

ÖĞRENCİLERİN EĞİTİMİ DEĞERLENDİRME VERİLERİNİN ANALİZİNDE ve YORUMLANMASINDA DİKKAT EDİLMESİ GEREKEN HUSUSLAR

Prof. Dr. Hüner Şencan

İÜ İşletme Fakültesi

huners@hotmail.com

ÖZET^a

Üniversitelerde dönem sonlarında öğrencilere uygulanan bir anketle eğitimin veya öğretim üyelerinin değerlendirilmesi uygulaması yaygınlaşmaya başlamıştır. Bu uygulamadan amaç eğitimin etkililiğini artırmaktır. Ancak doğru bir şekilde uygulanmadığı ve istatistik analizler doğru bir şekilde yapılmadığı zaman öğretim üyelerinde moral düşüklüğüne neden olabilmektedir. Bu yazıda geçerli bir anket formu oluşturulması, doğru istatistiksel analizlerin uygulanması ve sonuçların yarar sağlayacak bir şekilde sunulmasıyla ilgili konular üzerinde durulmuştur.

ABSTRACT

Turkish universities have started to utilize student evaluations of faculty and/or education at the end of academic terms, in order to improve the efficiency in university education. However incorrect implementation and incorrectly constructed statistical analysis may lead the faculty to develop low morale. In this paper, author focuses on issues such as; developing a valid evaluation form, implementing the correct statistical analysis, and presenting the outcomes effectively in order to create overall utility.

^a Bu yazının hazırlanması sırasında görüş ve önerilerinden yararlandığım Prof. Dr. İsmail Kaya'ya ve Prof. Dr. Kemal Kurtuluş'a teşekkür ederim.

GİRİŞ

Kamu üniversitelerinde öğrencilerin öğretim üyelerini değerlendirmeleri prosedürü geçmişte yaygın bir uygulamaya sahip olmamıştır. Son yıllarda bu üniversitelerde bağlı değerlendirme sistemine geçilmesiyle birlikte öğrencilerin öğretim üyelerini değerlendirmesi uygulaması gündeme gelmiştir. Bu tür uygulamalarda toplanan verilerin istatistiksel analizi ve analiz sonuçlarının yorumlanmasında bazı güçlüklerle karşılaşmaktadır. Bu makalenin amacı, geçerli bir ölçüm yapmak, verilerin istatistiksel analizinde doğru yöntemi seçmek ve elde edilen sonuçları doğru bir şekilde yorumlamak için dikkat edilmesi gereken hususları belirlemektir. Yazı, üç ana başlık çerçevesinde kurgulanmıştır. Öncelikle, uygulama öncesi dönemde yapılması gereken güvenilirlik ve geçerlilik analizleri üzerinde durulmuştur. İkinci ana başlıkta, uygun istatistikî analizlerin seçilmesi konusuna değinilmiş ve üçüncü ana başlıkta ise sonuçların sunumu, tablollaştırılması ve yorumlanmasında dikkat edilmesi gereken temel noktalar ele alınmıştır.

UYGULAMA ÖNCESİ

Uygulama öncesi dönemde, bir taraftan eğitimin değerlendirilmesine yönelik olarak geliştirilen ölçüm aracının geçerlilik ve güvenilirlik analizleri yapılırken diğer taraftan da anket uygulanacak sınıfların özellikleri incelemeye alınır. Ölçüm sonuçlarının geçerliliği, doğru ölçeğin/indeksin kullanılmasına ve değerlendirmelerin uygun sınıf büyüklüklerinden elde edilmesine bağlıdır.

Ölçüm Aracının Niteliği

Öğrencilerin öğretim üyelerini veya eğitimi değerlendirmeleri için geliştirilen ölçüm aracı ya bir tutum ölçeği veya bir indekstir. Ölçüm işlemine girişilmeden önce ölçüm aracının niteliği açık bir şekilde belirlenmelidir.

İndeksle tutum ölçeği, ortak yönleri olmasına karşın aynı şey değildir.¹ Ortak olan yönleri her ikisinin de *sıralı ölçek niteliğinde* olmaları ve bileşik

değişkenlere dayanmasıdır. Diğer bir deyişle belirli bir ölçüm alanıyla ilgili olarak birden fazla ifadeden yararlanılmasıdır. İndeksle tutum ölçeği arasındaki farklılık ise puanların atanma biçiminden kaynaklanır. İndekslerde belirli bir özelliğin bulunma veya bulunmama durumuna göre toplam puan hesaplanır. İndekslerde ölçüm noktaları *evet-hayır* şeklinde veya beşli, yedili ölçek dereceleri şeklinde düzenlenir ve indeksin puanı *evet* yanıtları veya işaretlenen dereceler toplanarak bulunur. İndeksler sıralı ölçek verilerine sahiptir. Beşli dereceleme ölçeğinin kullanıldığı durumda dahi arka planda belirli bir kavramsal yapıyı ölçmeye yönelik olarak oluşturulmadığından indeksler ölçek sayılmaz. Tutum ölçeklerinde ise, her bir maddede ölçülmek istenen özelliğin yoğunluğu dikkate alınır. Tutum ölçekleri, indeks puanlarıyla karşılaştırıldığında daha fazla bilgi yüküne sahiptir. Bir başka farklılık tutum ölçeklerinin aynı zamanda incelenen örnekleme bağımlı olmasıdır. Belirli bir örneklem için geçerli olan ölçek ifadeleri başka bir örneklem için geçerli olmayabilir. Tutum ölçeklerinde araştırmacı gizli yapıları ortaya çıkarmayı amaçlar.

Tutum ölçekleri standartlaştırılmış olan araçlardır. Tutum ölçeklerinde arka planda, gizli faktörler söz konusudur. Tutum ölçeklerinin geliştirilmesi, uygulanma biçimi ve puanlanması belirli prosedürlere tâbidir. Tutum ölçeğinde her bir ifade belirli bir *ölçek değerine* sahiptir. Ayrıca tutum ölçeği bir bütün olarak anlamlıdır, maddeler tek başına ele alınıp değerlendirilmez.² İndeksler veya bazı yazarlar tarafından “tepki ölçekleri” olarak adlandırılan ölçüm araçları ise, içerdiği maddelere gelen yanıtları belirlemek için kullanılır. Maddenin puanı, ölçeğin toplam puanıyla ilgili değildir. İndeksler veya tepki ölçekleri için kuramsal temel şartı yoktur. Araştırmacı gizli boyutları ortaya çıkarmak gibi bir amaç peşinde de değildir. Tepki ölçekleri kişilerin yanıtlarını 2 dereceli, 5 veya 7 dereceli bir ölçek üzerinde saptayacak şekilde düzenlenmiş olabilir. Tutum ölçeklerinde *arka plandaki faktörler* önemlidir. Ölçekler tek, iki, üç veya çok daha fazla faktörlü olabilir. Tek faktörlü ölçekler başlıca üç grupta incelenir: Likert ölçekleri, Thurstone ölçekleri ve Guttman ölçekleri. Anlamsal farklılık ölçekleri (semantic differential) ise çok faktörlü ölçek grubunda değerlendirilir (Cushman ve Rosenberg, 1991).³ Bu ölçeklerin hepsinde *toplam puan* temel alınır. İndekslerde veya tepki ölçeklerinde ise faktör aranmaz, ancak indeksler de tek bir konuyu ölçmeye yönelik olarak oluşturulur. Tepki ölçeklerinde toplam

puan hesaplanmış olsa bile araştırmacı isterse her bir maddenin merkezî dağılım ölçülerini ayrıca hesaplayabilir. Literatürde tepki ölçekleri değişik şekillerde adlandırılmıştır. Bazı yazarlar “Likert tipi ölçek” ifadesini kullanırken, bazıları “dizin (indices)” bazıları da “tek maddeli ölçekler (single item scales)” ifadelerini tercih etmişlerdir. Üniversitelerde öğrencilerin eğitimi değerlendirmeleri için bir ölçüm aracı geliştirilirken okul yöneticileri bilgi toplamak için bir tutum ölçeği mi yoksa tepki ölçeği mi geliştireceklerine karar vererek ölçüm aracının niteliğine uygun güvenilirlik ve geçerlilik analizlerini yapmalıdırlar.

