

Açık Uçlu Soruların Çok Yüzeyle Rasch Modeli ile Analizi

Şirin Çetin^{1*}

¹Biyoistatistik/Tıp Fakültesi, Gaziosmanpaşa Üniversitesi, Tokat, Türkiye

*Corresponding author: cetinsirin55@gmail.com

+Speaker: cetinsirin55@gmail.com

Type: Oral / Full Paper

Özet – Sınavlarda ölçme aracı olarak kullanılan açık uçlu sorulardan oluşan sınavlarda yansız puanlama önemlidir. Açık uçlu sorular, öğrencilere kendilerini özgür bir biçimde ifade etme olanağı sunmakta ve çeşitli becerilerin farkındalığına imkân sunmaktadır. Ayrıca kişilerin bazı konularda olan yetersizliklerini görmeyi sağlamaktadır. Ancak açık uçlu soruların birçok olumsuz noktaları da bulunmaktadır. Özellikle puanlayıcıların yanlılığı açık uçlu sınavların değerlendirilmesinde hataların ortaya çıkmasına sebebiyet vermektedir. Bu hataları ortadan kaldırmak yada açık uçlu sınav değerlendirmelerin de hatayı en aza indirmek için farklı analiz yöntemlerine ihtiyaç duyulmaktadır. Bu sebeple çalışmamızda klasik test teorisine alternatif olarak sunulmuş Madde Yanıt Teorisi ile açık uçlu soruların analizi amaçlanmıştır.

Anahtar Kelimeler: Madde Yanıt analizi, Açık Uçlu Soru, Çok Yüzeyle Rasch Modeli

Analysis of Open-Ended Questions with Many Facet Rasch Model

Abstract – Neutral scoring is important for open-ended questions used as a measurement tool in exams. Open-ended questions provide students to express themselves freely and give them opportunity to be aware of their various skills. Also it provides people to see their inadequacies in some subjects. But open-ended questions include a lot of negative points as well. Especially partiality of raters leads to errors in the evaluation of open-ended exams. Different methods of analysis are needed to eliminate these errors or to minimize the errors in open-ended exam evaluations. For this reason, in our study, it has been aimed to analyze open-ended questions with "Item Response Theory" presented as an alternative to "Classical Test Theory"

Keywords– item analysis, Open-ended questions, many facet Rasch model

I. GİRİŞ

Günlük yaşantımızda ve bilimsel çalışmalarda ölçme önemli bir yer tutmaktadır. Ölçme herhangi bir eğitim sürecinin değerlendirmesinde sık kullanılan bir araçtır. Eğitim sürecinde araştırılan ve ölçülmeye çalışılan değişkenler; başarı, ilgi, motivasyon ve yetenek gibi psikolojik değişkenlerdir. Bu değişkenlerin fiziksel özellikleri ve fiziksel boyutu tanımlanamadığından dolayı doğrudan ölçülemez. Eğitim sürecinde bu değişkenleri ölçülebilir duruma getirmek için değişkeni ölçülebilir özellikler yardımı ile tanımlamak amacı ile çeşitli ölçme araçları geliştirilmiştir. Günümüzde eğitimde sıklıkla kullanılan ölçme araçları, sözlü sınavlar, yazılı sınavlar ve çoktan seçmeli testlerdir. Çoktan seçmeli testler belirli sayıdaki yanıt seçeneği arasından sadece birini seçerek yanıtlanan maddelerden oluşan ölçme araçlarıdır. Çoktan seçmeli testlerde öğrenci cevabı kendisi oluşturamamaktadır. Öğrencinin sunulan seçeneklerden uygun olanı belirleyerek doğru cevaba ulaşması esastır [1,2]. Çoktan seçmeli sorulardan oluşan testler, daha kısa sürede daha fazla sayıda beceri ölçülebilmesi, puanlanma kolaylığı ve

objektifliği gibi üstünlüklere sahiptir [3]. Açık uçlu sorular, çoktan seçmeli testler gibi objektif bir biçimde puanlanamamaktadır [4]. Açık uçlu sorunun puanlanması, puanlamayı yapan kişiye göre farklılık gösterebilmektedir. Açık uçlu soruların en büyük dezavantajı çoktan seçmeli sorularda olduğu gibi objektif olarak puanlanmaması nedeniyle bir ölçme aracının en önemli özelliklerinden biri olan güvenilirliğinin puanlayıcıdan kaynaklanarak düşmesidir [5,6].

