

## 2011 Yılı Yükseköğretime Geçiş Sınavı Hakkında ODTÜ Eğitim Fakültesi Görüşü

Ülkemizde çoktan seçmeli test sorularına dayalı ölçme ve değerlendirme uygulamaları, eğitim sistemimizin çeşitli aşamalarında öğrenciler hakkında önemli kararlar vermek amacıyla kullanılmaktadır. Bu sınavlar arasında yer alan üniversite giriş sınavları, gerek etkilediği bireylerin sayısı gerekse ortaya koyduğu sonuçların çeşitli kurumlar ve süreçler üzerindeki etkisi bakımından en fazla göz önünde olan ve tartışılan sınav niteliğindedir. Üniversite eğitime olan talebin büyüklüğü ve üniversite kontenjanlarının bu talebi karşılayacak büyüklükte olmaması nedeniyle 40 yıla yakın bir süredir uygulanan bu sınavın ilk aşaması geçen yıldan bu yana Yükseköğretime Geçiş Sınavı (YGS), ikinci aşaması ise Lisans Yerleştirme Sınavı (LYS) adı altında yapılmaktadır. Bu sistem içinde seçme ve yerleştirme amacıyla kullanılan çoktan seçmeli testlerin önemli eksiklikleri mevcuttur.

Ölçme ve değerlendirme eğitim bilimleri içinde önemli bir teknik alandır ve uzmanlık gerektirmektedir. Genel olarak bir seçme sisteminin, testi alan tüm bireylerden doğru ve eksiksiz bilgi toplaması bireyler için verilen kararların isabetliliğini arttırma açısından oldukça önemlidir. Genel olarak özetlenecek olursa bir test programının ve uygulamaların aşağıdaki niteliklere sahip olması beklenmelidir:

1. Puanların mümkün olduğu ölçüde hatasız olması gerekir. Bu ölçme alanında güvenilirlik olarak tanımlanmaktadır. Güvenirliği etkileyen pek çok faktör vardır. YGS gibi geniş ölçekli testlerde öğrencilerden güvenirliliği yüksek ölçme sonuçları elde etmek için belli standartlara uyulması gerekmektedir. Bunların bir kısmı ölçme araçlarının niteliğine ilişkin standartlardır (Crocker, ve Algina, 1986)
2. Testin uygulama amacına hizmet etmesi gerekmektedir. YGS testlerindeki amaç öğrencilerin belirli bir kapsamda bilgi ve beceri düzeyini belirlemektir (ÖSYM, 2010). Bunlar kısmen öğretim programlarında ele alınan hedefler olarak tanımlanmaktadır. Dolayısıyla YGS sınavlarının ölçtüğü niteliklerin çok iyi irdelenmesi ve bu niteliklere sahip bireylerin üniversitedeki programlara seçilmesi, sahip olmayanların seçilmemesi sayılına dayanan sürecin doğru işletilmesi gerekmektedir. Ölçme alanında bu konu geçerlik olarak tanımlanmaktadır. Geçerlik, güvenirlilik gibi pek çok faktörden etkilenmektedir. YGS gibi geniş ölçekli ve önemli kararların verildiği sınavlarda belli standartlara ve koşullara dikkat edilmesi kaçınılmaz olarak önem kazanmaktadır. Bunlardan bir kısmı testlerin içerikleri ile ilgilidir. Örneğin test içerikleri, test planlarında hazırlanan konu boyutu ve düşünme süreçlerini kapsamak zorundadır. Aksi takdirde sınavların kapsam olarak geçerli olduğunu iddia etmek mümkün değildir (Crocker, ve Algina, 1986).
3. Testlerde kullanılan soruların farklı gruplar için yanlı sonuçlar ortaya koymaması beklenir. Örneğin testte yer alan her sorunun, içerik bakımından kız ve erkek öğrenciler için aynı anlama gelmesigerektenmektedir. Ya da özel okul ve devlet okulu