Tutum Ölçeğinin veya Tepki Ölçeğinin Güvenilirliği

Kullanılan ölçüm aracının niteliğine göre güvenilirlik analizinin yapılıp yapılmayacağına karar verilir. Tepki ölçeklerinde gizli yapıları ortaya çıkarma gibi bir amaç güdülmendiğinden güvenilirlik analizi yapılmaz. Bağozzi (1994) indekslerde iç tutarlılık analizi yapılmasının anlamlı olmadığını belirtmiştir.⁴ Bununla birlikte eğitimi değerlendirme ölçüm aracı gizli alt boyutları ortaya çıkaracak bir şekilde düzenlenmişse güvenilirlik analizi yapmak gerekir. Marsh ve arkadaşları yaptıkları faktör analizi araştırmaları sonucunda eğitimin çok boyutlu olduğunu bulmuşlar ve başlıca dokuz boyut belirlemişlerdir.⁵ Bu nedenle bu yazıda geliştirilen ölçüm aracının bir “ölçek” olduğu varsayımından hareket edilmiştir.

Güvenilirlik ifadelerin iç tutarlılığı, istikrarlılığı ve genellenebilirliği ile ölçülür. İç tutarlılığın Cronbach alfa katsayısı ile test edilmesi ve bu katsayının en az ,70’in üzerinde olması gerekir. İç tutarlılık analizleri, ölçeğin öğrencilere uygulanmasından önce bir pilot araştırma kapsamında yapılmalı ve daha sonra esas veriler üzerinde ikinci kez tekrar hesaplatılmalıdır. Maddelelere ait tüm yarılar arasındaki korelasyon ortalamalarını gösteren Cronbach alfa katsayısı güvenilirliği belirleyen önemli bir değerdir, ancak ölçeğin güvenilir sayılması için öğrenci algıları ölçekteki maddelerin toplam puanına veya ölçeğin aritmetik ortalama dizin değerine dayandırılmalıdır.⁶

İstikrarlılık, bir ölçüm aracının uzun zaman diliminde bir öğretim üyesiyle ilgili olarak hep benzeri sonuçları vermesidir. Bunun için anket formu birincisi eğitim dönemi içinde ve ikincisi ise dönem sonunda olmak üzere iki farklı zaman diliminde uygulanır ve elde edilen sonuçlar karşılaştırılır. Ayrıca istikrarlılık

paralel formlar yöntemiyle de test edilebilir. Öğretim üyesi farklı iki formdan da benzeri puanları almışsa ölçüm sonuçlarının istikrarlı olduğu söylenir. İki ölçüm sonucu arasında orta derecede veya güçlü bir ilişki bulunmalıdır.⁷

Genellenebilirlik, ölçeğin bir öğretim üyesinin öğretme etkinliği hakkında doğru bilgiler vermesidir. Ölçek sadece belirli bir programa ait olarak bir öğretim üyesinin etkili olduğunu yansıtmamalı, diğer derslerden ve diğer programlardan aldığı puanlar arasında da yüksek bir ilişki olduğunu ortaya koymalıdır. Gilmore, Kane ve Naccarato (1978) bir öğretim üyesi hakkında karar vermek için, bu öğretim üyesinin vermiş olduğu en az beş dersin puan ortalamalarının alınması gerektiğini belirtmişlerdir.⁸

Ölçüm Aracının Geçerliliği

Kullanılan ölçüm aracı bir tutum ölçeği veya tepki ölçeği ise her ikisi için de değişik düzeylerde geçerlilik analizleri yapılmalıdır. Güvenilirlik analizinin yapılması yeterli değildir. Güvenilirlik analizi yapmak, geçerlilik için yeter şartı oluşturmaz.⁹ Çünkü bir ölçüm aracı güvenilir olduğu halde geçerli olmayabilir. Öte yandan bir ölçüm aracı geçerli ise duruma göre, aynı zamanda güvenilir olabilir.

İndeks türü ölçüm araçları için sadece yüzey ve içerik geçerliliği analizleri yapılır. Gizli yapıları ortaya çıkarma gibi bir amaç güdülmediğinden yapısal geçerlilik analizi yöntemleri uygulanmaz.

Tutum ölçeği şeklinde oluşturulan ölçeklerde ise geçerlilik analizlerinin tümü uygulanır. Bu yazıda öğrencilerin eğitimi değerlendirmeleri için “tutum ölçeğinin” kullanıldığı varsayımından hareket edilmiştir.

Tutum ölçeğinin geçerliliği başlıca dört şekilde test edilir. Birincisi yüzey geçerliliğidir. Yüzey geçerliliği sorulan soruların veya kullanılan ifadelerin ölçüm amacına uygun olmasıyla ilgilidir. Bu geçerlilik analizinde değerleyiciler arası mutabakat aranır. Hakemlerin veya uzmanların değerlemeleri arasındaki korelasyon yüksekse mutabakatın sağlandığı söylenir.

İkincisi içerik geçerliliğidir. İçerik geçerliliği ölçeğin eğitim faaliyetlerine ilişkin gerekli olan boyut veya soruların tümünü yeterli ölçüde içermesidir. Bu konuda ölçeğin adı ile içerdiği boyutların veya maddelerin niteliğine bakılır. İçerik geçerliliği, ölçeğin ölçtüğünü iddia ettiği başlıca konu alanları olarak da

tanımlanabilir. Örneğin, öğretim üyesinin etkililiği ölçülüyorsa; öğretim üyesinin bilgisi, öğrencilerle ilişkileri, öğrencilere yardımcı olması, zaman kullanma alışkanlığı, not verme uygulamaları başlıca boyutlardır. Genel olarak programın etkililiği ölçülüyorsa bu kez; öğretim üyesinin başarısı, programın uzunluğu, kantin hizmetleri, yönetim ve sekreterlik hizmetlerinin yeterliliği gibi konular ölçeğin temel boyutları olarak belirlenir. Öğretim üyelerinin etkililiğini belirlemeyi amaçlayan bir ölçeğe; dersin gerekliliği, saat sayısının az veya çok olması, dersin sabah veya akşam saatlerine konması gibi doğrudan öğretim üyesiyle ilgili olmayan ifadeler konması ölçeğin içerik geçerliliğini bozar. Eğer bu ifadeler de konacaksa ölçeğin adını, Eğitim Etkililiği Ölçeği olarak değiştirmek gerekir.

