

Eğitim Psikolojisi Programının Değerlendirme Ögesinin Değerlendirilmesi

Evaluating the Evaluation Component of the Educational Psychology Curriculum

Gülen ONURKAN ALIUSTA*, Hasan ÖZDER**, Emre ÇETİN***

Öz: Bu çalışmada, Doğu Akdeniz Üniversitesi (DAÜ) Eğitim Fakültesi bünyesinde uygulanmakta olan EGİT112 kodlu Eğitim Psikolojisi Dersi Programı'nın değerlendirme ögesinin değerlendirilmesi amaçlanmıştır. Araştırmada 303 öğretmen adayının aldığı Eğitim Psikolojisi dersinde ara sınav ve final sınav puanlarının geçerlik ve güvenilirliği ile kullanılan ölçütler takımının uygunluğu incelenmiştir. Araştırmanın çalışma grubu, 2013-2014 öğretim yılının bahar döneminde DAÜ Eğitim Fakültesi Eğitim Bilimleri Bölümü'nün açtığı EGİT112 Eğitim Psikolojisi dersini alan 303 öğrenciden oluşturulmuştur. Dersin hedefleri, öğrencilerin ara sınav, proje, kısa sınav ve final sınavlarından elde ettikleri puanlar ile bu puanlardan oluşan Bileşke Puanı (BP) ve tüm bu puanların ağırlıklı katsayıları araştırmanın verileri olmuştur. Araştırma sonuçlarına göre, ara sınav ve final sınavlarına ilişkin hazırlanan başarı testlerinin kapsam geçerliklerinin ve Kuder-Richardson 21 (KR21) güvenilirlik katsayılarının düşük olduğu bulunmuştur. Öğretim elemanlarının yılsonu öğrencilerin başarılarını belirlemek için verdikleri harflendirmelerin ağırlıkları ile öğrencilerin bileşke puanları arasında ise yüksek bir korelasyon bulunmuştur. Bu bağlamda kullanılan ölçütler takımının güvenilir olduğu söylenebilir. Araştırmada ayrıca, ara sınav ve proje puanlarının final sınavı puanlarını yordamada oldukça etkili oldukları ortaya çıkmıştır.

Anahtar Kelimeler: Program, program değerlendirme, değerlendirme ögesinin değerlendirilmesi, eğitim psikolojisi programı

Abstract: This study aims to evaluate the evaluation component of EGİT112 Educational Psychology course curriculum offered at the Eastern Mediterranean University (EMU) Education Faculty. This study explores the validity and the reliability of the midterm and final exams administered and also the suitability of the assessment criteria used in Educational Psychology course. The participants of the study consisted of 303 students enrolled in EGİT112 Educational Psychology course at the Faculty of Education of EMU during spring semester of 2013-2014 academic year. The data of the study included the learning objectives of the course, the test scores obtained from the midterm, project, quiz and final exams, the total scores and weighted coefficient of these scores. The findings indicate that both midterm and final exams have low content validity. Kuder-Richardson 21 (KR21) reliability coefficient is also found rather low. The results demonstrate high correlation between letter grades assigned by the faculty at the end of the semester and the total achievement score of students which can be said that the assessment criteria are reliable in this respect. Grades obtained from the midterm and the research projects are reported to be rather effective in predicting final grades.

Keywords: Program, program evaluation, evaluating evaluation component, educational psychology curriculum.

Giriş

Eğitim sisteminin değerlendirme ögesinin değerlendirilmesi bir “program değerlendirme” çalışması olarak görülmektedir (Baykul, 1992). Öğrencilerin herhangi bir programda başarılı olup olmadıklarına ilişkin verilecek kararın isabetliliği önemlidir. Değerlendirme sonucunda varılan kararların doğruluğu, bir yandan ölçme sonuçlarının geçerlik ve güvenilirliğine, öte yandan da ölçütün yerinde bir ölçüt olmasına bağlıdır (Özçelik, 1981, 161). Ölçme araçlarının programda öngörülen hedefleri ne derecede karşıladığını belirlemek ve bu bağlamda ölçme

*Dr., Doğu Akdeniz Üniversitesi, Eğitim Fakültesi, Gazimağusa-KKTC, e-posta: gulen.onurkan@emu.edu.tr

** Yrd. Doç. Dr., Atatürk Öğretmen Akademisi, Lefkoşa-KKTC, e-posta: hasan.ozder@emu.edu.tr

*** Yrd. Doç. Dr., Doğu Akdeniz Üniversitesi, Eğitim Fakültesi, Gazimağusa-KKTC, e-posta: emre.cetin@emu.edu.tr

araçlarının geçerliğini arttırmak önemlidir. Öğrencilerin başarıları hakkında verilecek kararların isabetsizliği sistemdeki maliyeti arttırmaktadır.

Kararlardaki isabetsizlik, eğitim sisteminde, bu sistemin ürünü olan hedef davranışların, eğitime konu olan öğrencilerde beklenen düzeyde gelişmemiş olması halinde gelişmiş gibi, gelişmemiş olması halinde ise gelişmemiş gibi görünmesi şeklinde kendini gösterir. Böyle bir durumda eğitim sistemine, bir sistemin ürünlerinden sağlanan yanlış bilgiler yeni girdi olarak girer ve onarma işi güçleşir. Diğer taraftan, eğer, değerlendirme sonuçları öğrencilerin başarılarını saptama ve buna dayalı olarak geçti-kaldı kararlarını verme amacıyla kullanılacaksa, başarılı öğrenciler başarısız veya başarısız öğrenciler de başarılı olarak görülebilir (Baykul, 1983, 96-97).

Eğitimde program değerlendirme

Eğitimde program geliştirme bir programın tasarlanması, uygulanması, değerlendirilmesi ve yeniden düzenlenmesi aşamalarını kapsamaktadır (Çilenti,1984). Program geliştirmenin birinci aşaması programın tasarlanmasıdır. Bir program tasarısı hedef, içerik, eğitim durumları ve değerlendirme öğelerini kapsamaktadır (Taba, 1962; Tyler, 1946; Ertürk, 1972). Program tasarısında yer alan öğelerin her birinin uygulamada ne derece etkili olduklarını belirlemek hem programın aksayan ve aksamayan yönlerini hem de aksaklık varsa kaynağını ortaya çıkartıp gerekli iyileştirmelerin yapılmasını hedeflemektedir (MartoneveSireci, 2009). Program tasarısında yer alan değerlendirme öğesinin değerlendirilmesi programın yeniden düzenlenmesi hakkında bilgi verdiği için bu yaklaşım bir program değerlendirme yaklaşımı olarak kullanılabilir. Ölçme ve değerlendirme sisteminin değerlendirilmesi öğrenci başarısını belirlemek için kullanılan araç ve yöntemlerden elde edilen ölçmesonuçlarının yeterli geçerlik ve güvenilirlikte olup olmadığı ile ölçütün uygunluğunun kontrol edilmesiyle yapılabilir (Ananda, 2003a; Baykul, 1992; Erden, 1993; Tan, 1990).

Ölçme sonuçlarının geçerliğinin test edilmesi

Ölçme sonuçlarının geçerliği, hedeflenen ölçmenin ne derece gerçekleştiği ile ilgilidir. Ölçme işleminde kullanılan aracın ve o araçta kullanılan maddelerin ölçülmesi hedeflenen davranışları ölçmede yeterli olması büyük önem taşımaktadır (Sireci, 2007). Bu yüzden, ölçme araçlarının ve buna bağlı olarak ölçme sonuçlarının ölçülmek istenen değişkenin ölçüsü olabilecek bir puan vermesi ve bu ölçüyü başka değişkenlerle karıştırmaması gerekir (Turgut, 1997, 38). “Geçerliği olmayan sınavlarla yapılan değerlendirmeler sonucunda öğrenciler hakkında doğru kararların verilmesi mümkün değildir” (Berberoğlu, 2012, 10).

