

Doğan, C. D. (2016). Biçimlendirici değerlendiriminin üniversite öğrencilerinin değerlendirme tercihleri üzerindeki etkisi: Bir ölçekleme çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(2), 413-431.

Geliş Tarihi: 25/04/2016

Kabul Tarihi: 17/06/2016

DOI:

BİÇİMLENDİRİCİ DEĞERLENDİRME ÜNİVERSİTE ÖĞRENCİLERİNİN DEĞERLENDİRME TERCİHLERİ ÜZERİNDEKİ ETKİSİ: BİR ÖLÇEKLEME ÇALIŞMASI

Celal Deha DOĞAN*

ÖZET

Araştırmanın amacı üniversite öğrencilerinin değerlendirme tercihlerinin biçimlendirici değerlendiriminin kullanıldığı ve kullanılmadığı durumlarda iki karşılaştırma yöntemi ile ölçeklenmesi ve karşılaştırılmasıdır. Araştırma yarı deneysel desende tasarlanmıştır. Çalışma grubunu bir devlet üniversitesinin eğitim fakültesinde öğrenim gören ve bilimsel araştırma yöntemleri dersini almakta olan 208 öğrenci oluşturmaktadır. İlgili verilerin toplanmasında araştırmacı tarafından geliştirilen ve değerlendirme yöntemlerinin ikili olarak karşılaştırıldığı maddeleri içeren bir anket formu kullanılmıştır. Veri toplama aracı araştırmanın kontrol grubuna herhangi özel bir yönerge verilmeden uygulanırken deney grubuna biçimlendirici değerlendirimin kullanıldığı durumları dikkate alarak yanıtlamalarını belirten özel bir yönerge ile uygulanmıştır. Elde edilen veriler ikili karşılaştırma yöntemi ile ölçeklenmiştir. Araştırma sonucunda kontrol grubunda en yüksek ölçek değerine sahip olan ve daha fazla tercih edilen değerlendirme yöntemleri doğru yanlış testleri ve çoktan seçmeli testler iken deney grubunda araştırmaya dayalı performans görevleri ve birden fazla olası yanıtı olan sorulardan oluşan yazılı yoklamalardır. Bu bulgular ışığında değerlendirimin yapılaş amacının öğrencilerin değerlendirme tercihleri üzerinde bir etkiye sahip olduğu belirtilebilir.

Anahtar Kelimeler: Değerlendirme Tercihleri, Durum Belirleme Tercihleri, İkili karşılaştırmalarla ölçekleme

EFFECT OF FORMATIVE ASSESSMENT ON ASSESSMENT PREFERENCES OF THE UNIVERSITY STUDENTS: A SCALING STUDY

ABSTRACT

The aim of this study is scaling of university students' assessment preferences when formative assessment is used and is not used. Semi experimental design was used in this study. Participants of the study include 205 Faculty of Education students in a state university located in Ankara, Turkey. In order to collect data, a questionnaire was developed by the researcher. The questionnaire included a pair of assessment method on each line and the respondents were required to select one of two assessment methods. Assessment methods were arranged in pairs so that every one of them was paired with every other one. The questionnaire was administered to the control group without any additional instruction. On the other hand the questionnaire was administered to the experimental group with an additional instruction. In the instruction it was asked the respondents to answer the questionnaire considering formative assessment was applied. To scale the respondents' assessment preferences, method of paired comparison was used. Results showed us that in the control group true-false and multiple choice tests were preferred rather than other methods. On the other hand in the experimental group research based performance tasks and the essay type exams which consisted of questions which had more than one possible right answers were preferred rather than other methods. Results showed us that the aim of the assessment might affect students' assessment preferences.

Key Words: Assessment Preferences, Scaling, method of paired comparison

* Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitimde Ölçme ve Değerlendirme Bölümü, e-mail: dehadogan@gmail.com

1. GİRİŞ

Öğrenme-öğretme sürecinin niteliği; öğrenme ortamı, öğretmenin bilgi ve becerileri, izlenen öğretim programı ve öğretim araç gereçleri gibi pek çok değişkenden etkilenmektedir. Ancak bunların yanı sıra öğrenenin sürece yaptığı katkı ve sahip oldukları öğrenmeye yönelik bireysel farklılıklar da göz ardı edilmemesi gereken bir boyuttur. Öğretmenlerin bireysel farklılıklar hakkında bilgi sahibi olmaları ve onları öğrencinin öğrenme potansiyelini geliştirecek biçimde kullanabilmeleri öğrenme-öğretme sürecinin niteliğini önemli düzeyde etkilemektedir (Kuzgun ve Deryakulu, 2004).

Dikkate alınması gereken özelliklerden bir tanesi de öğrencilerin değerlendirme tercihleri (assessment preference)'dir. Öğrencilerin başarıları belirlenirken kullanılan değerlendirme yöntemlerine ve o yöntemlerin özelliklerine ilişkin görüş, tutum ve tercihleri, değerlendirme tercihleri (assessment preferences) olarak tanımlanmaktadır (Birenbaum, 1997). Başka bir ifade ile öğrenciler belirli değerlendirme yöntemlerine daha yatkın olabilir ve o değerlendirme yöntemi kullanıldığında performansını daha iyi düzeyde sergileyebilir. Bu da öğretim sürecinde dikkat edilmesi gereken bir durumdur.

Öğretim kurumlarında öğretmenler farklı değerlendirme yöntemleri kullanarak öğrencilerinin başarıları hakkında kararlar vermektedirler. Öğrencilerin çalışma alışkanlıkları, öğrenme yaklaşımları, öğrenme düzeyleri, bilgiyi kullanma ve uygulamaya dönüştürme becerileri gibi pek çok değişken kullanılan değerlendirme yöntemine göre farklılık göstermektedir. Başka bir ifade ile kullanılan değerlendirme yöntemleri ile öğrencilerin öğrenmelerine ilişkin pek çok özellik arasında sıkı bir bağlantı vardır (Gülbahar ve Büyüköztürk, 2008). Bu nedenle öğrencilerin değerlendirme tercihlerinin (assessment preference) belirlenmesi öğretim süreci hakkında verilen kararlarda önemli bir yer tutmaktadır.

Değerlendirme etkinlikleri, öğrencilerin öğrenmeleri ve başarıları üzerinde önemli bir etkiye sahiptir. Öğrencilerin değerlendirme yöntemlerine ilişkin tercihleri son yıllarda pek çok araştırmanın odağı olmuştur. Yapılan çalışmalar öğrencilerin öğrenme ve değerlendirme tercihlerindeki farklılıkların başarıları üzerinde etkiye sahip olduğunu ve öğrenme yaklaşımı, çalışma stratejisi, öğrenme stili gibi öğrenmeye ilişkin özelliklerle ilişkili olduğunu ortaya koymuştur (Birenbaum ve Feldman, 1998; Philips, 1999; Wilson ve Fowler, 2005; Birenbaum ve Rosenau, 2006). Bu nedenle öğretim sürecinde öğrenmeye ilişkin diğer bireysel farklılıklarla birlikte öğrencilerin değerlendirme tercihlerinin dikkate alınması gerekmektedir. Öğrenme-öğretme sürecinde, öğrencilerin değerlendirme tercihlerinin dikkate alınması değişen dünyanın gerektirdiği yeterliklere sahip bireylerin yetiştirilmesine katkı sağlayacaktır.

İlgili konuda yapılan ilk çalışmalar öğrencilerin hangi sınav türünü tercih ettiği ve bu tercihlerinin cinsiyete göre farklılaşıp farklılaşmadığı üzerine odaklanmıştır (Grandt, 1987; Zoller ve Ben-Chaim, 1990.) 1994 yılından itibaren yapılan çalışmaların çoğunluğunda Birenbaum (1994, 1997, 2007) tarafından geliştirilen "Değerlendirme Tercihleri Ölçeği" (Assessment Preference Inventory)" kullanılmıştır. Bu çalışmalarda öğrencilerin öğrenmeye ilişkin özellikleri ile değerlendirme tercihleri arasındaki ilişkiler incelenmiştir. Çalışmalar sonucunda elde edilen bulgular, öğrencilerin değerlendirme tercihleri ile öğrenmeye ilişkin özellikleri arasında güçlü ilişkilerin olduğunu ortaya koymuş ve öğretim sürecinde değerlendirme tercihlerinin dikkate alınması gerektiğinin önemini vurgulamıştır (Birenbaum 1997, 2003, 2007; Philips, 1999; Biggs, 2003;

Struyven, Dochy ve Janssens, 2005; Wilson ve Fowler, 2005; Birenbaum ve Rosenau 2006; Watering, Gijbels, Dochy ve Rijt, 2008; Doğan & Kutlu 2010; Doğan, Atmaca & Aslan 2010; Batram ve Bailey, 2010, Doğan, 2011).

Alanda yapılan çalışmalara göz atıldığında çalışmaların büyük çoğunluğunda değerlendirme tercihleri ölçeğinin kullanıldığı dikkat çekmektedir. Birenbaum (1997) geliştirdiği ölçme aracında öğrencilerin durum belirleme tercihlerini; “durum belirleme yöntemlerine ilişkin boyutlar”, “öğrencilere ilişkin boyutlar” ve “notlandırma ve raporlaştırma” olmak üzere üç bölümde ele almıştır.

Değerlendirme tercihleri ölçeğinde, öğrencilerin alternatif-geleneksel, karmaşık oluşturmacı-basit seçmeli sınav türlerine ilişkin tercihleri, ilgili alt boyutlarda ne düzeyde puana sahip olduklarına dayalı olarak belirlenmektedir. Ölçekten elde edilen sonuçlar neticesinde, örneğin öğrencilerin alternatif değerlendirme yöntemlerini mi yoksa geleneksel değerlendirme yöntemlerini mi daha çok tercih ettiğine ilişkin belirlemeler yapılabilmektedir.

Ancak ilgili ölçek, öğrencilerin geleneksel ve alternatif yöntemlerden hangilerini diğerlerine göre daha çok tercih ettiğine ilişkin bir bilgi sunmamaktadır. Bu bağlamda örneğin öğrencilerin iki geleneksel yönteminden hangisini daha çok tercih ettiğine ilişkin belirlemeler değerlendirme tercihleri ölçeği ile yapılamamaktadır. Bu boyuta ilişkin belirlemelerin yapılabilmesi için öğrenci yargılarına dayalı bir ölçekleme çalışmasının yapılması gerekmektedir.

Alan yazına bakıldığında öğretim sürecindeki boyutlara ilişkin ölçekleme çalışmalarının olduğu görülmektedir (Çüm & Doğan, 2016; Acar Güvendir & Özer Özkan 2011; Bal, 2011; Koğar & Gelbal, 2015). Altun ve Gelbal (2014) öğrenci başarısının belirlenmesinde kullanılan değerlendirme türlerine ilişkin yaptıkları çalışmalarında ölçekleme işleminin öğretmen yargılarına göre gerçekleştirmişlerdir. Tezbaşaran (2015) ise öğretmen adaylarının yargılarına dayalı olarak bir çalışma gerçekleştirmiş ancak çalışmada uyarıcı olarak sadece geleneksel değerlendirme yöntemlerine yer vermiştir. Bu noktada hem alternatif hem de geleneksel değerlendirme yöntemlerini içeren uyarıcı gruplarının öğretmen adaylarına sunulduğu bir ölçekleme çalışması bulunmamaktadır. Bunun yanı sıra eğitim fakültelerinde yer alan araştırma yöntemleri dersinin içeriği ve yapısı gereği farklı değerlendirme türlerinin kullanımına uygun olduğu düşünüldüğünden çalışma bilimsel araştırma yöntemleri dersi ile sınırlandırılmıştır.

Öğrencilerin başarıları belirlenirken tercih ettikleri değerlendirme yöntemleri üzerinde değerlendirmenin hangi amaçla yapıldığı da önemli rol oynayabilmektedir. Değerlendirmenin düzey belirleme amaçlı (summative assessment) veya biçimlendirme amaçlı (formative assessment) olması öğrencilerin tercihleri üzerinde önemli etkiye sahip olabilir. Biçimlendirici değerlendirme öğrenci öğrenmesine ilişkin kanıtların toplanması, öğrencilere geri bildirim verilmesi ve başarıyı artırmak amacıyla öğretim stratejilerinin ayarlanması sürecidir (McMillan, 2015). Düzey belirlemeye dayalı değerlendirmede ise temel amaç öğrenci başarısını bir not ile belirlenmek ve öğrenci hakkında geçti-kaldı, başarılı-başarısız gibi kararları vermektir. Ulusal alanyazın incelendiğinde, değerlendirme farklı amaçlar ile yapıldığında öğrencilerin değerlendirme tercihlerinin farklılaşp farklılaşmadığına ilişkin deneysel bir çalışma bulunmadığı görülmüştür. Uluslararası alanyazında ise ilgili konuda oldukça az çalışmanın olduğu belirlenmiştir (Gijbels ve Dochy (2006).

Bu bağlamda araştırmanın genel amacı eğitim fakültesi öğrencilerinin araştırma yöntemleri dersindeki başarıları belirlenirken kullanılmasını tercih ettikleri değerlendirme (durum belirleme) yöntemlerinin, biçimlendirici değerlendirme (formative assessment) yapılacağına ilişkin bir yönergenin verildiği ve değerlendirmenin amacına ilişkin herhangi özel bir yönergenin verilmediği durumlarda ikili karşılaştırma yöntemiyle ölçeklenmesidir. Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır.

- 1- Uygulama öncesinde, değerlendirme etkinliklerinin amacına ilişkin herhangi bir yönergenin verilmediği kontrol grubunda, öğrencilerin tercih ettikleri değerlendirme yöntemlerine yönelik ölçek değerleri nedir?
- 2- Uygulama öncesinde, değerlendirme etkinliklerinin biçimlendirme (formative assessment) amacı ile kullanılacağına yönelik yönerge verilen deney grubunda, öğrencilerin tercih ettikleri değerlendirme yöntemlerine ilişkin ölçek değerleri nedir?
- 3- Öğrencilerin tercih ettikleri değerlendirme yöntemlerine ilişkin elde edilen ölçek değerleri deney ve kontrol grubunda farklılaşmakta mıdır?

2. YÖNTEM

2.1 Araştırma Modeli

Çalışmada öğrencilerin başarıları belirlenirken kullanılacak 9 farklı yöntemden hangisinin daha çok tercih edildiğinin öğrencilerin görüşlerine dayalı olarak ikili karşılaştırma yöntemi ile ölçeklenmiştir. Bu süreçte öğrenciler ikili olarak kendilerine sunulan değerlendirme yöntemlerinden daha çok tercih ettiklerine 1 diğerine 0 vererek kodlama yapmışlardır.

Araştırmada öğrencilere değerlendirmenin amacına ilişkin bir yönergenin verilmediği ve değerlendirmenin gelişime katkı sağlama ve geri bildirim verme amacı ile gerçekleştirileceğine ilişkin yönergenin verildiği 2 farklı durumda, değerlendirme tercihlerinin ölçeklenmesi ve karşılaştırılması amaçlanmaktadır. Bu bağlamda araştırma deneysel desende tasarlanmıştır. Bu süreçte lisans düzeyinde araştırma yöntemleri dersini alan iki grup öğrenci deney ve kontrol grubu olarak belirlenmiştir. Deney ve kontrol gruplarının belirlenmesinde seçkisiz atama pratik nedenlerden dolayı gerçekleştirilememiş ancak; her iki grup, araştırma yöntemleri ara sınavından aldıkları puan ortalamalarına dayalı olarak eşleştirilmiştir. Her iki grubun ara sınav notları bağımız örneklem t testi ile karşılaştırılmış ve iki grup arasında evrende bir farklılık olmadığının belirtildiği yokluk hipotezi kabul edilmiştir. Bu nedenle çalışma yarı deneysel desen olarak tasarlanmıştır.