Üçüncü yöntem kriter geçerliliğidir. Tahmin geçerliliği adı da verilen bu yöntemde ölçek sonuçları kriter puanlarla karşılaştırılır. Kriter geçerliliğinde kullanılan ölçek, eğitimi etkili bir şekilde sunan veya etkili bir şekilde öğreten öğretim üyelerini gerçekten ortaya çıkarmalıdır. Ancak bunun ölçümü veya saptanması oldukça zordur. Etkili öğretimi ortaya koyacak tek bir kriter yoktur. Araştırmacı karşılaştırma yapmak için birden fazla kriterden yararlanabilir.¹⁰ Bu kriterlerin başlıcaları şunlardır: Öğrencilerin öğrenme durumlarını gösteren yazılı sınav sonuçları, öğretim üyesinin kendisini değerlendirmesi, okul yöneticisinin öğretim üyesini değerlendirmesi, meslektaşlarının öğretim üyesini değerlendirmesi, mezunlar derneği yöneticilerinin öğretim üyesini değerlendirmesi, dış gözlemcilerin öğretim üyesini değerlendirmesi, açık uçlu sorulara gelen öğrenci yorumları ve son olarak diğer yan faktörler. Puanları etkileyebilecek yan faktörler ise öğretim üyesine, öğrenciye, programın düzeyine ve yönetim uygulamalarına ilişkin olabilir. Örneğin, öğretim üyesinin unvanı, yöneticilik statüsünün bulunup bulunmadığı,^b fakülte öğretim üyesi olması veya dışarıdan gelmesi, öğrencilerin motivasyonu, programın zorluk derecesi, değerlendirme sırasında öğretim üyesinin salonda bulunması, değerlendirmenin yapılma amacının öğrencilere bildirilmesi, öğrencilere değerlendirme formuna ad

^b Bazı araştırmalarda bu faktörlerin öğrencilerin öğretim üyelerini değerlendirmelerinde çok az veya hiç bir etkisinin olmadığı bulunmuştur. Yurt dışında yapılan başka araştırmalarda ise, asistanların verdikleri derslerle profesörlerin verdikleri derslerin değerlendirilmesinde önemli ölçüde farklılıklar olduğu ortaya çıkmıştır.

ve soyadlarının yazdırılması bunlar arasındadır. Bu faktörler, anketin doldurulmasında yanlılığa neden olur.

Dördüncü yöntem yapısal geçerliliktir. Bu yöntemde ölçeğin faktöriyel bir yapıya sahip olup olmadığı ve ortaya çıkarılan faktöriyel yapıların literatürdeki yapılarla benzerlik gösterip göstermediği araştırılır. Bazı araştırmacılar ölçeğin sadece faktöriyel bir yapıya sahip olmasını yeterli görürken^c başka araştırmacılar bu yapıların aynı zamanda daha önceki araştırma sonuçlarıyla da tutarlı olup olmadığının kontrol edilmesini savunmuşlardır.¹¹ Trochim'e göre (2000) yapısal geçerlilik için *ayrılma* ve *birleşme* analizinin birlikte yapılması gerekir. Bu analizlerden sadece birisinin yapılması yeterli değildir.¹² Birleşme analizi, farklı noktaları ele alan çok sayıda ifadenin tek bir yapıya işaret edip etmediğini belirlemek amacıyla yapılır. Bunun için ölçeğin maddeleri arasında faktör veya korelasyon analizi yapılır ve çıkan sonuçlara bakılır. Birleşme analizinde korelasyon değerlerinin yüksek çıkması gerekir. Bir diğer yöntem, aynı yapıyı ölçen farklı bir anketin kullanılması ve iki anket sonuçları arasındaki korelasyonun araştırılmasıdır.¹³ Ayrılma analizinde ise araştırmacı, geliştirilen Eğitimi Değerlendirme Ölçeği'ni pilot araştırma sırasında bir veya daha fazla farklı yapıyı ölçen anket sorularıyla birlikte uygular. Karma nitelikteki bu ölçeğin maddeleri arasında yine korelasyon analizi yapılır. Korelasyon analizi sonucunda farklı yapılar arasında ayrışmanın ortaya çıkması gerekir. Aynı yapıyı ölçen ölçek maddeleri arasındaki korelasyon yüksek ve farklı yapıları ölçen maddeler arasındaki korelasyon düşük olmalıdır. Bu ayrışma net bir şekilde görülebiliyorsa ölçeğin geçerli olduğuna karar verilir.

^c Ölçeğin faktöriyel yapısını ortaya çıkarmak için, pilot araştırma verileri üzerinde keşfedici faktör analizi yöntemi uygulanır. Öz değeri 1'den büyük olan değişkenler tek bir yapıyı ölçen maddeler olarak değerlendirilir ve nihai ankete sadece bu ifadeler alınır. Bir faktöre ait değişkenler belli bir yapıyı ölçen ifadeler olarak değerlendirilecekse bu ifadelerin faktör ağırlığı en az %40 olmalıdır. Bu konuda ayrıca bk., D. Garson, "Factor Analysis [Faktör Analizi]," t.y., <<http://www2.chass.ncsu.edu/garson/pa765/factor.htm>> (29.07.2002), Overview; MEİ, "Measurement Theories [Ölçüm Kuramları]," t.y., <<http://www.MeasurementExperts.org/ClassicalTestTheory.htm>> (29.07.2002), Factor Analysis.

Verilerin Sınıf Temsil Etme Derecesi

Veriler sınıf mevcudunun en az %75'inden sağlanmış olmalıdır. Daha düşük katılıma ait sonuçlar güvenilir değildir.¹⁴ Sonuçlar sınıfın dörtte üçünden daha az öğrenciye aitse, böyle bir durumda sonuçların sadece sınıftaki belirli bir öğrenci grubunun görüşlerini yansıttığı söylenebilir. Ana kütlenin tamamına ulaşamadığı durumda veri toplanan küme, örneklem olarak değerlendirilir. Bazı üniversiteler sınıf büyüklüklerine göre belirli cevaplama oranları belirlemişlerdir. Örneğin 5-20 kişilik sınıflarda öğrencilerin %80'inin, 20-30 kişilik sınıflarda öğrencilerin %75'inin; 30-50 kişilik sınıflarda öğrencilerin %66'sının 50-100 kişilik sınıflarda öğrencilerin %60'nın ve 100'den büyük sınıflarda öğrencilerin en az %50'sinin değerlendirmeye katılmış olması önerilmiştir.¹⁵

İstatistiksel analiz sonuçlarında verilerin ana kütlenin yüzde kaçını temsil ettiği açık bir şekilde belirtilmelidir. Ayrıca veri toplanamayan öğrencilerin hangi nedenlerle anketi doldurmadıkları da açıklanmalıdır. Çünkü kendisine anket verilen öğrencinin bu anketi geri iade etmemesiyle, kendisine anket ulaştırılamaması farklı şekillerde yorumlanabilecek iki ayrı olgudur.

Verilerin Devam Eden Öğrencilere Ait Olması

Kendilerine anket uygulanan öğrenciler yine en az %75 oranında derse devam etmiş olmalıdırlar. Derse devam etmeyen bir öğrencinin her hangi bir şekilde yapacağı değerlemeler sonuçları çarpıtacaktır. Öğrencilerin belge üzerinde derse devam etmiş gözükmeleriyle fiilen devam etmeleri farklı iki olgudur. Öğretim üyeleri, %75 oranında derse devam etmedikleri halde öğrencilere devam hakkı vermiş olabilirler. Ancak bu kişiler değerlendirmeye katılmamalıdırlar. Anket formunda öğrencilere derse ne ölçüde devam ettiklerine ilişkin bir soru sormanın da fazla bir anlamı yoktur. Çünkü bu tür sorulara gelen yanıtlara güvenilmez. Anket sadece %75 devam hakkını fiilen gerçekleştiren öğrenciler tarafından doldurulmuş olmalıdır.