Öğrenci başarısını ölçen ölçme araçlarının geçerliğinin test edilmesi için en uygun geçerlik türü kapsam geçerliğidir. Bu nedenle programda kazandırılması öngörülen hedefler ve içerikle ölçme aracındaki soruların tutarlı olması beklenir. (Case, JorgensenveZucker, 2004). Günümüzde yapılan tutarlılık çalışmaları hedefler ve içerik ile ölçme aracındaki soruların ilişkisini ortaya koymayı hedeflemektedir (MartoneveSireci, 2009). Öğretilmek istenen hedefler ile değerlendirme öğelerinin bir biriyle tutarlı olması eğitim programının etkililiğinin önemli bir göstergesidir.

Belirtke tablosunun hazırlanması bu tutarlılığın sağlanması için önemli bir araçtır. Belirtke tablosunun hazırlanmadığı zamanlarda, programda yazılı olan hedeflerin görmezden gelindiği, programda olmayan hedeflerin ise önemli görüldüğü gibi durumlar ortaya çıkabilir. Bu durumda en önemli kaynak olan öğretmenden bilgi alınabilir. “Başarı değerlendirmeye odaklı bir sınavın derste yapılanlarla uyumlu olması gerekmektedir” (Berberoğlu, Arıkan, Çalışkan, İkinci, İpekçioğlu, ve Sevgi, 2012, 11). Programda öngörülen hedefler, öğretmenin sınıfta yaptığı öğretim ve öğretmenin ölçmek istedikleri arasında sıkı bir tutarlılık olmalıdır (Kubiszyn, veBorich, 2007; Case ve diğ., 2004; Popham, 2003). Bu nedenle, özellikle başarı testleri için konu-davranış karşılaştırmasını içeren belirtke tablosu hazırlamak büyük önem taşımaktadır. Ancak yapılan araştırmalar göstermektedir ki, öğretim elemanları hedeflerin düzeylerini belirlemede ve bu düzeye uygun soru yazmada yeterli düzeyde değildirler (Acar, 2007; Algan, 2008; Gümüş ve Aykaç, 2012; Kanatlı, 2009; Okur, 2008).

Ölçme sonuçlarının geçerliğinin yüksek olması kullanılan ölçme araç ve yöntemlerinin ölçülmek istenen değişkene uygun olması ile de ilişkilidir (MartoneveSireci, 2009). Sosyal bilimlerle ilgili disiplinlerde çalışan öğretim elemanları yazılı sınavlar, kısa cevaplı testler, sözel sunumlar, araştırma kâğıtları ve projelerle dönüt vermeyi daha uygun bulmaktadırlar (Doğan, 2009). Bir diğer anlatımla, birçok farklı ölçme araçlarının kullanılması ölçme sonuçlarının geçerliğini arttırmaktadır, denilebilir.

Yapılan araştırmalar incelendiğinde ilköğretim (Kaya, 2003; Kelecioğlu, Atalay ve Öztürk, 2010), ortaöğretim ve yükseköğretim (Tan, 1990) kademelerinde uygulanan başarı testlerindeki soruların kapsam geçerliğini sağlamadığı ve aynı zamanda ölçülmek istenen hedeflerin düzeyleri ile eşleşmediği anlaşılmaktadır. Ayrıca, Türkiye’de uygulanan geniş ölçekli sınavların geçerliklerinin de düşük olduğu söylenebilir (Afacan ve Nuhoglu, 2008; Berberoğlu, Arıkan, Çalışkan, Ekinci, İpekçioğlu ve Sevgi, 2012; Birinci, 2014; Çoban, Uludağ ve Yılmaz, 2006).

Uygulanan program analizi modeli

Bir testte yer alan soruların geçerliğinin belirlenmesi için birçok model kullanılabilir (MartoneveSireci, 2009; Case vd, 2004). En yaygın bilinen ve kullanılan Model ise Porter ve Smithson tarafından 2001 yılında geliştirilmiş olan “Uygulanan Program Analizi Modeli”(UPAM)’dir. Programda kazandırılması öngörülen öğrenme hedefleri ile ölçme araçlarındaki soruların ne ölçüde tutarlı olduğunu ortaya koymayı hedefleyen bir modeldir (MartoneveSireci, 2009; Case ve diğ., 2004). Bu model kapsam geçerliği üzerine inşa edilmiş olup öğrenme hedefleri ve ölçme araçlarındaki soruların içerik ve bilişsel düzey olarak kategorize edilmesini içerir. Diğer bir anlatımla, bu modelle hem öğrenme hedeflerinin hem de ölçülen maddelerin hangi konuya ait olduğu ve ayrıca bilişsel alanın hangi öğrenme basamağı ile ilgili olduğu uzmanlar tarafından belirlenir. Öğrenme hedefleri ve ölçülen maddelerin kategorize edilmesinden sonra hedefler ile maddelerin ne ölçüde tutarlı olduğu tablo ve şema gibi görsellerle ortaya konulur. Diğer modellerden farklı olarak bu model eğitim programının öğretim boyutunu, yani öğretmenlerin sınıflarında ne öğrettiklerini de göz önünde bulundurur. Böylece kullanılan bu model ile eğitim programının üç ögesi olan hedefler, kapsam (ne öğretildiği) ve değerlendirme arasındaki tutarlılık belirlenmiş olur (Sireci ve Faulkner-Bond, 2014). Bu araştırmada kapsam geçerliğini değerlendirilmesi için bu model kullanılmıştır.

Ölçme sonuçlarının güvenilirliğinin test edilmesi

“Güvenirlilik herhangi bir ölçme veya yönteminin ölçtüğü değişkeni ne derece duyarlılıkla ölçebildiği; başka bir ifadeyle, ölçme sonuçlarının tesadüfi hatalardan ne derece arınık olduğu” şeklinde tanımlanabilir (Turgut, 1997, 27).

Ölçme sonuçlarının güvenilirliğinden söz edilebileceği gibi, sadece ölçme aracının veya sadece ölçmenin güvenilirliği de tanımlanıp kestirilebilir. Eğitimdeki uygulamalar için, ölçme araç ve yöntemlerden elde edilen puanların güvenilirliğini tanımlamak ve onunla uğraşmak çoğu halde yeterlidir. Çünkü, araç, yöntem, ölçmecisi ve ölçme ortamı gibi etkenlerin ortak etkileri puanlara yansır (Turgut ve Baykul, 2010, 123).

Güvenirligi hesaplamak için birçok yöntem vardır. Tek bir uygulamaya dayalı olarak uygulanan testlerin güvenirligini kestirmede Kuder-Richardson 20 (KR20), Kuder-Richardson 21 (KR21) ve Cronbach Alpha (α) formülleri kullanılabilir.

Osborne (2003) psikoloji dergilerinde yayınlanan araştırmalarda kullanılan ölçeklerin ortalama olarak .83 güvenirlilik katsayısına sahip olduklarını belirtirken, Lou ve arkadaşları ise başarı testlerinin ise ortalama olarak .75 olduğu belirtmişlerdir. Bu bağlamda başarı testlerinin güvenirliginin psikolojik testlerden daha düşük olduğu söylenebilir. Bu nedenle başarı testlerinin güvenirlilik düzeylerinin sınanması önemlidir. Türkiye’de yapılan araştırma sonuçlarına göre çeşitli eğitim kademelerindeki derslerde kullanılan ölçme araçlarının güvenirliginin düşük olduğu söylenebilir (İğci, 1987; Karaca, 1990; Tan, 1990).

Ölçütün uygunluğu

Eğitimde öğrenciler değerlendirilirken öğretmen kanısı, sınıftaki diğer öğrencilerin başarısı, ülke çapındaki diğer normlar, öğrenci yeteneği, öğrencinin programa girişteki başarısı, programın hedefleri ölçüt olarak kullanılabilir (Turgut, 1997). Eğitimde temel amaç hedeflerin kazandırılması olduğuna göre kuşkusuz programın hedeflerinin ölçütler takımı olarak kullanılması gerekmektedir (Erden, 1993, 35).