Araştırmanın deney grubunda öğrencilerden başarılarının not ile değerlendirilmediği ve sadece geri bildirim verme amacı taşıdığı durumları düşünerek formu doldurmaları istenmiştir. Kontrol grubunda ise değerlendirmenin amacına ilişkin bir yönerge verilmemiştir. Ancak eğitim öğretim süreci içerisinde öğrencilerin dersi geçme ve not kaygısı duyacakları varsayılmıştır. Bu bilgiler ölçeklemede kullanılacak formların yönergelerinde yer almakla beraber araştırmacı tarafından uygulama öncesinde öğrencilere iletilmiştir. Araştırmada sönest kontrol gruplu eşleştirilmiş desen kullanılmıştır. Bu bağlamda araştırmada kullanılan yarı deneysel desene ilişkin açıklama tablo 1 'de verilmiştir.

Tablo 1.*Deneysel Desen*

GRUP	İşlem	Sontest
Deney M	Biçimlendirici değerlendirilenin (formative assessment)O1 kullanılacağına ilişkin yönerge	
Kontrol M	Değerlendirilenin amacına ilişkin bir yönerge yok	O2

2.2 Çalışma Grubu

Araştırmanın çalışma grubunu bir devlet üniversitesinin eğitim fakültesinde öğrenim gören ve bilimsel araştırma yöntemleri dersini alıyor olan 205 öğrenci oluşturmaktadır. Öğrenciler çalışmaya gönüllülük esasına dayalı olarak katılmışlardır. Araştırmanın deney grubunda 97 kontrol grubunda ise 108 öğretmen adayı yer almaktadır. Tüm öğrenciler araştırmanın gerçekleştirildiği dönemde bilimsel araştırma yöntemleri dersini almaya devam etmektedirler ve dersin ara sınavına girmişlerdir.

2.3 Veri Toplama Araçları

Çalışmada araştırmacı tarafından geliştirilen veri toplama aracı kullanılmıştır. Bu süreçte öncelikli olarak öğrenci başarısının belirlenmesinde kullanılan mevcut yöntemler, uygulamanın yapıldığı üniversitenin Bologna uyum süreci kapsamında belirlenen değerlendirme yöntemleri ve Birenbaum (2007) tarafından geliştirilen Değerlendirme Tercihleri Ölçeği'nde yer alan maddeler incelenmiştir. Literatürde var olan ve ölçek maddelerinde yer alan değerlendirme yöntemleri arasından öğrencilerin öğrenim yaşantıları boyunca deneyim etmiş oldukları ve yeterli bilgiye sahip oldukları düşünülen 9 yöntem seçilmiştir. Yöntemlerin seçilme aşamasında eğitim fakültesinde görev yapan ve araştırma yöntemleri dersini veren 3 öğretim üyesi, 1 ilkökul, 1 ortaokul ve 1 lise öğretmeni olmak üzere 6 uzmanın görüşüne başvurulmuştur. İlgili uzmanlar görev yaptıkları öğretim kademesinde kullanılan değerlendirme yöntemlerini dikkate alarak öğrencilerin deneyim edebilecekleri yöntemlerin belirlenmesinde katkı sağlamışlardır. Bu aşamada elektronik portfolyolar, kavram haritaları gibi öğrencilerin deneyim etmemiş olabilecekleri yöntemler çalışma kapsamı dışında tutulmuştur.

Sonuç olarak çoktan seçmeli testler, kısa yanıt ve tamamlama testleri, tek bir doğru yanıt olan sorulardan oluşan yazılı yoklamalar, birden fazla olası yanıt olan sorulardan oluşan yazılı yoklamalar, sınırlandırılmış performans görevleri, araştırmaya dayalı performans görevleri, ev ödevleri ve bireysel -grup sunumları, olmak üzere 9 değerlendirme yöntemini içerecek şekilde anket formu oluşturulmuştur. Ankette öğrencilerden kendilerine ikili olarak sunulan yöntemlerden daha çok tercih ettiklerini 1 ile kodlamaları istenmiştir. Anketler uygulamadan önce öğrencilere ilgili değerlendirme yöntemleri hakkında bilgiler verilmiş ve uygulama sürecinin nasıl olacağına ilişkin örnekler paylaşılmıştır. Bu sayede öğrencilerin eğer varsa ilgili değerlendirme yöntemlerine ilişkin eksik bilgilerinin giderilmesi veya hatalı bilgilerinin düzeltilmesi amaçlanmıştır.

Araştırmada deney ve kontrol grubunda uygulanmak üzere 2 farklı form geliştirilmiştir. Deney grubu için oluşturulan formda değerlendirilenin amacına yönelik bir yönerge bulunurken kontrol grubu için geliştirilen formda değerlendirilenin amacına ilişkin bir yönerge bulunmamaktadır.

2.4 Verilerin Analizi

İlgili verilen çözümlenmesinde psikolojik özelliklerin ölçeklenmesi sürecinde sıklıkla kullanılan ve Thrustone (1927) tarafından geliştirilen ikili karşılaştırma yöntemiyle ölçekleme tekniğinden faydalanılmıştır. Öğrencilerin ikili olarak yaptıkları karşılaştırmalara yönelik frekanslar belirlenmiş ve frekans matrisi oluşturulmuştur. Akabinde frekans matrisinde yer alan her bir değer toplam kişi sayısına bölünerek oranlar matrisi elde edilmiştir. Daha sonra ise oranlar matrisindeki hücre değerlerine karşılık gelen (z) standart değerleri belirlenerek normal sapmalar matrisi oluşturulmuştur. Z matrisindeki sütun toplamları ve ortalamaları hesaplanarak ölçek değerleri (S_j) elde edilmiştir. Ölçek değerlerinin pozitif olması için hesaplanan en küçük değer sıfıra eşitlenmiş ve bu değer diğer ölçek değerlerine eklenerek ötelenmiştir. Bu süreçte öncelikli olarak V. Hal yöntemi ile ölçekleme gerçekleştirilmiştir.

Elde edilen ölçek değerlerinin iç tutarlılığının belirlenmesi sürecinde gözlenen oranlar teorik oran arasındaki farkın ortalama değeri hesaplanmıştır. Bu değer başka bir ifade ile gözlenen değerlerle ampirik değerler arasındaki uyumun bir ölçüsünü verir (Turgut & Baykul, 1992). Bu nedenle ki kare testi ile uygunluk derecesinin anlamlı olup olmadığı sınıanmıştır. Ki kare değerinin anlamlı çıktığı durumlarda ise III. Hal denklemine dayalı olarak ölçekleme çalışması tekrar edilmiştir.

3. BULGULAR

Bu bölümde araştırmanın bulgularına dair sonuçlar sunulmuştur.

3.1 Kontrol Grubuna ilişkin Bulgular

Araştırmanın kontrol grubunda öğrencilere herhangi bir yönerge verilmediği durumda (mevcut eğitim öğretim sürecindeki algılarına dayalı olarak) öğrencilerin değerlendirme tercihleri iki karşılaştırma yöntemine göre ölçeklenmiştir. Bu süreçte öncelikli olarak öğrencilerin değerlendirme tercihlerine ilişkin frekans matrisi oluşturulmuş ve tablo 1’de sunulmuştur. Tabloda yer alan “ÇS” çoktan seçmeli testler, “TDY” Tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalar, “ÇDY” olası birden fazla doğru yanıtı olan sorulardan oluşan yazılı yoklamalar, “PS” sınırlandırılmış performans görevleri, “PA” araştırmaya dayalı performans görevleri, “KST” kısa yanıtı ve tamamlama testleri, “DY” doğru yanıt testleri, “EV” ev ödevleri, “SUN” ise bireysel veya grup sunumları ifade etmek için kullanılan kısaltmalardır.

Tablo 2.

Kontrol Grubu için F Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	24	53	40	57	43	58	49	47
TDY	84	*	73	61	65	84	87	62	58
ÇDY	55	35	*	40	54	62	67	48	47
PS	68	47	68	*	72	71	52	59	52
PA	51	43	54	36	*	47	66	57	48
KST	65	24	46	37	61	*	72	55	45
DY	50	21	41	56	42	36	*	38	36
EV	59	46	60	49	51	53	70	*	53
SUN	61	50	61	56	60	63	72	55	*

Tablo 2’de her hücredeki eleman f_{ij} ile gösterilmek üzere $S_j > S_i$ değerini veren frekansları ifade etmektedir. İlgili matrisin i . satır j sütun elemanı f_{ij} ; satır yer alan değerlendirme yönteminin, sütunda yer alan değerlendirme yönteminden ne kadar daha fazla tercih edildiğini göstermektedir. Matriste yer alan simetrik elemanların toplamı toplam gözlemci sayısı olan 108’i vermektedir. Daha sonra matrisin her bir hücresindeki değer toplam kişi sayısına (N) bölünerek tablo 3’te yer alan oranlar matrisi elde edilmiştir.

Tablo 3.

Kontrol Grubu için Oranlar Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	0,22	0,49	0,37	0,53	0,40	0,54	0,45	0,44
TDY	0,78	*	0,68	0,56	0,60	0,78	0,81	0,57	0,54
ÇDY	0,51	0,32	*	0,37	0,50	0,57	0,62	0,44	0,44
PS	0,63	0,44	0,63	*	0,67	0,66	0,48	0,55	0,48
PA	0,47	0,40	0,50	0,33	*	0,44	0,61	0,53	0,44
KST	0,60	0,22	0,43	0,34	0,56	*	0,67	0,51	0,42
DY	0,46	0,19	0,38	0,52	0,39	0,33	*	0,35	0,33
EV	0,55	0,43	0,56	0,45	0,47	0,49	0,65	*	0,49
SUN	0,56	0,46	0,56	0,52	0,56	0,58	0,67	0,51	*

Esas köşegene göre simetrik olan elemanlarının toplamının bire eşit olduğu oranlar matrisinden faydalanılarak normal birim sapmalar matrisi (Z matrisi) elde edilmiştir. Bu süreçte oranlar matrisinde yer alan her bir hücre değerine karşılık gelen standart değerler belirlenerek normal birim sapmalar matrisi elde edilmiş ve tablo 4’te sunulmuştur.

Tablo 4.

Kontrol Grubu için Normal Birim Sapmalar Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	-0,772	-0,025	-0,332	0,075	-0,253	0,100	-0,126	-0,151
TDY	0,772	*	0,468	0,151	0,253	0,772	0,878	0,176	0,100
ÇDY	0,025	-0,468	*	-0,332	0,000	0,176	0,305	-0,151	-0,151
PS	0,332	-0,151	0,332	*	0,440	0,412	-0,050	0,126	-0,050
PA	-0,075	-0,253	0,000	-0,440	*	-0,151	0,279	0,075	-0,151
KST	0,253	-0,772	-0,176	-0,412	0,151	*	0,440	0,025	-0,202
DY	-0,100	-0,878	-0,305	0,050	-0,279	-0,440	*	-0,385	-0,440
EV	0,126	-0,176	0,151	-0,126	-0,075	-0,025	0,385	*	-0,025
SUN	0,151	-0,100	0,151	0,050	0,151	0,202	0,440	0,025	*
TOPLAM	1,484	-3,570	0,596	-1,391	0,716	0,693	2,777	-0,235	-1,070
S _j	0,165	-0,397	0,066	-0,155	0,080	0,077	0,309	-0,026	-0,119
S _c	0,562	0,000	0,463	0,242	0,477	0,474	0,706	0,371	0,278

Bu aşamada öncelikli olarak normal birim sapmalar matrisinde yer alan her bir sütunun toplamı alınmış ve ilgili değerler toplam uyarıcı sayısına (9) bölünerek S_j değerleri elde edilmiştir. Hesaplanan en küçük S_j değeri Tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalardır (TDY). İlgili S_j değerinin mutlak değeri alınarak her bir S_j değerine eklenmiş ve eksenin başlangıç noktası kaydırılarak S_c değerleri elde edilmiştir.

Ancak Z matrisinden faydalanılarak hesaplanan uyum ölçüsü ki kare değeri 28 serbestlik derecesinde 0,01 düzeyinde anlamlı bulunmuştur ($X^2_{(28)} = 72,77$, $p < 0,01$). Uyum ölçüsünün manidar çıkması durumunda Guilford (1954) III. hal denkleminin kullanılmasını tavsiye etmektedir (akt. Turgut ve Baykul, 1992). Bu nedenle ölçekleme işlemine III. hal eşitliği kullanılarak devam edilmiştir.

Bu süreçte z matrisinin her bir hücrenin; toplamı ($\sum Z_j$), karelerinin toplamı ($\sum Z_j^2$), toplamlarının karesi ($(\sum Z_j)^2$), Kareler toplamının uyarıcı sayısı ile çarpımı ($K\sum Z_j^2$) hesaplanmıştır. Daha sonra Z matrisinin sütun elemanlarının standart sapması (V_j), standart sapmaların uyarıcı sayısı ile çarpımı (KV_j), bu değer 1 'e oranı ($1/KV_j$) ve tüm sütunların toplamı ($\sum 1/KV_j$) elde edilir. Bu değerlerden faydalanılarak $K.C$ sabiti hesaplanır. Akabinde aşağıdaki eşitlik yardımı ile ayırt etme yargılarına ilişkin standart sapmalar (σ_j) ve bu standart sapmaların karesi alınarak ayırt etme yargılarına ilişkin varyans değerleri (σ_j^2) elde edilmiştir.

$$\sigma_j = \frac{K.C}{K.V_j} - 1$$

Tablo 4'te yer alan birim sapmalar matrisine dayalı olarak hesaplanan bahsi geçen değerler tablo 5'te sunulmuştur.

Tablo 5.

Z Matrisine Dayalı Olarak Gözlemci Varyanslarının Kestirilmesinde Kullanılan Değerler

$\sum Z_j$	1,484	-3,570	0,596	-1,391	0,716	0,693	2,777	-0,235	-1,070
$\sum Z_j^2$	0,802	2,310	0,499	0,627	0,392	1,119	1,490	0,241	0,316
$(\sum Z_j)^2$	2,202	12,745	0,355	1,935	0,513	0,480	7,712	0,055	1,145
$K\sum Z_j^2$	7,221	20,787	4,495	5,647	3,531	10,067	13,407	2,166	2,843
V_j	0,249	0,315	0,226	0,214	0,193	0,344	0,265	0,161	0,145
$K.V_j$	2,240	2,836	2,035	1,927	1,737	3,096	2,386	1,453	1,303
$1/KV_j$	0,446	0,353	0,491	0,519	0,576	0,323	0,419	0,688	0,767
$\sum 1/KV_j$	4,582								
$K.C$	2,928								
C	0,436								
σ_j	0,753	0,385	0,930	1,039	1,261	0,269	0,646	1,704	2,014
σ_j^2	0,567	0,148	0,866	1,079	1,590	0,072	0,417	2,903	4,055

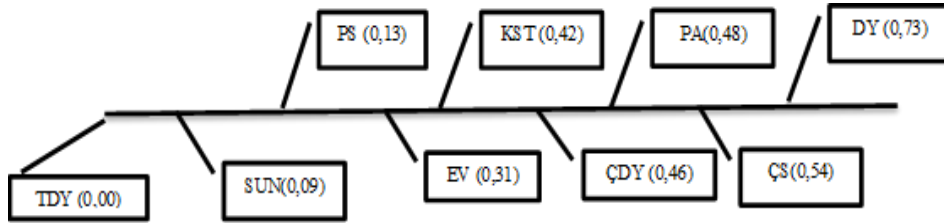
Bir sonraki aşamada ayırt etme yargılarının varyansları ikiye ikiye toplanarak varyans toplamları matrisi oluşturulmuştur. Esas köşegenin üstünde kalan elemanları Z

matrisinin ilgili elemanları ile çarpılarak da S matrisi oluşturulmuştur. S matrisi tablo 6'da sunulmuştur.