Sınıf Büyüklüğü

Değerlendirme yapılacak sınıfın büyüklüğü sonuçların güvenilirliğini büyük ölçüde etkiler. Bazı araştırmacılar yargıya varmak için en küçük sınıf

büyükliğini 10 olarak belirlemişlerdir. Sınıf büyüklüğü 30'un altında olduğunda özel istatistik yöntemlerin kullanılmasının gerekli olduğu vurgulanmıştır. Öte yandan bu rakamın 30'un üstünde olması halinde dizin değerlerinden hareket ederek parametrik istatistikî analizleri kullanma olanağı ortaya çıkar.

Öğrencilerin Bazı İfadelere Yanıt Vermemeleri

Öğrenciler çoğunlukla bazı ifadelere yanıt vermemişlerse yanıt verme oranları takip edilmeli ve analiz buna göre yapılmalıdır. Eğer bir ifadeye anketi dolduran öğrencilerin %50'sinden fazlası yanıt vermemişse bu soru veya ifade analizden çıkarılmalıdır. Bazı ifadeler/sorular belirli dersler veya öğretim üyeleri için geçerli olmayabilir. Bunun için beş dereceli ölçeğin dışında ayrı bir şık belirlenmeli ve bu şıkka gelen yanıtlar açık bir şekilde gözükmelidir. Değerlemeye katılım, %75'in üzerindeyse bazı sorulara gelen yanıt sayısının bu oranın bir miktar altında kalması önemli bir sakınca oluşturmaz. Ancak soruyu/ifadeyi yanıtlayan öğrenci sayısı tam olarak gösterilmelidir. Boş değer içeren değişkenlerde yorumlar SPSS'deki "geçerli yüzde" dağılım sütunu dikkate alınarak yapılmalıdır.

UYGULAMA SONRASI

Anket formunun öğrencilere uygulanmasından sonra verilerin kotlanması ve istatistikî analizlerin yapılması aşamasına geçilir. Sonuçların geçerli olabilmesi için bu evrede kotlamanın ve anket sorularının yapısal düzenlemesinin dikkatli bir şekilde incelenmesi gerekir.

Soruların Doğru Kotlanması

Özellikle dizin aritmetik ortalama değerlerinin doğru hesaplanabilmesi için ifadelerin olumlu veya olumsuz olma durumları dikkate alınarak gerektiğinde ters yönlü puanlama yapılması gerekebilir. Ters yönlü puanlama yapılmadığı durumda örneğin sadece iki madde yüzünden öğretim üyelerinin puan ortalamaları olduğundan daha düşük gözükür. İfadelerin olumlu veya olumsuz olma durumu cümle yapılarına bakılarak değil, anlamlarına bakılarak araştırılmalıdır. Doğru kotlama yapılması konusunda dikkat edilmesi gereken bir

diğer nokta, ifadelerin birbirleriyle ilgili olmasıdır. Örneğin “Bu ders için programda ayrılan süre yeterliydi” ifadesine gelen *Kuvvetle Kabul* yanıtı ile diğer ifadelere gelen *Kuvvetle Kabul* yanıtları arasında aynı yönde bir ilişki olmalıdır. Ters yönlü bir ilişki ortaya çıkıyor ve puanlar değiştirilmiyorsa kotlama hatası yapılıyor demektir.

En Önemli Soru

Anket formunun yapısal düzenlemesinde, sonucu bir bakışta görmek için “en önemli soru” olarak nitelendirilen bir ifade yer almış olabilir. Öğrencilerin öğretim üyelerini değerlendirmelerinde belirli bazı eksikliklere ve yetersizliklere rağmen genelde bir bütün olarak nasıl bir kanaate sahip oldukları bu som ile öğrenilmeye çalışılır. Bu soruya genel değerlendirme sorusu (global item) denir. Örneğin, şöyle bir som sorulmuş olabilir: “Her şeyi bir bütün olarak düşündüğünüzde bu öğretim üyesinin vermiş olduğu dersten ne ölçüde yararlandığınızı düşünüyorsunuz?” Tek bir somdan oluşması nedeniyle bu somya ait verilerin analizinde frekans dağılımı, mod ve medyan değerleri alınır. Literatürde, beş dereceli ölçek değerleri kullanılmış olmasına karşın bazı araştırmacıların bu değerlerin aritmetik ortalamalarını hesaplatıp yorumu buna göre yaptıkları görülmektedir. Bu tür bir uygulamada global som, bir çok ifadeyi içeren bir dizin değeri gibi ele alınmaktadır. Ancak global sorunun aritmetik ortalama değerlerine bakarak bir öğretim üyesinin 3,00 ve diğerinin 4,50 alması halinde ikinci öğretim üyesinden öğrencilerin %50 daha fazla yararlandıkları söylenemez.¹⁶ Kaldı ki bu tür aritmetik ortalama puan değerleri, eğitimin etkililiğini değil, öğretim üyesinin popüler olduğunu veya öğrencinin memnuniyetini gösteriyor olabilir.¹⁷ Bu nedenle, istatistikî sonuçları yorumlarken ihtiyatlı hareket etmek gerekir. Anket formlarındaki global soruların, eğitimi iyileştirmedeki katkısının çok az olduğu belirtilmiştir. Bu som ile, öğretim üyesine hangi alanda iyileştirme yapması gerektiği konusunda kesin bir bilgi verilemez.¹⁸ Eğitimin etkililiğini iyileştirme açısından faktör puanlarının aritmetik ortalamaları daha anlamlıdır. Ayrıca tek bir soruyla elde edilen veriler için güvenilirlik ve geçerlilik analizi söz konusu değildir. Bu tür sorular sadece yıllar arasında karşılaştırma yapmak açısından bir anlam ifade edebilir.

Frekans Analizi

Değerlendirme ölçeği 5 veya 7 dereceli Likert ölçeği şeklinde tasarlanmışsa her bir ifadenin yüzde dağılımları verilir. Ancak yüzde dağılımları verilirken beş öğrencinin %20'si ile 100 öğrencinin %20'si aynı değildir. Bunun için yüzde değerleri yanında öğrenci sayısı da verilmelidir.¹⁹ Yüzde dağılımlarının gösterilmesinde üç farklı yöntem uygulanabilir. Birincisinde her bir derecenin yüzde dağılımı gösterilir. Bazı derecelere hiç bir işaretleme yapılmamış bile olsa bu dereceler de tabloda verilmelidir. İkinci yöntemde 5 veya 7 dereceli ölçek üç dereceli ölçeğe indirgenerek tablo yapılır. *Kuvvetle Katılıyorum* şıkkı ile *Katılıyorum* şıkkı birleştirilir ve örneğin *Katılıyorum* gibi yeni bir isim belirlenir. Bir sonraki şık *Kararsızım* ve diğeri de *Katılmıyorum* şıkkıdır. Böylece okurlar bir bütün olarak belirli ifadeye katılanları, kararsız kalanları ve katılmayanları topluca görme olanağına sahip olurlar. Üçüncü yöntem, özellikle karşılaştırmalı tabloların verilmek istenmesi durumunda uygulanır. Bu yöntemde sadece 4 ve 5 derecesini işaretleyen öğrencilerin puanı verilir ve yorum buna göre yapılır. Çünkü hedef bu puanları işaretleyen öğrencilerin oranının yüksek olmasıdır. Tablonun diğer sütunlarında ise aynı branşta ders veren diğer öğretim üyelerinin frekans dağılımları, geçen dönemin frekans dağılımı, norm veya kriter puanlarına göre yüzde dağılımları verilir.

