

İNGİLİZCE ÖĞRETMENLİK SERTİFİKA GRUPLARINDA YÜRÜTÜLEN ÖLÇME VE DEĞERLENDİRME DERSİNİN ETKİLİLİĞİNİN DEĞERLENDİRİLMESİ

Yrd. Doç. Dr. Ömer KUTLU

Ankara Üniversitesi, Eğitim Bilimleri Fakültesi
kutlu@education.ankara.edu.tr

Öğr. Gör. Dr. Deniz GÜLLEROĞLU

Ankara Üniversitesi, Eğitim Bilimleri Fakültesi
denizgulleroglu@yahoo.com

ÖZET

Bu araştırmanın amacı Ankara Üniversitesi Eğitim Bilimleri Fakültesinde yürütülen İngilizce sertifika gruplarına verilen ölçme ve değerlendirme dersinin etkililiğini belirlemektir. Araştırma, geliştirilmiş olan başarı testine ilişkin bir ön deneme uygulaması ve aynı zamanda “eğitimin etkililiği” için bir örnek çalışma özelliği de taşımaktadır. Bu çalışma, 2008-2009 öğretim yılı güz döneminde Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi İngiliz Dili ve Edebiyatı, İngiliz Dil Bilimi ve Amerikan Kültürü ve Edebiyatı bölümlerinin 4. sınıfında okuyan 97 öğrenci üzerinde yürütülmüştür. Araştırmada öğrencilerin ölçme ve değerlendirme dersinde edindikleri bilgileri yoklamayı amaçlayan 45 maddelik bir bilgi testi ile veri toplanmıştır. Bilgi testi ile toplanan verilere madde analizi uygulanarak, öntest ve sontest uygulamalarından elde edilen madde istatistikleri değerlendirilmiştir. Bunun yanında, bilgi testi ile elde edilen veriler bağımlı gruplar için t-testi istatistiği tekniği kullanılarak çözümlenmiştir. Bağımlı gruplar için t- testi sonuçlarına göre bu gruplara verilen ölçme ve değerlendirme dersinin etkili olduğu sonucuna ulaşılmıştır. Çalışmadan elde edilen bulgular tartışılarak, önerilerde bulunulmuştur.

Anahtar Sözcükler: Eğitimde etkililiğin değerlendirilmesi, etkililik, verimlilik, program değerlendirme

THE EVALUATION OF THE EFFECTIVENESS OF MEASUREMENT AND EVALUATION COURSE CONDUCTING ON ENGLISH LANGUAGE TEACHING CERTIFICATION GROUPS

ABSTRACT

The purpose of this study is to determine the effectiveness of Measurement and Evaluation Course conducting on English Certification Groups at Faculty of Educational Sciences at Ankara University. At the same time, the pilot application of the test was made. The study has a model fort he effectiveness of education. The research was carried on 97 senior students of departments of English Language and Literature, Linguistics and American Culture and Literature at Ankara University in 2008-2009 fall semester. The data were collected by a-45 item test which aims to examine the level of students' achievement in “Measurement and Evaluation” course. By applying an item analysis to the data which is collected by the test the item statistics of pre-test and post-test were evaluated. Besides, the data collected by the achievement test were analyzed. According to the results of the t-test statistics, the measurement and evaluation course given to these groups is effective. The results of the study was discussed and the suggestions were given.

Key Words: Effectiveness in education, evaluating the effectiveness of education, productivity in education.

GİRİŞ

Geleneksel tanımı ile eğitim; bireyin davranışlarında kendi yaşantısı yoluyla, kasıtlı olarak istendik yönde davranış değişikliği meydana getirme sürecidir (Ertürk, 1986). Eğitim ile bireylerin bilgi ve yeteneklerinin geliştirilerek; bilişsel, duyuşsal ve devimsel davranışlarında değişiklik olması amaçlanmaktadır. Bu değişikliği sağlayacak kurumlar okullardır. Okullar, planlı eğitim etkinlikleriyle davranış değişikliklerinin gerçekleştirildiği yerlerdir. Buna göre, öğrenme-öğretme sürecinde düzenlenen etkinlikler bireylere yeni davranışların kazandırılması, kazanılmış olanlardan bazılarının geliştirilmesi ve pekiştirilmesi, bazılarının ise tamamen değiştirilmesi ve güncelleştirilmesi amacıyla gerçekleştirilir.

Eğitim etkinlikleri sonunda, amaçlanan bazı davranışlar istenen düzeyde kazanılmış, bazıları ise beklenen düzeyde kazanılmamış, bazıları da hiç kazanılmamış; hatta bazı istenmeyen davranışlar dahi meydana gelmiş olabilir. Bu durum, eğitim süreci sonunda sonuçların kontrolü ihtiyacını doğurur. Eğitim süreci dört temel öğeden meydana gelmektedir. Birincisi, eğitimde geliştirilmesi amaçlanan bazı davranışların bulunması, ikincisi bu amaçla bir dizi eğitim etkinliğinin düzenlenmesi, üçüncüsü bireylerin yeni davranışlar kazanmalarının sağlanması ve dördüncüsü bunların beklenen düzeyde kazanılıp kazanılmadığının denetlenmesidir. Eğitim sürecinin öğeleri birbirinden bağımsız düşünülemez. Bu öğeler sürekli birbirleriyle etkileşim halindedir (Baykul, 2000).