Tablo 6. S matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	-0,653	-0,030	-0,426	0,110	-0,202	0,099	-0,235	-0,325
TDY	0,653	*	0,471	0,167	0,334	0,362	0,660	0,307	0,205
ÇDY	0,030	-0,471	*	-0,463	0,000	0,170	0,345	-0,293	-0,335
PS	0,426	-0,167	0,463	*	0,719	0,442	-0,064	0,251	-0,113
PA	-0,110	-0,334	0,000	-0,719	*	-0,195	0,395	0,159	-0,359
KST	0,202	-0,362	-0,170	-0,442	0,195	*	0,308	0,043	-0,410
DY	-0,099	-0,660	-0,345	0,064	-0,395	-0,308	*	-0,701	-0,931
EV	0,235	-0,307	0,293	-0,251	-0,159	-0,043	0,701	*	-0,066
SUN	0,325	-0,205	0,335	0,113	0,359	0,410	0,931	0,066	*
TOPZ	1,662	-3,159	1,017	-1,956	1,162	0,637	3,376	-0,402	-2,334
S _j	0,185	-0,351	0,113	-0,217	0,129	0,071	0,375	-0,045	-0,259
S _c	0,536	0,000	0,464	0,134	0,480	0,422	0,726	0,306	0,092

S matrisinin sütunlar boyunca ortalamaları alınmış ve S_j değerleri elde edilmiştir. Bu değerlerin en küçüğü (-0,351) aynı V. Hal denkleminde olduğu gibi mutlak değeri alınarak her bir S_j değerine eklenmiş ve eksenin başlangıç noktası kaydırılarak S_c değerleri elde edilmiştir. III. Hal denkleminde ilişkin ölçek değerleri şekil 1'de sayı doğrusu üzerinde gösterilmiştir.



Şekil 1: Kontrol grubu için hesaplanan ölçek değerleri

Kontrol grubu için elde edilen bulgular incelendiğinde, değerlendirimin amacına ilişkin herhangi bir yönerge verilmediği durumda en çok tercih edilen yöntemin doğru yanlış testleri (0,726) olduğu görülmektedir. İkinci en çok tercih edilen yöntem ise çoktan seçmeli testler olmuştur (0,536). Daha sonra sırası ile araştırmaya dayalı performans görevi (0,480), birden fazla olası doğru yanıtı olan sorulardan oluşan yazılı yoklamalar (0,464), kısa yanıt ve tamamlama testleri (0,422), ev ödevleri (0,306), sınırlandırılmış performans görevleri (0,134) ve bireysel ve grup sunuları (0,092) gelmektedir. En az tercih edilen değerlendirme yöntemi ise (0,000) tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalardır

3.2 Deney Grubuna İlişkin Bulgular

Araştırmanın deney grubunda öğrencilere başarılarının not ile değerlendirilmediği ve sadece geri bildirim verme amacının olduğu durumları düşünerek formu doldurmaları yönünde bir yönerge verilmiştir. Öğrencilerin değerlendirme tercihleri iki karşılaştırma yöntemine göre ölçeklenmiştir. Bu süreçte elde edilen frekans matrisi, oranlar matrisi ve normal birim sapmalar matrisi kontrol grubu için hesaplanan adımlar tekrar edilerek oluşturulmuştur. İlgili değerler tablo 7, 8 ve 9'da sırası ile sunulmuştur. Çalışmanın deney grubunda 97 katılımcı bulunmaktadır. Çizelgelerde kontrol grubu için belirtilen kısaltmalar aynı şekilde kullanılmıştır.

Tablo 7.

Deney Grubu için F Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	20	56	47	64	45	49	45	48
TDY	76	*	78	64	77	70	69	68	52
ÇDY	40	18	*	44	61	43	48	42	42
PS	49	32	52	*	58	47	46	51	49
PA	32	19	41	39	*	41	44	29	36
KST	51	26	53	49	55	*	46	48	49
DY	47	27	48	50	52	51	*	45	53
EV	51	28	54	45	67	48	51	*	52
SUN	48	38	54	47	50	47	43	46	*

Tablo 8.

Deney Grubu için Oranlar Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	0,208	0,583	0,490	0,667	0,469	0,510	0,469	0,500
TDY	0,792	*	0,813	0,667	0,802	0,729	0,719	0,708	0,542
ÇDY	0,417	0,188	*	0,458	0,635	0,448	0,500	0,438	0,438
PS	0,510	0,333	0,542	*	0,604	0,490	0,479	0,531	0,510
PA	0,333	0,198	0,427	0,406	*	0,427	0,458	0,302	0,375
KST	0,531	0,271	0,552	0,510	0,573	*	0,479	0,500	0,510
DY	0,490	0,281	0,500	0,521	0,542	0,531	*	0,469	0,552
EV	0,531	0,292	0,563	0,469	0,698	0,500	0,531	*	0,542
SUN	0,500	0,396	0,563	0,490	0,521	0,490	0,448	0,479	*

Tablo 9.

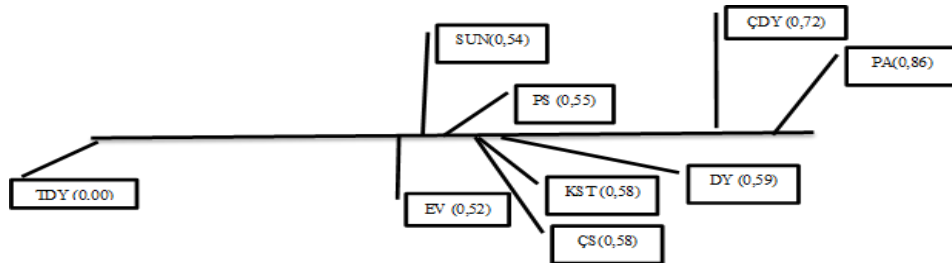
Deney Grubu için Normal Birim Sapmalar Matrisi

	ÇS	TDY	ÇDY	PS	PA	KST	DY	EV	SUN
ÇS	*	-0,813	0,210	-0,025	0,432	-0,078	0,025	-0,078	0,000
TDY	0,813	*	0,889	0,432	0,849	0,610	0,580	0,548	0,105
ÇDY	-0,21	-0,885	*	-0,105	0,345	-0,131	0,000	-0,156	-0,156
PS	0,025	-0,432	0,105	*	0,264	-0,025	-0,053	0,078	0,025
PA	-0,432	-0,849	-0,184	-0,238	*	-0,184	-0,105	-0,519	-0,319
KST	0,078	-0,610	0,131	0,025	0,184	*	-0,053	0,000	0,025
DY	-0,025	-0,580	0,000	0,053	0,105	0,078	*	-0,078	0,131
EV	0,078	-0,548	0,159	-0,078	0,519	0,000	0,078	*	0,105
SUN	0,000	-0,264	0,159	-0,025	0,053	-0,025	-0,131	-0,053	*
TOPLAM	0,274	-4,981	1,469	0,039	2,751	0,245	0,341	-0,258	-0,084
Sj	0,030	-0,553	0,163	0,004	0,306	0,027	0,038	-0,029	-0,009
Sc	0,583	0,000	0,716	0,557	0,859	0,580	0,591	0,524	0,544

Bu aşamada normal birim sapmalar matrisinde yer alan her bir sütunun toplamı alınmış ve ilgili değerler toplam uyarıcı sayısına (9) bölünerek Sj değerleri elde edilmiştir. Hesaplanan en küçük Sj değeri kontrol grubunda olduğu gibi tek doğru yanıt olan sorulardan oluşan yazılı yoklamalardır (TDY). İlgili Sj değerinin mutlak değeri alınarak her bir Sj değerine eklenmiş ve eksenin başlangıç noktası kaydırılarak Sc değerleri elde edilmiştir.

Z matrisinden faydalanılarak hesaplanan uyum ölçüsü ki kare değeri 28 serbestlik derecesinde 0,01 düzeyinde anlamlı bulunmamış ($X^2_{(28)} = 47,40$, $p > 0,01$) ve yokluk hipotezi reddedilememiştir. Başka bir ifade ile ölçek değerlerinin iç tutarlılığı istenilen düzeydedir. Bu nedenle V. Hal denklemine dayalı olarak yapılan ölçekleme dikkate alınmış ve III. Hal denklemine kullanarak yeniden ölçekleme yapılmasına gerek duyulmamıştır. Uyum ölçüsünün manidar bulunmaması durumunda ikili karşılaştırmalar yönteminin aksiyomlarının ve V. Halin varsayımlarının karşılandığı kabul edilir ve III. Hal eşitliği kullanılarak ölçekleme işleminin tekrar yapılmasına gerek duyulmaz (Guilford,1954).

Deney grubu için V. Hal denklemine ilişkin elde edilen ölçek değerleri şekil 2’de sayı doğrusu üzerinde gösterilmiştir.



Şekil 2: Deney grubu için hesaplanan ölçek değerleri

Deney grubu için elde edilen bulgular incelendiğinde, en çok tercih edilen yöntemin araştırmaya dayalı performans görevleri (0,859) olduğu görülmektedir. İkinci sırada birden fazla olası yanıtı olan sorulardan oluşan yazılı yoklamalar (0,716) yer alırken üçüncü sırada doğru yanlış testleri (0,591) yer almıştır. Daha sonra sırası ile çoktan seçmeli testler (0,583), kısa yanıtı ve tamamlama testleri (0,580), sınırlandırılmış performans görevleri (0,557), bireysel veya grup sunuları (0, 544) ve ev ödevleri (0,524) gelmektedir. En az tercih edilen değerlendirme yöntemi ise kontrol grubunda olduğu gibi (0,000) tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalardır.

Elde edilen bulgular dikkate alındığında, değerlendirmenin not verme amacı gütmekten geri bildirim verme amacı ile yapılacağına ilişkin yönerge verildiği zaman öğrencilerin değerlendirme tercihlerinde bir farklılık yarattığı belirtilebilir. Tablo 10'da deney ve kontrol grubu için gerçekleştirilen ölçekleme sonucunda değerlendirme yöntemleri için belirlenen sıralar ve ölçek değerleri sunulmuştur.

Tablo 10.

Deney ve Kontrol Grubu İçin Elde Edilen Ölçek Değerleri Ve Sıralar

Değerlendirme Yöntemi	Deney grubu için sıra ve ölçek değeri	Kontrol grubu için sıra ve ölçek değeri
Araştırmaya Dayalı Performans Görevleri	1 (0,859)	3 (0,477)
Birden fazla olası yanıtı olan sorulardan oluşan yazılı yoklamalar	2 (0,716)	5 (0,463)
Doğru yanlış testleri	3 (0,591)	1 (0,706)
Çoktan seçmeli testler	4 (0,583)	2 (0,562)
Kısa yanıtı ve tamamlama testleri	5 (0,580)	4 (0, 474)
Sınırlandırılmış performans görevleri	6 (0,557)	8 (0,242)
Bireysel veya grup sunuları	7 (0,544)	7 (0,278)
Ev Ödevleri	8 (0,524)	6 (0,371)
Tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalar	9 (0,000)	9 (0,000)

Tablo 10 incelendiğinde kontrol ve deney grubu için değerlendirme yöntemlerinin tercih edilme sıralarının farklılaştığı görülmektedir. Kontrol grubunda ilk iki sırada alt düzey zihinsel süreçlerin ölçülmesinde yaygın olarak kullanılan doğru yanlış testleri ve çoktan seçmeli testler yer alırken deney grubunda üst düzey zihinsel süreçlerin ölçülmesinde kullanılan araştırmaya dayalı performans görevleri ve birden fazla olası yanıtı olan sorulardan oluşan yazılı yoklamalar yer almaktadır. Deney ve kontrol grubu için elde edilen sıra puanları arasındaki ilişki Spearman Brown Sıra Farkları Korelasyon katsayısı ile hesaplanmış ve yüksek düzeyde pozitif anlamlı bir ilişki olduğu belirlenmiştir ($r=0,75$ $p<0,01$).

Sıra puanları arasındaki farklılığa deney grubunda değerlendirmenin amacına ilişkin verilen yönergenin etkili olduğu belirtilebilir. Ancak araştırmada yarı deneysel desen kullanılması ve dolayısı ile grupların seçkisiz atanmaması, ön testin pratik nedenlerden dolayı yapılamaması nedenleri ile iç geçerliğinin çok yüksek olmadığı göz ardı edilmemelidir. Belirtilen bu yöntemsel sınırlılıklara rağmen değerlendirmenin yapılış amacının öğrencilerin değerlendirme tercihleri üzerinde bir etkiye sahip olduğu belirtilebilir.

4. TARTIŞMA VE ÖNERİLER

Kontrol grubunda elde edilen bulgular incelendiğinde öğrencilerin en çok tercih ettikleri iki yöntemin doğru yanlış testleri ve çoktan seçmeli testler olduğu görülmektedir. Bu durumun altında yatan temel etmen öğrencilerin seçeneklerden faydalanarak doğru yanıtı daha kolay bulabileceklerini düşünmeleri olabilir. Dikkat edilmesi gereken bir diğer önemli nokta ise kontrol grubunda öğrencilere değerlendirmenin amacına ilişkin herhangi bir yönerge verilmemiş olsa bile devam eden eğitim öğretim sürecinde öğrencilerin ders geçme veya daha yüksek not alma kaygısını duyabilecekleri gerçeğidir. Bu durum öğrencilerin öncelikli olarak doğru yanlış ve çoktan seçmeli testleri tercih etmelerinde önemli bir rol oynamış olabilir. Zeinder (1987) lise öğrencileri üzerinde yaptığı bir çalışmada öğrencilerin çoktan seçmeli sınavları, daha basit, anlaşılır, ilgi çekici ve adil buldukları için açık uçlu sınavlara kıyasla daha çok tercih ettiklerini belirlemiştir. Anderson (1987) üniversite öğrencileri ile yaptığı çalışmada da benzer sonuçlara ulaşmıştır. Başka bir çalışmada ise öğrencilerin daha az heyecanlandıkları ve daha yüksek not alabildikleri için kısa-yanıtlı ve çoktan seçmeli gibi yüzeysel durum belirleme yöntemlerini tercih ettikleri vurgulanmıştır (Sambell, McDowell ve Brown, 1997). Watering ve diğerleri (2008) ise çalışmalarında öğrencilerin yeni öğrenme ortamları ile karşılaştıkları durumlarda mümkün olduğu kadar kapalı uçlu yanıtlar içeren değerlendirme yöntemlerini tercih ettiklerini belirtmiştir. Araştırmanın bu boyutuna ilişkin elde edilen bulguların alayazın ile örtüştüğü belirtilebilir.

Kontrol grubunda elde edilen veriler öğrencilerin üçüncü sırada araştırmaya dayalı performans görevlerini tercih ettiklerini göstermektedir. Bu durum farklı nedenlerden kaynaklanabilir. Öğrenciler, ders dışında yapacakları çalışmada, zorlandıkları noktada yardım alabileceklerini ve birbirleri ile iletişim içinde olacakları için derste başarılı olmalarını sağlayacak nitelikte bir çalışmayı gerçekleştirebileceklerini düşünmüş olabilirler. Ancak diğer yandan, öğrenciler aynı şekilde ders dışında gerçekleştirilen ev ödevlerini altıncı sırada tercih etmişlerdir. Bu durum öğrencilerin araştırmaya dayalı performans görevlerini başka gerekçeler ile tercih ettiklerini düşündürebilir. Bu durum performans görevlerinin daha geniş bir zamanda gerçekleştirildiği için sınav kaygısını azaltmasından veya öğrencilerin bilinçli bir şekilde kendi gelişimlerine önem vermelerinden ve öğrenmelerinin sorumluluğunu almalarından kaynaklanabilir. Zoller ve Chaim, (1990) öğretmen adayları üzerinde gerçekleştirilen başka bir çalışmada ise, öğretmen adaylarının zaman kısıtlaması olmayan ve bilgilerin ezberlenmesini daha az gerektiren sınav türlerini tercih ettikleri sonucuna ulaşılmıştır.

