

DEĞİŞİMİN ÖLÇÜLMESİ

W.R.Borg and M.D.Gall*

Çev: Nilgün Köklü**

Bütün deneyler bağımsız bir değişkenin bağımlı bir değişken üzerindeki etkisini tayin etmek için yapılan girişimlerdir. Eğitim araştırmalarında, bağımsız değişken genelde bir eğitim metodu veya ürünü ve bağımlı değişken öğrencinin başarı, tutum veya kişiliğinin ölçümüdür. Eğer bir etki varsa, bağımsız değişken deneysel işlemde öntest ve sontest olarak uygulanan ölçümlerdeki öğrenci puanlarındaki bir değişim olarak yansıtılmalıdır. Böylece, deneysel desenin önemli bir yanı öntestten sonteste değişimin ölçümüdür.

Değişimin ölçülmesi için öğrencinin öntest puanlarını sontest puanlarından çıkarılmasının yeterli olduğu düşünülebilir. Örneğin, eğer öğrencinin ilk başarı ölçümü 50 ise ve puanı deneysel işlemde sonra 65'e yükseldiyse, kazanılan puan - değişme ya da fark puanı- 15 olacaktır. Bununla birlikte, deneysel bir işlemin etkilerini belirlemede değişme puanının kullanılmasında ciddi güçlükler vardır.

Bu güçlükler Paul Dressel ve Lewis Mayhew tarafından yapılan kolej 1. sınıfta başından sonuna kadar kazanılan başarı ile ilgili çalışmada ele alınmıştır. Tablo 1 dokuz kolejin öğrencilerinin çeşitli başarı testlerinde kazandığı puanları vermektedir.

* Walter R.Borg and Meredith D.Gall. Measurement of Change. **Educational Research**. New York: Pitman Pub. 1989.p. 728-733.

**Doç. Dr., A.Ü.Eğitim Bilimleri Fakültesi Eğitim Yönetimi ve Planlaması Bölümü Öğretim Üyesi.

Tablo 1. Öntest Durumlarına Göre Sınıflandırılan Öğrencilerin Sontestte Ortalama Kazançları

TEST	DÜŞÜK GRUP	DÜŞÜK ORTA GRUP	ORTA GRUP	YÜKSEK ORTA GRUP	YÜKSEK GRUP
Sosyal bilimlerde kritik düşünme	6.89	5.48	3.68	4.20	2.26
Bilimsel düşünme ve anlama	6.26	5.16	2.93	2.04	0.31
Sosyal bilimlere katılım envanteri	18.00	5.05	4.94	1.39	-2.07
Okuma ve yazma analizi	5.33	2.89	1.81	1.22	0.25
Kritik düşünme	6.68	4.65	3.47	2.60	1.59
İnançlar envanteri	9.09	5.31	4.65	3.32	1.01
İnsan ilişkilerinde problemler	3.19	1.67	1.31	1.51	-0.36

Kaynak: University of Chicago Yayınlarının izniyle, Paul B. Diederich."Pitfalls in the Measurement of Gains in Achievement" in yeniden basımı. Telif hakkı : 1956 University of Chicago. School Review 64 (1956) , tablo 1, s.60 'dan yeniden basımı.

Kazanılan puanlar her bir testte öntest puanlar esas alınarak oluşturulan alt gruplara göre sunulmaktadır (Tablo 1). Öntest puanı ile kazanılan başarı arasında kuvvetli ters ilişki vardır. Örneğin, "Sosyal bilimlerde kritik düşünme" testinde yılın başında en düşük puanı alanlar (6.89) yüksek puan alanlara göre (2.26) daha fazla kazanmışlardır.

Bu veriler nasıl yorumlanabilir? Bunlar ilk başarısı düşük olanların ilk başarısı yüksek olanlara göre daha fazla öğrendikleri (değişme puanlarıyla ölçüldüğü gibi) anlamını verebilir mi? Her ne kadar bu yorum doğru da olsa, öntest puanları ve başarı kazanma puanları arasındaki ters ilişkinin öntest ve sontestteki ölçme hatalarından kaynaklanması ihtimali vardır.

Aşağıda, tabii bir büyüme sürecinin sonucu olarak kişilerde oluşan değişimin miktarını ölçmede ham kazanma puanı (son puan eksi ön puan) kullanıldığında ortaya çıkan yorumlama problemlerinden söz edilmektedir.