Eğitim daima bir amaca yönelik olarak yapılır. Sistem içinde eğitimin amaçları hedeflerle ifade edilir ve bu hedefler doğrultusunda plan yapılarak, program geliştirilir. Programın hedeflerine ulaşp ulaşmadığını belirlemek amacıyla da ölçme ve değerlendirmeye başvurulur. Bu nedenle ölçme ve değerlendirme eğitim sürecinde önemli bir yere sahiptir. Ölçme ve değerlendirme bilimi aracılığıyla dersler ve diğer eğitsel etkinliklerle elde edilen başarıyı, öğrenci ve öğretmenin programın amaçlarına ne kadar yaklaşabildiğini, bir başka deyişle verilen eğitimin etkililiğini görmek mümkündür. Etkililik, materyallerin çeşitliliği, ders ortamının niteliği, öğrenci tutumları gibi pek çok açıdan incelense de özellikle son ürün olan öğrenci başarı düzeyinin belirlenerek yıllar içerisinde izlenmesi her nitelik çalışmasında önemli bir boyut olarak ortaya çıkmaktadır (Berberoğlu ve Kalender, 2005). Etkililik, insan çabasının çıktısı ve kalite için konulan hedefleri gerçekleştirme derecesidir (Kutlu, 1998). Baykul (1992) ise

etkililiği, sistemin amaçlarına ulaşabilme derecesi olarak tanımlamıştır. Eğitim sürecinde yer alan girdilerle çıktılar arasındaki farklar eğitimin etkililiği ve verimliliği hakkında ipucu verir. Belli bir programa dayalı olarak yürütülen bir eğitime katılan bireylerin, programda yer alan öğrenmelere ilişkin bilgilerinin, programa dayalı olarak geliştirilmiş bir ölçme aracıyla hem programa katılmadan önce (girişte) öntestle hem de programa katıldıktan sonra (çıkışta) sontestle ölçülmesi gerekmektedir (Kutlu ve Çok, 2002a). Çünkü yürütülen eğitimin verimliliği ve etkililiği hakkında birşeyler söyleyebilmek, girişteki ve çıkıştaki öğrenme düzeyi farklılıklarını incelemekle olanaklı olacaktır. Buna göre, öğretimin başlangıcında ön testle ve öğretimin bitiminde son testle ölçülen öğrenci davranışlarından elde edilen gözlem sonuçlarına dayanarak girişteki ve çıkıştaki öğrenme düzeyi farklılıklarına bakarak, öğretimin verimi hakkında şunlar söylenebilir (Kutlu, 1998):

- Öğrenciler programda yer alan davranışlara girişte sahip olabilirler. Bu durumda eğitim verimsiz demektir.
- Öğrenciler programda yer alan davranışları girişte kazanmamış olarak gelebilirler ve çıkışta da yetersiz kazanmış olabilirler.
- Öğrenciler programda yer alan davranışları girişte kazanmış olarak gelebilirler ve çıkışta da kazanmamış olabilirler.
- Girişte öğrenciler programda yer alan davranışları kazanmamış; ancak çıkışta yeterli düzeyde kazanmış olabilirler. Bu durumda eğitim verimlidir ve istenilen durum da budur.

Bir eğitim programının uygulanmasından sonra beklenen, eğitime katılanların tümüne, eğitim programında yer alan ve programın özel hedeflerinde belirlenen kazanımların kazandırılmasıdır. Bu özel hedeflere dayalı kazanımların katılımcılara ne derece kazandırıldığının belirlenmesi eğitim sürecinin bitiminde ölçme ve değerlendirme yapmayı yani programı değerlendirmeyi gerektirir (Kutlu, 2003).

Verilen bir eğitimin değerlendirilmesinde kullanılan yöntemlerden biri olan test-tekrar yönteminde eğitime katılacak olan gruba eğitim verilmeden önce bir test (ön test) uygulanır. Daha sonra eğitim programı uygulanır. Eğitim sonrası aynı test (son test) aynı gruba aynı koşullar altında tekrar uygulanarak, arada bilgi, beceri ve yeteneklerin geliştirilmesi yönünde olumlu bir farkın olup olmadığı ve bunun öğrenci

performanslarına yansiyıp yansımadığı belirlenir (Uğur ve Kutlu, 2010). Bu çalışmada da öğrencilere dönem boyunca verilen ölçme ve değerlendirme dersinin etkili olup olmadığını belirlemek amacıyla test-tekrar yöntemi uygulanmıştır.