Merkezî Dağılım Değerleri

Dereceleme ölçekleri, bir öğretim üyesinin vermiş olduğu ders hakkında genel bir değerlendirme yapmak amacıyla kullanılır. Tek tek, her bir ifadeye ait puanların toplamını alıp sonra bunları öğrenci sayısına bölerek o ifadenin aritmetik ortalama puanını bulmak geçerli bir yöntem değildir.²⁰ Cuseo'ya göre bu yöntem sadece etkisiz değil, her bir ifadenin ağırlığı ve önemi diğer ifadelerin öneminden farklı olduğundan değişkenleri aritmetik ortalama değerlerini dikkate alarak karşılaştırma yapma açısından da sakıncalıdır. İfadelerin aritmetik ortalama değerlerine bakarak bir öğretim üyesine söylenen “ders anlatmada iyisin, ama yardımcı ders malzemesi hazırlamada pek iyi değilsin” sözlerinin bir anlamı yoktur. Çünkü o ders için yardımcı ders malzemesi hazırlama pek önemli olmayabilir. Öğrencilerin gözünde ifadelerin önemi ve ağırlıkları eşit değildir. Bununla birlikte literatür incelendiğinde bazı araştırmacıların her bir ifadenin

aritmetik ortalama ve standart sapma deęerlerini hesaplattıęı g r l r. Bu uygulama ancak yanında frekans deęerleri, mod ve medyan deęerlerinin verilmesi halinde hoř g r lebilir. Tek bir ifadeye gelen yanıtlara baęlı olarak bir konu hakkında karar vermek,  rneęin “ ęretim  yesinin derse hazırlıklı gelme durumu” gibi,  l  m hatasına neden olur. Tek bir ifadeye gelen yanıtlar  ęrencinin ger ek tutumunu yansıtmaz, sadece belli bir g r ře yakın olduęu hakkında bilgi verir. Eęer aynı konuyla ilgili olarak d rt veya beř ifade kullanılırsa daha az  l  m hatası yapılır.²¹ Bu nedenle okul y neticileri  l  k oluřtururlarken  l  ęin fakt r yapısını, madde sayısını, sonu ları nasıl deęerlendirip yorumlayacaklarını daha bařtan belirlemelidirler.

Nonparametrik nitelikteki bir  l  kte tek tek madde deęerleri i in merkez  daęılım  l  s  olarak medyan veya mod deęerleri kullanılır. Gerekiyorsa tabloda bu deęerlerin her ikisi de verilebilir. Medyan, deęerlendirmede u  deęerlerden daha az etkilendir. Puanların daęılımında eęer  ok az 3 varsa veya hi  yoksa iki mod var demektir. B yle bir durumda,  ęrencilerin s z konusu maddeyi deęerlendirmelerinde iki gruba ayrıldıkları s ylenir. Medyan puanı y ksek olduęu halde  ęrencilerin bir b l m  4 ve 5’i; b y k bir  oęunluęu da 1 ve 2’yi iřaretlemiřlerse belirli sayıda  ęrencinin y ksek puanlarına karřı olduęu s ylenir. Bu durum  zellikle  ęretim  yesinin sınıfın %10’una y nelik bir ders anlatma ve  ęretme politikası g tt ę  durumlarında ortaya  ıkar.²²

Aritmetik ortalama deęeri,  l  kteki t m ifadeleri temsil ettięi zaman anlamlı olabilir. Bu deęer artık bir dizin deęeridir. Aritmetik ortalama deęeri, merkez  odaklařmayı g sterirken standart sapma puanların ortalama etrafında ne kadar farklılařabileceęini ortaya koyar. Beř dereceli bir  l  kte dizin deęerine ait standart sapmanın sıfır  ıkması  ęrencilerin fikir birlięi i inde olduęunu g sterirken standart sapmanın 1,2 ve daha y ksek olması,^d aritmetik ortalama deęerinin  ęrencilerin kanaatlerini ortaya koymak i in iyi bir g sterge olmadığı anlamına gelir.²³ Standart sapmanın 1’in altında olması deęerlendiricilerin belirli bir fikir etrafında k melendięini g sterir. Bazı arařtırmacıların standart sapma yerine varyans deęerini hesaplatmaları g r lmektedir. Varyansın hesaplatılması,

^d K. Kurtuluř 1,2’lik bir standart sapmanın 5 dereceli  l  kler i in y ksek olduęunu belirtmiř bu deęerin 7 dereceli  l  kler i in kullanılabileceęini ifade etmiřtir. Kurtuluřa g re 5 dereceli  l  kler i in standart sapma 1 olarak d ř n lmelidir.

örneklem verilerinden hareket edilerek ana kütlede ortaya çıkabilecek değişkenliği tahmin etmek için başvuru bir yöntemdir. Sadece bir üniversite bünyesinde az sayıda öğrencinin örneklem olarak belirlenmesi ve bu örneklemde elde edilen sonuçların ana kütleye genellenmek istenmesi durumunda uygulanabilir. Tek bir sınıfın verileri temel alınarak yapılan analizde varyans değerleri bir anlam ifade etmez. Öte yandan aritmetik ortalama ve standart sapma değerleri ise, pratik bir yorum yapmaktan çok istatistikî testlerle belirli bir hipotezi kabul etme veya ret etmek için daha uygundur.

Dereceleme ölçeklerinde esas olarak tüm ifadelerin toplam puanlarından hareket edilir.²⁴ Öğretim üyelerindeki gelişmeleri pratik olarak izleyebilmek için daha doğru olanı toplam puanların kullanılmasıdır. Toplam puanların kullanılması halinde yıllar veya dönemler arasında karşılaştırma yapma ve gelişmeleri görme imkanı ortaya çıkar. Toplam puan, ölçeğin tamamına ait olarak veya eğer belirli faktörleri içeriyorsa faktörlerin toplam puanı ve ölçeğin genel toplam puanı alınarak hesaplanabilir. Ancak bu yöntemin sakıncası, kullanılan değerlendirme formlarının her yıl değiştirilmesi ve farklı programlarda farklı formların uygulanması nedeniyle toplam puanları arasında karşılaştırma yapmaya imkan vermemesidir. Kullanılan formun programlar itibariyle, yıllar veya dönemler itibariyle madde sayısında değişiklik yapılma ihtimali varsa dizin değeri olarak bu kez aritmetik ortalama ve standart sapma değerlerinin hesaplanması daha uygundur.²⁵ Ayrıca Likert ölçeğinde eğer bazı ifadeler için işaretleme yapılmamışsa yine aritmetik ortalama puanının hesaplanması daha doğrudur.²⁶ Tek tek maddelerin aritmetik ortalama puanına nazaran dizin değerinin aritmetik ortalama puanı daha anlamlıdır. Öğrencilerin dağılımı hakkında kesin bir fikir vermemekle birlikte karşılaştırma yapmak için elverişli bir rakamdır.