1. Kazanılan puanlar bir tavan etkisine (**ceiling effect**) tabii olmuştur. Tavan etkisi kavramı test maddelerinin güçlük ranjının sınırlı olduğu anlamına gelir. Bunun için, test ölçülen boyutta mümkün olan tüm başarı ranjını ölçmez. Örneğin, bir öğrenci 100 maddelik bir öntestte 90 maddeyi doğru cevaplırsa, bu kişi sontestte puanını sadece 10 puan artırabilir.. Buna karşılık, bir öğrenci öntestte 40 puan alırsa, olası kazancı 60 puan olacaktır. Böylece, tavan etkisi ilk duruma göre kazanılan puanların dağılımında suni bir sınırlanmaya yer verir. Dressel ve Mayhew' in çalışmasında görüldüğü gibi, orta yüksek ve yüksek grupta olanlar sontesti aldıklarında en düşük kazanılan puanlara sahip olmaktadır.

2. Kazanılan puanların yorumu ortalamaya doğru regresyon etkisi (**regression effect**) ile karıştırılabilir. Bu etki ile, öntestte yüksek puan alanlar, sontestte daha düşük puan alabilirler. Buna karşılık, düşük öntest puanlı öğrenciler sontestte daha yüksek puan kazanacaklardır. Regresyon etkisi, testler birbirleriyle ilişkili olduğundan ve öntest ile sontestteki ölçme hatalarından ortaya çıkar.

Regresyon etkisini istatistiksel olmayan terimlerle açıklamak için, bir öğrencinin bir çoktan seçmeli başarı testinde çok yüksek bir puan kazandığını farzedin. Bu belkide öğrencinin “gerçek” puanı değildir. Yüksek puanın en az bir determinanı belkide şans faktörlerinin etkisidir. Örneğin, öğrenci emin olmadığı çoktan seçmeli maddelerden bazılarını şansla tahminde bulunmuş olabilir; veya test şansla öğrencinin bildiği yüksek oranda maddeleri kapsamaktadır. Şimdi öğrenci belkide testin paralel formunu (bir deneyde sonteste eşdeğer) aldığı anda aynı iyi şansa sahip olmayacaktır. Böylece , öğrenci belkide biraz daha düşük puan alacaktır. Ters durum başlangıçtaki düşük puana uygulanır. Bu öğrenci belkide aksi tesadüfle kötü şansından dolayı kısmen düşük puan kazanmıştır. Olasılık kanunlarından dolayı , öğrenci şansı belkide sontest uygulandığı zaman düzelecektir. Böylece, biraz daha yüksek puan alacaktır. Şimdi kazanılmış puanların regresyon etkisiyle kısmen yıkıldığını görebiliriz. Tablo 1’de ilk düşük başarılı öğrenciler daha büyük kazanılmış puanı yapmada onlar için çalışan regresyona sahip olurken, buna karşılık ilk yüksek başarılı öğrenciler onlara karşı çalışan regresyona sahip olurlar.

3. Kazanma puanları testin bütün puanlarında eşit aralıklı olduğunu varsayar, ancak bu sayıltı eğitim ölçmeleri için asla geçerli değildir. Örneğin, 100 maddelik testte, 90’dan 95’ e olan puan kazanımınının 40’ dan 45’ e olan puan kazanımına eşit olduğu varsayılır. Gerçekte ilk puanı 90 olan kişinin (tavan ve regresyon etkilerinden dolayı) 40 olan kişiye göre 5 puan kazanması daha zordur. Eğer test kelime tanımları bilgisini ölçüyor ise, örneğin, ilk puanı 45 olan öğrenci kolay, yani sık sık kullanılan kelimelerin anlamını öğrenerek belki de 5 puan kazanabilir, buna karşılık ilk puanı yüksek olan öğrenci puanını artırmak için zor, yani seyrek kullanılan kelimelerin anlamını öğrenmek zorunda kalabilir.

4. Kuramsal faktöryel testlerin dışında bir test puanı farklı öğrenciler için farklı yetenek seviyelerini yansıtabilir. Örneğin, bir matematik başarı testi toplama, çıkarma, cebir gibi farklı alt testler içerebilir. Aynı puanı alan iki öğrenci için, bu puan farklı bir zayıflık ya da kuvvetlilik örüntüsü yansıtabilir. Örneğin, bir öğrenci çıkarmada zayıf olabilir, fakat soyut problemlerde kuvvetlidir, bunun yanı sıra diğer bir öğrenci çıkarmada kuvvetli soyut problemlerde zayıftır. Belli bir süre sonunda, iki öğrencide aynı puana sahip olabilirler, çünkü kendi eksikliklerini yenmişlerdir. Böylece, birinci öğrenci için kazanılan puan çıkarmadaki ilerlemeyi yansıtırken, ikinci öğrenci için kazanılan puan matematiksel düşünmedeki ilerlemeyi yansıtır. Kazanılan puanlar anlamda eşit olmadığından, istatistiksel olarak karşılaştırılması problemi vardır.