Eğitim-öğretim sürecinde çeşitli alanlarda verilen eğitimlerin etkililiğini ve verimliliğini araştıran çalışmalara yurtiçi ve yurtdışı alanyazında sıkça rastlanmaktadır (Kelecioğlu, 1989; Baykul, 1992; Kutlu, 1998; Kutlu, 2002b; Doyle, 2002; Coşgun ve Ilgar, 2004; Mink ve Fraser, 2005; Butakın ve Özgen, 2007; Bay ve Karakaya, 2009; Barchard ve Pace, 2009). Türkiye’de bu konuda yapılan araştırmalar incelendiğinde Eğitim fakültelerinde okutulan ölçme ve değerlendirme dersinin etkililiğine ilişkin bir araştırmaya rastlanmamıştır. Bu araştırma 2008-2009 eğitim-öğretim yılı güz döneminde Ankara Üniversitesi Eğitim Bilimleri Fakültesi’nde yürütülen İngilizce sertifika grubu öğrencilerine verilen ölçme ve değerlendirme dersinin etkililiğinin belirlenmesini amaçlamaktadır. Bu genel amaç çerçevesinde araştırmada aşağıdaki problemlere yanıt aranacaktır:

- 1) Ölçme ve Değerlendirme Bilgi Testi’ nin ön test ve son test uygulamalarından elde edilen madde güçlük ve madde ayırt edicilik değerleri nasıldır?
- 2) Ölçme ve Değerlendirme Bilgi Testi’ nin ön test ve son test uygulamalarına ilişkin puanların KR 20 güvenilirlikleri nasıldır?
- 3) Ölçme ve Değerlendirme Bilgi Testi’ nin ön test ve son test uygulamalarından elde edilen puanların ortalamaları arasında manidar bir fark var mıdır?

YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama araçları, verilerin toplanması ve verilerin çözümlenmesi ile ilgili bilgilere yer verilmiştir.

Araştırma Modeli

Bu araştırma; 2008-2009 eğitim-öğretim yılı güz döneminde, Ankara Üniversitesi Eğitim Bilimleri Fakültesi’nde İngilizce sertifika grubu öğrencilerine verilen ölçme ve değerlendirme dersinin etkililiğini belirlemeyi” amaçlamaktadır. Araştırma aynı

zamanda bir ön deneme uygulaması niteliği de taşımaktadır. Bu amaç için geliştirilmiş olan bilgi testinden ön test ve son test olarak elde edilen veriler gerekli istatistiksel tekniklerle çözümlenmiş ve bulgular betimlenmiştir. Araştırma daha çok deneme öncesi (tek grup öntest-sontest) model özelliği taşımaktadır (Karasar, 2009; s.96).

Çalışma Grubu

Çalışma grubu, 2008-2009 eğitim-öğretim yılı güz döneminde Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi İngiliz Dili ve Edebiyatı, İngiliz Dil Bilimi ve Amerikan Kültürü ve Edebiyatı bölümlerinin 4. sınıfında okuyan 97 öğrenciden oluşmaktadır. Çalışma grubuna, ilgili programa devam eden öğrencilerin tümü dahil edilmiştir. Ölçme ve Değerlendirme dersi, öğrenciler iki farklı gruba ayrılarak, iki farklı öğretim üyesi tarafından yürütülmüştür. İki öğretim üyesi ölçme ve değerlendirme dersini, gruplar için ortak ders programına ve ortak kaynak kitaplara dayalı olarak yürütmüşler ve gruplarda yer alan öğrencilere ortak çalışmalar vermişlerdir. Öğrencilerin gruplara dağılımı Çizelge 1’de verilmiştir.

Çizelge 1. İngilizce sertifika grubunda yer alan öğrenci sayısı

Gruplar	Öğrenci Sayısı
A Grubu	49
B Grubu	46
Toplam	97

Araştırmanın çalışma grubunu oluşturan ve derse kayıtlı olan öğrencilerin sayısı 97 olmasına rağmen, araştırmanın ön test uygulaması sırasında, bazı öğrenciler derse gelmemişlerdir. Bu nedenle, son testi aldıkları halde ön testleri bulunmayan öğrencilerin yanıt kâğıtlarından elde edilen puanlar çözümlene dışı bırakılmıştır. Dolayısıyla, çözümlenmelerden elde edilen bulgular 73 öğrenci üzerinden yorumlanmıştır.

Veri Toplama Araçları

Araştırmada kullanılan ölçme aracı ve geliştirme süreçleri aşağıda anlatılmıştır.

Bilgi Testi. Program sonunda öğrencilerin, ölçme ve değerlendirme konusunda gerekli davranışları karşıladıkları yeni durumlarda kullanabilmeleri, program kapsamında yer alan temel bilgi ve becerileri kazanmış olmalarına da bağlıdır. Bunlar

her bir kazanımla ilişkili olan kavramlar, ilkeler, yöntemler gibi öğrencinin mutlaka bilmesi gereken temel bilgiler ile bu bilgilerin kullanılma yollarını içeren durumlardır. Bilgi testinde, programda yer alan her bir bilgi ve beceri çoktan seçmeli soru türüyle ölçülmüştür (Kline, 1986; Tekin, 1987; Özçelik, 1989; Linn ve Gronlund, 1995; Haladyna, 1996; Turgut, 1997).

