



Hacettepe Üniversitesi Sosyal bilimler Enstitüsü Eğitim Bilimleri
Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**DERECELİ PUANLAMA ANAHTARI VE PUANLAMA
ANAHTARINDAN ELDE EDİLEN PUANLARIN
KARŞILAŞTIRILMASI**

Yusuf Kasap

Yüksek Lisans Tezi

Ankara, 2008

**DERECELİ PUANLAMA ANAHTARI VE PUANLAMA
ANAHTARINDAN ELDE EDİLEN PUANLARIN
KARŞILAŞTIRILMASI**

Yusuf Kasap

**Hacettepe Üniversitesi Sosyal bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı**

Yüksek Lisans Tezi

Ankara, 2008

KABUL VE ONAY

Yusuf KASAP tarafından hazırlanan "Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Karşılaştırılması" başlıklı bu çalışma 02.06.2008 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.



Doç. Dr. Selahattin GELBAL (Başkan, Danışman)

Doç. Dr. Şener BUYUKOZTURK

Öğr. Gör. Dr. Nuri DOĞAN



Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. İrfan ÇAKIN

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- ☒ Tezimin tamamı her yerden erişime açılabilir.
- ☐ Tezim sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- ☐ Tezimin.....yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

02.06.2008

Yusuf KASAP



ÖZET

YUSUF, KASAP. Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Karşılaştırılması, Yüksek Lisans Tezi, Ankara, 2008

Bu araştırmada, puanlama anahtarı ile son yıllarda yaygın olarak kullanılan dereceli puanlama anahtarının karşılaştırılması amaçlanmıştır.

Bu amaçla, araştırma bu alandaki geleneksel değerlendirme yaklaşımı olarak puanlama anahtarıyla değerlendirme, yeni değerlendirme yaklaşımı olarak dereceli puanlama anahtarıyla değerlendirme, öz değerlendirme, akran değerlendirme ve anketle sınırlandırılmıştır. Dereceli puanlama anahtarıyla öğrencilerin başarılarının değerlendirilmesi durumu ve puanlama anahtarıyla öğrencilerin başarılarını değerlendirilmesi durumu karşılaştırılmıştır. Bu karşılaştırma yapılırken, öğrencilerin akademik başarıları dikkate alınmıştır. Kullanılan puanlama anahtarı ve dereceli puanlama anahtarından elde edilen puanlar arası tutarlılık, duyarlılık ve uyuma bakılarak, bu iki yöntemin güvenilirlikleri karşılaştırılmıştır. Ayrıca, bazı anket sorularıyla öğrencilerden öz değerlendirme, akran değerlendirme ve öğretmen değerlendirmesiyle ilgili düşüncelerini yazmaları istenmiştir. Araştırmanın bulgularına göre, puanlama anahtarıyla değerlendirme sonuçlarıyla dereceli puanlama anahtarından elde edilen değerlendirme sonuçları arasında anlamlı fark vardır. Buna karşın, akran değerlendirme sonuçlarıyla öz değerlendirme sonuçları arasında anlamlı fark bulunmamıştır. İkili karşılaştırılan değerlendirmelerden elde edilen puanlar anlamlı ölçüde farklılaşmaktadır. Araştırmanın bulgularından diğeri ise; puanlama anahtarı ve dereceli puanlama anahtarıyla elde edilen puanların diğer puanlama tekniklerinden elde edilen puanların ortalamasıyla korelasyonlarının hesaplanması sonucunda dereceli puanlama anahtarından elde edilen korelasyon biraz yüksek çıkmıştır. Son bulgu olarak da bireylere uygulanan anketler sonucunda kullanılan dereceli puanlama anahtarındaki ölçütlerin öğrenciler tarafından bilinmesi öz değerlendirme, akran değerlendirme yaparak sorumluluk almalarını sağlamakla birlikte, öğrenciler sınavda sorulan soruları cevaplarken sorudaki cevaplama ölçütlerini bilmelerinden kaynaklı olarak soruyu daha doğru algıladıklarını ve cevapladıklarını belirtmektedirler.

Anahtar Sözcükler

Puanlama Anahtarı, Dereceli puanlama anahtarı (Rubric), Öz Değerlendirme, Akran Değerlendirme, Anket

ABSTRACT

KASAP, YUSUF. Comparison of the scores emerged by scoring rubric and answer key master's thesis, Ankara, 2008.