Ölçme araçlarının birbirleri ile benzer ve birbirlerinden üstün yanları bulunmaktadır. Öğrencinin bildiklerini ne zaman, nasıl, nerede ve niçin uygulayacağı ile yeterlilikleri çoktan seçmeli testler yapılarak ölçmek mümkün değildir [7]. Açık uçlu sorularda ise puanlamanın objektif olarak gerçekleştirilmesi çok mümkün olmamaktadır. Eğitim ve meslek hayatında bireylerin biyoistatistik bilgilerini kullanmalarını gerektiren koşullar, sunulan seçeneklerden herhangi birini seçmesini gerektiren durumlardan çok daha farklıdır. Bu sebeple çoktan seçmeli testler, bireylerin günlük hayatta karşılaşacağı biyoistatistik

problemlerine çözüm üretebilecek donanıma sahip bireyler yetiştirmede yetersiz kalmaktadır [8].

Her ne kadar açık uçlu soruların da ölçmede eksik yönleri bulunsada öğrencinin sunulan seçeneklerden herhangi birini seçmesi yerine; yanıtı kendisinin oluşturması bireylere daha çok katkıda bulunmaktadır [9].

Bu araştırmada açık uçlu biyoistatistik sorularında puanlayıcı etkilerinin Çok Yüzeyle Rasch modeli ile incelenmesi amaçlanmıştır.

Çok Yüzeyle Rasch Modeli'nin eğitimle ilgili pek çok araştırmada kullanıldığı görülmektedir. [10]

Madde yanıt teorisi (MYT) ile klasik test teorisi (KTT) Ölçme sonuçlarını analiz etmede kullanılan iki temel teoridir. MYT, KTT'nin kısıtlılıklarını öne sürerek ortaya çıkmıştır. Bu teoriye modern test teorisi de denilmektedir. MYT'nin, KTT'ye göre pek çok avantajı bulunmaktadır. KTT'de madde ve test parametreleri örneklemeğe bağımlıdır. MYT'de ise madde parametreleri bireylerin incelenen özellik düzeylerinin dağılımından bağımsız olup; bireylerin incelenen özellik düzeyleri de ölçme aracındaki maddelerden bağımsızdır [11].

MYT analizinde yetenek kestirimlerini elde etmek amacıyla analize dâhil edilen puanlayıcı, madde, ve kişi yüzeyleri logit olarak adlandırılan ortak bir eksen üzerine yerleştirilmeye çalışılmaktadır. Klasik test teorisinde doğrudan ham puanlar üzerinden işlem yapılırken; MYT analizinde her bir yüzeye ilişkin ölçümler eşit aralıklı logit ölçeğine dönüştürülmekte ve tüm yüzeylerin bileşenlerine ait ölçümler de dikkate alınarak, öğrencilerin yetenek düzeyleri hakkında kestirimler yapılmaktadır. Kısacası: kişilerin yetenek kestirimleri, bütün puanlayıcıların maddelerin tümüne verdikleri puanları tamamı üzerinden hesaplanmaktadır. Birey, puanlayıcı ve madde şeklinde üç yüzeyin bulunduğu MYT analizinde, yetenek kestirimlerinin analizde işlem gören üç yüzeyden elde edilen fonksiyon yardımıyla elde edildiği söylenebilir. [12,13,14]

Çok yüzeyle Rasch modeli, tek parametrelili bir model olan madde yanıt teorisinin bir uzantısıdır.

Çok yüzeyle Rasch modeli için n. değerlendiriciden i. kişinin j. maddeden x puan alma olasılığı aşağıdaki formülle verilebilir.

$$P(X = x_{nijx}) = \frac{\exp \left[x(\theta_n - \beta_i - C_j) - \sum_{k=0}^x F_k \right]}{\sum_{r=0}^m \exp \left[r(\theta_n - \beta_i - C_j) - \sum_{k=0}^r F_k \right]}$$

Burada $x = 0, 1, \dots, m$

θ_n : n. kişinin incelenen özellik düzeyi

β_i : i. maddenin zorluk düzeyi

C_j : j. Değerlendiricinin tutumu

F_k : (k-1). kategoriye göre k. kategoriye gözlemlemenin zorluğu

Çok Yüzeyle Rasch (ÇYRM) analizinde puanlayıcılar arası farklılığın olması bir sınırlılık olmamakta; bu farklılıklar kontrol altında tutularak birey ve maddelere ilişkin istatistikler bu şekilde hesaplanmaktadır [14].

ÇYRM, yansız ve etkili bir ölçme elde etmek için yüzeyleri ortak bir düzlemde birleştirerek standart hale getirmekte ve bireylerin bilimsel araştırma becerilerini, soruların güçlüğü ve puanlayıcı katılık ya da cömertlik davranışlarını aynı anda karşılaştırma olanağı sunmaktadır.

II. YÖNTEM

Araştırmanın çalışma grubunu Tokat Gaziosmanpaşa Üniversitesi'nde, 2019–2020 eğitim öğretim yılında "Biyostatistik" dersini alan 100 lisans öğrencisi oluşturmaktadır.