öğrencilerinin soruları aynı yanıt örüntüsü içinde yanıtlanması beklenir. Bu kapsamda yapılacak analizler , farklı gruplara ait (örneğin kız ve erkek öğrenciler) aynı yetenek düzeyindeki bireylerin her bir soruya doğru yanıt verme olasılıklarının aynı olması ilkesine dayanmaktadır (Berberoğlu, 1996; Camilli ve Shepard, 1994). Eğer bu olasılıklar farklılaşırsa, test soruları belli grupların lehine çalışacak ve seçme sürecinde bu gruplara avantaj sağlayacaktır. Buna test geliştirme sürecinde madde yanlılılığı ve test yanlılığı denmektedir. Madde yanlılığının da pek çok farklı nedeni vardır. Bunların önemli bir bölümü, test sorularının içeriklerinin iyi hazırlanmamasından ve testin genel olarak iyi planlanmamasından kaynaklanmaktadır (Bakan Kalaycıoğlu ve Berberoğlu, basımda).

4. İyi bir ölçme sürecinin, testi alan bireylerin yetenek dağılımları ile uyumlu olması gerekir. Yetenek dağılımı ile uyumlu olmayan bir ölçme sonucu kişiler hakkında bilgi toplanmamasına, ya da toplanan sınırlı bilgiler çerçevesinde verilecek kararların yanlış olmasına neden olabilir. Örneğin herkesin doğru yanıtladığı sorularla verilecek seçme kararları, ya da büyük çoğunluğun doğru yanıtlamadığı sorularla verilecek seçme kararları doğru olmayacaktır. İlgili literatürde bu teknik sorunlara tavan ve taban etkisi denilmektedir (Allen ve Yen, 1979).

YGS kişiler hakkında önemli kararların verildiği bir testtir. Bu nedenle yukarıda sıralanan niteliklere dikkat edilerek hazırlanmalıdır. Bu niteliklere uyulmadan hazırlanacak sınavlarda sorun yaşanacağı açıktır. YGS sınavlarında kullanılan testler uluslararası normlara göre standart test kapsamına girmemektedir. Çünkü soruların ön uygulaması yapılmamakta, testlerin normları hazırlanmamakta, her sene farklı sorulardan oluşan sınavlarla önemli kararlar alınmakta, ancak bu sınavlarda kullanılan testler eşleştirilmediği için bir yıldan diğerine karşılaştırma yapılamamaktadır. Ancak yine de verilen kararların önemi dikkate alındığında, YGS testlerinin belli standartlara göre hazırlanması gerekmektedir.

Yukarıdaki sıralanan nitelikleri sağlamak amacı ile bir testin hazırlanmasında dikkat edilmesi gereken temel ilkeler bulunmaktadır. Bunlardan en önemlisi testlerin bir bireyden diğerine aynı standart bilgiyi topluyor olmasıdır. Bu standartlık, yukarıda sıralanan niteliklerin sağlanması için gerekli olan kimi daha karmaşık kimi de en temel ve yaklaşık yüz yıldır kullanılan uygulamaları da kapsamaktadır. Bu ilkelere uymak yukarıdaki nitelikleri sağlamak için bir ön koşuldur, ancak yeterli değildir. YGS gibi bir sınavın daha profesyonel bir yaklaşımla hazırlanması gerekmektedir. Bu nedenle yıllar içinde uygulamalardaki niteliklerin artırılması beklenen bir durumdur. Ancak bir test hazırlama sürecinin ayrılmaz bir parçası olan ve ÖSYM'nin daha önceki yıllarda dikkate aldığı kimi temel ilkeleri de göz ardı etmemek gerekir. Şöyle ki;

1. Seçenekler algısal hatalar nedeniyle doğru yanıt vermeye engel yaratmamak için sayısal büyüklük sırası dikkate alınarak, sözelse ve harfler kullanılıyorsa alfabetik sıra dikkate alınarak sıralanır (Haladyna, 2004; Linn ve Miller, 2005).

2. Sorular öğrencilerin anlayacağı düzeyde belli kurallara uygun olarak hazırlanmalıdır. Özellikle çoktan seçmeli soru yazmak ciddi bir uzmanlık gerektirmektedir. Yukarıda da belirtildiği gibi soru ile neyin ölçüldüğü test planları ile tanımlanmalı ve sorular kapsam açısından belli ölçütlere uygun hazırlanmalıdır (Downing, 2006; Millman ve Greene, 1989). Test soruları kapsam olarak belli bir sıra içinde öğrencilere sunulmalıdır. Genelde ön uygulamaya dayanarak elde edilen soru güçlükleri, bu sıralamanın yapılması için dikkate alınır. Ancak YGS gibi ön uygulama yapılmadığı durumlarda da sorunun kapsamına dayanarak yapılacak değerlendirme ile belli bir sistematik sıraya dikkat edilmelidir.