In this research, it is purposed to compare the scoring scale and rubric which is widely used in recent years.

Therefore, this research is limited to assessment by answer key as a traditional assessment in this field, assessment by scoring rubric as a new assessment approach, self-assessment, peer assessment and survey. The assessment (state) of the achievement of students by scoring rubric and the assessment (state) of the achievement of students by answer key are compared. Academic achievements of students are taken into consideration when doing this comparison. The reliability of these two methods is compared considering the coherence, sensitiveness and harmony between the scores emerged by scoring rubric and answer key used. Besides, the students are required to express their thoughts about self-assessment, peer assessment and teacher assessment through a poll taken. According to the results of the research, there is a meaningful difference between the assessment results by scoring rubric and answer key. On the other hand, a meaningful difference is not found between the results of peer assessment and self assessment. The scores emerged by double compared assessments differentiate meaningfully. Another result of the research is, the average of the scores emerged by different scoring methods are calculated. The average of the scores and the scores emerged by each method are compared one by one. At the end of the comparison, the correlations are calculated. The correlation emerged by rubric is a bit higher than the scoring scale. As a last result, if the measurements used in scoring rubric are known by the students, in addition to taking responsibility by making self-assessment and peer assessment, they state that they can also perceive and answer the questions in the exams much easily.

Key Words

Scoring Rubric, Answer Key, Self-Assessment, Peer Assessment, Survey

İÇİNDEKİLER

	Sayfa
TEŞEKKÜR.....	i
ÖZET.....	ii
ABSTRACT.....	iv
İÇİNDEKİLER.....	v
TABLolar.....	viii

BÖLÜM I

GİRİŞ.....	1
Problem Durumu.....	1
Psikometrik Paradigma.....	4
İçeriksel(contextualized) Paradigma.....	8
Bireyselleştirilmiş(personalized) Paradigma.....	10
Paradigmaların Karşılaştırılması.....	14
Ölçme Araçlarında Aranan Başlıca Nitelikler.....	16
Bireyselleştirilmiş Paradigmaya Dayalı Olarak Yapılan Performans Değerlendirme.....	18
İçeriksel Paradigmaya Dayalı Olarak Yapılan Performans Değerlendirme.....	27
Problem Cümlesi.....	32
Alt Problemler.....	32
Sayıtlar.....	33
Sınırlılıklar.....	33
Tanımlar ve Kısaltmalar.....	34
Araştırmanın Amacı ve Önemi.....	34
İlgili Araştırmalar.....	36

BÖLÜM II

YÖNTEM.....	43
Araştırmanın Türü.....	43
Çalışma Grubu.....	43
Veri Toplama Araçları.....	43
Verilerin Toplanması.....	46
Verilerin Çözümlemesi.....	48

BÖLÜM III

BULGULAR VE YORUMLAR.....	50
1.Alt Problem.....	50
2.Alt Problem.....	53
3.Alt Problem.....	56
4.Alt Problem.....	57
5.Alt Problem.....	59
6.Alt Problem.....	60

BÖLÜM IV

SONUÇ VE ÖNERİLER.....	64
Sonuçlar.....	64
Öneriler.....	66
KAYNAKÇA.....	68
EKLER.....	74
Ek1: Yazılı Sınav soruları.....	74
Ek2: Puanlama Anahtarı.....	76
Ek3: Dereceli Puanlama Anahtarı(rubric).....	78

Ek4: Anket Soruları.....	80
ÖZGEÇMİŞ.....	81

TABLOLAR

	Sayfa
Tablo 1: Yeni Değerlendirme Sistemleri için Bakış Açıları ve Örnek Durumlar.....	13
Tablo 2: Vermont'un Analitik Değerlendirme Rehberinde Belirlenen Standartlar ve Ölçütler.....	20
Tablo 3: Performans Dereceleme Ölçeği Geliştirmede İzlenecek Adımlar.....	25
Tablo 4: Geçerlilik Türlerine Uygun Kanıt Toplamak İçin Sorulabilecek Sorular.....	26
Tablo 5: Çalışma grubuna alınan şubeler ve bu şubelerde uygulanan değerlendirmelere ve ankete katılan öğrenci sayıları.....	47
Tablo 6: Puanlama anahtarı ve dereceli puanlama anahtarından elde edilen sonuçların Karşılaştırılması ve bu araçların kullanımı sonucu elde edilen betimsel istatistikler.....	50
Tablo 7: Farklı puanlama yöntemleriyle puanlama sonucunda elde edilen puanların ortalamaları arasında anlamlı bir farklılık olup oluşmadığını belirleme.....	51
Tablo 8: Uygulanan Farklı Puanlama Yöntemleri İçin İkili Karşılaştırmalar.....	52
Tablo 9: Değerlendirme Yöntemleri Arasındaki Korelasyonlar.....	53
Tablo 10: Cronbach Alfa (α) güvenirliklerinin incelenmesi.....	56
Tablo 11: puanlayıcı güvenirliklerinin incelenmesi.....	58
Tablo 12: Farklı puanlama yöntemlerinin ortalamalarla korelasyonunun incelenmesi...	59