Ölçme ve Değerlendirme Bilgi Testi, test geliştirme süreçleri dikkate alınarak geliştirilmiştir. Bir testin, güvenilirliği ve geçerliği test oluşturulduktan sonra hesaplanıyor olmasına rağmen, yüksek kalitede bir test hazırlamanın en iyi yolu, testin uygun ve doğru bir yöntemle geliştirilmiş olmasıdır (Walsh ve Betz, 1995). Bir ölçme aracı olarak kullanılması düşünülen testin işe yaraması geliştirilmesinden, uygulanmasına ve puanlanmasına kadar geçen süreçte doğru adımların atılmış olmasına bağlıdır. Bu süreç alanyazında “test geliştirme” olarak nitelendirilmektedir. Test geliştirme, bir takım özellikleri önceden bilinen bir ölçme aracının hazırlanması işidir (Baykul, 2000). Thorndike (1982)’ye göre, envanter, test ya da diğer ölçme araçlarının gelişimi, belirli adımlar boyunca ilerleyen ve birçok kararı içeren, testin basımına kadar uzanan karmaşık bir süreçten oluşur.

Test geliştirme, geliştirilen testin gerçek amaç (asıl uygulama) için kullanılmasına kadar bir takım aşamaların yerine getirilmesini gerektiren bir süreçtir. İnsan davranışlarının ölçülmesinde kullanılacak ölçme aracının geliştirilmesi, dikkate alınan ölçme kuramına göre değişebilmektedir. Bu çalışmada kullanılan test, araştırmacılar tarafından “Klasik Test Kuramı”nın sayılıları dikkate alınarak geliştirilmiştir. Bu araştırma ayrıca bilgi testi için bir ön deneme uygulaması niteliği taşımaktadır.

Bu çalışmada kullanılan çoktan seçmeli soruların yazımında “Ölçme ve Değerlendirme Dersi”nin konuları ve kapsamı dikkate alınmıştır. Çoktan seçmeli sorular, programın kazanımlarına dayalı olarak oluşturulan ölçülebilir zihinsel özelliklere göre yazılmıştır. Ölçülebilir zihinsel özelliklerin soruya dönüştürülebilir olmasına özen gösterilmiştir. Soruların hangi zihinsel düzeyi ölçtüğünü belirlemede Haladyna (1997) tarafından geliştirilen aşamalı sınıflama kullanılmıştır. Bu sınıflamanın genel özellikleri Çizelge 2. de sunulmuştur.

Çizelge 2. Haladyna'nın Zihinsel süreçler için Aşamalı Sınıflaması

Anımsama	Herhangi bir nesne ve olguyla ilgili bazı özelliklerin görünce tanınması, sorunca söylenmesi ya da olduğu gibi tekrar edilmesidir.
Anlama	Anımsamadan daha karmaşıktır ve aşağıdaki özellikleri içerir. 1. Olguları, kavramları, prensipleri ve süreçleri (prosedür) tanımlar. 2. Olgu, kavram, ilke ve işlemlerin ilgili özelliklerini seçer veya ortaya çıkarır. 3. Olgu, kavram, ilke ve işlemlere örnek verir.
Problem Çözme	Problem çözme sürecinin adımlarını izleyerek çözüme ulaşma.
Eleştirel Düşünme	Yansıtma, karşılaştırma, değerlendirme ve onlardan karar üretmedir. Zihinsel becerilerin bir birleşimidir. Test maddeleri hazırlamak açısından iki yönlü önemlidir. 1. <i>Değerlendirme</i> : Ölçüt seçme, ölçüt kullanma, ölçüt seçme ve kullanma 2. <i>Öngörüle bulunma</i> : Verilen bir durumda, ne olacağını ve etki eden faktörleri bilme

Araştırmacılar tarafından geliştirilen test, kapsam ve görünüş geçerliği açısından incelenmek üzere program ve belirtke tablosuyla birlikte iki dil uzmanı ile bir program geliştirme uzmanına verilmiştir. Bu kapsamda uzmanlara;

- Soruların Türkçe dil ve anlatım özelliklerine sahip olup olmadığı,
- Soruların yalın, açık ve anlaşılır bir dille yazılıp yazılmadığı,
- Programdaki bilgi ve becerilerle çelişen bir yönünün bulunup bulunmadığı ve
- Bilimsel hata taşıyıp taşımadığı soruları sorulmuş ve bu sorulara yanıt vermeleri istenmiştir.

Uzmanların görüşlerine dayanarak, Ölçme ve Değerlendirme Bilgi Testi'nin beklenen özellikleri taşımaya çalışılmıştır. Ayrıca testte yer alan sorular bir ölçme ve değerlendirme uzmanına da gösterilmiş maddelerin taşınması gereken teknik özelliklere ne derece uygun olup olmadığı yönünde görüşleri alınmıştır. Uzman görüşlerine dayalı olarak düzeltmeler yapılmış ve bu yolla teste uygulama öncesi son biçimi verilmiştir. Gerekli düzeltmelerden sonra 50 olan testin soru sayısı 45 olarak belirlenmiştir

Verilerin Toplanması

Çalışmaya, aynı zamanda dersin yürütücüsü de olan araştırmacıların "Ölçme ve Değerlendirme Dersi"nin amaçlarını ve kazanımlarını öğrencilere tanıtmadan sonra başlanmıştır. Geliştirilen ölçme aracı ilgili gruplara dönem başında ön test, dönem

sonunda da son test olarak uygulanmıştır. Bu uygulama ayrıca bilgi testi için bir ön deneme uygulaması özelliği de taşımaktadır.