Değerlendirme işlemi sonunda doğru bir karara varabilmek için, eldeki ölçülerin geçerlik ve güvenilirliklerine ek olarak, kullanılan ölçütün de yerinde bir ölçüt olması gerektiği açıktır. Her ölçüt, ölçme sonucu ile belirlenen miktar veya derecedeki bir özelliğin belli bir amaçla işe yarama-yaramama koşulunu belirler. Ölçülerle betimlenen özelliği ölçüte uygun olan bir nesne söz konusu amaçla işe yarar, uygun değilse işe yaramaz demektir. Bu nedenle, değerlendirme işleminde kullanılacak ölçüt dikkatle seçilmeli; onun ilgili amaçla işe yarama koşulunu olabildiğince yüksek bir doğrulukla belirlemekte olup olmadığı sıkı bir biçimde denetlenmeli; gerektiğinde ölçütte iyileştirmeler yapılmalıdır (Özçelik, 2010, 193).

Ölçütler takımının belirlenmesi oldukça önemlidir. Kesme puanlarının çok hassas olması öğrencilerin sınav başarısını yansıtması için gereklidir (Marzano,2006, 122). Kabaca yapılan kesme puanları aralıkların adaletsiz olmasına neden olabilir. Guskey (1996, 17) bu durumu şöyle açıklamaktadır:

Eğer B harfi 80-90 arasındaki puanları kapsıyor ise, 89 alan bir öğrenci de B alacak 80 alan da B alacak. Aralarında 9 puan fark olmasına rağmen. Oysa ki, 79 puan bir öğrenci ise – sadece bir puan fark var - C notunu alacak. Bunun durumun ortaya çıkmasının nedeni, C ile B harflerinin kesme puanlarından kaynaklanmaktadır.

Not ölçeğindeki kategori sayısı arttıkça, notun dayandığı ölçmeler yeterince güvenilir olmak kaydıyla, değerlendirmenin güvenilirliği de artar (Turgut ve Baykul, 2010, 366). Ölçme sonuçlarının güvenilirliği ile değerlendirmenin güvenilirliği arasında kategorilere bağlı bir ilişkinin olduğu söylenebilir (Ebel, 1969).Öğrencilerin bir derse ilişkin başarıları belirlenirken tek bir ölçmeye dayalı not verme ve birden çok ölçmeye dayalı not verme yöntemleri vardır (Turgut ve Baykul, 2010). Not verme sürecini kolay, acısız ve genel olarak tatmin edici yapabilecek tek bir notlandırma sistemi bulunmamaktadır (EbelandFrisbie, 1991). Birden çok ölçmeye dayalı not verme yönteminde önemli olan bileşke bir puanın hesaplanmasıdır. Test veya bölüm puanlarından bileşke hesaplanırken bazı noktaların göz önünde tutulması gerekir. Bileşke, genellikle bir ağırlıklı ortalamadır. Ağırlıklar belirlenirken, ağırlıklandırılacak olan puanın, ölçüsü olduğu hedefin veya hedef grubunun önemi ve bu puanın güvenilirlik ve geçerliği göz önünde bulundurulur (Özçelik, 2010). Test sonuçlarına verilecek ağırlıkların nispi büyüklükleri bu sonuçlar üzerinde yapılacak değişik amaçlı regresyon analizleri ile daha net bir biçimde ortaya konabilir (Özçelik, 2010, 210).

Çalışmanın amacı

Araştırmanın temel amacı “Eğitim Psikolojisi dersinin değerlendirme çalışmaları ne tutarlıkta yapılmaktadır?” sorusuna yanıt aramaktır.

Araştırma soruları

1. Eğitim Psikolojisi dersinin ara ve final sınavlarında yer alan sorular dersin hedefleri ile tutarlı mıdır?
2. Eğitim Psikolojisi dersinin ara sınav ve final sınavının puanlarının KR-21 güvenilirlik katsayıları ne düzeydedir?
3. Eğitim Psikolojisi dersinin ara sınav, final sınavı, kısa sınav, proje ve bileşke puanları ile öğretim elemanlarının verdiği harf notlarının ağırlıklandırılmış katsayı puanları arasında anlamlı bir korelasyon var mıdır?

Yöntem

Araştırmada DAÜ Eğitim Fakültesi Eğitim Bilimleri Bölümünde yer alan Eğitim Psikolojisi dersinin değerlendirme ögesinin değerlendirilmesi amaçlandığından betimsel yöntem kullanılmıştır.

Çalışma grubu

Araştırmanın çalışma grubu, 2013-2014 öğretim yılının bahar döneminde Doğu Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü'nün açtığı EGIT112 Eğitim Psikolojisi dersini alan 303 öğrenciden oluşmaktadır.

İşlem

Araştırmanın verileri EGIT 112 Eğitim Psikolojisi dersinin hedefleri ile ara ve final sınavı sorularından oluşturulmuştur. Ayrıca, öğrencilerin ara sınav, proje, kısa sınav ve final sınavlarından elde ettikleri puanlar ile bu puanlardan oluşan Bileşke Puanı (BP) ve tüm bu puanların ağırlıklı katsayıları olmuştur. Eğitim Psikolojisi dersi DAÜ Eğitim Fakültesi Eğitim Bilimleri Bölümü Programı altında yürütülen ve fakültenin çeşitli bölümlerinde okuyan öğretmen adaylarının aldığı bir derstir. 2013-2014 yılında 303 öğrenci tarafından alınan ders 7 grupta 3 öğretim elemanı tarafından yürütülmüştür. Bir öğretim elemanı 1., 4. 5. ve 7. gruptardan, bir öğretim elemanı 6. gruptan diğer bir öğretim elemanı ise 2. gruptan sorumlu olmuşlardır. Ders programında *Eğitim Psikolojisinin Alanı, Gelişme ve Büyümenin Temelleri, Bedensel Büyüme ve Gelişme, Bilişsel gelişim, Kişilik gelişimi, Ahlak gelişimi, ve Davranışçı yaklaşım* konuları dersin kapsamı içindedir. Bu derste öğrenci başarısının belirlenmesi için ara sınavı %30, proje %20, kısa sınav %10, dönem sonu sınavı ise %40 olarak ağırlıklandırılmıştır.

Doğu Akdeniz Üniversitesi'nde, öğrencilerin başarıları değerlendirilirken uyulacak ölçütler DAÜ Sınav, Öğretim ve Başarı Tüzüğü'nün 16. maddesinde belirtilmiştir. Aşağıda belirtilen not aralıklarına karşılık gelen harflere göre başarılı veya başarısız sayılırlar.

Tablo 1: DAÜ Öğrenci Başarılarının Belirlenmesindeki Ölçütler Takımı

Puan aralığı	Harf	Katsayı	Başarı durumu
90-100	A	4.00	Başarılı
85- 89	A-	3.70	Başarılı
80- 84	B+	3.30	Başarılı
75-79	B	3.00	Başarılı
70-74	B-	2.70	Başarılı
65-69	C+	2.30	Başarılı
60-64	C	2.00	Başarılı
56-59	C-	1.70	Koşullu başarılı
53-55	D+	1.30	Koşullu başarılı
50-52	D	1.00	Koşullu başarılı
45-49	D-	0.70	Başarısız
44 ve altı	F	0.00	Başarısız
Devamsızlık	NG	0.00	Başarısız

Araştırma verilerinin elde edilişi için ayrıntılı açıklamalar aşağıda yapılmıştır.