BÖLÜM I

GİRİŞ

Problem Durumu

Dünyada 1975–2000 yılları arasında toplumsal, sosyo-kültürel, ekonomik, politik ve teknolojik alanda birçok sarsıcı değişiklikler olmuştur. Bu alanlardaki sarsıcı değişimler toplumsal hayatı geri dönülmez bir sürece sokmuştur. Bu sürecin çok boyutlu ve çok geniş çaplı olduğu görülmektedir. Bu değişimler toplumların hayata bakış açılarını, değerlerini, maddi yaşamdan beklentilerini köklü bir şekilde etkilemektedir.

Toplumsal hayattaki bu değişimler toplumu oluşturan yapılardaki değişimi de zorunlu kılmıştır. Bu yapılar kendi içinde bu hızlı değişimle meydana gelen çağın sorunlarına çözüm bulma adına, evrimleşmeye gitmiştir. Bu yapılardan biri olan eğitim, bu toplumsal değişim dinamiklerini toplum hayatına adapte etme ya da toplumu bu değişimlere adapte etme adına en önemli alanlardan biri olarak günümüzde kabul görmektedir.

İnsanlığın yaşamını eğitim yoluyla geliştirdiği kabul edilen evrensel bir gerçektir. Eğitim insanın bireysel, çevresel ve sosyal yönlerden başarıya ulaşmasında; barış, özgürlük, sosyal, adalet ve evrensel bütünlük ideallerine erişmesinde temel araçtır. Ayrıca eğitim; toplumsal ve ekonomik kalkınmanın da itici bir gücü olarak tüm sektörleri etkilemektedir (Arslan ve Eraslan, 2003).

İçinde bulunduğumuz dönüşüm evresinde çağın gereklerine ve beklentilerine cevap verecek bireyler yetiştirmek için doğru adımlar atılması ve kaynağı insan olan bir yapıda hatalardan mümkün olduğunca arınık, güvenilir sınamalar bütünü oluşturulması gerektiği açıktır. Bilgi toplumu olgusu hayatımıza hızla girerken,

bireylerden beklenen mesleki, teknik, eğitsel bilgi ve becerileri de kökten değiştirmiştir (Şimşek, 1997; Aşkar, 2004).

Bireyin nasıl eğitileceği, gerçekleştirilecek eğitim durumlarının istendik ve beklendik davranışları kazandırmadaki etkililik derecelerinin araştırılması yani değerlendirmesi eğitimin önemli bir ögesini oluşturmaktadır. Bireylerin nitelikli eğitim ürünleri ortaya koyma süreçlerinin ve bireylerin ortaya koydukları ürünlerin ve performanslarının değerlendirmesinde seçilecek yaklaşımların yerinde olup olmamasının bireyin başarısını etkileyeceği açıktır (Ertürk, 1998).

Eğitim, yaygın bir biçimde “insanın kişiliğini besleme süreci” ve “insan sermayesine yapılan yatırım” olarak kabul edilmektedir. En genel anlamda “istendik davranış oluşturma ya da istendik davranış oluşturma süreci” olarak tanımlanan eğitim, toplumun süzgeçten geçirilmiş değerlerinin, ahlak standartlarının bilgi ve beceri birikimlerinin yeni nesillere aktarılmasıyla ilgilidir. Bu anlamda eğitim “Bireyi istendik biçimde kültürleme sürecidir.”(Senemoğlu, 2004, s86).