YGS'nin hazırlanmasında geliştirilmesi gereken pek çok durum mevcuttur. Bu geliştirme sürecinin yukarıda sıralanan dört niteliğe uygun olması kaçınılmazdır. Bunun yanı sıra soruların test içindeki sırasına dikkat edilmesi, seçeneklerin büyüklük sıralarına dayanarak verilmesi gibi en temel ve artık üzerinde hiçbir şekilde tartışılmayan ve herkesçe bilinen ve uygulanan ilkeler de mutlaka kullanılmalıdır. Bu ilkelere önceki yıllarda ÖSYM tarafından özellikle dikkat edilmiştir.

ÖSYM, Şubat 2011 tarihinden itibaren tüm sınavlarında “hak ve adaleti” sağlamak amacıyla “adaya özgü soru kitapçığı” uygulamasına başlamıştır. 21 Mart 2011 tarihli ÖSYM basın duyurusunda, 27 Mart 2011 tarihinde yapılan YGS soru kitapçıklarıyla ilgili şu açıklama yapılmıştır: “Soru kitapçıklarındaki sorular birbirinin tamamen aynı olacak, ancak soruların veya sorulardaki seçeneklerin sıralanışı ile doğru cevap seçeneğinin yeri değişebilecektir. Bu nedenle her adayın cevap anahtarı da farklı olacaktır.” Bu uygulama, yukarıda verilen genel ölçme ve değerlendirme nitelik ve ilkelerine ters düşmektedir. Şöyle ki;

1. En temel ölçme ilkelerinden biri soruların test içinde zorluk sırasına göre verilmesidir. Ancak 1.700.000 aday için ayrı ayrı hazırlanan formlarda bu ilkeye ters düşüldüğü için adaylar arasında adil bir ölçme yapılmasını tehlikeye atacak bir uygulama ortaya çıkmaktadır. Bu kadar çok test formunun, soruların güçlük sırası dikkate alınarak hazırlanması mümkün değildir. Soruların test formları içinde farklı yerlerde olması belli istatistiksel parametreleri değiştirdiğinden test formları arasındaki benzerlik bozulmaktadır. Test soruları, formlar içinde farklı yerlerde oldukları için öğrenciler tarafından daha zor ya da daha kolay algılanabilmektedir (Balch, 1989; Impara ve Foster, 2006; Laffitte, 1984; Pettijohn ve Sacco, 2007). Ya da formlar arasında yanlılık gösterebilmektedir. En doğru uygulama tek bir formun olmasıdır ve böylece her aday için aynı ölçme koşulları sağlanmış olacaktır. Ancak ÖSYM yıllar içinde 4 ya da 5 form uygulamış, son yıllarda da form sayısını 10'a çıkarmıştır. Form sayısı arttıkça farklı formları alan bireyler arasında adaletsizlik artacaktır. Büyük olasılıkla sorular yanlı çalışacaktır. Farklı test formları kullanmanın geçerlik ve güvenilirlik ölçülerini nasıl etkilediğini istatistiksel olarak test etmek, başka bir deyişle formlar arasında farklılıkların olup olmadığını anlamak mümkündür, ancak var olan veri üzerinden 1.700.000 form için her forma yanıt veren birey sayısı 1 olduğundan ne ölçüde bir

yanlılığın olduğunu istatistiksel olarak analiz etmek ve değerlendirmek mümkün değildir. Ancak yapılan araştırmalardan hareketle formlar arasında soru parametrelerinin belli ölçüde değiştiğini biliyor olmamız, 1.700.000 form için kontrolü ve etkisi belirlenemeyen bir durum yaratacağı için 10 formda ortaya çıkabilecek adaletsizlikten daha büyük sorunlar yaşanacağını söylemek mümkündür. Eğer araştırmalara dayanarak bir karar verilmesi gerekiyorsa YGS testlerindeki form sayılarını arttırmak değil, tam tersine azaltmak gerekmektedir.