Dersin hedefleri ile ara sınav ve final sorularının uyumu

1. Dersin hedeflerinin belirlenmesi için Eğitim psikolojisi dersinin ders tanımı ve kullanılan ders kitabı incelenip dersin hedefleri belirlenmiştir. Belirlenen hedeflerin ders öğretim elemanları tarafından ders sürecinde işlenip işlenmediği ve başka hangi

- hedefleri iřledikleri sorulmuřtur. Öęretim elemanlarının tümünün hemfikir olduęu hedefler listelenmiřtir. Bu iřlemden sonra dersin 46 tane hedefi belirlenmiřtir.
2. Belirlenen hedefler ve ierik eřleřtirilip hedeflerin hangi üniteye ait olduęu belirlenip hem ara sınav hem de final sınavı için belirtke tabloları hazırlanmıřtır.
 3. Öęretim programlarında kazandırılması öngörülen öęrenme hedefleri ile ölçme araçlarındaki soruların tutarlılıęını ortaya koymak için Uygulanan Program Analizi Modeli'nden yararlanılmıř olup öęrenme hedefleri ile ölçme araçlarındaki soruların hem ierik hem de biliřsel düzey olarak uyumuna bakılmıřtır.
 4. Ara ve final sınavında yer alan soruların hangi konuya ait olduęu ve Bloom Taksonomisi'ne göre hangi öęrenme basamaęını (Bilgi-Kavrama-Uygulama-Analiz-Sentez-Deęerlendirme) temsil ettięinin ortaya ıkartılmasında bir program geliştirme ve bir ölçme-deęerlendirme uzmanından yararlanılmıřtır (Ananda, 2003a, 2003b).
 5. Belirtke tablolarına baęlı olarak ara sınav ve final sorularının dersin hedefleri ile tutarlılıęına ve aęırlıęına bakılmıřtır. Bu iřlemlerle ara sınav ve final sınavlarının kapsam geçerlięi yoklanmıřtır.

Ara sınav ve final sınavlarının güvenilirlięi

1. Ara sınav ve final sınavlarının güvenilirliklerini belirlemek için KR-21 güvenilirlik katsayıları hesaplanmıřtır.
2. Ara sınav puanları ile final sınavı puanları arasındaki korelasyona bakılmıřtır.

Ölçütün uygunluęu

1. *Ara sınav puanı (ASP)*: 30 oktan seçmeli sorudan oluřan testten öęrencilerin aldıkları puanlardır.
2. *Final Sınavı Puanı (FSP)*: 40 oktan seçmeli sorudan oluřan testten öęrencilerin aldıkları puanlardır.
3. *Kısa Sınav*: 10 açık uçlu sorudan oluřan testten öęrencilerin aldıkları puanlardır.
4. *Arařtırma Projesi*: Derste okutulan kuramsal bilgilerden yola ıkarak öęrencilerin ilgi alanları doęrultusunda bir konu, kavram, özellik vs belirleyip, yaptıkları gözlemler sonucunda öęrencilerin hazırlayacakları rapordan aldıkları puanlardır.
5. *Bileřke Puanı (BP)*: Öęrencilerin ASP'lerinin yüzde 30'u, arařtırma projesi puanlarının yüzde 20'si, kısa sınavın puanlarının yüzde 10'u ve FSP'lerinin yüzde 40'ının alınmasından oluřan aęırlıklı puandır. BP'nin hesaplanmasına iliřkin formül řöyle yazılabilir: $(ASP \times 0.30) + (APP \times 0.20) + (QSP \times 0.10) + (FSP \times 0.40)$
6. *Bileřke Puanlarının Katsayısı (BPK)*: Öęrencilerin BP'lerinin yüzdelięe dönüřtürüldükten sonra ortaya ıkan harfin DAÜ Öęretim ve Sınav Tüzüęüne göre katsayı puandır.

Ölçme araçları

Eđitim psikolojisi dersinin ara sınav, kısa sınav, final sınavlarında kullanılan başarı testleri ve arařtırma projesi arařtırmanın ölçme araçlarını oluřurmaktadır.

Ara sınav: 30 adet oktan seçmeli maddeden oluřmuř olup *Eđitim Psikolojisinin Alanı, Geliřme ve Büyümenin Temelleri, Bedensel Büyüme ve Geliřme ve Biliřsel Geliřme* konularını kapsamaktadır.

Final Sınavı: 40 adet oktan seçmeli maddeden oluřmuř olup ara sınav konularına ek olarak *Kiřilik Geliřimi, Ahlak Geliřimi ve Davranıřçı Yaklařım* konularını içermektedir.

Arařtırma Projesi: Bu projede her öęrenciden bireysel veya grup olarak (en fazla 4 öęrenci) sınıfta iřlenen kuramsal bilgiden yola ıkarak (biliřsel geliřim, zihinsel geliřim, davranıřçı yaklařım vb.) ilgi alanları doęrultusunda bir konu, kavram, özellik vs. belirleyip buldukları evrede eřitli gözlemler yapıp bilimsel bir rapor hazırlamaları istenmiřtir. Hazırlanan raporlar

içerik, sunum, dil ve özgünlük açısından öğretim elemanları tarafından değerlendirmeye alınmıştır.

Kısa sınav: 10 adet açık uçlu sorudan oluşup dönem boyunca okutulan tüm konuları kapsamaktadır. Puanlanması her öğretim elemanının kendisi tarafından yapılmıştır.

Verilerin analizi

Birinci araştırma sorusu için elde edilen veriler “Uygulanan Program Analizi Modeli” ile çözümlenmiş, ikinci araştırma sorusu için elde edilen verilerin analizi KR21formülü ile yapılmıştır. Ara ve final sınavının güvenilirliğinin belirlenmesi için KR21 formülünden yararlanılmıştır. Her iki sınava ilişkin madde istatistiklerinin elde edilemeyeşinden dolayı KR20 formülü kullanılamamıştır. Test istatistiklerinin elde olması nedeniyle KR21 formülü kullanılmıştır. Üçüncü araştırma sorusu için elde edilen verilerin analizi için ise, PearsonMomentler Çarpım Korelasayon ve çoklu regresyon tekniğinden yararlanılmıştır

Bulgular

Araştırmanın sorularına ilişkin elde edilen bulgular tablolar yardımıyla aşağıda sunulmuştur.

Dersin hedefleri ile ara sınav ve final sınavlarındaki sorular arasındaki tutarlılık

Tablo 2. *Ara Sınavın Kapsadığı Hedefler ve Sorular*

Konu	Bilgi basamağı		Kavrama basamağı	
	Eğitim programında yer alan hedef sayısı	Sınavda yer alan sorular ve öğrenme basamakları	Eğitim programında yer alan hedef sayısı	Sınavda yer alan sorular ve öğrenme basamakları
1. Eğitim psikolojisinin alanı	5	Bilgi (26, 29)	1	-
2. Gelişme ve büyümenin temelleri	5	Bilg (3, 5) Kavrama (1,2, 6)	1	Kavrama (4)
3. Bedensel büyüme ve gelişme	4	Bilgi (7, 21)	1	Kavrama (8 öğretimi yapılmadı)
4. Bilişsel gelişim	6	Bilgi (13, 23,18, 20) Kavrama (9,12 öğretimi yapılmadı, 22, 24, 27, 30)	1	Kavrama (10,11,14, 15,17, 19, 25, 28) Bilgi (23)
Toplam	17		4	

Tablo 2’den de anlaşılacağı üzere, öğretim elemanlarının uzlaştıkları hedeflerin ve konuların büyük çoğunluğunun ara sınavda yer almadığı anlaşılmaktadır. “Eğitim Psikolojisinin Alanı” ünitesine ait 6 hedefin 5’i bilgi basamağında 1’i ise kavrama basamağındadır. Ara sınavda bu ünitenin bilgi basamağında yer alan 5 hedefinin 2’si ölçülürken 3’üne ilişkin ise soru sorulmuştur. Kavrama düzeyinde yer alan 1 hedef ile ilgili ara sınavda herhangi bir soruya yer verilmemiştir. “Gelişme ve Büyümenin Temelleri” ünitesine ait 6 hedefin 5’i bilgi basamağında 1’i ise kavrama basamağındadır. Ara sınavda bu ünitenin bilgi basamağında yer alan 5 hedefin 2’si ölçülürken diğer 3’üne ait soru sorulmamıştır. Kavrama düzeyinde yer alan 1 hedef ise 1 soru ile ölçülmüştür. Üniteye ilişkin kavrama basamağına ait soruların 3 soru ise incelendiği zaman bu sorulara ait hedef yazılmamasına karşın ölçüldükleri anlaşılmaktadır. “Bedensel büyüme ve gelişme” ünitesine ait 5 hedefin 4’ü bilgi basamağında 1’i ise kavrama

basamağındadır. Ara sınavda bu ünitenin bilgi basamağında yer alan 4 hedefin 2'si ölçülürken 2'si ölçülmemiştir. Kavrama düzeyinde yer alan 1 hedefe ait bir soru ara sınavda yer almamıştır. Ancak kavrama düzeyinde yer alan bir sorunun ise bu üniteye ait olmasına rağmen öğretim elemanları tarafından öğretimi yapılmayan bir hedefe ait olduğu belirlenmiştir. “Bilişsel Gelişim” ünitesine ait bilgi basamağındaki 6 hedefin 3üne ilişkin soru ara sınavda sorulmuş ancak bilgi düzeyinde yer alan ama kavrama basamağında yer almayan sorular da sorulmuştur. Bir sorunun ise öğretimi yapılmamıştır. Kavrama basamağında yer alan 1 soruya ilişkin ise 8 soru sorulmuştur.