Verilerin Çözümlemesi

Verilerin çözümülenmesi işlemlerinde madde analizinden, KR 20 güvenilirlik katsayısına dayalı hesaplamalardan ve t-testinden yararlanılmıştır. Bu çözümleme işlemleri için SPSS 13.0 ve ITEMAN 3.5 istatistik paket programları kullanılmıştır. Elde edilen bulgular programın etkililiğini değerlendirmek amacıyla yorumlanmıştır. Bilgi testinden elde edilen verilerin çözümülenmesi yolları aşağıda kısaca açıklanmıştır.

Bilgi testinden elde edilen puanların madde güçlük indeksleri (p_j) ile madde ayırt edicilik indeksleri (r_{jx}) hesaplanmıştır. Madde güçlük indeksi değeri ile öğrencilerin her bir soruya verdikleri doğru yanıtların yüzdelere bakılmış ve bu yüzdelere öğrenmeler arasında ilişkiler kurulmuştur. Bu yolla soruların ölçülen özelliğe sahip öğrencileri ne derece belirleyebildiği konusunda yorumlar yapılmıştır.

Araştırma kapsamında ilk alt probleme cevap verebilmek için; ön test ve son test uygulamaları sonucunda bilgi testinde yer alan maddelere ilişkin madde güçlük indeksleri (p_j) çift serili korelasyon katsayısına dayalı madde ayırt edicilik gücü indeksleri (r_{jx}) elde edilmiştir. İkinci alt probleme cevap bulabilmek amacıyla, bilgi testinin öntest ve sontest olarak uygulanmasıyla elde edilen puanların KR 20 güvenilirlikleri kestirilmiştir. Üçüncü alt probleme cevap bulabilmek amacıyla ise bilgi testinden elde edilen öntest ve sontest puanlarının ortalamaları arasındaki farkın manidarlığı bağımlı gruplar için t-testi ile test edilmiştir. Elde edilen bulgular 0.01 manidarlık düzeyine göre yorumlanmıştır.

BULGULAR VE YORUMLAR

Bu bölümde; toplanan verilerin, uygun yöntem ve teknikler kullanılarak çözümülenmesi sonucunda elde edilen bulgular, araştırmada yanıt aranmak üzere ele alınan soruların sırası dikkate alınarak sunulmuştur. Bu çerçevede önce yanıtlandırılmak üzere ele alınan araştırma sorusu tekrarlanmış, sonra bu soruya ilişkin bulgulara ve yorumlara yer verilmiştir.

1. Alt Problem: “Ölçme ve Değerlendirme Bilgi Testi’ nin ön test ve son test uygulamalarından elde edilen madde güçlük ve madde ayırt edicilik değerleri nasıldır?

Çizelge 3. Ön test ve son test uygulamasına ilişkin madde analizi sonuçları

Madde No	Ön Test		Son Test	
	Madde Güçlük İndeksi (p_i)	Ayırtedicilik İndeksi (r_{ix})	Madde Güçlük İndeksi (p_i)	Ayırtedicilik İndeksi (r_{ix})
1	0.22	-0.03	0.78	0.38
2	0.47	0.21	0.95	0.83
3	0.31	0.12	0.79	0.49
4	0.90	0.39	0.96	0.79
5	0.53	0.28	0.74	0.23
6	0.83	0.19	0.99	0.83
7	0.88	0.67	0.92	0.24
8	0.38	0.32	0.73	0.07
9	0.05	0.35	0.31	0.06
10	0.54	0.28	0.86	0.05
11	0.53	0.40	0.66	0.29
12	0.40	0.41	0.65	0.50
13	0.78	0.33	0.96	0.25
14	0.77	0.35	0.96	0.64
15	0.45	0.30	0.60	0.11
16	0.67	0.25	0.86	0.58
17	0.23	0.38	0.70	0.67
18	0.55	0.35	0.91	0.30
19	0.33	0.21	0.64	0.46
20	0.24	0.14	0.60	0.32
21	0.35	0.16	0.58	0.41
22	0.54	0.11	0.85	0.33
23	0.22	0.04	0.25	0.21
24	0.37	0.33	0.69	0.44
25	0.33	0.29	0.89	0.57
26	0.06	0.38	0.26	0.38
27	0.08	0.17	0.64	0.46
28	0.21	0.20	0.73	0.66
29	0.74	0.28	0.92	0.77
30	0.35	0.44	0.85	0.46
31	0.06	0.43	0.45	0.25
32	0.45	0.47	0.82	0.26
33	0.36	0.38	0.80	0.24
34	0.32	0.37	0.75	0.61
35	0.19	0.41	0.92	0.62
36	0.13	0.62	0.69	0.30
37	0.01	0.52	0.76	0.37
38	0.08	0.67	0.97	0.42
39	0.04	0.28	0.84	0.34
40	0.01	0.14	0.66	0.62
41	0.31	0.30	0.80	0.26
42	0.09	0.33	0.24	0.44
43	0.18	0.40	0.52	0.21
44	0.18	0.41	0.15	0.11
45	0.05	0.44	0.40	0.55