SONUÇLARIN RAPORLANMASI

Sonuçların raporlanmasında önemli olan konu, bulguların kuruma ve öğretim üyelerine yarar sağlamasıdır. Yöneticiler bulgulara bakarak karar verebilmeli ve eğitimciler de kendilerini hangi alanlarda geliştirmeleri gerektiği konusunda bir fikir sahibi olabilmelidirler.

Puanların Karşılaştırılması

Hazırlanan rapor karşılaştırma yapmaya imkan tanınmalıdır. Bir öğretim üyesine ait etkililik puanının, eğitim branşı farklı bir öğretim üyesinin puanıyla karşılaştırılması doğru sonuç vermez. Karşılaştırma yapılan dersler arasında benzerlik olmalıdır.²⁷ Çünkü her bir dersin işleniş biçimi, ödev verilmiş biçimi ve içeriği farklıdır. Diğer taraftan puanların aynı programda ders veren öğretim üyelerinin puan ortalamalarıyla karşılaştırılması da yanıltıcı olabilir. Örneğin bir öğretim üyesinin aritmetik puan ortalaması 3,94 ve tüm öğretim üyelerinin ortalaması da 3,92 olsun. Böyle bir durumda bu öğretim üyesinin 0,02 puan yüksek alması ortalamadan daha iyi olduğu anlamına gelmeyebilir. Puanların dağılımı önemlidir. Dağılım dikkate alındığında ortalamadan yüksek gibi gözükken bir öğretim üyesinin durumu daha zayıf çıkabilir. Bu nedenle gerek öğretim üyeleri arasında karşılaştırma yaparken ve gerekse grup ortalamalarıyla karşılaştırma yaparken frekans değerlerini kullanmak daha anlaşılır sonuçlar verecektir.

Puan karşılaştırmaları yıllar ve dönemler arasında yapılıyorsa bu kez de sınıf büyüklüklerinin dikkate alınması gerekecektir. Çünkü yapılan araştırmalar sınıf büyüklüğü arttıkça puanların düşme eğilimi gösterdiğini, küçüldükçe ise büyüme eğilimi taşıdığını ortaya koymuştur. Sınıf büyüklüğü 20'nin altındaysa öğretim üyelerinin yüksek puanlarla değerlendirildikleri, 40 ilâ 100 arasındaysa düşük puan ve 100'den fazlaysa öğretim üyelerinin orta derecelerde değerlendirildiği bulunmuştur.²⁸ Böyle bir durumda puanların düşük veya yüksek olduğu değerlendirmesi de ihtiyatlı bir biçimde yapılmalıdır.

Puan karşılaştırmasında dikkat edilmesi gereken bir diğer konu da seçmeli derslerin puanlarının zorunlu derslerden daha yüksek olduğudur. Öğrenciler bu dersleri gönüllü olarak seçtiklerinden ve ayrıca bu sınıflarda öğrenci sayısı daha az olduğundan bu dersleri veren öğretim üyelerinin puanları daha yüksek çıkabilmektedir.

Karşılaştırma Kriteri

Öğrencilerin değerlendirmeleriyle elde edilen puanların karşılaştırılmasında değişik yöntemlerden yararlanılabilir. Bunlardan birincisi bir öğretim üyesinin

puanlarının aynı programda değişik dersleri veren diğer öğretim üyelerinin puanlarıyla karşılaştırılmasıdır. İkincisi, sadece aynı branşta ders veren öğretim üyelerinin puanlarıyla karşılaştırma yapmaktır. Üçüncüsü kriter temelli karşılaştırmadır. Bu uygulamada örneğin 4,5 değerinin veya 4 ve 5 puanlarının yüzde ağırlığının %70 olması gibi bir değer, *kriter* olarak kabul edilir. Bu kriteri aşan öğretim üyeleri gözde olan simalardır. Norm temelli yöntemde ise, öğretim üyelerinin başarısı meslektaşlarının ortalamasına göre değil, örneğin ilk %10'luk dilime girip girmediğine göre yapılır.²⁹ İlk %10'luk dilimde yer alan öğretim üyeleri gözde olanlar sınıfında değerlendirilir.

Puanların Bir Kaç Dönem İzlenmesi

Hazırlanan raporlar sürekliliği yansıtmalıdır. Öğrenci değerlendirmesine ilişkin yüzde dağılımlarından ve dizin aritmetik puan ortalamalarından bir öğretim üyesi hakkında sağlıklı bir bilgi edinilebilmesi için bu puanların en az üç dönem, sömestr veya yıl izlenmesi gerekir. Değerlendirme raporları sadece ilgili branştaki öğretim üyelerini kapsamalı ve aynı zamanda daha önceki yıllardaki verileri de içermelidir. Bu raporlarda tek tek maddelerin aritmetik ortalama değerleri çok fazla bir anlam ifade etmeyeceğinden yıllar ve dönemler arasındaki karşılaştırmalar toplam dizin puan ortalamalarına göre yapılırsa öğretim üyesindeki gelişmeleri net bir şekilde görmek mümkün olur. Belirli sayıda dönemden önce herhangi bir öğretim üyesiyle ilgili olarak *iyi* veya *kötü* değerlendirmeleriyle bir yargıya varılmamalıdır.

Puan Dilimleri veya Ortalamaları Arasındaki Fark

Rapor, rakamsal bulguları abartmamalıdır. Öğrencilerin değerlendirmelerine bağlı olarak ortaya çıkan küçük puan dilimleri veya ortalamaları arasındaki fark aslında çok önemli değildir. Örneğin, 30 kişilik bir sınıfta 3 kişinin farklı bir yanıtı işaretlemesi sonucunda öğretim üyeleri arasında %10'luk bir fark varmış gibi gözükür. Yüzde oranlarının anlamlı olarak değerlendirilebilmesi için persentil değerleri yerine bu farkın en az %25'lik dilimler halinde olması gerekir ki daha sağlıklı bir yorum yapılabilsin.³⁰ Aynı şey puan ortalamaları açısından da söz konusudur. Dizin puan ortalamaları arasında ,50 puandan daha az olan farklılıkların çok fazla bir anlamı yoktur. Örneğin 4,17 ile 4,52 puanları

arasındaki farkı pratik olarak gözlemlemek, ölçmek ve değerlendirmek mümkün değildir. Sadece belirli ağırlıklarda davranış ve tutum farklılıkları ortaya çıkabilir. Bu açıdan öğretim üyelerinin değerlendirme raporlarında bu hususun belirtilmesi gerekir.

Başarı Puanlarıyla Değerleme Sonuçları Arasındaki İlişki

Raporda, ders başarı puanlarıyla öğrencilerin değerlendirmeleri arasında ilişki kurulmamalıdır. Çünkü ders başarı puanlarıyla öğrencilerin değerlendirmeleri arasındaki ilişkinin zayıf olduğu saptanmıştır. Yapılan bir araştırmada iki puan dizisi arasındaki korelasyonun %18 ilâ %20 arasında değiştiği bulunmuştur.³¹ Araştırmalarda ders başarısından çok, yapılan değerlendirmenin öğrenmeyle olumlu yönde ilişkili olduğu görülmüştür. Kendilerinin iyi öğrendiklerini ve dersin kendilerine bir şeyler kazandırdığını düşünen öğrenciler öğretim üyelerine daha yüksek puan verme eğilimi içinde olmuşlardır. Bu açıdan belirli bir dersten öğrencilerin önemli bir kısmının kalmış olması veya dersi geçemeyecekleri beklentisinin değerlendirmeleri büyük ölçüde olumsuz etkileyeceği düşüncesi tam olarak doğru değildir.