5. Kazanılan puanlarla ilgili diğer bir problem ise onların güvenilir olmadığıdır. Öntest ve sontest puanları arasındaki korelasyon ne kadar yüksekse, değişme puanlarının güvenilirliği o denli düşüktür. Hatta değişme puanlarının güvenilirliği öntestin ve sontestin güvenirsizlik derecesinden etkilenmektedir.

DEĞİŞİMİN İSTATİSTİKSEL ANALİZİ

Değişme puanlarının anlamını ve güvenilirliğini pek çok faktör etkilese de , eğer araştırmacı farklı deneysel işlemlerin etkilerini karşılaştırmak isterse, değişmenin ölçülmesi zorunlu olmaktadır. Her ne kadar değişme puanlarının sınırlılıkları ortadan tam olarak kaldırılamasa da, bu sınırlılıkların bazılarını yenmek için bazı istatistiksel yöntemler mevcuttur.

Birinci olarak, araştırmacının öğrencilerin bireysel olarak değişme puanları ile ilgilenmesi durumu ele alınsın. Örneğin, bazı öğrencilerin sosyal bilimlerde kritik düşünme (SBKD) de diğer öğrencilere göre niçin daha fazla puan kazandıkları bilinmek

istensin. Farzedin ki, kazanılan puanlar öğrencilerin lisedeki başarı puanı ortalaması (BPO) ile ilişkili olduğu hipoteziniz var. Bu hipotezi test etmek için, basitce BPO ile SBKD kazanılan puanlarının korelasyonu alınamaz, çünkü, SBKD puanları regresyon ve tavan etkileri ile bozulmuştur. Öğrencilerin lisedeki BPO puanlarının SBKD yeniden düzeltilmiş kazanılan puanlarıyla olan ilişkisini bulmada daha etkin bir yöntem (13. konuyu gör) kısmi korelasyon kullanmaktır böylece ilk puan sabit tutulur.

Eşdeğer bir yöntem ise öğrencilerin sıntest SBKD puanlarının bağımlı değişken, öntest SBKD puanları ve lise BPO puanlarının yordayıcı (kestirici) değişken olarak alınması durumunda çoklu regresyondur (14.konuyu gör). Öntest SBKD puanları yordama denkleminde önce girer. Önemsiz bir değişkenlik, sadece öntest SBKD puanlarını bir yordayıcı değişken olarak kullanıp **çoklu regresyon** yapmaktır. Sonuçtaki çoklu regresyon eşitliği, her bir öğrencinin sıntest SBKD puanlarını yordamada kullanılabilir. Bu puanlara kalan (residual) veya düzeltilmiş (adjusted) kazanılan puanları denir. Öğrencilerin lise BPO puanları ile kazanılan düzeltme puanları arasındaki korelasyon alınır.

Değişme puanları kullanıldığında diğer bir durum ortalama değişiminin analizidir. İki gruba başarı öntesti veriliyor olsun- bir grup deneysel işlem alıyor, diğer grup ise kontrol grup olarak kullanılıyor olsun. İki grup random olarak oluşturuluyorsa, deneklerin öntest puanlarında eşdeğer ortalamaları olacaktır. Böyle bir durum söz konusu ise, o zaman ortalama değişme puanlarının istatistiksel anlamlılığı için **t-test** kullanılabilir.

Bazen denekler şans yoluyla gruplara random atansa da ortalama öntest puanları arasındaki fark anlamlı çıkabilir. Hatta yarı (quasi) deneysel desenler kullanıldığı zaman öntest ortalamaları farklı olabilir. Öntest ortalamalarında ilk farklılıkları düzeltmede, **kovaryans analizi** kullanılabilir. Bu istatistiksel teknikle, ortalama değişme puanlarına ilk puanlardaki farklılıklardan ziyade deneysel işlemin etkisi olduğu yorumu gerçekleştirilebilir. Örneğin, kovaryans analizi Tablo 1 verilerine uygulanabilir. Sanki birinci

yılın başında aynı ortalama başarı puanını kazanmışlar gibi, bu istatistiksel tekniği kullanarak beş alt grubun herbirinin ortalama kazanılan başarı puanlarını karşılaştırabiliriz . 13. konuda tartıştığımız gibi, kovaryans analizinin gerektirdiği bazı varsayımları karşılayıp karşılamadıklarını belirlemede verilerin açıklanmaya ihtiyacı vardır.