Tablo 3. Final Sınavın Kapsadığı Hedefler ve Sorular

Konu	Bilgi		Kavrama	
	Eğitim programında yer alan hedef sayısı	Sınavda yer alan sorular ve öğrenme basamakları	Eğitim programında yer alan hedef sayısı	Sınavda yer alan soru sayısı
1.Eğitim psikolojisinin alanı		Kavrama (2, 22)		
2.Gelişme ve büyümenin temelleri		Bilgi (13, 19 öğretimi yapılmadı)		
3.Bedensel büyüme ve gelişme		Kavrama (26)		Kavrama (24)
4.Bilişsel gelişim		Bilgi (33, 39) Kavrama b(9,4, 25)		Kavrama (3, 5,18, 23, 29, 34, 38, Kavrama (20)
5.Kişilik gelişimi	9	Bilgi (10,14, 32) Kavrama (1, 12, 30)	1	
6.Ahlak gelişimi	4	Kavrama (6,7,8, 28 öğretimi yapılmadı)	4	Kavrama (15, 16)
7.Davranışçı yaklaşım	3		3	Kavrama (11, 21,17, 27, 31, 35,36, 37,40)

Tablo 3'den de anlaşılacağı üzere, “Kişilik Gelişimi” ünitesinin bilgi basamağında yer alan 9 hedefine ilişkin 3 soru sorulmuş; ayrıca, 3 soru da kavrama basamağında sorulmuştur. Ünitenin kavrama düzeyine ilişkin yer alan 1 hedefine karşılık ise 1 soru sorulmuştur. “Ahlak Gelişimi” ünitesine ilişkin ise bilgi basamağında yer alan 4 hedefe karşılık hiç bir soru sorulmamıştır. 4 kavrama basamağına ilişkin hedefe ise 2 soru sorulmuştur. Kavrama düzeyinde 3 soru ise, bilgi basamağına ilişkin konuları ölçmek için sorulmuştur. Kavrama basamağında yer alan bir soru ise öğretim elemanları tarafından öğretilmemiştir. “Ahlak Gelişimi” ünitesine ilişkin 4 bilgi basamağında 4 de kavrama basamağında öngörülen hedeflere ilişkin bilgi basamağına yönelik hiç soru sorulmamış kavrama basamağına ilişkin ise 2 soru sorulmuştur. “Ahlak Gelişimi” ünitesine ilişkin bilgi düzeyinde öğretimi yapılan hedeflerin ölçülmesi için 3 tane kavrama basamağında soru sorulmuştur. Aynı zamanda, bir soru ise, öğretim elemanları tarafından öğretimi yapılmayan bir konuya ait olduğu belirlenmiştir. “Davranışçı Yaklaşım” ünitesinin bilgi düzeyinde 3 hedefi olmasına rağmen bunların hiçbiri ölçülmemiş kavrama düzeyinde yer alan 3 hedef ise 9 soru ile ölçülmüştür.

Tablo 4'ten de anlaşılacağı üzere, ara sınavda “Gelişimi belirleyen etmenleri söyleyebilme” ölçme amacıyla sorulan 1. soru hedefin düzeyi ile aynı değildir. Hedef incelendiği zaman bilgi düzeyinde olmasına rağmen soru kavrama düzeyindedir. Hedef “etmenlerin söylenmesini” bir başka deyişle, “ezbere yazılması veya söylenmesini” içerir; oysa

ki soruda istenilen, öğrencinin verilen örneklerin hangisinin doğru olduğunu bulmasıdır. Bu nedenle sorunun düzeyi bilgi basamağının bir üstü olan kavrama basamağıdır. “Bilişsel gelişim dönemlerinin temel özellikleri bilgisi” hedefinin ölçülmesi amacıyla sorulan 9. soru sözkonusu hedefle aynı düzeyde değildir. Hedef, bilişsel olarak bir öğrencinin bilişsel gelişim dönemlerinin özelliklerini ezbere bilmesini ve bunu ya yazarak ya da söyleyerek göstermesidir. Oysa ki, soruda istenilen verilen bir örneğin hangi gelişim dönemine ait olduğunu ortaya koymaktır. Bu nedenle soru ezberin üzerinde olan ”örneklendirme” davranışını kapsadığından kavrama basamağında bir öğrenmedir.

Tablo 4. Ara ve Final Sınavlarında Yer Alan Bazı Soruların Sahip Oldukları Hedef Basamağı ile Tutarlılıkları

<p>Konu: Gelişime ve büyümenin temelleri Hedef: Gelişimi belirleyen etmenleri söyleyebilme</p>
<p>1. Aşağıdakilerden hangisi gelişim sürecinde kritik dönemlerin bulunduğu görüşünü doğrulayan bir örnek olabilir?</p> <p>a) Ters doğan Ayşe, doğum sırasında beyinine bir süre oksijen taşınmaması sonucu ileriki yıllarda normalin altında zekaya sahip olmuştur.</p> <p>b) Prematüre doğan İlker, iyi beslenerek birkaç ay içinde normal doğan çocukların kilosuna erişmiştir.</p> <p>c) Başından geçen olumsuz bir olay sonrasında kekelemeye başlayan Sabri, tüm çabalamalara karşın kekeleme kalmıştır.</p> <p>d) Doğuştan hipermetrop olan Zehra'nın bu görme bozukluğu yaşla birlikte giderek artmıştır.</p> <p>e) Gebeliğin ilk üç ayında ‘X’ ilacı kullanan Fatma, kolsuz bir bebek dünyaya getirmiştir.</p>
<p>Konu: Bilişsel gelişim Hedef: Bilişsel gelişim dönemlerinin temel özelliklerini söyleyebilme (Bilgi basamağı)</p>
<p>9. Piaget'e göre, bir çocuğun okula giderken izlediği yoldan evine geri dönebilmesi, hangi evrede kazanılan bir bilişsel beceridir? (Kavrama basamağı)</p> <p>a) İşlem öncesi</p> <p>b) Somut işlemler</p> <p>c) Duyusal motor</p> <p>d) Soyut işlemler</p> <p>e) Soyut ötesi işlemler</p>
<p>Konu: Kişilik gelişimi Hedef : Psiko-seksüel gelişim dönemlerinin özelliklerini söyleyebilme</p>
<p>30. 3-7 yaşları arasına denk gelen dönemdir. Bu dönemde cinsel kimlik gelişmeye başlar. Çocuk cinsiyet farklılıklarını keşfeder, sorular sorar. Merakı yüzünden cezalandırılan çocuklar, yetişkinlik döneminde uygun cinsel kimliği benimsemekte sorun yaşayabilirler. Yukarıda Freud'un Psikoseksüel gelişim dönemlerinden hangisinden bahsedilmektedir?</p> <p>a) Oral dönem</p> <p>b) Anal dönem</p> <p>c) Fallik dönem</p> <p>d) Latans dönem</p> <p>e) Genital dönem</p>
<p>Konu: Kişilik gelişimi Hedef : Psiko-seksüel gelişim dönemlerinin özelliklerini söyleyebilme</p>
<p>10. Aşağıdakilerden hangisi Erikson'un Psiko-sosyal gelişim kuramıyla ilgili <u>değildir</u>?</p> <p>a) Sosyal çevre çocuğun psiko-sosyal gelişimi için önemlidir.</p> <p>b) İlk altı yılda geçirilen yaşantılar yetişkinlik yıllarında da bireyi etkiler.</p> <p>c) Kişilik gelişim dönemleri hiyerarjik bir sıra ile ortaya çıkar.</p> <p>d) Benlik gelişimi belirli zaman dilimleri içinde aşamalı olarak biçimlenir.</p> <p>e) Dönemlerin her birinde farklı bir karmaşa durumu yaşanır.</p>