Çizelge 3 incelendiğinde, son test uygulamasından elde edilen madde güçlük indekslerinin maddelerin büyük çoğunluğu için, ön test uygulamasından elde edilen madde güçlük indekslerinden daha yüksek olduğu görülmektedir. Bu durum öğrencilerin ölçme ve değerlendirme konusunda çok az bir bilgiyle programa katıldıklarını, ancak programın sonunda bu konudaki bilgilerini daha çok artırdıklarını Testte yer alan maddelere verilen yanıtlar incelendiğinde 23. sorunun doğru cevaplandırılma yüzdesinin ön testte çok düşük olduğu, eğitimden sonra uygulanan son testte bu yüzdenin fazla yükselmediği dikkat çekmektedir. Eğitimden önce öğrenmelerin olmaması ya da az olması doğaldır. Ancak eğitimin bitiminde öğrenmelerin artması ve madde güçlüklerine ilişkin yüzdelerin büyümesi gerekmektedir. Bu anlamda sözü edilen sorunun ölçtüğü kazanımlara ilişkin öğretim etkinliklerinin yeniden gözden geçirilmesi yerinde olacaktır. Bunun yanında, maddeler doğru yanıtlanma yüzdeleri bakımından incelendiğinde, hem eğitimin başında hem de eğitimin sonunda yüksek yüzdelere sahip maddeler de bulunmaktadır (6. , 7. ve 13. maddeler gibi). Eğitimin başında öğrencilerin çoğunluğu tarafından doğru yanıtlanan soruların ölçtüğü özelliklerle ilgili öğretim etkinliklerinin, öğrenmeleri daha çok artırıcı yönde düzenlenmesi doğru olacaktır.

Madde güçlük indeksleri, aynı zamanda o maddeyle ölçülen kazanımın öğrenilme düzeyi hakkında da bilgi vermektedir. Bir maddenin güçlük düzeyinin artması o maddeyle ölçülen özelliğin öğrenildiğini, düşmesi ise öğrenilemediğini gösterir. Bir eğitim programında beklenti, program uygulanmadan önce öğrencilerin programda yer alan kazanımlar hakkında az bilgiye, program uygulandıktan sonra daha çok bilgiye sahip olmalarıdır. Bu açıdan soruların son test ve ön test doğru yanıtlanma yüzdelere bakıldığında bazı soruların ölçtüğü özelliklerin, programın etkililiği açısından ideal durumda olduğu görülmektedir. Örneğin 2., 19., 27., 28. ve 40. sorularda olduğu gibi, arzu edilen öğrencilerin eğitimden önce, programda yer alan kazanımlarla ilgili daha az bilgiye, eğitimden sonra daha fazla bilgiye sahip olmalarıdır. Bu, öğretim programının başarısını yansıtan bir durumdur (Baykul, 1992; Kutlu, 1998).

Bu bilgilere göre, ön test ve son test yüzdeleri arasında az fark bulunan maddelerle, son test yüzdesi ön test yüzdesinden düşük olan maddelerle ilgili kazanımların ve etkinliklerin gözden geçirilmesi yerinde olacaktır.

Testteki maddelerin ayırt edicilikleriyle ilgili şu yorumlar yapılabilir. Eğer uygulanacak bir programın içeriği hakkında, uygulama grubunun yeteri kadar bir bilgisi

yoksa özellikle çoktan seçmeli maddelerdeki şans başarısının da etkisiyle madde ayırt ediciliklerinin olması gerekenden daha düşük sonuçlar vermesi mümkündür. Program uygulandıktan sonra ise, madde ayırt ediciliklerinin daha iyi değerler vermesi yerinde olacaktır. Bu çalışmada, ön teste ait madde ayırtedicilikleri ile son teste ait madde ayırtedicilikleri arasında -birkaç madde dışında- önemli bir artış olduğu görülmektedir. Bu bulgudan hareketle, dönem boyunca verilen eğitimin etkili olduğu sonucuna ulaşılabilir.

3. *Alt Problem: “Ölçme ve Değerlendirme Bilgi Testi”nin ön test ve sontest uygulamalarına ilişkin puanların KR 20 güvenirlikleri nasıldır?*

Çizelge 4. “Zaman yönetimi bilgi testi”nin ön test ve son test uygulamalarına ilişkin puanların KR 20 güvenirlik katsayıları

Uygulama	KR 20 Güvenirliği
Öntest	0.54
Sontest	0.70

KR 20 güvenirlik katsayısı, bir testte yer alan maddelerin iç tutarlılığına ilişkin bilgi veren bir istatistiktir. Güvenirlik katsayısının yüksek değerleri maddelerin aynı özelliği ölçtüğü, düşük değerleri ise testin ölçmek istediği özelliği yeterince ölçemediği anlamına gelir. Bu durumda KR 20 güvenirlik katsayısı, bir içtutarlılık katsayısı olarak da kabul görür. Bir testin güvenirliğini düşüren en büyük etken, test puanlarına karışan tesadüfi hatalardır. Bu çalışmada, Ölçme ve Değerlendirme Bilgi Testi'nin ön test ve son test uygulamalarına ilişkin güvenirlik katsayıları özellikle son test puanları açısından daha kabul edilebilir düzeydedir. Ön teste ait güvenirlik düzeyi daha düşükken, son teste ait güvenirlik katsayısı oldukça yükselmiştir. Ön test uygulamasında öğrenciler, soruların ölçtüğü özelliğe yeterince sahip olmadıkları için raslantısal yanıtlar vermiş olabilirler. Bu durum puanlara karışan tesadüfi hatayı artırarak, güvenirliğin düşmesine neden olmuş olabilir.