Sunum Biçimi

Bulgular yönetim raporları ve öğretim üyesi raporları şeklinde farklı iki biçim içinde sunulur. Yönetime sunulan raporların içeriği daha geneldir ve öğretim üyelerini yıllar itibarıyla izlemeye yöneliktir. Ardışık üç yıla ait veriler bir öğretim üyesi hakkında belirli bir kanaatin oluşması için yeterlidir. İkinci üç yılda ise, ilk sütun önceki üç yılın ortalaması için ayrılır. Yönetime sunulan raporlarda aynı bölümde benzer dersleri veren öğretim üyelerinin sonuçları bir grup olarak değerlendirilir. Öğretim üyelerine sunulan raporlar ise onların kendilerini geliştirmelerine yöneliktir ve bu nedenle bu raporlarda tek tek ifadeler yerine bu ifadelerin gruplandırılmış sonuçları olan faktör puanlarının ortalamaları ve yüzde ağırlıkları verilir.

Yönetime sunulan raporlarda öğretim üyesinin aritmetik ortalama cinsinden genel dizin puanı, mod veya medyan değerleri ile olumlu boyutu simgeleyen 4 ve 5 puanlarına gelen toplam yüzde dağılımları gösterilir. Mod ve medyan değerleri öğretim üyeleri arasında anlamlı bir farklılık yaratmıyorsa tabloda sadece

aritmetik ortalama ve yüzde deęerleri verilir. Standart sapma deęerlerinin tabloda gsterilmesine gerek yoktur. Ancak standart sapma 1,2'den yksek ıkmyısa bu durum yıldız iřareti kullanılarak tablo dipnotunda gsterilir. Yapılacak bir arařtırma sonucunda bulguların sunumu Tablo 1'deki gibi dzenlenebilir.

Tablo 1. ğretim yelerinin Blm Bazında Etkililik Deęerlendirme Sonuları

ğretim yesinin Adı Soyadı	Puan ortalaması			4 ve 5 puan yzdesi		
	2002	2003	2004	2002	2003	2004
X	4,55	4,78	4,65	,69	,71	,72
Y	4,16*	4,42	4,52	,70	,73	,72
Z	4,69	4,56	4,72	,69	,72	,74

* Standart sapması 1,2'nin zerinde olduęundan bu ortalama ğrencilerin kanaatlerini tam olarak yansıtmıyor olabilir.

ğretim yelerinin kendilerine sunulacak raporlarda yıllar itibariyle gelişme rakamlarının gsterilip gsterilmeyeceęi deęerlendirme komisyonunun alacağı karar erevesinde belirlenir. Deęerlendirme komisyonu raporun sunum biimini belirlerken ğretim yelerinin tepkilerini ve yaklaşımlarını gz nnde bulundurmalıdır. nk bazı ğretim yelerinin bu tr deęerlendirmelere ok fazla nem vermedięi bilinmektedir. Ancak karřılařtırma rakamları verilmese bile, bu raporlarda leęin dizin deęerinin aritmetik ortalamalarıyla birlikte her bir ifadenin 4 ve 5 puanlarına gelen frekans deęerleri gsterilmelidir. Mod ve medyan deęerlerinin gsterilmesi ise, ğretim yelerinin puanları arasında nemli farklılıklar bulunduęu durumda dřnlebilir.

Tablolar basit, yalın ve kolay anlaşılr bir tarzda hazırlanmalı ve gereęinden fazla veriyle yklenmemelidir. En iyi tablo, grsel aıdan basit ve kolay algılanabilir olandır. Tabloda aritmetik ortalama ve yüzde deęerlerinde virglden sonra iki hane gsterilir. Deęiřkenler stnuna sadece faktr isimleri alınır ve son satıra ise dizin deęerinin aritmetik ortalamaları ve frekans daęılımları yazılır. rnek olarak verilen Tablo 2 ve Tablo 3'de ğretim yelerine blm bazında deęerlendirme yapmaya imkan saęlayacak bir dzenleme biimi nerilmiřtir.

Tablo 2. Öğretim Üyelerinin Bölüm Bazında Etkililik Değerlendirme Sonuçları (Aritmetik Ortalama Değerleri)

	Öğretim Üyeleri		
	X Öğretim üyesi	Y Öğretim üyesi	Z Öğretim üyesi
Öğretim üyesinin bilgisi	4,7	4,8	4,5
Öğrencilerle ilişkileri	4,7	4,2	2,3
Öğrencilere yardımcı olması	4,8	4,6	2,8
Not verme uygulamaları	4,8	4,8	4,1
Genel dizin değeri	4,7	4,5	3,5

Tablo 3. Öğretim Üyelerinin Bölüm Bazında Etkililik Değerlendirme Sonuçları (4 ve 5 Puanlarının Yüzde Değerleri)

	Öğretim Üyeleri		
	X Öğretim üyesi	Y Öğretim üyesi	Z Öğretim üyesi
Öğretim üyesinin bilgisi	,76	,79	,70
Öğrencilerle ilişkileri	,78	,74	,52
Öğrencilere yardımcı olması	,75	,74	,50
Not verme uygulamaları	,77	,78	,68

SONUÇ ve ÖNERİLER

Öğrencilerin öğretim üyelerini değerlendirmeleri sonucunda elde edilen verilerin doğru bir şekilde analiz edilmesi ve değerlendirilmesi için okul yöneticileri tarafından temel kriterlerin ortaya çıkarılması gerekir. Bu kriterler; ölçeğe ait tek tek maddelerin frekanslarının alınması, merkezî dağılım için ifadelerin medyan ve mod değerlerinin hesaplanması, dizin değeri olarak aritmetik ortalama veya toplam puanların kullanılması, ölçeğin geçerlilik ve güvenilirlik analizinin yapılması, karşılaştırma kriterlerinin öğretim üyelerinin görüşleri alınarak tespit edilmesi ve sonuçların program bazında değil, yıllar arasında karşılaştırma

yapmaya imkan verecek şekilde sunulmasıdır. Öte yandan küçük sınıflarda, seçmeli derslerde ve sosyal derslerde puanların yüksek çıkma olasılığının bulunması nedeniyle kriter ve norm değerler bu olgu göz önünde bulundurulmuş olarak saptanmalıdır. Başarı kriteri, her bir bölüm için ayrı olmak üzere birden fazla değer olarak belirlenebilir. Öğrencilerin eğitimi değerlendirme prosedüründe, öğretim üyelerinin pasif gözlemciler olmaktan çıkarılmasıyla sistem çok daha başarılı sonuçlar ortaya koyacaktır.

ALINTI YAPILAN KAYNAKLAR

¹ Babbie, "Indexes, Scales, and Typologies [İndeksler, Ölçekler ve Sınıflandırmalar]," <http://www2.chass.ncsu.edu/Judge/PS371/Lecture6a_part1.htm> (07.10.2002), Index vs. Scales.

² W.M.K. Trochim, "General Issues in Scaling [Ölçek Geliştirmede Genel Sorunlar]," <<http://trochim.human.cornell.edu/kb/scalgen.htm>> (07.10.2002), Scaling.