Öğrenciler dördüncü sırada ise birden fazla olası doğru yanıtı olan yazılı yoklamaları tercih etmişlerdir. Bu durumda öğrenciler, birden fazla yanıt olduğu için doğru yanıt bulma veya kısmen de olsa puan alma olasılıklarının yüksek olduğunu düşünmüş olabilirler. Ancak bu durum öğrencilerin derinlemesine öğrenme yaklaşımını benimsemiş olmalarından da kaynaklı olabilir. Doğan, Atmaca ve Aslan (2010) çalışmalarında derinlemesine öğrenme yaklaşımını benimseme ve alternatif değerlendirme yöntemlerini tercih etme arasında orta düzeyde pozitif bir ilişki olduğunu belirlemişlerdir.

Öğrencilerin değerlendirme tercihlerinde son üç sırada ise ev ödevleri, bireysel veya grup sunumları ve tek doğru yanıtı olan sorulardan oluşan yazılı yoklamalar gelmektedir. Öğrencilerin tek doğru yanıtı olan sorulardan oluşan yazılı yoklamaları son sırada tercih etmelerinin nedeni olarak doğru yanıtı bulma olasılıklarının düşük olduğunu düşünmeleri

olabilir. Bireysel veya grup sunumlarını son sıralarda tercih etmelerinin nedeni olarak öğrencilerin sunumları gerçekleştirmek için hem okul dışında çalışmak hem de derste bu çalışmalarını sunmak zorunda olmaları düşünülebilir.

Ancak bahsi geçen gerekçeler araştırmacının gözlem ve deneyimlerine dayalı olmakla beraber güçlü bir neden sonuç ilişkisi kurmak için yeterli değildir. Bu noktada derinlemesine nitel araştırmaların gerçekleştirilmesi ve bu araştırmanın bulguları sonucunda eksik kalan boyutların doldurulması yerinde olacaktır. Kontrol grubundan elde edilen verilere ilişkin yapılabilecek en güçlü yorum öğrencilerin sınav ve ders geçme kaygısı duydukları varsayımına dayalı olarak yüksek not alabileceklerini düşündükleri değerlendirme yöntemlerini ilk sıralarda tercih etmiş olabilecekleridir. Bu bulgu güçlü bir nedensellik sağlamamakla beraber bir nedenselliğin olabileceğine ilişkin ipuçları vermektedir.

Araştırmanın deney grubunda ise öğrencilere değerlendirme sonucunda başarılarının not ile değerlendirilmeyeceği, geri bildirim verilerek öğrenmelerinin gelişimine katkı sağlanacağı (biçimlendirici değerlendirmenin uygulandığı) durumları düşünerek anketi yanıtlamalarına yönelik bir yönerge verilmiştir. Elde edilen bulgular incelendiğinde deney grubu öğrencilerinin değerlendirme tercihlerinin kontrol grubu öğrencilerinininkinden farklılaştığı görülmüştür.

Kontrol grubunda en yüksek ölçek değerlerine sahip değerlendirme yöntemleri doğru yanlış testleri ve çoktan seçmeli testler iken deney grubunda araştırmaya dayalı performans görevleri ve birden fazla olası yanıtı olan sorulardan oluşan yazılı yoklamalar ön plana çıkmıştır. Bu bulgu biçimlendirici değerlendirme (formative assessment) kullanıldığında öğrencilerin üst düzey zihinsel süreçleri ölçen, yanıtını kendilerinin yapılandıracakları, mevcut bilgileri olduğu kullanmanın yanı sıra gerçek yaşamla ilişkilendirebilecekleri değerlendirme yöntemlerini tercih ettikleri şeklinde yorumlanabilir. Gijbels ve Dochy (2006) sınıflarda uygulanan biçimlendirici değerlendirme (formative assessment) uygulamalarının, öğrencilerin değerlendirme tercihleri ve öğrenme yaklaşımlarına etkisini belirlemeyi amaçlayan bir çalışma gerçekleştirmiştir. Araştırma sonucunda öğrencilerin değerlendirme tercihleri ve öğrenme yaklaşımları arasında ilişki olduğu ve derinlemesine öğrenme yaklaşımını benimseyen öğrencilerin üst düzey zihinsel süreçleri ölçen soruları tercih ettikleri belirlenmiştir. Bu bağlamda öğrencilerin not ve dersten geçme kaygısı hissetmedikleri durumlarda, öğrenmelerinin sorumluluklarını alıp kendi gelişimlerine katkı sağlayan değerlendirme yöntemlerini tercih ettikleri belirtilebilir. Bartam ve Carol (2010) üniversite öğrencileri üzerinde yaptığı çalışmasında katılımcıların sınav kaygısı ve stres içermediği için alternatif değerlendirme yöntemlerini tercih ettiği sonucuna ulaşmıştır.

Ancak dikkat çekici olan bir diğer nokta ise deney grubunda gerçekleştirilen ölçekleme çalışmasında üçüncü, dördüncü ve beşinci sırada doğru yanlış testlerinin, çoktan seçmeli testlerin ve kısa yanıtı veya tamamlama testlerinin yer almasıdır. Başka bir ifade ile öğrenciler not ve dersten geçme kaygısı hissetmedikleri durumlarda bile bu değerlendirme yöntemlerinden vazgeçememektedirler. Bunun altında yatan temel etmen öğrencilerin eğitim yaşantıları boyunca ilgili yöntemler kullanılarak değerlendirilmeleri ve bu yöntemler ile elde edilen sonuçlara dayalı olarak önemli kararların (seçme ve yerleştirmeye dayalı geniş ölçekli testler) alınmış olması olabilir. Bu noktada kapsamlı bir nitel çalışmanın gerçekleştirilmesi ve bu durumun psikolojik ve sosyolojik nedenlerinin açığa çıkarılması yerinde olacaktır.

Önemli diğer bir bulgu ise hem deney hem de kontrol grubunda tek doğru yanıtı olan sorulardan oluşan yazılı yoklamaların en az tercih edilen yöntem olmasıdır. Bu noktada kontrol grubundaki öğrenciler doğru yanıtı bulma olasılıklarının düşük olması deney grubundaki öğrenciler de öğrenme süreçlerine önemli katkı sağlamadığını düşündükleri için tek doğru yanıtı olan sorulardan oluşan yazılı yoklamaları tercih etmemiş olabilirler.

Araştırmanın deney grubunda öğrencilere biçimlendirici değerlendirme yapılacağı durumları dikkate alarak anketi yanıtlamaları istenmiştir. Bu durum biçimlendirici değerlendirme gerçekte yaşamda uygulandığı durumlardaki öğrenci tepkilerini tam olarak yansıtmayabilir. Pratik nedenlerden dolayı biçimlendirici değerlendirme etkin olarak uygulandığı durumlar elde edilemediği için öğrencilere kurgusal olarak biçimlendirici değerlendirme yapıldığı durumları düşünerek anketi yanıtlamaları istenmiştir. Bu durum araştırmanın bir sınırlılığıdır. Bu bağlamda biçimlendirici değerlendirme etkin olarak uygulandığı durumlarda benzer çalışmaların yapılması önerilebilir.

Araştırmanın diğer bir sınırlılığı ise deney ve kontrol gruplarının seçkisiz olarak atanmadığı için yarı deneysel bir desen kullanılmasıdır. Bu durum iç geçerliliği olumsuz etkilemekle beraber benzer konuyu çalışmayı planlayan araştırmacıların gerçekte deneysel desen kullanmaları önerilmektedir.

Uygulama sürecinde öğretmenlerin mümkün olduğu kadar biçimlendirici değerlendirme etkinliklerine yer vermeleri önemli bir noktadır. Buna dayalı olarak öğrencilerin değerlendirme tercihleri farklılaşmaktadır. Yapılan çalışmalar öğrencilerin değerlendirme tercihleri ve öğrenmeye ilişkin diğer özellikler arasında ilişki olduğunu ortaya koymuştur. Değerlendirmenin öğretim sürecinin pek çok aşamasında önemli bir rolü olduğu düşünüldüğünde öğrencilerin üst düzey zihinsel süreçleri ölçmeyi amaçlayan değerlendirme yöntemlerini tercih etmelerini sağlamak, öğrenmelerinin sorumluluğunu almalarında ve öğrendiklerini yaşamla ilişkilendirmelerinde önemli bir etkiye sahip olabilir.

KAYNAKÇA

- Acar Güvendir, M. ve Özer Özkan, Y. (2013). İki ölçekleme yönteminin karşılaştırılması: İkili karşılaştırma ve sıralama *Eğitim Bilimleri Araştırmaları Dergisi – Journal of Educational Sciences Research*, 3 (1), 105–119.
- Altun, A. ve Gelbal, S. (2014). Öğretmenlerin kullandıkları ölçme ve değerlendirme yöntem ve araçlarının iki karşılaştırma yöntemi ile belirlenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. 5(1) 1-11.
- Anderson, P. S. (1987). *Comparison of student attitudes about seven formats of educational testing, with emphasis on the MDT multi-digit testing technique*. paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago.
- Bal, Ö. (2011). Seviye belirleme sınavı (SBS) başarısında etkili olduğu düşünülen faktörlerin sıralama yargıları kanunıyla ölçeklenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2 (2), 200-209.
- Bartram, B. & Bailey, C. (2010). Assessment preferences: a comparison of uk/international students at an English university', *Research in Post-Compulsory Education*, 15(2), 177 - 187

- Biggs, J. (2003). *Teaching for quality learning at university* (2nd edition). United Kingdom: The Society for Research into Higher Education and Open University Pres.
- Birenbaum, M. (1994). Toward adaptive assessment - the student's angle. *Studies in Educational Evaluation*, 20, 239-255.
- Birenbaum, M. (1997). Assessment preferences and their relationship to learning strategies and orientations. *Higher Education*. 33, 71-84.
- Birenbaum M., & R. A. Feldman. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research* 40(1), 90-7.
- Birenbaum, M. (2003). *New insights into learning and teaching and their implications for assessment*, in Segers, M., Dochy, F. and Cascallar, E. (eds.), *Optimizing New Methods of Assessment: In Search of Qualities and Standards*. Dordrecht, The Netherlands: Kluwer, pp. 13-36.
- Birenbaum, M. & Rosenau, S. (2006). Assessment preferences, learning orientations, and learning strategies of pre-service and in-service teachers. *Journal of Education for Teaching*, 32(2), 213-225.
- Birenbaum, M. (2007). Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education*, 53, 749-768.
- Çüm, S. ve Doğan, N. (2016). Öğretmenlerin çalışma yaşamında güdülenme düzeylerini artıran faktörlerin önem sırasının belirlenmesi. *Kastamonu Eğitim Dergisi*, 24(1), 119-132
- Doğan, D., Atmaca, S. ve Aslan, F. (2012). The correlation between learning approaches and assessment preferences of 8th grade students. *Elementary Education Online*, 11(1), 264-272,
- Doğan, D. ve Kutlu, Ö. (5-7 Mayıs 2010). *Değerlendirme tercihleri ile öğrenme stratejileri arasındaki ilişki: İngilizce öğretmen adayları örneği*. 2. Ulusal Eğitimde Ölçme ve Değerlendirme Sempozyumu sunuldu. Mersin Eğitim Fakültesi Mersin. 5-7 Mayıs. Mersin Üniversitesi Eğitim Fakültesi.
- Doğan, C.D. (2011). *Öğretmen adaylarının başarıları belirlenirken tercih ettikleri durum belirleme yöntemlerini etkileyen faktörler ve bu yöntemlere ilişkin görüşleri*. Yayımlanmamış Doktora Tezi. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Gijbels, D. & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference? *Educational Studies*, 32(4), 399-409.
- Guilford, J.P. (1954). A least square solution for successive intervals assuming unequal standard deviations. *Psychometrika*, 19,117-137.
- Gülbahar, Y. ve Büyüköztürk, Ş. (2008). Değerlendirme tercihleri ölçeğinin Türkçeye uyarlanması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 35, 148-161.
- Grandt, J. (1987). Characteristics of examinees who leave questions unanswered on the general test rights-only scoring. *ETS Research Report* 87- 83, Princeton, NJ: Educational Testing Service
- Koğar, H. ve Gelbal, S. (2015). Üniversite öğrencilerinin sosyal beğenirlik tercihlerinin yargıcı kararları ile ölçeklenmesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*. 4(1), 136-152
- Kuzgun, Y. ve Deryakulu, D. (2004). *Bireysel farklılıklar ve eğitime yansımaları*. (Editörler: Y. Kuzgun ve D. Deryakulu). Eğitimde Bireysel Farklılıklar içinde (ss.1-11). Ankara: Nobel Yayın-Dağitim.

- MicMillan, J.H., (2015). *Sınıfıçı deęerlendirme: Etkili ölçütlere dayalı etkili bit öğretim için ilke ve uygulamalar*. Konya: Eğitim Yayınevi
- Phillips, F. (1999). 'Business students' learning preferences and associated task performance', *Journal of Education for Business* 75(1), 27–32.
- Sambell, K., McDowell, L. & Brown, S. (1997). 'But is it fair?': An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371.
- Struyven, K., Dochy, F. & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A Review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Tezbaşaran, E. (2015). Öğretmen adaylarının görüşlerin göre sınav türlerinin ikili karşılaştırma yöntemi ile ölçeklenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi* 4(3), 283-288
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Turgut, M. F. ve Baykul, Y. (1992). *Ölçekleme Teknikleri*. Ankara: ÖSYM Yayınları.
- Wilson, K. & Fowler, J. (2005) Assessing the impact of learning environments on students' approaches to learning, *Assessment and Evaluation in Higher Education*, 30(1), 85–99
- Watering, G. van de, Gijbels, D., Dochy, F., & Rijt, J. van der (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education*, 56, 645-658.
- Zeinder, M. (1987). Essay versus multiple-choice type classroom exams: the student's perspective. *Journal of Education Research*. 80, 352-358.
- Zoller, U. & Ben-Chaim, D. (1990). Gender differences in examination-type performance test anxiety, and academic achievements in college science education - a case study- *Science Education*, 74, 597-608.

EXTENDED ABSTRACT

1. Introduction

The quality of learning depends on many factors such as learning environment, skills of the instructor, curriculum, learning materials. However, the students' individual differences should not be ignored (Kuzgun & Deryakulu, 2004). Learning is permanent changes in behaviors as a result of life experiences. So, effective learning can be provided by constructing learning area effectively. In order to construct learning area effectively, students' individual differences should be taken into consideration in schools (Doęan, Atmaca Aslan, 2012)

Assessment is an important component of learning environments as learning approaches. It explains quantity of students' achievement and development by measuring both learning process and products. While planning the learning environment, students' individual differences should be considered both in instruction and assessment process. For these reasons, recently, the concept of "assessment preferences" has gained importance. The concept of "assessment preferences" refers to students' opinions, attitudes, and preferences of assessment methods and their properties (Birenbaum, 1997).

So many factors such as learning approaches, learning strategies, studying strategies are correlated to students' assessment preferences and students' achievement (Birenbaum & Feldman, 1998; Philips, 1999; Wilson & Fowler, 2005; Birenbaum & Rosenau, 2006). Students' assessment preferences should be taken in to consideration during the instruction process (Gülbahar & Büyüköztürk, 2008).

In the majority of those studies "Assessment Preference Inventory" developed by Birenbaum (1994, 1997, 2007) was used. Unfortunately there is no scaling study about assessment preferences of university students. Moreover it is probable that type of the assessment may influence students' assessment preferences. Students may prefer different assessment methods in conditions where formative assessment or summative assessment is used. Thus, in this study, it is aimed at scaling of university students' assessment preferences methods in conditions where formative assessment is used and not used.