Veri toplama aracı geliştirilirken öncelikle Biyoistatistik eğitimi alanına ilişkin literatürdeki açık uçlu çalışmalar incelenmiştir. Çalışma kapsamında ele alınması düşünülen her bir stratejiyi temsil edecek şekilde 12 adet açık uçlu sorulardan oluşan sınav soruları hazırlanmıştır.

III. BULGULAR

Araştırmanın verileri öncelikle araştırmacılar tarafından uygulamadan önce belirlenmiş olan puanlama kriterlerine uygun olmak üzere bağımsız iki araştırmacı tarafından ele alınmıştır. İki araştırmacı birbirinden bağımsız olarak, öğrencilerin sorulara verdikleri yanıtları 1-5 puan aralığında puanlamıştır. Bu puanlamaya göre (1) çözüm veya seçilen strateji tamamıyla yanlış, (2) çözüm veya seçilen strateji yanlış (3) çözüm ve seçilen strateji kısmen doğru, (4) çözümün büyük bir kısmı doğru ve seçilen strateji uygun, (5) çözüm tamamıyla doğru ve seçilen strateji uygun şeklinde sınıflama yapılmıştır. Normal dağılım göstermeyen puanlar için Spearman-Brown korelasyon katsayısı hesaplanmış ve puanlayıcılar arası güvenirliliğin 0.89 olduğu bulunmuştur. Rasch analizinde güvenirlilik kat sayısı, tıpkı KR-20 veya Cronbach Alpha gibi aynı anlamda hesaplanarak yorumlanır (Gözlenen varyansın gerçek varyansa oranı).

Araştırmanın veri toplama aracında kullanılan problemlerden 3. soru hariç diğer tüm sorularda değerlendirme açısından 1. puanlayıcının 2. puanlayıcıya göre daha hoşgörülü davrandığı görülmüştür.

4. soru her iki puanlayıcı için de en kolay madde iken, 2. soru puanlayıcı 1 için, 8. soru da puanlayıcı 2 için en zor sorulardır. puanlayıcılar arasındaki en büyük farklılık 7. ve 10. sorularda gözlemlenmiştir.

IV. TARTIŞMA

Araştırmada lisans öğrencilerinin biyoistatistik açık uçlu problemlerine ilişkin 12 sorudan oluşan bir problem çözme sınavından aldıkları puanlar iki farklı puanlayıcı açısından çok yüzeyle Rasch analizi modeli ile incelenmiştir. Araştırmanın sonucunda, puanlama işlemine geçilmeden önce puanlama ölçütleri belirlenmiş olmasına rağmen puanlayıcılar arasında puanlama sürecinde farklılıklar olduğu görülmüştür. 1. puanlayıcının sınavda yer alan bir soru haricindeki tüm sorularda 2.puanlayıcıya göre daha hoşgörülü davrandığı tespit edilmiştir. Bununla birlikte puanlayıcılara göre problem çözme testinde yer alan

soruların en zor ve en kolay olanları da değişiklik göstermektedir.

V. ÖNERİLER

Yüksek öğrenimde yapılan sınav değerlendirmelerinde Çok-Yüzeyle Rasch ölçme modelinin etkin olarak kullanılabileceği önerilebilir.

KAYNAKLAR

- 1) Birenbaum, M., & Feldman, R.A. (1998) Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98.
- 2) Simkin, M.G., & Kuechler, W.L. (2005). Multiple-choice tests and student understanding: What is the connection. *Decision Sciences Journal of Innovative Education*, 3(1), 73-97.
- 3) Haladyna, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*. USA: Allyn & Bacon.
- 4) Tuckman, B.W. (1991). Evaluating the alternative to multiple-choice testing for teachers. *Contemporary education*, 62(4), 299-300.
- 5) Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher*, 94(1), 31-37.
- 6) Takunyacı, M. (2016). Çoktan seçmeli sorulara dayalı olmayan bir kitle matematik sınavı sürecinin değerlendirilmesi: Grup uyumu değerlendirme modeli, Doktora Tezi, Marmara Üniversitesi.
- 7) Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.
- 8) Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG*. New York, NY: Taylor and Francis.
- 9) Zhu, X. (2009). *Assessing fit of item response models for performance assessments using bayesian analysis*. Unpublished Doctoral Thesis, University of Pittsburgh, Pittsburgh, ABD.
- 10) Şahin, M. G., Taşdelen Teker, G. ve Güler, N. (2016). An analysis of peer assessment through many facet rasch model, *Journal of Education and Practice*, Vol 7, No 32.
- 11) Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Sage Publications; 1991).
- 12) Linacre, J.M. (1989, *Many-facet Rasch measurement*. Chicago: MESA Press.;
- 13) Lunz, M.E., & Wright, B.D. (1997), *Latent trait models for performance examinations*. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 80-88). Münster, Germany: Waxmann.).
- 14) (Linacre, J. M. (1989). *1989: Many-facet Rasch measurement*. Chicago, IL: MESA Press.