³ Oulu University, "Methods and Procedures [Yönetim ve İşlemler]," <<http://herkules oulu.fi/ISBN9514259378/html/i1070866.html>> (07.10.2002).

⁴ A. Diamantopoulos ve H Winklhofer "Index Construction with Formative Indicators: An Alternative to Scale Development [Oluşturucu göstergelerle İndeks Oluşturma: Ölçek Geliştirmeye Alternatif]," <<http://www.lboro.ac.uk/departments/bs/research/1999-41.doc>> (14.10.2002).

⁵ E.P. Sheehan, T. DuPerey, "Student Evaluations of University Training [Öğrencilerin Eğitimi Değerlendirmeleri]," *J. Instructional Psychology*, 26, (3).

⁶ Dawn M. VanLeeuwen, T.J. Tormedy ve B.S. Seever, "Assessing The Reliability of Student Evaluations of Teachings With Generalizability Theory [Öğrencilerin Eğitimi Değerlendirmelerinin Genelleme Kuramıyla Birlikte Ele Alınması]," t.y., <<http://pubs.aged.tamu.edu/jae/pdf/Vol40/40-04-01.pdf>> (07.07.2002), s. 2.

⁷ J.E. Hoyt, "Course Evaluation Results: Their Reliability and Validity [Kurs Değerlendirme Sonuçları: Geçerlilik ve Güvenilirlik]," t.y., <<http://www.admin.utah.edu/ia/reports/courseevaluations.pdf>> (07.07.2002), Introduction.

⁸ H. W. Marsh, "Students' Evaluations of University Teaching [Öğrencilerin Üniversite Eğitimlerini Değerlemeleri],"

<http://216.239.51.100/search?q=cache:hX6RSfavs38C:www.uws.edu.au/uws/edc/seeq/SETs_HerbMarsh_presentation_2001.pdf+%22student+evaluation+of+teaching%22+Reliability+median&ht=tr&ie=UTF-8> (07.07.2002), Generalizability

⁹ DSS Research, "Validity and Reliability [Geçerlilik ve Güvenilirlik]," t.y., <<http://www.dssresearch.com/library/general/validity.asp>> (30.07.2002), Validity.

¹⁰ William E. Cashin, "Student Ratings of Teaching: The Research Revisited [Öğrencilerin Eğitimi Değerlemeleri: Araştırmaların Gözden Geçirilmesi]," Eyl 1995, <<http://lohman.tamu.edu/summarypapersrewards/cashin.htm>> (07.07.2002), Approach One.

¹¹ E. Burdet, "ACE Pilot Test: Rating Reliability and Validity Analysis [ACE Pilot Testi: Geçerlilik ve Güvenilirlik Analizi]," <<http://www.mtroyal.ab.ca/mrfa/pubs/acepilot.htm>> (07.07.2002), Factorial Validity Evidence.

¹² William M.K. Trochim, "Discriminant Validity [Ayrıt Etme ve Bağdaşırlık Analizi]," 2000, <<http://trochim.human.comell.edu/kb/convdisc.htm>> (29.07.2002), par. 1.

¹³ Lee A. Becker, "Reliability and Validity [Güvenilirlik ve Geçirlilik]," 1999, <http://web.uccs.edu/lbecker/Psy590/relval_II.htm> (29.07.2002), Convergent Validity.

¹⁴ University of North Carolina, "Guide for Department Chair's Report [Bölüm Başkanı Raporları İçin Rehber]," <<http://ctl.unc.edu/3setcr.html>> (05.07.2002), Enrollment/Respondents.

¹⁵ University of Arizona, "Guide to Student Ratings (TCEs) at the University of Arizona [Arizona Üniversitesinde Öğrenci Değerlemeleri İçin Rehber]," <<http://aer.arizona.edu/Assessment/guide/guidc.html>> (05.07.2002), Sample Quality.

¹⁶ "Student Evaluation of Teaching: A Methodological Critique of Conventional Practices [Öğrencilerin Eğitimi Değerlemeleri: Mevcut Uygulamaların Yöntem Bilimi Açısından Eleştirisi]," 02 Jfm 2000, <<http://epaa.asu.edu/epaa/v8n50.html>> (07.07.2002), Instructor Based

¹⁷ Robert Sproule, Student Evaluation of Teaching: A Methodological Critique of Conventional Practices [Öğrencilerin Eğitimi Değerlemeleri]," t.y., <<http://epaa.asu.edu/epaa/v8n50.html>> (05.07.2002),

¹⁸ Smith College, "Recommendations for Student Evaluation of Courses at Smith College [Smith Kolejinde Yapılan Öğrenci Değerlemeleri İçin Öneriler]," <<http://www.smith.edu/deanoffaculty/At.html>> (26.07.2002), Overall Ratings.

¹⁹ "Interpreting Ratings [Dereceleri Yorumlama]," <http://aer.arizona.edu/Assessment/Evaluations/interpreting_ratings.htm>

²⁰ Joe Cuseo, "Evaluating New -Student Seminars & Other First -Year Courses via Course -Evaluation Surveys, [Yeni Öğrenci Seminerlerinin Değerlendirilmesi]," <http://216.239.51.100/search?q=cache:TsWo_NRL0uoC:www.brevard.edu/fyc/1ya/Joe%2520Cuseo.PDF+%22Course+evaluation%22+student+teacher+statistics+mean&hl=tr&ie=UTF-8> (05.07.2002), s. 5.

²¹ Report Of The Academic Senate Student Evaluation Of Teaching [Öğrencilerin Eğitimi Değerlendirmeleri Üzerine Akademik Senato Raporu], t.y., <<http://216.239.51.100/search?q=cache:YrG2mdwP-SkC:www.otl.wayne.edu/pdf/asset.pdf+%22Course+evaluation%22+student+statistics+teacher+results+mean+set&hl=tr&ie=UTF-8>> (05.07.2002), s. 15.

²² University of North Carolina, "Guide for Department Chair's Report [Bölüm Başkanı Raporları İçin Rehber]," <<http://ctl.unc.edu/3setcr.html>> (05.07.2002), General Guidelines.

²³ Cuseo, "Evaluating New." Students Raters.

²⁴ Charles Emery, Tracy Kramer, Robert Tian, "Return to Academic Standards: Challenge the Student Evaluation of Teaching Effectiveness [Akademik Standartlara Dönüş]," <http://www.bus.lsu.edu/academics/accounting/faculty/lcmnbley/stu_rat_of_%20instr.htm> (05.07.2002), Summative set.

²⁵ University of Minnesota, "Student Evaluation of Teaching Usage [Eğitimi Öğrencilerin Değerlendirmeleri]," 2002, <<http://www.ucs.umn.edu/oms/setstats.html>> (05.07.2002), Questionnaire Count.

²⁶ John FitzGerald, "Stems and Scales [Gövde ve Ölçekler]," t.y., <<http://www.coolth.com/likert.htm>> (26.07.2002).

²⁷ Emery, Kramer, Tian, "Return to Academic," When making judgments.

²⁸ University of Arizona, "Guide to Student."

²⁹ Aynı.

³⁰ University of North Carolina, "Guide for Department Chair's Report [Bölüm Başkanı Raporları İçin Rehber]," <<http://ctl.unc.edu/3setcr.html>> (05.07.2002), General Guidelines.

³¹ University of North Carolina, "Guide for."