Ara sınav ve final sınavı puanlarının güvenilirlik düzeyleri ve merkezi dağılım ölçüleri**Tablo 5:** Ara, Final ve Kısa Sınavlarına ve Ödev İlişkin Öğrenci Puanlarının Merkezi Dağılım Ölçüleri ve KR21 Katsayıları

	K	N	\bar{X}	S	Çarpıklık katsayısı	Tepe değer	Ortanca	KR21
Ara Sınav	30	302	18,53	4,19	-,28	20,00	19,00	0.60
Final	40	302	22,07	6,85	-1,58	25,00	23,00	0.60
Kısa Sınav	-	302	4,66	2,44	,11	4,00	4,40	-
Ödev	-	302	15,63	6,33	-1,67	20,00	18,00	-

Tablo 5'ten de anlaşılacağı üzere, eğitim psikolojisi dersinde öğrencilerin başarılarını belirlemek için kullanılan vize, final, kısa sınav ve proje puanlarına ilişkin merkezi dağılım ölçüleri verilmiştir. Ara ve final sınavlarına ilişkin madde istatistiklerinin elde edilemeyeşinden dolayı KR20 formülü kullanılamamıştır. Test istatistiklerinin elde olması nedeniyle KR21 formülü kullanılmıştır. Ara sınav ve final sınavı sorularına ilişkin istatistiki bilgi elde edilebildiğinden dolayı bu iki sınavın KR21 güvenilirlik katsayıları hesaplanmıştır; ancak, proje ve kısa sınav için her öğretim elemanı ayrı ayrı çalışma yaptığı için ve ayrıntılı bilgi elde edilemediğinden bu iki sınavın güvenilirlik katsayıları hesaplanamamıştır.

Ara sınavın ortalaması ortanca ve tepe değerinden düşük olduğu için öğrencilerin elde ettikleri başarının yüksek olduğu söylenebilir. Sola çarpık negatif bir durum sözkonusudur. Aynı durum final sınavı için de söz konusudur. Ortalama (22,07) hem tepe değer hem de ortanca değerden daha düşüktür. Aynı durum ödev notları için de geçerlidir. Kısa sınav için ise ortalama hem tepe değerden hem de ortancadan daha yüksektir. Bu da kısa sınavda öğrencilerin başarılarının oldukça düşük olduğunu göstermektedir. KR21=0.60 güvenilirlik ortanın biraz üzerinde bulunmuştur. Genellikle, güvenilirlik katsayısının 0.7 veya üzeri olması istenir (Boyle ve Fisher, 2007). Sınavın zorluk derecesi ise kolaya yakın bulunmuştur ($p=0.61$).

Ara sınav, final ve öğretim elemanlarının verdikleri puanlar arasındaki korelasyon katsayıları**Tablo 6.** Eğitim Psikolojisi Dersinin Ara Sınav ve Final Sonuçlarına Göre Meydana Gelen Sıralama ile Öğretim Elemanlarının Ortaya Koyduğu Sıralamalar Arasındaki Korelasyon

	1	2	3	4	5
1. Ara sınav					
2. Final	.421**				
3. Proje	.148*	.414**			
4. Kısa Sınav	.282**	.247**	.215**		
5. Toplam	.603**	.836**	.732**	.478**	
6. Öğretim elemanı ders sonu ağırlıklı notu	.632**	.754**	.719**	.467**	.960**

Tablo 6'dan da anlaşılacağı üzere, öğretmen adaylarının eğitim psikolojisi dersine ilişkin elde ettikleri toplam başarı ile bu başarının harfe ve ağırlıklandırılmış puana dönüşmesi ile edilen puanlar arasındaki korelasyon .960 olarak hesaplanmıştır. Final sınav sonuçları ile en düşük korelasyon proje ($r=.148$) ve kısa sınav sonuçları ($r=.282$) ile hesaplanmıştır. Ara sınav ve final arasındaki hesaplanan korelasyon katsayısı 0.421'dir.

Tablo 7. Diğer Puanların Toplam Puanı Açıklama Gücüne İlişkin Çoklu Regresyon Analizi Sonuçları

Puanlar	B	St. Hata	β	T	P
Sabit	4,643	1,621		2,865	,004
Vize	,572	,083	,349	6,929	,000
Proje	,375	,054	,346	6,988	,000
Kısa Sınav	,208	,143	,074	1,448	,149

**0.01’de anlamlı. TAP= 4,643+ ,572 vize + ,375 proje + ,208 kısa sınav [F = 44,336 ; p<0,00 ; R=0,556; R^2 =0,302]

Tablo 7’den de anlaşılacağı üzere, regresyon analizinden elde edilen F değeri 0,01 düzeyinde manidar olduğundan, regresyon modelinin uygun olduğu sonucuna ulaşılmıştır. Eğitim Psikolojisi dersi başarısının % 30’u vize, ödev, final ve kısa sınav puanları ile açıklanmaktadır. Regresyon katsayılarının anlamlılığı t-testi ile test edilmiştir. Buna göre, ele alınan değişkenlerden vize (t=6,929, p=.00) ve projenin (t=6,988, p=.00) regresyona katkısı anlamlı çıkarken kısa sınavın değeri (t=1,448, p=.149) anlamlı çıkmamıştır.

Tartışma

Bu araştırmada toplanan veriler ışığında, ara sınav ve final sınavında yer alan soruların programda öngörülen hedefleri dengeli bir biçimde karşılamadığı görülmektedir. Programdaki hedeflere ilişkin gerek arasınv gerekse final sınavında yer alan sorular orantısız olarak dengeli bir biçimde hedefleri temsil etmemiştir. Oysaki öğretilmek istenen hedefler ile sınavlarda yer alan soruların bir biriyle tutarlı olması gerekmektedir. Araştırmadan elde edilen bulgular literatürde yer alan araştırma sonuçları ile benzerlik göstermektedir. Alanda yapılan çalışmalar incelendiği zaman ilköğretim, ortaöğretim, yükseköğretim kademelerinde ve ayrıca Türkiye’de uygulanan geniş ölçekli sınavlarda kullanılan maddelerin kapsam geçerliğini sağlamadığı ve çoğu zaman da ölçülmek istenen hedeflerin düzeyleri ile eşleşmediği anlaşılmaktadır (Afacan ve Nuhoğlu, 2008; Berberoğlu ve diğ., 2012; Birinci, 2014; Çoban ve diğ., 2006; Kaya, 2003; Tan, 1990)

Belirtke tablosunun hazırlanması bu tutarlılığın sağlanması için önemli bir araçtır. Bu bağlamda bulgu şöyle yorumlanabilir: soruların yazılmasında öğretim elemanlarının belirtke tablosu hazırlamadan soruları hazırladıkları söylenebilir. Belirtke tablosu sınavların kapsam geçerliğini kontrol altında tutmada önemli bir araçtır. Belirtke tablosu hazırlamadan, öğretim programında her ünite için yer alan öğrenme hedeflerini net bir şekilde görebilmek mümkün değildir. Öğretim elemanlarının programda yer alan üniteleri ve hedeflerin düzeylerini net bir biçimde görmeleri hem öğretim açısından hem de ölçme açısından önemlidir. Belirtke tablosunun hazırlanması, öğretilmek istenen hedefler ile değerlendirme öğelerinin tutarlılığının sağlanmasında önemli bir rol oynadığından, sınav sorularının hazırlanma sürecinde mutlaka yer alması gereken bir öğedir.