Bireyin başarısını çok yönlü ve değişik açılardan ölçen yeni paradigmlar, bireyin kendi yaşantısı yoluyla davranışlarında meydana getirilen değişiklik sonucunda ortaya çıkan ürünün niteliğini yeni paradigmların anlayışına uygun bir yaklaşımla değerlendirmek gerektiğini belirtir. Böylece ölçmenin farklı yaklaşımlarla ve çok boyutlu olarak yapılması, elde edilen verilerin daha zengin oluşu çeşitli tür ve biçimlerde değerlendirme yapmayı olanaklı hale getirirken bireylerin başarı, yetenek, tutum, zeka vb.’nin ölçme ve değerlendirmesinde hatayı minimize ettiği düşünülmektedir. Bu iddialar güvenilirlik ve geçerlik kapsamı içinde değerlendirilecek araştırma alanlarına girmektedir.

Belli bir zaman dilimi içinde bir grubun ya da topluluğun düşünme biçimi ve davranışlarını belirleyen bir dünya görüşü, bir algı dayanağı, bir izlenceler bütünü, bir perspektif ya da modele paradigma denir (Şimşek, 1997). Bazı uygulamaların uygun, bazı uygulamalarında uygun olmadığına karar verilmesinde paradigma neyin doğru olduğu ile ilgili bir anlayış veya genel bir düşüncedir (Marby, 1999).

Belli bir alandaki bilimsel etkinlikleri, bilim insanlarının dünyayı nasıl yorumladıklarını ve nelerin kabul edilebilir ya da nelerin kabul edilemez olduğunu bu düşünce ya da düşünceler belirler (Şimşek, 1997).

Tabii olduğumuz değerlendirme yaklaşımı öğrencilerin öğrendiklerini anlamaya çalışırken yapmaya değer olduğuna inandığımız etkinlikleri ortaya çıkarır. Farklı değerlendirme yaklaşımlarını anlamak, öğrencilerin neyi bildiğini ve neyi yapabileceklerini tespit etmeye yönelik girişimimiz konusunda dikkatli tercihler yapmamıza yardım edebilir (Marby, 1999).

Yaklaşımlar arasındaki uyumsuzluklar ya da değerlendirme sistemlerindeki kavram karmaşası ve bireylerde uyandırdığı soru işareti, seçilecek yaklaşımdan elde edilecek yararı istemeden de olsa kısıtlayacak ve özgün değerlendirme sisteminin amaçlarına bütünüyle ulaşmasına engel olacaktır. Bu sebeplerden dolayı değerlendirme sisteminin yaklaşımları arasındaki farklılıklar kavranırsa, doğru ve zamanında değerlendirme yaklaşımları kullanılır. Bütün yönleriyle eksiksiz bir değerlendirme sistemi geliştirilirse, değerlendirme ve öğretim arasında bir bağ kurulabilir. Bunun nedeni değerlendirme ile eğitim arasındaki uyumun önemidir. Değerlendirmelere, eğitim süreci kaynaklık etmelidir. Yani öğrenmelerin saptanması değerlendirme faaliyetlerinin en temel kaynağıdır. Değerlendirme öğrenmeleri belirleme konusunda yetersiz kaldığı durumda değerlendirme faaliyetleri geçerliklerini ve güvenilirliklerini kaybederler.

Marby (1999) eğitimde ölçme literatüründe, değerlendirme ile ilgili üç değişik paradigma bulunduğunu belirtmektedir. Bunlar:

1. Psikometrik Paradigma
2. İçeriksel Paradigma
3. Bireyselleştirilmiş Paradigma

Psikometrik Paradigma

Bu yüzyılda, dışsal ölçümler kademeli olarak eğitimsel değerlendirmeleri değişik eğitimsel kararlar için temel olarak aldığından, zeka, yetenek ve başarı ölçümü, belli bir yaklaşım şeklini almış ve bu yaklaşıma psikometrik paradigma denilmiştir. Süreçten çok ürünle ilgilidir. çünkü ürün gözlenebilir (MEB, 2003). Bu tür ölçmenin temel yöntemi, standardize edilmiş çoktan seçenekli testlerdir (Mabry, 1999). Çoktan seçmeli testlerin kullanışlılık yönünden dezavantajları bilişsel çaba düzeyine bağlı olarak, daha fazla zaman harcanması ve daha zor hazırlanmasıdır. Diğer yandan, çoktan seçmeli sorular düşük bilimsel çaba düzeyini talep eden, önceden ezberlenmiş bilgiyi hatırlama düzeyinde soruları test etmede etkili olarak kullanılırken, diğer yandan, öğrencilerin analiz etme ve sentezleme gibi daha yüksek türde düşünme becerilerini ölçen soruları test etmede etkili olarak kullanmak daha zordur.