2. Çoktan seçmeli testlerde uyulması gereken diğer bir temel ilke sayısal cevap seçeneklerinin büyükten küçüğe ya da küçükten büyüğe göre sıralanmasıdır. Bu sıraların değiştirilmesi yukarıda dile getirilen endişeyle paralel olarak bireyler arasında adil olmayan bir durumun doğmasına neden olacaktır. Bu sınav bir ölçme aracıdır ve öğrenciyi yanıltmak ya da ipucu vermek gibi bir amacı olamaz. Bu nedenle en yalın şekilde hazırlanmalıdır. Aksi takdirde öğrenciler üzerinde yeteri kadar sınav baskısı varken bir de soruların ve seçeneklerin sıralamasının uygun olmamasının hem dikkat hem de konsantrasyon açısından sorun yaratacağı yapılan pek çok bilimsel araştırma ile de doğrulanmıştır (Hambleton ve Traub, 1974; Neely, Springston ve McCann, 1994)
3. Sınavın hazırlanması, çoğaltılması, dağıtılması ve uygulanması sırasında alınabilecek önlemlerle sınav güvenliğini sağlamak mümkün iken sınavın geçerliğini ve güvenilirliğini olumsuz etkileyecek, ÖSYM tarafından da bu yıl ilk defa kullanılan şifreleme sistemi bu tür geniş ölçekli sınavlarda bilimsel dayanağı olmayan bir uygulamadır. ÖSYM daha önceki yıllarda seçenekleri karıştırmadan uygulama yapmakta idi. Seçeneklerin karıştırılmasının soruların parametrelerine olumsuz etki yaratacağı kaçınılmazdır. Hem soru yerlerinin farklı olması hem de seçeneklerin bireyden bireye karıştırılarak verilmesi bireyler arası adaletsizliği daha da arttıran bir etkiyi kaçınılmaz olarak yaratacaktır. Kaldı ki işin en sorunlu yanı her formu bir aday yanıtladığı için bu olumsuz etkinin ne ölçüde büyük olduğunu kestirmek ve anlamak istatistiksel olarak mümkün olmayacaktır.
4. Standartlardan uzaklaşma bu yıl sınavı iptal edilen YGS ve ALES öğrencilerine yeni bir sınav uygulaması ile daha da büyümektedir. Test formlarındaki basım hatalarından dolayı sınava giremeyen öğrencilere soruları farklı yeni testler uygulanacaktır. Ancak bu formlardan elde edilen sonuçları daha önceki uygulamalarla karşılaştırmak ve birlikte yorumlamak teknik olarak mümkün değildir. Burada bir başka önemli ölçme ilkesine ters düşülmektedir. O da soruları farklı olan ve puanları eşleştirilmeyen sınav sonuçlarının (test formlarının) hiçbir koşul altında karşılaştırılmaması ilkesidir (Kolen ve Brennan, 2004). Ancak ne yazık ki ÖSYM bir kez daha ölçme kuramlarına ve uygulamalarına ters düşen ciddi bir hata yapmaktadır.

Sonuç olarak, geniş ölçekli testlerin hazırlanmasında ve uygulanmasında ÖSYM'nin yaptığı gibi her adaya özgü soru kitapçığı hazırlama bilimsel olarak psikometrik ölçütlere dayanan, kabul görmüş ve dünyada Educational Testing Service (ETS), Cito, Pearson gibi test geliştiren

ve uygulayan kurumlar tarafından uygulanan bir sistem deęildir. Yukarıda sıralanan ölçme sonuçlarının niteliklerine yönelik kaygılar nedeni ile böyle bir uygulama, geniş ölçekli testlerde kullanılmamaktadır.

