. An evaluation of the theory test in the Swedish driving-license test Umeå University.
- Young, John W. (1991). Gender bias in predicting college academic performance: a new approach using Item Response Theory. *Journal of Educational Measurement*. 28(1), 37-47.

Yazarlar Hakkında

Yrd. Doç. Dr. Murat AKYILDIZ



Psikoloji lisans eğitimi aldıktan sonra yüksek lisansını ve doktorasını ölçme ve değerlendirme alanında tamamladı. Çok boyutlu ve çok parametrelili lojistik modellerle bireylerin bilişsel ve duyuşsal özelliklerinin kestirilmesi temel ilgi alanıdır. Diğer ilgi alanları arasında psikometri, istatistiksel analiz, bilim yöntemi ve felsefesi yer almaktadır.

Tel (İş): +90 222 3350580/2410
Eposta: muratakyildiz@anadolu.edu.tr

Dr. Murat Doğan ŞAHİN



Yüksek Lisans ve Doktora eğitimini Hacettepe Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalında tamamladı. Temel ilgi alanları arasında bireyselleştirilmiş testler ile nicel araştırma yöntemleri ve psikometri yer almaktadır. Halen Anadolu Üniversitesinde görev yapmaktadır.

Tel (İş): +90 2223350580/5874
Eposta: mdsahin@anadolu.edu.tr