2.Method

2.1. Research model

This is an experimental study. In the control group the questionnaire forms used for scaling assessment preference of students were administered without any additional instruction about the aim of the assessment. On the other hand, in the experimental group the questionnaire forms were administered with an additional instruction about the aim of the assessment. The instruction in the experimental group was asked participants to answer the questionnaire considering the conditions when formative assessment was applied. Because the participants in the experimental and control group were not assigned randomly, semi experimental design was used. Experimental and control groups were matched according to the midterm exam results of scientific research methodology course.

2.2 Participants

Participants of the study consisted of 205 undergraduate students at the Faculty of Education of a state university stated in Ankara Turkey. Out of 205 participants, 108 of them were in control group, and 97 of them were in experimental group.

2.3 Data collection tools

The data collection tool used to scale the assessment preference of students were developed by the researcher. The questionnaire included a pair of assessment method on each line and the respondents were required to select one of two assessment methods. Assessment methods were arranged in pairs so that every one of them was paired with every other one.

The questionnaire included assessment methods which were experienced by the participants before. Those assessment methods were multiple choice tests, true-false tests, close tests, essay types exams which consisted of questions which had more than one possible correct answer, essay types exams which consisted of question which had one correct answer, research based performance tasks, limited answered performance tasks, home works and individual or group presentations

The study included two different forms. One of them did not include an instruction about the aim of the assessment. On the other hand, the other form included an additional

instruction which asked the participants to answer the questionnaire considering the conditions when formative assessment was applied

2.4. Data analysis

To scale the respondents' assessment preferences, method of paired comparison was used. Thurstone (1927) first introduced a scientific approach to using pairwise comparisons for measurement which he referred to as the law of comparative judgment. Paired comparison method can be used to order items along a dimension such as preference or importance using an interval-type scale.

3. Findings, Discussion and Result

In the control group, students' assessment preferences were scaled in the condition where no additional instruction were given to the respondents about the aim of the assessment. Students' assessment preferences were scaled according to III. case equation.

Results showed that in the control group students mostly prefer true-false items and multiple choice tests which aims to assess lower order thinking skills. Students might have preferred those assessment methods because they thought that it is easy to find the correct answer. The assessment methods that the students preferred in control group and their scale values were respectively; 1. Multiple choice tests (0.73) 2. True-false items (0.54) 3. Research based performance tasks (0.48) 4. Essay types exams which consisted of question which had more than one possible correct answer (0.46) 5. Close tests (0.46) 6. Home works (0.31) 7. Limited answered performance tasks (0.13) 8. Individual and group presentations (0.09) 9. Essay types exams which consisted of question which had one correct answer (0.00).

In the experimental group assessment preference of the students were scaled according to V. case equation. Results showed that students in experimental group mostly prefer research based performance tasks and essay types exams which consisted of question which had more than one possible correct answer which aim to assess higher order thinking skills. So it can be said that the instruction about the aim of the assessment might have affected students' assessment preference. In other words, students' assessment preference may vary according to the aim of assessment. The assessment methods that the students preferred in the experimental group and their scale values were respectively; 1. Research based performance tasks (0.86) 2. Essay types exams which consist of question which has more than one possible correct answer (0.72) 3. True-false tests (0.59) 4. Multiple choice tests (0.58) 5. Close Tests (0.58) 6. Limited answered performance tasks (0.55) 7. Individual and group presentations (0.54) 8. Homeworks (0.52) 9. Essay types exams which consisted of question which had one correct answer (0.00).

When formative assessment is used students have tendency to prefer assessment methods which aim to assess higher order thinking skills whereas when summative assessment is used students have anxiety for passing or fail from the course and have tendency to prefer assessment methods which is easy to find correct answer.

On the other hand, in this study it is not possible to highlight very strong cause and effect relationship between the assessment preference of the students and the usage of formative assessment because the experiential design is not include random assignment of the participants to the groups. This is the limitation of the study. For the researchers who aims to study similar subjects, it is advised to conduct a qualitative analysis to explore the reasons lay behind the students' assessment preference.

Döş, İ., Atalmış, E. H. (2016). OECD verilerine göre PISA sınav sonuçlarının değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, Cilt 16(2), 432-450.

Geliş Tarihi: 28/12/2015

Kabul Tarihi: 09/06/2016

DOI:

OECD VERİLERİNE GÖRE PISA SINAV SONUÇLARININ DEĞERLENDİRİLMESİ*

İzzet DÖŞ**
Erkan Hasan ATALMIŞ***

ÖZET

Bu çalışmanın amacı, OECD raporlarında yer alan ülkelere ait eğitim verilerinin, ülkelerin PISA skorlarını nasıl yordadığını araştırmaktır. Bu amaç doğrultusunda bağımsız değişken olarak ülkelerin sınıf başı öğrenci sayısı, öğretmen başı öğrenci sayısı, öğrenci başı yıllık harcama, öğretim süresi ve öğretmen maaşları olarak alınırken, bağımlı değişken olarak da, PISA matematik, okuma ve fen skorları kullanılmıştır. OECD'ye üye olan ülkelerden 26 ülkenin 2006, 2009 ve 2012 yılı verileri kullanılmıştır. Verilerin analizinde regresyon analizi ve ANOVA kullanılmıştır. Yapılan doğrusal regresyon analizi sonucunda, PISA matematik, okuma ve fen skorlarını anlamlı olarak yordayan tek ortak değişkenin öğretmen maaşı olduğu ortaya konulmuştur. Öğretmen maaşına ek olarak sınıf başı öğrenci sayısı (SBÖS) matematik ve okuma skorlarını anlamlı olarak yordamaktadır. Bu sonuçlara göre öğretmen maaşının artması PISA'nın tüm skorlarını pozitif olarak etkilerden, SBÖS ise PISA matematik ve okuma skorlarına negatif bir etki oluşturmaktadır. **Anahtar Kelimeler:** OECD ülkeleri, ülkelerin eğitim verileri, PISA

EVALUATING OF PISA RESULTS BASED ON OECD DATA

ABSTRACT

The purpose of this study is to explore to what extent PISA scores are predicted by the educational investments of countries which are available in OECD reports. Accordingly, teaching time, teacher salary, annual expenditure on educational institutions per student (AEPS), number of students per teaching staff (SPTS), and estimated class size were set as independent variables while PISA scores of countries (mathematics, reading, and science scores) were accepted as dependent variables. The data obtained from the reports prepared for a total of 26 out of around 40 OECD member countries (i.e. PISA scores of the countries in 2006, 2009, and 2012) were analyzed. Linear regression method was employed to predict the related PISA scores. Findings of the study indicated that teacher salary is a common variable that statistically influences predict students' achievement in mathematics, reading, and science. Class size has revealed another variable that has a significant impact on predicting their achievement in mathematics and reading. All in all, teacher salary has proved a positive impact on students' performance in mathematics, reading, and science while class size negatively influences their performance in mathematics and reading.

Key Words: OECD countries, Education datas, PISA.

* Bu çalışma X. Ulusal Eğitim Yönetimi kongresinde bildiri olarak sunulmuştur.

** Kahramanmaraş Sütçü İmam Üniversitesi, Eğitim Fak., Eğitim Yönetimi Teftişi Planlaması ve Ekonomisi Bölümü, e-mail: id27tr@gmail.com

*** Kahramanmaraş Sütçü İmam Üniversitesi, Eğitim Fak., Psikolojik Danışmanlık ve Rehberlik e-mail: erkanatalmis@gmail.com

1.GİRİŞ

Eğitim hem bireye hem de topluma yönelik bir yatırımdır. Eğitim uzun vadeli, getirisi en çok olan yatırımlardan biri olup, özde bireye, genelde topluma hizmet üreten bir süreçtir. Eğitim alanında yapılan harcamalar, yatırımlar aynı zamanda eğitim çıktıları üzerinde de etkili olup, ülkelerin kalkınmalarında önemli roller oynamaktadır. Bu bağlamda gelişmiş ve gelişmekte olan ülkeler eğitim harcama ve yatırımlarını devamlı artırmaktadırlar.

İktisatta iki tür harcamadan bahsedilir. Eğer bir harcama gelecekte fayda sağlıyorsa, elde edilen mal veya hizmetin getirisi süreklilik arz ediyorsa, buna yatırım harcaması denilmektedir. Bir harcamanın faydası süreklilik arz etmiyorsa bu harcamada tüketim harcaması olarak kabul edilmektedir (Özdemir, 2013). Eğitim bağlamında bu tanım ele alındığında bireyler, eğitimle hem anlık bazı faydalar elde ederken, elde ettiği birikimlerle gelecekte de fayda elde etmeye devam ederler. Örneğin yabancı dilini ve matematik bilgisini kullanma gibi. Dolayısıyla eğitim hem yatırım hem de tüketim özelliği olan bir yapıdadır ama yatırım özelliği daha çok ön plandadır (Gümüş ve Şişman, 2012).

Ülkeler yaptıkları yatırımların verimliliği ölçmek için bazı araçlar geliştirmişlerdir. Her ülke eğitim durumunu ve eğitim çıktılarını kendi içerisinde bazı değerlendirme kriterlerine göre ölçmektedir. Ama bu ölçümler çoğu zaman anlamlı yorumlar ortaya koyamamaktadır. Bundan dolayı özellikle, Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) ülkeleri bu ölçümleri kendi içlerinde karşılaştırmalı olarak yaparak daha anlamlı ve gerçekçi çıkarımlara ulaşabilmektedirler. OECD'ye üye olan ülkeler, eğitim alanında çok farklı değişkenlere ilişkin çıktılara ve bilgilere sahiptirler. OECD merkezi ise tüm üye ve ortak ülkelerden bu bilgileri elde edip çıktıları karşılaştırmalı olarak raporlarında yayınlamaktadır. Bu verilerin başında PISA sınavları gelmektedir.

Dünya genelinde, politika belirleyicileri kendi ülkelerindeki öğrencilerin bilgi ve beceri düzeylerini, PISA'ya katılan diğer ülkelerdeki öğrencilerin bilgi ve beceri düzeyleriyle karşılaştırmak, eğitim düzeyinin yükseltilmesi amacıyla standartlar oluşturmak, eğitim sistemlerinin güçlü ve zayıf yönlerini belirlemek için sınav sonuçlarını kullanmaktadırlar (MEB, 2013). Yine gelişmiş ve gelişmekte olan ülkeler bu verileri kullanarak, kendi eğitim seviyelerini diğer ülkelerle karşılaştırarak nerede oldukları, eğitim durumları, başarı seviyeleri hakkında bilgi sahibi olmaktadır. Ülkeler, bu eğitim çıktılarını program geliştirme, öğretmen yetiştirme ve etkin eğitim politikası geliştirmede kullanmaktadırlar.

Çok amaçlı ve büyük bir organizasyon olan PISA, "Uluslararası Öğrenci Değerlendirme Programı" adıyla, OECD tarafından üçer yıllık dönemler hâlinde, 15 yaş grubundaki öğrencilerin eğitim durumlarını ölçen bir araştırmadır. PISA en az 7 yıl öğrenim görmüş olması koşuluyla 15 yaş öğrencilerine uygulanır. Bu yaştaki öğrenciler, birçok ülkede geçerli olan zorunlu öğrenim süresini doldurmak üzere olduklarından, PISA tarafından öğrenciliğin ardından yetişkin bir vatandaşlığın başlayabileceği en küçük yaş olarak kabul edilirler (MEB, 2005, 2013).

PISA'nın temel amacı, çocukları daha iyi tanımak ve geleceğe ilişkin projeler ve yatırımlar için ön bilgi sahibi olmaktır. PISA sonuçları ile gençleri daha iyi tanımak; onların öğrenme isteklerini, derslerdeki performanslarını ve öğrenme ortamları ile ilgili tercihlerini daha açık bir biçimde ortaya koymak mümkün olmaktadır. PISA sınavlarında

15 yaş grubundaki öğrencilerin; problem çözme, matematik ve fen bilimleri okuryazarlığı ile okuma becerileri alanlarındaki bilgi düzeyleri yoklanmaktadır (MEB, 2015).

PISA sınavlarında matematik alanında, uzay ve şekil, değişme ve ilişkiler, sayı ve belirsizlik alanlarında sorular çıkmaktadır. Fen alanında fiziksel sistemler, canlılar, yerküre ve uzay sistemleri, teknoloji sistemleri, bilimsel sorgulamalar içeriği verilmektedir. Okuma alanında öyküleme, yorumlama, tartışma, düz yazılar, grafik formlar ve listeler vb. alanlardan sorular çıkmaktadır (MEB, 2005, 2010). Problem çözme yine sınavlardaki konulardan olup öğrencilerin bu alandaki becerileri sınanmaktadır. Okulda verilen temalarda iyi olan bir öğrencinin problem çözmede başarılı olacağı düşünülmemelidir. Avustralya, Brezilya, İtalya, Japonya, Kore, Çin (Macao), Sırbistan, İngiltere ve ABD problem çözme becerisinde (matematik, fen ve okumada benzer performans gösteren öğrenciler arasında) anlamlı bir şekilde diğer ülkelerden daha iyidirler.

En son sınav olması nedeniyle 2012 yılı sınav sonuçlarına ilişkin bazı sonuçları vermekte fayda vardır. 2012 sınavlarına göre problem çözme alanında 562 puanla Singapur, birinci; Kore, ikinci ve Japonya üçüncü olmuştur (OECD, 2014a). 2012 PISA matematik sonuçlarında Çin (Şangay), 613 puanla en yüksek skoru almıştır. Singapur, Hong-Kong, Kore, Japonya diğer takip eden en yüksek puanlı ülkelerdir. Yine okuma alanında Çin (Şangay), Hong-Kong, Singapur ve Japonya en yüksek puan alan ülkelerdir. Çin (Şangay), Hong-Kong, Singapur, Japonya ve Finlandiya fen bilimleri testinde 2012 yılında en yüksek beş ülke içerisinde yer almışlardır. Çin (Şangay), 580 puanla birinci olmuştur. Okuma yazma alanında, Çin (Şangay), Hong-Kong, Singapur, Japonya ve Kore okuma alanında ilk beştedir. 570 puanla Çin (Şangay) birinci sıradadır. (OECD, 2014b).

2012 PISA sonuçları 65 ülkenin katılımı ile ülkelerin ekonomik kriz ile uğraştıkları bir zamanda gerçekleşmiş, sonuçta ülkeler arasında ve ülkelerin kendi içlerinde büyük birçok farklılıkların olduğu görülmüştür. Bu sınavda önceki sınava göre 40 ülke en az bir konuda ortalamanın üzerinde bir performans göstermiştir. Şangay, Singapur başarılarını daha da yukarıya çıkarmayı başarmıştır. Brezilya, Meksika ve Tunus ve Türkiye gibi düşük seviyede bulunan ülkeler ise büyük ilerleme kaydetmişlerdir (OECD, 2013a).

PISA sınavlarında bilgi ve beceriler, global anlamda 70'ten fazla ülkenin katılımıyla ölçülmektedir. Ayrıca öğrencilerin motivasyonları, kendileri hakkındaki görüşleri, öğrenme biçimleri, okul ortamları ve aileleri ile ilgili veriler de toplanmaktadır. 2012 yılında 65 ülkedeki, 15 yaş çağlarında 28 milyon öğrenciden yaklaşık 510000 öğrenci okuma, matematik ve fen alanında test almıştır. Bu ülkelerin 44 tanesi problem çözme becerisi ve 18 tanesi de finansman alanında testler almışlardır. 70'ten fazla ülke de 2015 sınavları için kaydolmuştur. 2012 sınavlarında 15 yaş grubu çocukların sınavları 3 ayda, 16 yaş grubundaki çocukların sınavları da 2 ay içerisinde tamamlanmıştır. (OECD, 2014b, 2014c).