Yukarıda sıralanan ölçme ilkelerinden sonuncusu yetenekle uyumlu test formlarının hazırlanmasına yöneliktir. ÖSYM'nin bu anlamda da eksiklikleri vardır. Özellikle sayısal bölümdeki test formlarının çok az bir kitle tarafından yanıtlanıyor olması bu alt testlerde 'taban etkisi' yaratmakta, tüm kitleden bilgi toplanamamaktadır. ÖSYM zaman içinde daha doğru ölçme yöntemlerini kullanarak bireyler hakkında daha isabetli kararlar vermek için çaba harcayan bir kurum olmalıdır. Çalışmaların bu yönde geliştirilmesi gerekirken, ilgili alanda hiç kullanılmayan, bilimsel olarak sakıncalı olan yöntemlere zaman ve enerji harcanarak, sınavın nitelięi düşürülmektedir.

Eęitimde ölçme ve deęerlendirme uygulamaları uzmanlık gerektiren, bilimsel doęrularla yönlendirilmesi gereken önemli bir süreçtir. ÖSYM'nin eęitimde ölçme ve deęerlendirme alanındaki çalışmalarda hiçbir şekilde ele alınmayan bireye özgü test kitapçığı hazırlama uygulaması ile alandaki bilimsel doęrular dışında hareket edilmektedir. YGS testlerinin eęitim sistemimize dersane, soru üzerinden eęitim yapılması gibi son derece olumsuz yansımaları bilinen bir gerçektir. Bu sınavda kullanılan soru tarzı ve formatlarının sistemde yarattığı olumsuz etkilerin giderilmesi yönünde tedbirler almak yerine, uygulamada geniş ölçekli testlerde kullanılmayan yöntemlere zaman, para ve emek harcanmasını sistemin geçerliğini ve güvenilirliğini olumsuz yönde etkilemesi nedeni ile doğru bulmuyoruz. ÖSYM'nin bir an önce yaptığı yanlışlıktan geri dönmesini, bundan sonra yapılacak uygulamalarda ölçme ve deęerlendirme çalışmalarının gereęi olan bilimsel doęrulara göre hareket etmesini beklemekteyiz.

Saygılarımızla.

Orta Doęu Teknik Üniversitesi

Eęitim Fakültesi

Kaynaklar:

Allen, M.J., ve Yen, W.M. (1979). Introduction to measurement theory. Belmont, CA: Wadsworth.

Bakan Kalaycıoğlu, D., ve Berberoğlu, G. (basımda). Differential item functioning analysis of the science and mathematics items in the university entrance examinations in Turkey. *Journal of Psychoeducational Assessment*.

Balch, W.R., (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology*, 16(2), 75-77.

Berberoğlu, G. (1996). The university entrance examinations in Turkey. *Studies in Educational Evaluation*, 22(4), 363-373.

Camilli, G., ve Shepard, L.A. (1994). Methods for identifying biased test items. Thousand Oaks: Sage.

Crocker, L., ve Algina, J. (1986). Introduction to Classical and Modern Test Theory (pp. 217 – 242). New York: CBS College Publishing

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing ve T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum.

Hambleton, R. K., ve Traub, R. E. (1974). The effects of item order on test performance and stress. *Journal of Experimental Education*, 43, 40-46.

Impara, J.C. ve Foster, D. (2006). Item and test development strategies to minimize test fraud. In S. M. Downing ve T. M. Haladyna (Eds.), *Handbook of test development* (pp. 91-114). Mahwah, NJ: Lawrence Erlbaum.

Kolen, M.J. ve Brennan, R.L. (2004). *Statistics for social science and public policy: Test equating, scaling, and linking methods and practices*. Springer.

Laffittee, R.G., (1984). Effects of item order on achievement test scores and students' perceptions of test difficulty. *Teaching of Psychology*, 77(4), 212-214

Linn, R.L., ve Miller, M.D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson Merrill/Prentice Hall

Millman, J., ve Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 335-366). New York: Macmillan.

Neely, D. L., Springston, F. J., ve McCann, S. J. H. (1994). Does item order affect performance on multiple choice exams? *Teaching of Psychology*, 21(1), 44-45.

ÖSYM. (2010). Ortaöğretim kurumlarına göre 2010 öğrenci seçme sınavı sonuçları kitabı. Ankara: Yüksek Öğretim Kurumu, Öğrenci Seçme ve Yerleştirme Merkezi.

Pettijohn, T. F. II, ve Sacco, M. F. (2007). Multiple-choice exam question order influences on student performance, completion time, and perceptions. *Journal of Instructional Psychology*, 34(3), 142-149.