3. *Alt Problem: Ölçme ve Değerlendirme Bilgi Testi'nin ön test ve son test uygulamalarından elde edilen puanların ortalamaları arasında manidar bir fark var mıdır?*

Öğrencilerin “Ölçme ve Değerlendirme Bilgi Testi”nden elde ettikleri öntest ve sontest puanlarının ortalamaları arasındaki farkın manidarlığı için yapılan bağımlı gruplar için t-testi sonuçları Çizelge 5’te verilmiştir.

Çizelge 5. Ölçme ve Değerlendirme Bilgi Testi ön test ve son test puan ortalamalarına ait t-testi sonuçları

Ölçüm (ZYBT)	N	Aritmetik Ortalama	Standart Sapma	sd	t	P
Ön test	73	32.14	8.20	72	27.65	.000*
Son test	73	71.14	9.90			

* p< .01

Çizelge 5’ten de görülebileceği gibi öğrencilere ölçme ve değerlendirme eğitimi verilmesi sonrasında, bilgi testi puanlarında manidar bir yükselme olmuştur($t_{(72)}= 27.65$, $p< .01$). Öğrencilerin eğitim öncesi bilgi testi toplam puanlarının aritmetik ortalaması 100 puan üzerinden $\bar{X}= 32.14$ iken, program uygulandıktan sonra $\bar{X}= 71.14$ ’e yükselmiştir. Bu bulgu, dönem boyunca okutulan ölçme ve değerlendirme dersinin öğrencilerin bilgilerini artırmada etkili olduğunu göstermektedir. Ayrıca öğrencilerin son testten aldıkları puanların standart sapmasının yüksek bir değer olmaması, bireylerin alınan puanlar açısından benzer özelliklere sahip olduğunu da ortaya koymaktadır. Bu durum öğrencilerin son testte yüksek puanlarda daha çok toplandığı hakkında bilgi vermektedir.

SONUÇ VE ÖNERİLER

Ankara Üniversitesi Eğitim Bilimleri Fakültesinde 2008-2009 eğitim-öğretim yılı güz döneminde İngilizce sertifika gruplarında yürütülen ölçme ve değerlendirme dersinin etkililiğini belirlemeyi amaçlayan, aynı zamanda da bir ön deneme uygulaması niteliği taşıyan bu çalışmanın sonucunda, verilen eğitimin etkili olduğu bulgusuna ulaşılmış, programın bitiminde öğrencilerin öğrenmelerini önemli ölçüde artırdıkları gözlenmiştir.

Bu araştırmadan elde edilen bulgular Kelecioğlu (1989) ve Mink ve Fraser (2005)’in araştırma bulguları ile paralellik göstermektedir. Madde analizi sonuçlarına

bakıldığında öğrencilerin ölçme ve değerlendirme konusunda çok az bir bilgiyle programa katıldıkları, ancak programın sonunda bu konudaki bilgilerini artırdıkları sonucuna ulaşılmıştır.

Ölçme ve Değerlendirme dersinin etkililiğini belirlemek için kullanılan “Ölçme ve Değerlendirme Bilgi Testi”nden elde edilen puanların bağımlı gruplar için t-testi sonuçları incelendiğinde, programın uygulanan öğrenciler üzerinde etkili olduğu sonucuna ulaşılabılır.

Bu çalışma, öğretmenlik sertifikası programına devam eden üniversite dördüncü sınıf öğrencileri üzerinde yürütülmüştür. Araştırma sonuçları göstermektedir ki bu ve benzer çalışmalar (Kelecioğlu, 1989; Baykul, 1992; Kutlu, 1998; Kutlu, 2002b; Doyle, 2002; Coşgun ve Ilgar, 2004; Mink ve Fraser, 2005; Butakın ve Özgen, 2007; Bay ve Karakaya, 2009; Barchard ve Pace, 2009) alanyazına önemli katkılar getirmektedir. Benzer çalışmaların farklı eğitim basamaklarına devam eden öğrenciler üzerinde yürütülmesi, eğitimin etkililiğine ilişkin birtakım önemli bilgilerin ortaya konmasına olanak tanıyabilir.

Araştırmanın bulguları göstermektedir ki, eğitimin verimliliğini konu alan araştırmalar birçok eğitim sorununa çözüm önerisi getirebilecek sonuçlar ortaya koymaktadır. Bu nedenle, okullarda, her sınıf düzeyinde verilen eğitimlerin etkililiğini ve verimliliğini konu alan araştırmalar yapılabilir.