PISA sınavları; sırasıyla 2003 yılında 40 ülkede uygulanmıştır. 2006 yılında programa 17 yeni ülke eklenmiş ve ülke sayısı 57'ye yükselmiştir. 2009 yılında uygulanan sınava dâhil olan ülke sayısı ise 65 olmuştur. Türkiye, PISA'ya söz konusu üç yılda da sırasıyla 4855, 4942 ve 4996 öğrenci ile katılmıştır.

PISA sınavları benzersizdir. Çünkü bu sınavlar doğrudan okul eğitim programlarına bağlı olarak geliştirilmemektedir. Bu testler zorunlu eğitimin sonunda bireyin belirlenen niteliğin hangi aşamasında olduğunu da belirlemektedir. Bu yüzden gerçek hayatta ve

toplumsallaşma sürecinde kullanabilecekleri bilgilerin niteliğini de bu testle ölçülmektedir.

PISA sınavları açık uçlu, kapalı uçlu ve çoktan seçmeli soruların kombinasyonlarından oluşan ve 2 saat süren bir sınavdır. Bu soru grupları paragraflar halinde özel hayattan uyarlanmış şekilde hazırlanmaktadır. Öğrenciler farklı testlerin farklı kombinasyonlarını alırlar. Öğrencilerin hazır bulunuşluk düzeyleri, öğrenme tecrübeleri, okul sistemleri ve öğrenme çevreleri hakkında da öğrenciler ve okul müdürleri teste tabi tutulurlar.

PISA sınavlarında elde edilen sonuçlar, OECD tarafından her dönem sonunda raporlaştırılmaktadır. Bu sonuçlar, eğitim-öğretim programlarının geliştirilmesinde karşılaşılan eksiklerin giderilmesinde ve eğitim alanında yapılan araştırmalara kaynak olarak kullanılmaktadır. OECD raporlarında kullanılan değerler, kullanışlı, eğitim açısından anlamlı ve eğitim ile ilişkisi olan verilerdir. Mesela verileri raporlarda mevcut olan ülkelere ait eğitim yatırımları değişkenleri ile ülkelerin PISA skorları arasında ilişki kurulabilir. Bu rapordaki eğitim yatırım değişkenlerinden bazıları öğrencilerin öğretim süresi, öğretmenlerin öğretim süresi, öğretmen maaşları, ortalama sınıf büyüklükleri ve öğrenci başı yıllık harcamadır (OECD, 2014c, 2013b, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004).

1.1.Araştırmanın Amacı

Eğitim, her toplumun tarihten beri önem verdiği ve yüksek düzeyde yatırımlar yaptığı bir alandır. Her ülke bir diğer ülkeden daha üstün olmak, devletlerarasında muvazene unsuru olmak için güçlenmek ister, güçlenmenin ve büyümenin temelinde ise eğitim en önemli araç olarak görülmektedir.

Bu çalışmanın amacı, OECD raporlarındaki ülkelerin eğitim değişkenlerinin, PISA sınav sonuçlarını ne kadar yordadığını ortaya koymaktır. Bu amaç etrafında aşağıdaki sorulara cevap aranmıştır:

- 1- Ülkelerin PISA skorları ve eğitim değişkenleri yıllara göre anlamlı bir şekilde değişmekte midir?
- 2- Ülkelerin eğitim değişkenleri, PISA matematik skorlarını yordamakta mıdır?
- 3- Ülkelerin eğitim değişkenleri, PISA okuma skorlarını yordamakta mıdır?
- 4- Ülkelerin eğitim değişkenleri, PISA fen skorlarını yordamakta mıdır?

2. YÖNTEM

Bu araştırma korelasyonel bir çalışmadır. Korelasyonel araştırma iki ya da daha çok değişken arasındaki ilişkinin herhangi bir şekilde bu değişkenlere müdahale edilmeden incelendiği araştırmalardır. Korelasyonel çalışmalarda değişkenler arasındaki değişimler incelenir (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2015). Daha sonra değişkenler arasında nedensellik araştırılmıştır. Bu bağlamda PISA sınav sonuçları ve OECD ülkelerinin eğitim değişkenleri modele dâhil edilmiştir.

Değişkenlerin belirlenmesinde PISA sonuçları ve bu sonuçların üzerinde etkisi olduğu düşünülen eğitim yatırım değişkenleri ele alınmıştır. Bağımsız değişkenler; öğrencilerin öğretim süresi, öğretmen maaşları, öğrenci başı yıllık harcama (ÖBYH), öğretim kadrosuna düşen öğrenci sayısı (ÖBÖS) ve sınıf başı öğrenci sayıları (SBÖS) şeklindedir. Bağımlı değişkenler ise ülkelerin okuma, matematik ve fen skorları olarak alınmıştır.

Verilerin analizlerinde PISA 2006, 2009 ve 2012 fen, matematik ve okuma skorları ele alınmıştır. Bu skorlarla ilişkisi olduğu düşünölen ve bir yıl öncesine ait (2005, 2008, 2011) eğitim verileri analize dâhil edilmiştir. Bir yıl öncesinin eğitim verilerinin ele alınmasının nedeni ise, bir sonraki yıl yapılacak PISA sınavı etkileyeceği düşüncesinden dolayıdır.

Analize alınan ölkeler belirlenirken bazı kriterler göz önüne alınmıştır. PISA 2006, 2009, 2012 yıllarında yapılan sınavlarda yer almış ve OECD raporlarına göre her üç dönemde de eğitim verilerine dair bilgileri tam olan ölkeler, çalışmaya dâhil edilmiştir. Kriterlere göre her üç dönemde de verileri tam olan 26 ölkenin PISA skorları ile 2005, 2008, 2011 yıllarına ait üç dönemlik toplam 78 eğitim verisi bu çalışmanın analizine dâhil edilmiştir.

Gerek PISA skorlarının gerekse eğitim yatırım değişkenlerinin yıllara göre değişip değişmediğini incelemek için ANOVA testi uygulanmıştır.

Değişkenler arasındaki ilişkiyi ortaya koymak için önce Pearson korelasyonu yapılmış, anlamlı ilişki çıkan durumlarda ise çoklu regresyon analizi yapılmıştır.

3. BULGULAR

3.1.Yıllara Göre Değişim

2006, 2009 ve 2012 yıllarına ait eğitim değişkenleri ve PISA skorlarına ilişkin 26 ölkenin PISA skor ortalamaları ve eğitim değişkenlerinin yıllara göre ortalamaları gösteren tek yönlü varyans analiz sonuçları (ANOVA) Tablo 1’de verilmiştir.

Tablo 1.

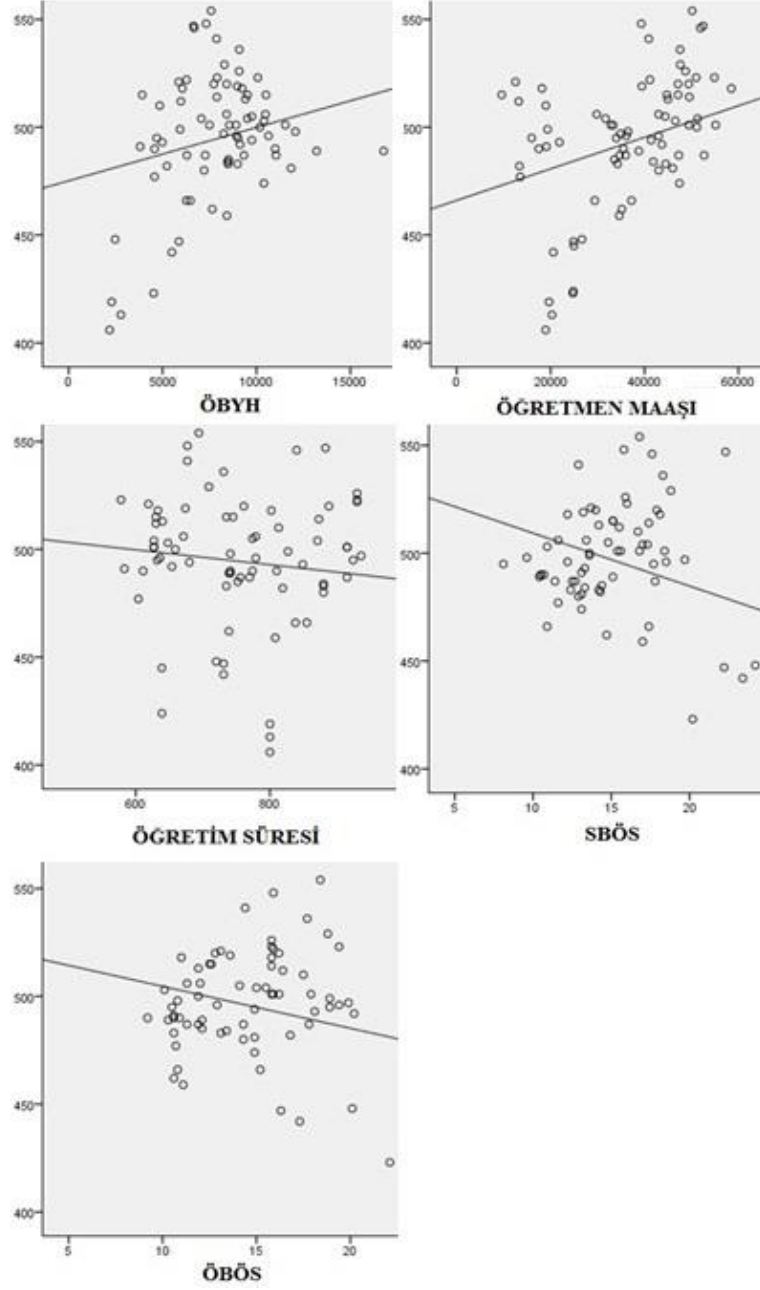
Ülkelerin Yıllara Göre Eğitim Değişkenlerinin ve Skorların Ortalamaları ve Tek Yönlü Varyans Analizi Sonuçları

Değişkenler	Yıl	μ	S	F	p
SBÖS	2005	17.6	6.0	0.64	0.53
	2008	16.3	6.0		
	2011	15.9	4.4		
ÖBÖS	2005	16.1	5.3	0.21	0.81
	2008	15.8	4.8		
	2011	15.2	4.0		
ÖBYH (\$)	2005	7,732.8	3,168.3	0.37	0.69
	2008	8,236.6	2,949.3		
	2011	8,494.3	3,499.1		
Öğretmen Maaşı (\$)	2005	37,409.7	13,967.4	0.03	0.97
	2008	37,770.5	13,852.5		
	2011	38,504.5	18,091.0		
Öğretim Zamanı	2005	762.1	125.4	0.30	0.74
	2008	767.5	123.0		
	2011	787.6	132.3		
Matematik	2006	492.6	34.8	0.04	0.96
	2009	495.0	28.2		
	2012	493.9	30.4		
Okuma	2006	487.1	32.0	1.33	0.27
	2009	493.0	23.4		
	2012	500.3	31.0		
Fen	2006	492.4	35.5	0.71	0.49
	2009	501.3	30.0		
	2012	501.8	29.9		

Tablo 1'e göre gerek eğitim değişkenleri gerekse PISA skorları ortalamaları yıllara göre hiçbir şekilde anlamlı olarak bir değişme göstermemiştir. Yıllara göre fen, okuma, kısmen matematik puanları tüm ülke ortalamalarında artış göstermesine rağmen bu artışlar yıllara göre anlamlı derecede değildir.

Matematik Skorları için Çoklu Regresyon Analizi

Ülkelerin eğitim değişkenlerinin PISA matematik skorlarına göre dağılım ve ilişkileri Şekil 1'de verilmektedir.



Şekil 1. Ülkelerin eğitim yatırım değişkenlerinin PISA matematik skorlarına göre dağılım ve ilişkileri

Şekil 1'e göre matematik skoru öğrenci başı yıllık harcama (ÖBYH) ve öğretmen maaşı ile anlamlı ve pozitif doğrusal ilişki göstermektedir (ÖBYH: $r = .27, p < .05$; öğretmen maaşı: $r = .36, p < .01$). Matematik skoru sınıf başı öğrenci sayısı (SBÖS) ve öğretmen başı öğrenci sayısı (ÖBÖS) ile anlamlı ve negatif doğrusal bir ilişki göstermektedir (SBÖS: $r = -.43, p < .01$; ÖBÖS: $r = -.29, p < .05$). Ancak matematik skoru ile öğretim zamanı arasındaki ilişki anlamsızdır ($r = -0.14, p = .22$).

Yukarıdaki sonuçları doğrultusunda ülkelerin eğitim yatırım değişkenlerinden ÖBYH, Öğretmen maaşı, SBÖS ve ÖBÖS bağımsız değişkenler olarak Tablo 2'de görüldüğü gibi regresyon analizine dâhil edilmiştir.

Tablo 2.*Matematik Skorlarını Yordama Yönelik Çoklu Regresyon Analizi Sonucu*

Değişkenler	B	SH	β	t	p
Sabit	504.49	20.18		25.00	0.00
SBÖS	-3.06	1.12	-0.48	-2.74	0.01
ÖBÖS	1.00	1.26	0.15	0.73	0.43
ÖBYH (\$1000)	-2.11	2.18	-0.21	-0.97	0.34
Öğretmen Maaşı (\$1000)	1.08	0.39	0.53	2.75	0.01
R ² =0.26		F = 6.06	p= 0.00		

Yapılan analiz sonucunda regresyon modeli istatistiksel olarak anlamlı olup ($F(5,67)=6.06, p < .001$), bağımsız değişkenler matematik sınav sonuçlarındaki değişimin %26'sını açıklamaktadır. Bu analiz sonuçlarına göre modelde öğretmen maaşı ve SBÖS anlamlı çıkarken ÖBÖS ve ÖBYH anlamsız çıkmaktadır.

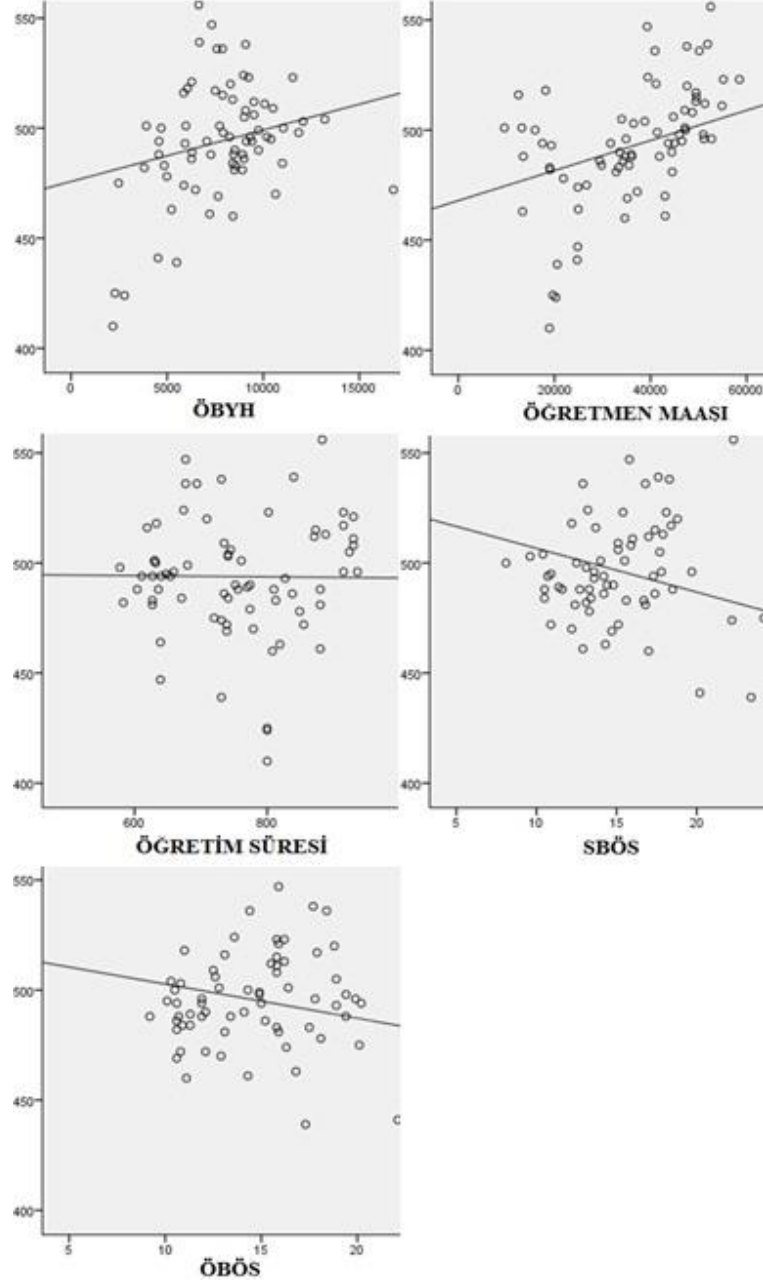
Çoklu regresyon analizine göre matematik skorlarının yordamasını gösteren ilişki aşağıda verilmektedir.