Benzer olarak aşağıdaki bakımlardan dolayı çoktan seçmeli testlere eleştiriler getirilmektedir (Dikli, 2003). Bu eleştiriler:

- 1-Sadece fark etme bilgisini ölçmesi
- 2-Şans başarısı olması
- 3-Ayırt ediciliği yüksek sorular yazmanın zorluğu
- 4-Dönüt vermenin zorluğu
- 5-Kopya çekmeyi kolaylaştırması

Farklı bir bakış açısı da çoktan seçmeli test maddelerinin, öğrenci başarısını gerçek dünyadan yalıtılmış biçimde, sınırlandırılmış zamanlarda ve sınırlı bir şekilde ölçtükleri, öğrencinin gerçek dünyayla ilişkilendirilmiş problemleri çözmeleri için yeterli olmadıkları yönünde ifade edilmiştir (MEB, 2003). Bu testler öğrencilere asıl öğretilmekte olanla çok yakından ilişkili değildir. Aynı zamanda zorluklar yaşayan öğrencileri belirlerken asıl problemin doğası hakkında çok az bilgi veririler. Bu sınavların amacı, çocukları birbirleriyle ve çoğunlukla yıllar öncesinden geliştirilmiş bazı sınavları almış olan çocuklardan oluşan örneklemelerle karşılaştırmaktır (Conner 1991'den Akt: İlhan, 2006).

Psikometrik paradigmada öğrenci başarısının anlaşılmasındaki temel strateji karşılaştırmadır. Öğrenci performansı, ya önceden belirlenmiş bir standart ile ya da

puan kesintisi veya diğer testi alanların performansları ile karşılaştırılır. Norm referanslı teste (NRT), öğrenci performansı, tipik olarak bir puan olarak değerlendirilir ve yaşlılarının olacağı varsayılan ile karşılaştırılır. Test soruları veya maddeleri, öğrencilerin neleri bildiklerini göstermek için öğrenci performansının maksimize edilmesi amacı ile seçilir. Puanlar büyüklüklerine göre sıraya konulur. Öğrenci başarısı, öğrencinin sırası ile belirlenir ve genellikle bu sıralama, öğrencilerin özel programlara veya özel okullara kabul için ayıklama veya seçme için kullanılır. Mabry (1999)'dan alınarak aşağıda norm dayanaklı değerlendirme özetlenmiştir.

Norma dayalı değerlendirmenin dayandığı temel denencelerden biri de öğrencinin başarılarının normal bir eğri ile tanımlanabilmesidir. Öğrencilerin birçoğunun başarı düzeyi olarak orta ranj ölçüsünde, az miktarda öğrencinin en yüksek ve en düşük puanlarda toplandığı görülür. Ama başarının her zaman merkeze eğilim göstermesi şart değildir (Mabry, 1999).

Norma dayalı değerlendirme eğitimsel sınırlılıklar içermektedir. Bu değerlendirme yaklaşımı öğrencilerin devamlı şekilde birbirleriyle rekabet etmelerini özendirir. Takım rekabeti ve karşılıklı etkileşimle öğrenme gibi öğrenme biçimlerinin üzerinde durmamaktadır. Bu yönden norma dayalı değerlendirme bağıl değerlendirmedir. Bununla birlikte norma dayalı sınavlarla elde edilen puanlar doğal olarak öğrencilerin yarı yarıya ortalamanın altında kalması ve standart puanların her öğrencinin puanlarının ortalama ile karşılaştırarak elde edilmesi bazı öğrencilerin hayal kırıklığına uğramasına neden olmaktadır. Öğrencilerin akranları arasında nasıl bir performansının olduğunun bilinmesi yararlı olsa da her bir öğrencinin grup içerisindeki yerlerini bilmek öğrencinin ne kadar öğrendiğiyle ilgili çok az bilgi verecektir. Norma dayalı sınavların öğrencilerin hangi tür öğrenmeleri ne oranda gerçekleştirdiği hakkında detaylı bilgi vermede eksikleri vardır.