Öntest-sontest değişiminin istatistiksel anlamlılığını belirlemede diğer bir yaklaşım ise, **tekrarlanmış ölçümler için varyans analizidir**. Bu tür bir varyans analizinde, öntest ve sontest olarak uygulanan bağımlı değişken ölçümü bir faktör olarak, deneysel ve kontrol işlemleri ise ikinci faktör olarak ele alınır.

İki faktör için F oranları (bazen esas etkiler olarak adlandırılan) bu varyans analizinin kapsamına girmez. Örneğin, puanların deney ya da kontrol grubu öğrencilerinden olup olmadığına aldırmadan, bütün öntest puanları ortalaması ile sontest puanları ortalamasının karşılaştırılması anlamlı değildir. İlgilenilen durum ölçme zamanı ve işlem arasındaki etkileşimdir; öyle ki deney grubunun öntest ve sontest arasındaki farkının kontrol gruba ait farka göre daha büyük ya da daha küçük olup olmadığı ile ilgilenmektedir.

Bu tür bir istatistiksel analiz Norbert Johnson, Jerome Johnson ve Coy Yates' in çalışmasında yapılmıştır. Deneylerinin amacı öğrencilerin mesleki olgunluğunu düzeltmede bir rehberlik programının (Mesleki İnceleme Grubu) etkililiğinin değerlendirilmesidir. 60 sekizinci ve dokuzuncu sınıf öğrencisi random olarak deney ve kontrol gruplarına atanmıştır. Mesleki Olgunluk Envanteri (MOE) bütün öğrencilere öntest, deneysel programın sonucunda sontest ve programdan 6 ay sonra izleme testi olarak uygulanmıştır. MOE altı ölçekten oluşmakta olup gerçekçi mesleki kararlar verebilmek için gerekli olan tutumların ve yeterliklerin olgunluğunu ölçmektedir

Tablo 2. Mesleki Olgunluk Envanteri (MOE) Öntest, Sontest ve İzleme Testi Ortalama Farkları

MOE ÖLÇEĞİ	Öntest		Sontest		İzleme		Ön-Son	Ön-
	Deney	Kontrol	Deney	Kontrol	Deney	Kontrol	Farkı	İzleme
	X	X	X	X	X	X	F	Farkı
Tutum	34.3	34.3	38.1	35.1	38.0	34.7	4.15*	4.57*
Yeterlik Ölçeği								
Kendini Tanıma	9.3	9.6	12.1	10.3	10.4	10.4	A.D	A.D
İşleri Tanıma	10.1	10.6	15.8	11.3	13.4	11.2	16.13*	12.92*
İş Seçme	11.3	11.6	13.9	12.2	11.9	12.2	A.D	A.D
İleriye Bakma	10.6	11.6	13.1	12.2	12.7	11.8	A.D	A.D
Ne Yapılmalı	9.9	10.1	11.4	10.8	11.4	10.8	A.D	A.D

Kaynak: Norbert Johnson, Jerome Johnson, ve Coy Yates, " A Six-Month Follow-up on the Vocational Exploration Group on Career Maturity," Journal of Counseling Psychology 28 (1981): 71.

Not: A.D.= anlamlı değil.

* $p < .05$. ** $p < .001$.

Tablo 2 MOE'ye ait deney ve kontrol gruplar için öntest , sontest ve izleme sonuçlarını vermektedir. Görüldüğü gibi , her bir ölçek için deneysel grubun öntest-sontest farkı kontrol grubunkine göre büyüktür. Örneğin, Tutum Ölçeğinde öntest-sontest farkı deney grubu için (3.8) kontrol grubundan (0.8) daha büyüktür.

Bu farklılıkların istatistiksel anlamlığı iki yönlü varyans analizi (İşlem Grubu x Test Zamanı) yapılarak tayin edilmiştir. Etkileşim etkisi için istatistiksel olarak anlamlı bir F değeri bir gruba ait öntest-sontest farkının, diğer grubun farkından güvenilir olarak daha büyük ya da daha küçük olduğunu göstermektedir. Tablo 2'nin son iki sütunu, her bir MOE ölçeğinde ön-son farkı ve ön-izleme farkı için F oranlarını vermektedir. Anlamlı F oranları Tutum ölçeği ve Kendini Yeterli Bulma ölçekleri için ortaya çıkmıştır. Çoklu karşılaştırma için t-testleri serisi (13.konuyu gör) deney grubunda anlamlı kazançlar olduğunu göstermiş, kontrol grupta ise bir değişiklik olmamıştır.