Matematik Puanı = 504,49 - 3.06 (SBÖS) + 1.00 (ÖBÖS) - 2.11 (ÖBYH) + 1.08 (Öğretmen Maaşı)

Tablo 3'e göre diğer değişkenler kontrol edildikten sonra SBÖS 1 kişi artığında matematik skoru yaklaşık 3,1 puan düşmektedir. Aynı şekilde yıllık öğretmen maaşı \$1000 artığında matematik skorlarının ortalaması yaklaşık 1,1 puan artığı görülmektedir.

Okuma Skorları için Çoklu Regresyon Analizi

Ülkelerin eğitim yatırım değişkenlerinin PISA okuma skorlarına göre dağılım ve ilişkileri Şekil 2'de verilmektedir.



Şekil 2. Ülkelerin eğitim yatırım değişkenlerinin PISA okuma skorlarına göre dağılım ve ilişkileri

Şekil 2'ye göre okuma skoru ile ÖBYH ve öğretmen maaşı arasındaki ilişki anlamlı ve pozitif doğrusal ilişki göstermektedir (ÖBYH: $r = .26$, $p < .05$; öğretmen maaşı: $r = .36$,

$p < .01$). Okuma skoru, SBÖS ve ÖBÖS ile anlamlı ve negatif doğrusal bir ilişki göstermektedir (SBÖS: $r = -.37, p < .01$; ÖBÖS: $r = -.25, p < .05$). Ancak matematik skoru ile öğretim zamanı arasındaki ilişki anlamsızdır ($r = -.01, p = .93$).

Yukarıdaki bulgular sonucunda ülkelerin eğitim yatırım değişkenlerinden ÖBYH, Öğretmen maaşı, SBÖS ve ÖBÖS bağımsız değişkenler olarak Tablo 3'te görüldüğü gibi regresyon analizine dâhil edilmiştir.

Tablo 3.
Okuma Skorlarını Yordama Yönelik Çoklu Regresyon Analizi Sonucu

Değişkenler	B	SH	β	t	p
Sabit	499.23	19.89		25.09	0.00
SBÖS	-2.90	1.11	-0.47	-2.62	0.01
ÖBÖS	1.13	1.25	0.17	0.91	0.37
ÖBYH (\$1000)	-1.66	2.15	-0.17	-0.97	0.44
Öğretmen Maaşı (\$1000)	1.00	0.39	0.51	2.59	0.01
R ² =0.25		F = 5.61	p= 0.00		

Yapılan analiz sonucunda regresyon modeli istatistiksel olarak anlamlı olup ($F(5,67)=5.61, p < .01$), okuma skorlarındaki değişimin %25'i bağımsız değişkenler tarafından açıklamaktadır. Bu analiz sonuçlarına göre modelde öğretmen maaşı ve SBÖS anlamlı çıkarken ÖBÖS ve ÖBYH anlamsız çıkmaktadır.

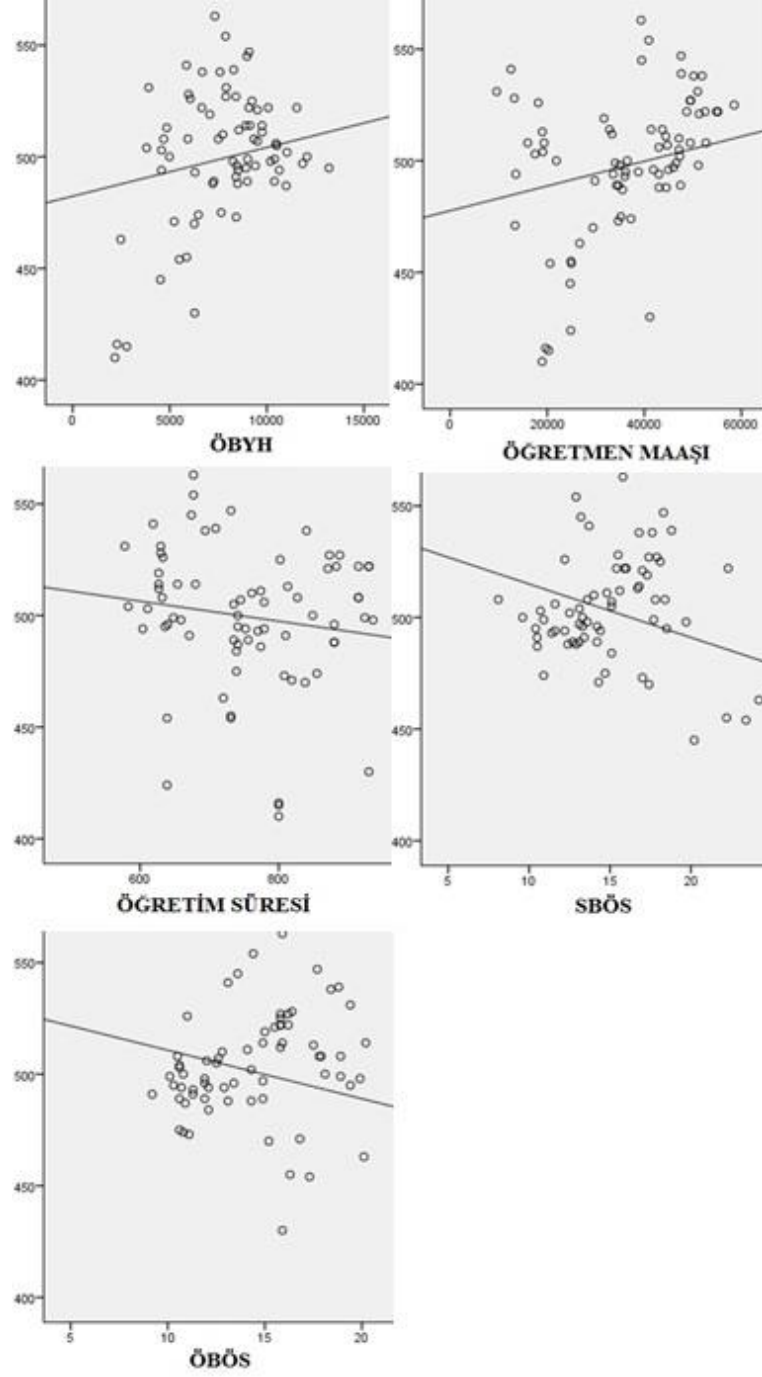
Çoklu regresyon analizine göre okuma skorlarının yordamasını gösteren ilişki aşağıda verilmektedir.

$$\text{Okuma Puanı} = 499,23 - 2,90 (\text{SBÖS}) + 1,13 (\text{ÖBÖS}) - 1,66 (\text{ÖBYH}) + 1,00 (\text{Öğretmen Maaşı})$$

Bu verilere göre diğer değişkenler kontrol edildikten sonra SBÖS 1 kişi artığında matematik puanı 2,9 puan düşmektedir. Aynı şekilde yıllık öğretmen maaşı \$1000 artığında okuma skorlarının 1 puan artığı görülmektedir.

Fen Skorları için Çoklu Regresyon Analizi

Ülkelerin eğitim yatırım değişkenlerinin PISA fen skorlarına göre dağılım ve ilişkileri Şekil 3'te verilmektedir.



Şekil 3. Ülkelerin eğitim yatırım değişkenlerinin PISA fen skorlarına göre dağılım ve ilişkileri

Şekil 3'e göre fen skoru ile ÖBYH ve öğretmen maaşı arasındaki ilişki anlamlı ve pozitif doğrusal ilişki göstermektedir (ÖBYH: $r=.23$, $p<.05$; öğretmen maaşı: $r=.27$, $p<.05$). Fen skoru, SBÖS ve ÖBÖS ile anlamlı ve negatif doğrusal bir ilişki göstermektedir (SBÖS: $r=-.43$, $p<.01$; ÖBÖS: $r=-.32$, $p<.01$). Ancak fen skoru ile öğretim zamanı arasındaki ilişki anlamsızdır ($r=-.18$, $p=.11$).

Yukarıdaki bulgular sonucunda ülkelerin eğitim yatırım değişkenlerinden ÖBYH, Öğretmen maaşı, SBÖS ve ÖBÖS bağımsız değişkenler olarak Tablo 4'te görüldüğü gibi regresyon analizine dâhil edilmiştir.

Tablo 4.*Fen Skorlarını Yordama Yönelik Çoklu Regresyon Analizi Sonucu*

Değişkenler	B	SH	β	t	p
Sabit	517.36	20.64		25.06	0.00
SBÖS	-2.21	1.14	-0.35	-1.94	0.06
ÖBÖS	0.05	1.29	0.01	0.04	0.97
ÖBYH (\$1000)	-1.91	2.23	-0.19	-0.86	0.40
Öğretmen Maaşı (\$1000)	0.89	0.40	0.45	2.21	0.03
$R^2=0.20$		$F=4.15$		$p=0.00$	

Yapılan analiz sonucunda regresyon modeli istatistiksel olarak anlamlı olup ($F(5,67)=4.15$, $p<.001$), fen skorlarındaki değişimin %20'si bağımsız değişkenler tarafından açıklamaktadır. Bu analiz sonuçlarına göre modelde sadece öğretmen maaşı anlamlı çıkarken, SBÖS, ÖBÖS ve ÖBYH anlamsız çıkmaktadır.

Çoklu regresyon analizine göre fen skorlarının yordamasını gösteren ilişki aşağıda verilmektedir.

Fen Puanı = 517,36 - 2.21 (SBÖS) + 0.05 (ÖBÖS) - 1.91 (ÖBYH) + 0.89 (Öğretmen Maaşı)

Tablo 4'e göre diğer değişkenler kontrol edildikten yıllık öğretmen maaşı \$1000 artığında fen skorlarının yaklaşık 0,9 puan artığı görülmektedir.

4.TARTIŞMA ve SONUÇ

Eğitim gerek bireylerin gerekse toplumların gelişmelerine önemli katkıda bulunan uzun vadeli bir yatırımdır. Dolayısıyla gelişmiş ve gelişmekte olan ülkeler eğitim yatırımlarına önem vermektedir. Bu çalışmada OECD ülkelerinin OECD raporlarında mevcut olan eğitim yatırım değişkenleri göz önüne alınmış ve bu eğitim yatırım değişkenlerinin eğitim çıktısı olarak ülkelerin PISA skorları arasındaki ilişkisine bakılmıştır. Eğitim değişkenleri, sınıf başı öğrenci sayısı, öğretmen başı öğrenci sayısı, öğrenci başı yıllık harcama, öğretim süresi ve öğretmen maaşları iken eğitim çıktısı olarak ise PISA matematik, okuma ve fen skorları kullanılmıştır.

Yapılan Pearson korelasyon analizine göre öğretim süresi hariç diğer tüm eğitim yatırım değişkenleri PISA skorlarıyla (matematik, okuma ve fen) anlamlı bir ilişki göstermektedir. Bu ilişkilerden öğrenci başı yıllık harcama (ÖBYH) ile öğretmen maaşı PISA skorlarının hepsiyle pozitif bir ilişki gösterirken sınıf başı öğrenci sayısı (SBÖS) ve öğretmen başı öğrenci sayısı (ÖBÖS) bu skorlarla negatif bir ilişki göstermektedir.

Bu sonuçlar doğrultusunda araştırma sorularını incelemek için bu eğitim değişkenleri kombinasyonlarından oluşan lineer regresyon modelleri oluşturulmuştur. Bu modellerde sadece anlamlı olan eğitim yatırımı değişkenleri yer almaktadır. Yapılan analizler sonucunda, PISA matematik, okuma ve fen skorlarını yordamaya yönelik olan her bir model anlamlı olarak bulunmuştur. Yatırım değişkenleri arasında PISA matematik, okuma ve fen skorlarını anlamlı olarak yordayan ortak değişken öğretmen maaşlarıdır. Öğretmen maaşına ek olarak SBÖS değişkeni de PISA matematik ve okuma skorlarını anlamlı olarak yordamaktadır. Bu sonuçlara göre öğretmen maaşının artması PISA'nın tüm skorlarını pozitif olarak etkilerden, SBÖS ise PISA matematik ve okuma skorlarına negatif bir etki oluşturmaktadır.

Ücret insanları motive eden unsurların başında gelmektedir. Öğretmenlerin öğretmenlik mesleğine ilişkin yapılan çalışmalarda ücret kavramı sıklıkla vurgulanmış ve maaşların öğretmenlerin niteliğini ve motivasyonunu etkilediği ifade edilmiştir (Çelikten, Şanal ve Yeni, 2005; Figlio, 1997). Ülkemizde de öğretmen maaşlarının Avrupa ülkelerin gerisinde kalmıştır (Tunçkaşık, 2007). Yine ülkemizin PISA sınavlarında geri kalması bu bilgi ile bağdaştırılabilir. Bu çalışmada elde edilen sonuçlar ışığında öğretmen maaşları ile başarı arasında bir ilişki görülmektedir. Yani bireyin performansı ve işini iyi yapması, ekonomik kazanımlardan önemli ölçüde etkilenmektedir (Erdem, 2010; Süngü, 2012; Sağlam ve Sağlam, 2005).

Yapılan birçok çalışmada sınıf mevcudu ile öğrenci başarısı arasında ilişki bulunmuştur. Özellikle sınıf büyüklüğünün düşük başarı gösteren öğrenciler, ilköğretim okul öğrencileri, düşük sosyo-ekonomik düzeye sahip öğrenciler için önemli ve anlamlı bir etkisi vardır (OECD, 2014b). Bu gruplar için küçük sınıfların daha etkin olduğu vurgulanmaktadır (Akarhielm, 1995). Sınıf mevcudunun 20-30 arasında olması tartışma ve eğitim açısından en uygun sayı olduğu belirtilmekte ve özellikle okuma başarısızlığının bir nedeni de sınıf büyüklüğü olduğu ifade edilmektedir (Hattie, 2006). Yine sınıf mevcudu ve öğrenci başarısı arasında düşük bir ilişki Hoxby (2000) tarafından ifade edilmiş ve matematik, okuma ve yazma becerisinde sınıf büyüklüğünün etkili olduğunu ama bu etkinin çok küçük olduğunu belirtmiştir.