Ölçüt referanslı değerlendirme, adayın beklenen iş performansı için bilgili ve becerili olup olmadığını ya da eğitimi ya da eğitim programını başarıyla tamamlama ile ilgili bilgi verir. Bu bilgi ve beceriler net bir iş analizi ve görev analizi ya da diğer

prosedürler göz önüne alınarak net bir şekilde tanımlanmalı ayrıca ölçütler ve standartlar belirlenmelidir. Ölçüt referanslı değerlendirme, mutlak değerlendirme olduğundan dolayı performans testindeki geçme notunu belirler. Bu notun altındakiler başarısız puanlardır. Bu da minimum yeterliliği belirleyen ölçüt referanslı teste yaklaşımın özüdür (Priestley, 1982).

Ölçüt referanslı değerlendirme, NRT’lerde karşılaşılan yaşıt karşılaştırma sorunlarından kaçınmak için bir dereceye kadar NRT’lerden ağırlıklı olarak etkilenme eğilimindedir. Yani, başarılı performansı başarısız performanstan ayıran ana çizgi genellikle neyin başarılı performans olduğunun değerlendirilmesinden ziyade öğrencilerin fiilen nasıl bir performans gösterdikleri esasına göre belirlenir. Dolayısıyla, öğrencinin minimum yeterlilik veya yetkinlik göstermek için ne yapması gerektiği ile ilgili içerik uzmanlarının değerlendirmesinde diğer öğrenci performansları hala “normu” bulmak için veya belirli bir performans düzeyine ilişkin puan kestirimini belirlemek için karşılaştırılmaktadır. Ölçütler vardır ama teste giren öğrencinin ölçütü karşılayıp karşılamadığı diğer sınav alanlarının nasıl bir performans gösterdiklerine göre tespit edilir (Mabry, 1999).

Norma dayalı değerlendirmelerde, öğrencilerin sınavdan elde ettikleri puanlar sıralanmakta, başarı puanları en yüksek olandan en düşük olana doğru çeşitli puanlar atanarak bireyler hakkında bir yargıya varılmaktadır. Ölçüt dayanaklı değerlendirmelerde ise sıralama önemli değildir. Sınavda başarıları belli bir puanın üzerinde olan bireyler için yeterli, altında olanlar için ise yeterli değil şeklinde bir yargıya varılır. Her iki değerlendirme yaklaşımında da bir takım sınırlılıklar bulunmaktadır. Norm dayanaklı değerlendirmelerde ölçüt, ölçüt dayanaklı değerlendirmelerde de karşılaştırma öğelerinin bulunmayışı, bu değerlendirme yaklaşımlarının yeni paradigmalara uyum sağlamakta yetersiz kaldığını göstermektedir (Mabry, 1999).

Psikolojik paradigma o kadar aşına hale gelmiştir ki, genel olarak zeka veya başarının ölçülmesinde tek yaklaşım olarak, tek geçerli, güvenilir, adil ölçme yaklaşımı olarak kabul edilmektedir. Ancak teknik olarak kusursuz olduklarında

yasal olarak doğrulanabilirler ve bazıları, gelecekteki performansların iyi bir tahmini olduğunda bile, standardize testlerin değerlendirme tekniği olarak yetersiz olduğunu söylemektedirler. Seçme amaçları için yasal olarak kabul edilebilir ama öğrencinin kendine özgü başarısının doğru yansıtılmasını sağlamaları zordur. Eğitim, Robert Wood'un ileri sürdüğü gibi, başarı sırasına koyup seçen farklılık psikolojisi için geliştirilen değerlendirmeye karşı çıkamamıştır (Mabry, 1999).

Eğitim, psikometrik saldırı ve etkiye karşı her zaman kırılgan olmuştur. Okullarda sınav amacı ve hangi testlerin uygun hangilerinin ise uygun olmadığıyla ilgili belirleyici ve özgüvene sahip görüş eksikliği, daha ziyade müşteri durumu gibi, psikometrik doktrin ve uygulamalar kuruldukça umutsuzca bakmıştır. Ama eğitim ve farklılık psikolojisi aynı amaçlara sahip değildir. Robert Stake'in kaydettiği gibi; Mabry (1999) eğitim değerlendirmelerinden olan psikolojik değerlendirme; öğrencinin gerçek başarısının genişliğini, derinliğini, ayrıntısını ve özelliğini olması gerektiğinin altında gösterir. Öğrencinin bilgisinin yeterliliğini öğrencilerin sıralanmasıyla açıklamak öğrencinin ne bildiğini bilmekle kesinlikle aynı şey değildir. Eğitim bazı sabit standartların başarılmaması değildir. Gerçek anlamda, her bir öğrenci için kişiye özgü ve kişisel bir tanım gerektirir. Eğitim, bir kişisel süreç ve kişiye özgü bir başarıdır.