Bu bulgunun bir nedeni ise öğretim elemanlarının soru hazırlamada yetkin olmamalarından kaynaklanabilir. Soru hazırlamak uzmanlık gerektiren, kuramsal temeli olan bir süreçtir. Tüm planlar doğru tasarlanmış olsa da soruların teknik olarak iyi hazırlanmaması sonuçta geçerliği düşük ölçme yapılmasına neden olmaktadır (Berberoğlu ve diğ., 2012). Elde edilen bulgular Berberoğlu (2012), Kaya (2003) ve Tan (1990)’ın araştırma bulguları ile paralellik göstermektedir.

Bir diğer bulgu ise, öğretim elemanlarının hedef ve sorular arasındaki öğrenme basamağı açısından paralelliklerini sağlayamamış olmalarıdır. Bu bulgu şöyle yorumlanabilir: Öğretim elemanlarının hedeflerin düzeylerini belirlemede ve bu düzeye uygun soru yazmada yetersiz oldukları veya bu konuda yeterince hizmetiçi eğitim almadıkları söylenebilir (Acar, 2007; Algan, 2008; Gümüş ve Aykaç, 2012; Kanatlı, 2009; Okur, 2008).

Ara sınav ve final sınavının güvenilirlik katsayıları orta düzeyde hesaplanmıştır. Bu durum şöyle yorumlanabilir: KR-21 güvenilirlik katsayısı hesaplanabildiğinden düşük çıkmış olabilir. KR20 hesaplanmış olsa daha yüksek sonuçlar elde edilecekti. Bir dięer neden ise ara sınavda 30 final sınavında ise 40 soru sorulmasından yani soru sayısının azlığından da kaynaklanabilir. Tan'ın (1990) araştırma sonuçlarına göre, yükseköğretimde kullanılan ölçme araçlarının güvenilirliği oldukça düşüktür.

Toplam puanlar ile harfe dönüştürülmüş ve ağırlıklandırılmış puanlar arasındaki korelasyonun yüksek çıkması deęerlendirmede kullanılan ölçütler takımının hassas olduđu, çok fazla esnek olmadığı ve toplam puanların harfe dönüştürülmesi esnasında çok fazla puan kaybının olmadığı anlaşılmaktadır.

Final sınavı ile en düşük korelasyon kısa sınavın puanları ile çıkmıştır. Aynı zamanda yapılan regresyon analizinde de kısa sınavın final sınavını yordamadığı ortaya çıkmıştır. Bu şöyle yorumlanabilir: Öğretim elemanlarının herhangi bir çalışma sonucuna dayandırılmadan kısa sınav, ara sınav ve proje deęerlendirmeye yönelik belirledikleri ağırlıklar ile regresyon denklemi sonucunda elde edilen ağırlıkların kabaca uyumlu olduđu söylenebilir.

Sonuç ve Öneriler

Bu çalışmada toplanan veriler ışığında, ara sınav ve final sınavında yer alan soruların programda öngörülen hedefleri dengeli bir biçimde karşılamadığı görülmektedir. Programdaki hedeflere ilişkin gerek arasınava gerekse final sınavında yer alan sorular orantısız olarak dengeli bir biçimde hedefleri temsil etmemiştir. Ara sınav ve final sınavının güvenilirlik katsayıları orta düzeyde hesaplanmıştır. Toplam puanlar ile harfe dönüştürülmüş ve ağırlıklandırılmış puanlar arasındaki korelasyonun yüksek çıkması deęerlendirmede kullanılan ölçütler takımının hassas olduđu, çok fazla esnek olmadığı ve toplam puanların harfe dönüştürülmesi esnasında çok fazla puan kaybının olmadığı anlaşılmaktadır. Final sınavı ile en düşük korelasyon kısa sınavın puanları ile çıkmıştır. Aynı zamanda yapılan regresyon analizinde de kısa sınavın final sınavını yordamadığı ortaya çıkmıştır. Bu sonuçlar doğrutusundaki aşağıdaki öneriler getirilebilir:

1. Dersin başında belirtke tablosunun hazırlanması gerekir. Dersin hedeflerinin açık net biçimde ortaya konması ve dersi yürüten öğretim elemanlarının ortak bir biçimde hedefleri ve hedeflerin düzeylerini anlamaları önemlidir. Bu amaçla öğretim elemanlarının işbirliği halinde olması önemlidir. Hazırlanan final ve ara sınavlardaki soruların belirtke tablosuna göre hazırlanması sınavların kapsam geçerliklerini artıracaktır.
2. Hedef yazma ve hedeflerin belli taksonomi içinde olmasını sağlama becerisinin gelişmesi için program geliştirmenin önemli bir konusu olan “hedef ve hedef yazma” konusunda hizmetiçi eğitim verilmelidir.
3. Öğretim elemanlarına yönelik hedeflerin düzeyine uygun soru yazma becerisinin gelişimi için ölçme ve deęerlendirme alanının önemli bir konusu olan “çoktan seçmeli maddelerin yazımı” konusunda hizmetiçi eğitim verilmelidir.
4. Öğretim elemanlarının soru yazma konusunda daha yetkin olmaları için hizmetiçi eğitim almaları gerekebilir.

İlerki araştırmalar için öneriler

1. Programın deęerlendirme öęesine ilişkin çalışmalar farklı eğitim kademesinde ve farklı derslerde yapılabilir.
2. Programın deęerlendirme öęesine ilişkin öğretmen ve öğrenci görüşleri alınabilir.
3. Programın deęerlendirme öęesinin deęerlendirilmesi için aynı programı uygulayan farklı okulların deęerlendirme çalışmaları ele alınabilir.

Kaynaklar

- Acar, H. (2007). *Yeni ilköğretim programlarının öğretmen görüşlerine dayalı olarak değerlendirilmesi*. Yayınlanmamış yüksek lisans tezi. Eskişehir. Osmangazi Üniversitesi Sosyal Bilimler Enstitüsü.
- Afacan, Ö. Ve Nuhoğlu, H. (2008). Canlılar bilimi konusunda TIMSS-R (1999) soruları ile LGS (1999) sorularının karşılaştırmalı analizi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD)*, 9 (1), 31-43.
- Algan, S. (2008). *İlköğretim 6 ve 7. Sınıf sosyal bilgiler dersi öğretim programının ölçme ve değerlendirme ögesinin öğretmen görüşleri açısından incelenmesi*. Yayınlanmamış yüksek lisans tezi. Adana. Çukurova Üniversitesi, Sosyal Bilimler Enstitüsü.
- Ananda, S. (2003a). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd. A.
- Ananda, S. (2003b). Achieving alignment. *Leadership*, 33 (1).18-37.
- Baykul, Y. (1992). Eğitimdesistemindedegerlendirme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 7, 85-94.
- Baykul, Y. (1983). Lisans düzeyindeki eğitimde ölçme, soru bankası. *Askeri Yatılı Okul Koşullarında Lisans Düzeyinde Eğitim Sempozyumu Bildiri Metinleri*. İstanbul: 3 Mayıs 1983.
- Berberoğlu, G. (2012). Kapsam geçerliği, *Cito Eğitim: Kuram ve Uygulama*, 15, 10-16.
- Berberoğlu, G., Arıkan, S., Çalışkan, M., Ekinci, P., İpekçioğlu, S. ve Sevgi, S. (2012). Geniş ölçekli sınavlarda kapsam geçerliği nasıl sağlanır? *Cito Eğitim: Kuram ve Uygulama*, 17, 10-22.
- Birinci, D. K. (2014). Merkezi system ortak sınavlarında ilk deneyim: matematik dersi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 3 (2), 8-16.
- Boyle, J. ve Fisher, S. (2007). *Educational Testing: A competence based approach*. USA: Blackwell
- Case, B., Jorgensen, M. A., Zucker, S. (2004). *Alignment in educational assessment*. San Antonio, TX: Pearson. www.pearsonassessments.com
- Çilenti, K. (1984). *Eğitim teknolojisi*. Ankara: Gül Yayınevi.
- Çoban, A., Uludağ, N. ve Yılmaz, A. (2006). Kimya dersinin lise programları ve ÖSS soruları açısından değerlendirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30, 102-109.
- Doğan, A. C. (2009). *Öğretim elemanlarının ölçme ve değerlendirme araçlarına yönelik görüşleri*. Yayınlanmamış yüksek lisans tezi. Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Doğu Akdeniz Üniversitesi (1986). *Öğretim sınav ve başarı tüzüğü*. Gazimağusa.
- Ebel, R.L. (1969). The relation of scale fineness to grade accuracy. *Journal of Educational Measurement* 6 (4), 217-221.
- Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of Educational Measurement* (5th ed). Englewood Cliffs, NJ: Prentice Hall.
- Erden, M. (1993). *Eğitimde program değerlendirme*. Ankara: Pegem, Personel Eğitim Merkezi Yayın No: 6.
- Ertürk, S. (1972). *Eğitimde "program" geliştirme*. Ankara: Yelken Tepe Yayınları.
- Guskey, T.R. (1996). Responding on student learning: Lessons from past-prescriptions for the future. In T.R. Guskey (Ed.), *Communicating student learning* (1996 ASCD Yearbook, pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development.
- Gümüş, M. ve Aykaç, N. (2012). Hayat bilgisi dersi öğretim programının değerlendirme ögesinin öğretmen görüşlerine göre değerlendirilmesi. *Elektronik Sosyal Bilimler Dergisi*, 11 (40), 59-68.
- İğci, C. (1987). *Test sınavlarının güvenilirliğinin araştırılması ile ilgili bir çalışma*. Yayınlanmamış yüksek lisans tezi. İzmir. Ege Üniversitesi Fen Bilimleri Enstitüsü.