KAYNAKÇA

- Barchard, K. A. ve Pace, L. A. (2009). Prevent data entry catastrophes: Use double entry. Paper Presented at the Western Psychological Association Conference, Portland, Oregon.
- Bay, E. ve Karakaya, Ş. (2009). Öğretmen eğitiminde yapılandırmacı yaklaşıma dayalı uygulamaların etkililiğinin değerlendirilmesi. *Elektronik Sosyal Bilimler Dergisi*, 8(28), 40-55.
- Baykul, Y. (1992). Productivity and evaluation in education, *Education*, 113(2), 307-311.
- Baykul, Y. (2000). Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması. Ankara: ÖSYM Yayınları.
- Berberoğlu, G. ve Kalender, İ. (2005). Öğrenci başarısının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi. *Eğitim Bilimleri ve Uygulama*, 22(4), 21-35.

- Butakın, V. ve Özgen, K. (2007). Yeni ilköğretim matematik dersi öğretim programının uygulamadaki etkililiğinin değerlendirilmesi (Diyarbakır İli Örneği), Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, 8(1).
- Büyüköztürk, Ş. (2007). Sosyal Bilimler İçin Veri Analizi El Kitabı. Ankara: Pegem Yayıncılık.
- Coşgun, S. ve Ilgar, Z. (2004). Rehberlik ve psikolojik danışmanlık deneyimi çalışmalarının adayların özyeterlilik algılarına etkisi. 13. Ulusal Eğitim Bilimleri Kurultayı, İnönü Üniversitesi Eğitim Fakültesi, Malatya.
- Crocker, L. ve Algina, J. (1986). Introduction to classical and modern test theory. Orlando: Harcourt Brace Jovanovich Inc.
- Doyle, G. (2003). Media ownership: The economics and politics of convergence and concentration in the U.K. and european media. Journal of Cultural Economics. 27(3-4), 290-293.
- Ertürk, S. (1986). Eğitimde program geliştirme. Yelken Tepe Yayınları, Ankara.
- Haladyna, T. M. (1997). Writing test items to evaluate higher order thinking, USA: Allyn & Bacon.
- Karasar, N. (2004). Bilimsel araştırma yöntemi. Ankara: Nobel Yayın Dağıtım.
- Kelecioğlu, H. (1989). İlkokul ikinci sınıf matematik programının hedef davranışlarının ulaşılabilirlik ve tutarlılık yönünden değerlendirilmesi. Yayımlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kimberly, A. B. ve Larry, A. P. (2009). Evaluating the effectiveness of collaborative computer-intensive projects in an undergraduate psychometrics course. Interactive Learning Environments, 1744-5191.
- Kline, P. (1986). A Handbook of Test Construction: Introduction to Psychometric Design. Methuen, New York.
- Kutlu, Ö. (1998). Eğitimde verimliliğin ölçülmesi üzerine bir araştırma. Yayımlanmamış Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kutlu, Ö. ve Çok, F. (Ocak-2002a). 12-14 yaş grubu ergenler için hazırlanmış olan cinsel eğitim programına dayalı cinsel bilgi testinin geliştirilmesi. Eğitim ve Bilim Dergisi, Cilt 27(123), 3-12.
- Kutlu, Ö. (2002b). Yapılandırılmış öğretim programlarının verimlilik ve etkililik düzeyinin saptanması. XI. Eğitim Bilimleri Kongresi, 23-26 Ekim, Yakın Doğu Üniversitesi, Lefkoşa, KKTC.
- Kutlu, Ö. (2003). 12-14 yaş grubu ergenlere yönelik cinsel eğitim programının değerlendirme modeli. (3. Bölüm), Figen Çok, Ergenlerin cinsel eğitimi: Bir program denemesi. Ankara: Ankara Üniversitesi Basımevi.

- Kutlu, Ö. ve Vefikuluçay, D. (Ekim-Kasım-Aralık 2003). Üreme bilgi testinin geliştirilmesi. HIV/AIDS Dergisi, HATAM Yayınları, Cilt 5, Sayı 4.
- Linn, R. L. (1989). Educational measurement (Third Edition), American Council on Education, NewYork: Macmillan Publishing Company.
- Linn, R. L. ve Gronlund, N. E. (1995). Measurement assessment in teaching. (7th Edition). New Jersey: Prentice-Hall Inc.
- Mink, D. ve Fraser, B. (2002). Evaluation of a K-5 mathematics program which integrates children's literature: classroom environment, achievement and attitudes. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Özçelik, D. A. (1989). Eğitim programları ve öğretim. Ankara: ÖSYM Yayınları.
- Stevens, J. (1992). Applied multivariate statistics for the social sciences. New Jersey: Lawrence Erlbaum Associates Publishers.
- Tabachnick, B. G. ve Fidell, L. S. (1996). Using multivariate statistics. NewYork: Harper Collins College Publishers.
- Tekin, H. (1987). Eğitimde ölçme ve değerlendirme. Ankara: Yargı Yayınevi.
- Thorndike, R. L. (1982). Applied psychometrics. Boston: Houghton Mifflin Company.
- Turgut, M. F. (1997). Eğitimde ölçme ve değerlendirme metodları. Ankara: Saydam Matbaacılık.
- Uğur, A. ve Kutlu, Ö. (2010). Hizmetiçi eğitimde verimliliğin ölçülmesi, Milli Prodüktivite Merkezi Verimlilik Dergisi, 1, 95-107.
- Walsh, W. B. ve Betz, N. E. (1995). Test and assessment. (Third Edition), New Jersey: Prentice Hall, Englewood Cliffs.