Eğitim ve öğrenmenin, geleneksel değerlendirme araçlarıyla değerlendirilmesinde karşılaşılan yukarıda anlatılan sorunlar, değerlendirme ile öğrenme teorisi arasındaki uyumsuzluğu içerir. Bu yargı sadece yapılandırmacı öğrenme teorisi değil, aynı zamanda insan ve bilişsel gelişim teorilerini de içerir. Bu tarz yaklaşımlar 1980'lerde popülerlik kazanmıştır. Ayrıca, son yıllarda, eğitimciler ve eğitim araştırmacıları aşağıdakiler gibi, öğrenci başarısının standardize değerlendirmesinin ciddi insani ve eğitimsel sonuçlarını görmüş ve kanıtlamışlardır (Mabry, 1999). Bunlar:

- Çoktan seçmeli test ile test edilen müfredatın konu, başlık ve becerilere daraltılması
- Müfredatın yüzeysel hale getirilmesi ve ilköğretim ile ilköğretim öncesi

sınıflarda uygunsuz şekilde akademik kılma

- Eğitimin "muhtelif ölü olguların" öğretimi ve ezberlenmesine indirgenmesi
- Sınav yapma ve test hazırlamaya fon ve öğretimsel zaman ayrılması
- Öğrencilerin demoralize edilmesi
- Öğretmenlerin uzmanlaşmalarının engellenmesi
- Sınav kaygısı olması nedeniyle öğrencilerin hatalı belirlenmesi ve eğitim kaynaklarının öğrencilere yanlış tahsisi
- Öğrenci başarısı ve okul programlarının kalitesi konusunda yanıltıcı olduğu ve uygun olmayan eğitim politika ve uygulamalarını teşvik ettiğinin bildirilmesi.

İçeriksel (Contextualized) Paradigma

Psikometrik paradigmalara standart hale getirilmiş testlerde karşılaşılan problemler, öğretmenleri daha kapsamlı, geniş ölçekli değerlendirmelere ve performans değerlendirmelere götürmüştür.

Psikometrik paradigma, öğrencilerin yaşantılarından bağımsız, tek tip, standart içeriği yansıtmasına karşın, içeriksel paradigma öğrencilerin yaşadığı gerçek deneyimleri yansıtmayı amaçlamıştır. Bu temel fark içeriksel değerlendirmenin sadece sıralama ve sınıflama yapmaktan öteye geçip, öğrencilerin gerçekte neler öğrendiklerini ortaya koymaktadır.

Puanlayıcı açısından puanlama zorluğu yaratırken, puanlama bakımından zaman alıcı ve kullanışlılık yönünden ekonomik değildir. Uygulamada standart testlere göre daha az soru sorulur bu da konunun kapsamını sınırlar ve kapsam geçerliliğiyle ilgili sorunlar meydana getirebilir (Burstein, 1991; Frechtling, 1991).

Öğretmenler, ölçme araştırmacıları ve psikometristlerin okullara uygun bir değerlendirme paradigması geliştirmelerini beklememişlerdir. Öğretmenlerin standardize test ile ilgili yaşadıkları sorunların ciddiyeti kendi değerlendirme uygulamalarında değişiklikler gerektirmiş ve bu uygulamalar, performans değerlendirmesine yönelik büyük ölçekli testleri teşvik etmiştir. Hangi

değerlendirme türünün öğrencilere ve öğretmenlere en uygun olduğu ile ilgili sezgisel yargılama esasında çalışarak, öğretmenler paradigma değişikliği, eyalet ve ulusal performans değerlendirmeleri ile ilgili teşvik edici çalışmalar yoluna girmişlerdir. Ancak sezgisel kararlar belirlenmesi güç yargılardır. Öğretmenler ile okul idarecileri yeni değerlendirme uygulamalarının neden daha iyi olduğu ile ilgili kanıtlar ileri sürme konusunda ve test geliştiricileri, politika belirleyicileri veya kamuyu ikna etme gerekçelerinde yetersiz kalmışlardır (Mabry, 1999).