Aslında eğitime yapılan yatırımların hemen hepsi doğrudan veya dolaylı olarak öğrencilere yapılmaktadır. Her şey, öğretmenlere, okulun fiziki yapısına hatta yönetmeliklerdeki değişiklikler bile öğrencilerin daha başarılı olmasını desteklemek içindir. Dolayısıyla eğitime yapılan yatırım öğrencinin başarısını da pozitif anlamda etkilemektedir. Sanders (1999) bu konuda yaptığı çalışmada öğrenci başı yapılan harcama ve ortalama öğretmen maaşlarının sekizinci sınıf matematik skorlarında olumlu anlamda az da olsa etkisi olduğunu ortaya koymuştur. Benzer bir bulguyu Hanushek (1997) ifade etmiş ve 400 civarı araştırmanın meta analiz sonucunda okul kaynakları ile öğrenci başarısı arasında tutarlı ve güçlü bir ilişki olmadığını belirtmiştir. Hanushek'e göre okula yapılan yatırımlardan önemli bir bileşeni de "öğrenci başı yapılan harcamalar"dır. Bu bileşen öğrencinin temel eğitim ihtiyaçları, yemeği, ulaşımı, barınma ihtiyacı, personel geliştirme ve eğitimin gelişimi için harcanan ücret dâhil çok geniş kapsamlı olarak tanımlanır (OECD, 2007). Bu çalışmada da Hanushek'e paralel olarak öğrenci başı yıllık harcama öğrencinin hiçbir skorunu anlamlı bir şekilde yordamamaktadır.

KAYNAKÇA

- Akarhielm, K. (1995). Does class size matter. *Economics of Education Review*, 14(3), 229-241.
- Çelikten, M., Şanal, M. ve Yeni, Y. (2005). Öğretmenlik mesleği ve özellikleri. *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 19(2), 207-237.
- Büyüköztürk, Ş., Çakmak, E.K., Akgün, Ö.E., Karadeniz, Ş. ve Demirel, F. (2015). *Bilimsel araştırma yöntemleri*. Ankara: Nobel Yay.
- Erdem, A. R. (2010). İlköğretim ve ortaöğretim öğretmenlerinin karşılaştığı ekonomik sorunlar ve bu ekonomik sorunların performanslarına etkisi konusundaki görüşleri. *Uluslararası İnsan Bilimleri Dergisi*, 7(1), 270-287.
- Figlio, D.N. (1997). Teacher salaries and teacher quality. *Economic Letters*, 55, 267-271.
- Gümüş, E. ve Şişman, M. (2012). *Eğitim ekonomisi ve planlaması*. Ankara: Pegem Akademi Yay.
- Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: an update. *Educational Evaluation and Policy Analysis*, 19(2), 141-164.
- Hattie, J. (2006). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43, 387-425.
- Hoxby, C. M. (2000). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics*, (2000a), 115(4), 1239-85.
- MEB (2015). *PISA nedir?* Alıntılanan web: http://pisa.meb.gov.tr/?page_id=18&lang=tr (alınma tarihi: 30/04/15; 23:00).
- MEB (2013). *Pisa 2012 ulusal ön raporu*. Ankara: Yenilik Ve Eğitim Teknolojileri Genel Müdürlüğü.
- MEB (2010). *Pisa 2006 projesi ulusal nihai rapor*. Ankara: Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- MEB (2005). *Pisa 2003 projesi ulusal nihai rapor*. Ankara: Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- OECD. (2004). *Education at a Glance 2004: OECD Indicators*. OECD Publishing
- OECD. (2005). *Education at a Glance 2005: OECD Indicators*. OECD Publishing
- OECD. (2006). *Education at a Glance 2006: OECD Indicators*. OECD Publishing
- OECD. (2007). *Education at a Glance 2007: OECD Indicators*. OECD Publishing
- OECD. (2008). *Education at a Glance 2008: OECD Indicators*. OECD Publishing
- OECD. (2009). *Education at a Glance 2009: OECD Indicators*. OECD Publishing
- OECD. (2010). *Education at a Glance 2010: OECD Indicators*. OECD Publishing
- OECD. (2011). *Education at a Glance 2011: OECD Indicators*. OECD Publishing
- OECD. (2012). *Education at a Glance 2012: OECD Indicators*. OECD Publishing
- OECD. (2013a). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II)*, PISA. OECD Publishing.
- OECD. (2013b). *Education at a Glance 2013: OECD Indicators*. OECD Publishing
- OECD (2014a). *PISA in Focus*. 2014/04 (April).
- OECD. (2014b). Strengthening Resilience through Education and Skills: PISA Results.
- OECD. (2014c). <http://www.OECD.org/pisa/aboutpisa/> (alınma tarihi: 30/04/15; 23:00).
- Özdemir, S. (2013). *Eğitim Yönetiminde Kuram ve Uygulama*. Ankara: Pegem Akademi.
- Sağlam, M. ve Sağlam, A. Ç. (2005). Öğretmenlik mesleğinin maddi yönüne ilişkin genel bir değerlendirme. *Türk Eğitim Bilimleri Dergisi*, 3(3), 317-328.

- Sander, W. (1999). Endogenous expenditures and student achievement. *Economics letters*, 64(2), 223-231.
- Süngü, H. (2012). Bazı OECD Ülkelerindeki Öğretmen Maaşlarının Karşılaştırmalı Bir Analizi. *21. Yüzyılda Eğitim ve Toplum*, 1(2), 21-45.
- Tunçkaşık, H. (2007). *Türkiye’de ve çeşitli ülkelerde öğretmen maaşları*. Ankara: TBMM Araştırma merkezi.

EXTENDED ABSTRACT

1. Introduction

Education is an investment to contribute individuals and society. Investment in education pays off in the long run. It plays an important role in educational output and countries’ social and economic development. Therefore, developed and developing countries continuously increase their education expenses and investments. It is referred by economists as investment and consumption. It is considered as consumption when it provides satisfaction in a particular period of time, and as investment when the continuous satisfaction is provided in a relatively long period of time. In accordance with this argument, education helps individuals improve their abilities and skills (i.e. foreign language skills and mathematics skills), which can have significant impact on their future career. Therefore, although education is considered as consumption and investment, the latter is more preferable (Gümüş and Şişman, 2012:55).

OECD collects the educational variables from the member countries and prepares official reports to compare the members’ performance. Such reports are considered to help developing and developed countries to identify strengths and weaknesses of their education systems by comparing their education system with those implemented in other countries. The Program for International Student Achievement (PISA) is a well-known international assessment which provides such countries with the opportunity to monitor outcomes of their education system.

It is conducted by OECD every three years with the aim of assessing 15-year-old students’ competence in mathematics, science, and reading literacy, which are the subjects offered during the last year of their compulsory education in most OECD countries (OECD, 2006). Thereby, it is specifically aimed to identify students’ characteristics and their educational background to develop future projects and investment in education. Accordingly, their interests, skills essential to adult life, academic motivation level, learning styles as well as characteristics of their school and family structures (OECD, 2014a).

PISA was started to be implemented in 43 countries in 2000. The number of countries participating in the assessment have increased over the past fifteen years (41 countries in 2003, 57 countries in 2006, 65 countries in 2009 and 2012). A total of 71 countries were signed up to participate in it for 2015.

It is a unique type of assessment in the sense that it is not directly created based on a particular school curriculum. It is developed to measure to what extent the goals of compulsory education have been achieved, and to what extent the students apply their knowledge in real-life situations. Therefore, the results of PISA are used not only for displaying the performance of countries across years but also leading educational policy-

makers in each country to develop policies which meets the educational needs of their own countries.

The program contains open-ended, closed-ended, and multiple-choice questions, which are designed taking real-life contexts into consideration. Students are allowed two hours to complete the test consisting of reading, mathematics, and science, and one hour to complete a questionnaire about their educational background, preparedness for higher education, learning experiences and environments. In addition to students, the principals of participating schools are supposed to complete a questionnaire about demographics characteristics and learning environment of their schools’.

The results of PISA are reported by OECD at the end of each circle. These reports provide considerable variables highly associated with educational investments, such as teaching time, teacher salary, annual expenditure on educational institutions per student, and number of students per teaching staff (OECD, 2014a, 2013, 2012).

Purpose: Each individual country pays attention to education throughout its history and makes remarkable educational investments. They hope, through education, to adapt to developments around the world such as new technologies and information systems. This enables them to compete with other countries in various platforms, such as science and technology.

The purpose of this study is to explore whether/to what extent PISA scores are predicted by the educational variables identified in OECD reports. The research questions are as follows:

- 1- Are PISA mathematics scores predicted by educational variables? If so, to what extent?
- 2- Are PISA reading scores predicted by educational variables? If so, to what extent?
- 3- Are PISA science scores predicted by educational variables? If so, to what extent?
- 4- Do PISA scores and the proportion of educational variables change over the time? If so, how?

2. Method

The current study is correlational in design due to the fact that it addresses causal relationship between two or more variables. It analyzes the relationship between PISA scores and educational variables of OECD countries. Teaching time, teacher salary, annual expenditure on educational institutions per student (AEPS), number of students per a teaching staff (SPTS), and estimated class size were set as independent variables, and PISA scores of countries in mathematics, reading, and science scores were set as dependent variables. Data were collected from the reports of 26 out of around 40 countries of which scores and educational variables are available in PISA reports from 2005 to 2014. Since their investment in education influence their educational outcomes of the following year (i.e. PISA scores), their educational investment per year was taken into consideration before PISA examination is held. To illustrate, a particular country’s educational investment in 2005 might have influenced its PISA scores in 2006. To predict PISA scores, multiple linear regression method was used in data analysis. Only educational variables which have a statistical significant correlation with PISA scores

were involved in the regression model. For example, the model used in predicting mathematics scores did not involve “teaching time” as an independent variable since the correlation between mathematics scores and teaching time was not found statistically significant.

In addition, one-way analysis of variance (ANOVA) was conducted to test whether PISA scores and the size of educational investment variables change over the time.

3. Findings, Discussion and Results

Change in Scores and the size of educational investment over the time

The results of one-way ANOVA indicated that mean of countries’ PISA scores in mathematics, reading, and science did not statistically change over time (Mathematics: $\mu_{2006}=492,6$, $\mu_{2009}=495,0$, $\mu_{2012}=493,9$, $F=.04$, $p=.96$; Reading: $\mu_{2006}=487,1$, $\mu_{2009}=493,0$, $\mu_{2012}=500,3$, $F=1.33$, $p=.27$; Science: $\mu_{2006}=492,4$, $\mu_{2009}=501,3$, $\mu_{2012}=501,8$, $F=.71$, $p=.49$).

ANOVA results also showed that the change in the size of educational investments across countries was not statistically significant (Class size: $\mu_{2005}=17,6$, $\mu_{2008}=16,3$, $\mu_{2011}=15,9$, $F=.64$, $p=.53$; SPTS: $\mu_{2005}=16,1$, $\mu_{2008}=15,8$, $\mu_{2011}=15,2$, $F=.21$, $p=.81$; AEPS (\$): $\mu_{2005}=7,732.8$, $\mu_{2008}=8236,6$, $\mu_{2011}=8,494.3$, $F=.37$, $p=.69$; Teacher Salary (\$): $\mu_{2005}=37,409.7$, $\mu_{2008}=37,770.5$, $\mu_{2011}=38,504.5$, $F=.03$, $p=.97$; Teaching Time (hours): $\mu_{2005}=762,1$, $\mu_{2008}=767,6$, $\mu_{2011}=787,6$, $F=.30$, $p=.74$).

Multiple Regression for Mathematics Scores

Prior to implementation, multiple regression model was tested to see whether the relationship between educational variables and PISA mathematics score was significant. It was found that the only correlation between teaching time and mathematics scores was not statistically significant at 0.05 level ($r=-.14$, $p=.22$). Based on these results, class size, AEPS, SPTS, and teacher salary were involved in the model as the predictors.

The findings of multiple regression analysis indicated that the model was statistically significant, $R^2=.26$, $F(5,67)=6.06$, $p<.001$. Approximately 26% of the variance of PISA mathematics scores in the sample can be accounted for by the linear combination of educational variables. In the model, only effects of teacher salary and class size on mathematics were statistically significant ($p<.05$). Mathematics scores were positively correlated with teacher salary, and negatively correlated with class size.

Multiple Regression for Reading Scores

It was measured whether every educational variable was significantly correlated with reading scores before conducting multiple regression model to predict reading scores. The findings pointed out that the correlation between reading scores and teaching time was not statistically significant ($r=-.01$, $p=.93$) while reading scores were significantly correlated with other variables ($p<.05$), which were involved in multiple regression model.

The results of multiple regression analysis showed that the model was statistically significant, $R^2=.25$, $F(5,67)=5.61$, $p<.001$. Approximately 25% of the variance of PISA reading scores in the sample could be explained by the linear combination of educational variables. In the model, reading scores were significantly positively correlated with

teacher salary whereas they were significantly negatively correlated with class size ($p < .05$). However, no significant correlation was found between reading scores and AEPS and SPTS.

Multiple Regression for Science Scores

Before multiple regression model conducted, it was tested whether the relationship between educational variables and PISA science scores was significant. The findings showed that only teaching time was not statistically correlated with science scores ($r = -.18$, $p = .11$). So, regression model did not involve teaching time as a predictor.

Based on the results of multiple regression analysis, the model was found statistically significant, $R^2 = .20$, $F(5,67) = 4.15$, $p < .001$. Approximately 20% of the variances of PISA science scores in the sample could be explained by the linear combination of educational variables. In the model, science scores were not significantly correlated with the educational variables but teacher salary ($p < .05$). Based on the results, teaching salaries have a positive impact on the average of PISA science scores.

Discussion and Results. Education is an investment which remarkably contributes to the improvement of individual's life and society. To this end, developed and developing countries pay too much attention to educational investments. The current study addressed the relationships between the educational investment variables of OECD countries available in OECD reports and PISA scores as an outcome of education. These variables were teaching time, teacher salary, annual expenditure on educational institutions per student (AEPS), number of students per teaching staff (SPTS), and estimated class size.

The findings of the study indicated that all educational variables were significantly correlated with PISA scores except teaching time. So, teaching time was excluded from the linear regression model to predict PISA mathematics, reading, and science scores. It was revealed that each regression model was statistically significant. The findings indicated that only teacher salary and class size were significantly correlated with mathematics and reading scores. Teacher salary has a positive impact on students' performance while class size negatively influenced it. They also pointed out that teacher salary was the only significant predictor, which is positively associated with the science scores.

Existing studies indicated that teacher salary plays an important role in increasing their motivation and recruiting higher quality teachers (Çelikten, Şanal and Yeni, 2005; Figlio, 1997), stressing that it is significantly associated with students' performance. This shows that teachers' effectiveness and performance are mostly influenced by their income (Erdem, 2010; Süngü, 2012; Sağlam and Sağlam, 2005).

A large body of literature reported strong relationship between class size and students' performance. Particularly, it is considered more beneficial for low achievers, students from lower socio-economic status, and elementary school students to be educated in relatively small size classes (Akarhielm, 1995). Classes of 20 and 30 students are regarded acceptable to create more effective environment by facilitating classroom discussion and group work. A recent study investigated by Hattie (2006) suggested that early reading difficulties are mainly attributed to the large class size. Another study by Hoxby (2000) reported that the class size has a slight impact on students' mathematics, reading, and writing scores.

In fact, investments in education can directly and indirectly contribute to students' academic achievement. For instance, any investment in teachers, physical environment of the schools and educational directions is likely to have positive impacts on students' performance. AEPS and teacher salary has a slight but statistically significant impact on 8th grade students' achievement in mathematics (Sanders, 1999). Hanushek (1997) found not a statistically significant relationship between school source and students' performance after conducting Meta-Analysis of approximately 400 studies. Since annual expenditure on educational institutions per student (AEPS) includes educational core services, meals, transportations, housing as well as teaching and personal training services (OECD, 2007), AEPS did not reveal a statistically significant predictor of students' achievement in mathematics, reading, and science in the current study.