- Kanathı, F. (2008). *Alternatif Ölçme ve Değerlendirme Teknikleri Konusunda Sınıf Öğretmenlerinin Görüşlerinin Değerlendirilmesi*. Yayınlanmamış yüksek lisans tezi. Hatay. Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü.
- Karaca, S. (1990). *Yabancı dil sözlü sınavlarının geçerlik ve güvenilirliği*. Yayınlanmamış yüksek lisans tezi. Ankara. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Kaya, N. (2003). *6. Sınıf "demokratik hayat" ünitesinde sosyal bilgiler öğretmenlerinin hazırladıkları yazılı sınav sorularının kapsam geçerliği ve taksonomik boyutunun incelenmesi*. Yayınlanmamış yüksek lisans tezi. Trabzon. Karadeniz Teknik Üniversitesi, Sosyal Bilimler Enstitüsü.
- Kelecioğlu, H., Atalay, K. ve Öztürk, N. (2010). Seviye belirleme sınavı 7. sınıf matematik alt testinin madde hedef uyumu açısından incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1 (1), 37-43.
- Kubiszyn, T., & Borich, G. (2007). *Educational testing and measurement: classroom application and practice* (8th Ed.). Wiley Jossey-Bass.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment and instruction. *Review of Educational Research*, 79 (3), 1-76.
- Marzano, R.J. (2006). *Classroom assessment and grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Okur, M. (2008). *4. ve 5. Sınıf Öğretmenlerinin Fen ve Teknoloji Dersinde Kullanılan Alternatif Ölçme ve Değerlendirme Tekniklerine İlişkin Görüşlerinin Belirlenmesi*. Yayınlanmamış yüksek lisans tezi. Zonguldak. Karaelmas Üniversitesi Sosyal Bilimler Enstitüsü.
- Osborne, J.W. (2003). Effect sizes and the disattenuation of correlation and regression coefficients: Lessons from educational psychology. *Practical Assessment, Research and Evaluation*, 8 (11).
- Özçelik, D. A. (1981). *Okullarda ölçme ve değerlendirme*. Ankara: ÜSYM Yayınları.
- Özçelik, D.A. (2010). *Yeni öğretim programlarına göre okullarda ölçme ve değerlendirme öğretmen el kitabı*. Ankara: Pegem Yayınları.
- Popham, W.J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Sireci, S. G. (2007). On validity theory and the validation. *Educational Researcher*, 36, 477-481.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26 (1), 136-144.
- Taba, H. (1962). *Curriculum development theory and practice*. New York: Harcourt Brace Jovanovich.
- Tan, Ş. (1990). *Eğitimde değerlendirme ögesinin değerlendirilmesi*. Yayınlanmamış yüksek lisans tezi. Ankara. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Turgut, M. F. (1997). *Eğitimde ölçme ve değerlendirme* (10. Baskı). Ankara: Gül Yayınevi.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme metodları*. Ankara: Pegem Yayınları.
- Tyler, R.W. (1946). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

Extended Abstract

Introduction

The evaluation of the evaluation component of an education system can be considered as a 'program evaluation' study (Baykul, 1992). The accuracy of the decision made regarding student achievement within a program plays a significant role. The accuracy of the decisions made depends on the validity and the reliability of the measurement results as well as the criteria used (Özçelik, 1981: 161). It is of utmost importance to assess the extent the learning objectives are reflected in the measurement tools used, and improve the reliability of these tools

accordingly. The inaccuracy regarding the decision made on student achievement increases the cost of the system.

Evaluating the evaluation component in curriculum design can be used as a curriculum evaluation approach due to the fact that it gives information regarding the redesign of the curriculum. Assessing the extent to which each component is effective in the curriculum reveals the strengths and the weaknesses of the curriculum and also the source of the weaknesses, if any. Evaluating the appropriateness of the measurement and evaluation component necessitates assessment of the tools and methods used to assess student success (Baykul, 1983, 95-97). Evaluating the measurement and evaluation component can be done through exploring the reliability and the validity of assessment scores obtained from the tools and methods used to assess student success and also analysing the suitability of the criteria used (Baykul, 1992; Erden, 1993; Tan, 1990). This study aims to evaluate the evaluation component of EGIT112 – Educational Psychology course offered in education faculty of the Eastern Mediterranean University. The study addresses the following research questions:

1. Are the questions used in the midterm and final exams of Educational Psychology course consistent with the course objectives?
2. What are the KR-21 reliability coefficients of scores obtained from the midterm and the final exams of the Educational Psychology course?
3. Is there a significant correlation between midterm, final, quiz, project, and total scores and weighted coefficient scores of letter grades given by the faculty members?

Method

Since this study aims to evaluate the evaluation component of EGIT112 – Educational Psychology course offered in education faculty of the Eastern Mediterranean University, descriptive method is used. The data of the study include EGIT112-Educational Psychology course objectives and midterm and final exam questions. Also, the scores students obtained from midterm and final exams, total scores and weighted coefficient of these scores are used. The participants of the study consist of 303 students enrolled in EGIT112-Educational Psychology course in the faculty of education, Eastern Mediterranean University in 2013-2014 spring semesters. The data obtained for the first research question were analysed through ‘Surveys of Enacted Curriculum (SEC) Model’. The data obtained from the second research question were analysed by using KR-21 formula. KR-21 formula was used to assess the reliability of midterm and final exams. KR-21 formula was chosen to be used in the study due to lack of item statistics. Pearson Moment Correlation Coefficient technique and multiple regression analysis were used for the analysis of the data obtained from the third research question.

Results and Discussion

In light of the data obtained in this study, it is found out that the questions used in the midterm and final exams do not meet the objectives of the course. The questions used both in midterm and final exams do not represent the course objectives proportionally and in a balanced way. The reliability coefficients of midterm and final exams are calculated as moderate level. High correlation between total scores, letter grades and weighted scores indicate that the criteria used in assessment are sensitive, not very flexible and there is not much score lost in transferring total scores to letter grades. The lowest correlation is between final exam scores and quiz scores. Also, the regression analysis indicates that the quiz does not estimate the final exam. Based on the results obtained, the following recommendations can be made:

1. It is necessary to prepare a specification table. Identifying course objectives clearly and common understanding of objectives and their levels by the faculty members are important. Therefore, it is important for the faculty members to work in cooperation with each other. Preparing midterm and final questions based on specification table would increase the content validity of the exams.

2. In-service training on ‘objectives and writing objectives’ should be organized for faculty members to develop the skill of writing objectives and making sure that objectives are in a taxonomy.
3. In-service training on ‘writing multiple choice test items’- an important area of measurement and evaluation- should be offered to develop faculty members’ ability to write test items based on the level of course objectives.
4. Faculty members may need in-service training to be more competent in writing test items.