Berlak (1992) testlere karşı empirik kanıtlar ile tutarlı bir teorik eleştiri getirerek standardize testlere alternatif olan testlere destek vermiştir. Bazı geleneksel test varsayımlarına aşağıdaki gerekçelerle karşı çıkmıştır.

- Sadece dolaylı olarak başarıyı ölçen testlerden alınan puanlar çelişkili olsa da, farklı insanlar ile ilgili farklı şeyler ifade edebilir (Yani bir test puanının anlamı ne şeffaf ne de evrenseldir).
- Testlerin ahlaki olarak nötr olmadığı daha ziyade insan karar mekanizması ve ilişkileri üzerinde güçlü bir etkiye sahip olduğu söylenebilir.
- Testler sadece bilişsel sonuçları ölçmeyi iddia edebilir ama bilişsel, duyuşsal ve devinişsel başarının güvenilir ayrımı olamayabilir.
- Testler sıralama oluşturabilse de öğrencilerin bildiği ve yapabildiği konusunda çok az bilgi verir. Geri bildirim araçları olarak değil, demokrasi ve özgür irade ilkeleri ile tutarsız kontrol ve takip araçları olarak işlev görebilir.

Berlak (1992) testler veya Psikometrik paradigma yerine bir içeriksel (durumsal) paradigma önermiştir. İçeriksel paradigmanın özelliklerini aşağıdaki gibi açıklamıştır.

- Standardize edilmeyen ama müfredata duyarlı ve orijinal içerik şeklindeki değerlendirme özellikleri (yani öğrencilerin öğrenme şansına sahip olabileceği ve okul dışındaki dünyada yararlı olabilecek içerik)
- Testin çoktan seçmeli formatla sınırlanmadan, öznel olarak değerlendirilebilen yapıcı cevaplanan soru, problem ve görevler içerebilmesi

- Öğrencilerin aşına oldukları, öğrenme ortamının doğal sınav ortamına uyduğu kendi sınıflarında test edilmeleri
- Öğretmenlerin öğrenci çalışmalarını değerlendirmede kendi mesleki yargılamalarını kullanma yetkisine sahip olmaları
- Öğrencilerin kendi kendilerini değerlendirme olanaklarına sahip olmaları ve değerlendirme sonuçlarının, öğrencilerin öğrendikleri konusunda sadece öğretmene ve diğerlerine bilgi vermemesi ve aynı zamanda müfredat planlamaya ve gelecekteki öğrenme imkânlarının tasarlanmasına da bilgi vermesi

Bireyselleştirilmiş (Personalized) Paradigma

Öğrenci başarısının değerlendirilmesine yönelik bireyselleştirilmiş yaklaşım yapılandırmacı öğrenme teorisi ve bireyselleştirilmiş eğitime daha iyi uyar. Literatürde yapılandırmacılık temel alınarak bireysel paradigmayla ilgili açıklamalar özellikle Wiggins (1993) tarafından yapılmıştır.

Yapılandırmacı öğrenme teorilerinin bireysel paradigmayla çok fazla uyum sağladığı ifade edilir. Bu yapılandırmacı perspektife göre eğitim öğrencilerin bilgiyi nasıl elde edeceğini öğrenmelerine yardımcı olmaktadır. Bu teorinin içinde aynı zamanda bunu dönüştürürken nesnel gerçekliğin yönünü yapılandırdığımız için insan olarak bu gerçekliğe ulaşamayız düşüncesi vardır. Eğitimin odak noktası öğrenmenin yer aldığı kültür ve konuları göz önünde bulundurur (Anderson,1998).

Yapılandırmacılık bir bilme kuramı olduğu için; bilme, bilen, bilinen, bilgiyi yapılandırma süreci, bu süreci etkileyen etkenlerle ilgili birçok açıklama içerdiğinden bir bilme kuramı olarak öğretme ve öğrenme süreçleri üzerinde önemli doğurguları vardır. Yapılandırmacılığa göre bilgi, duyularla ya da çeşitli iletişim kanallarıyla edilgin olarak alınan ya da dış dünyada bulunan bir şey değildir. Tersine; bilgi, bilen (öğrenen) tarafından yapılandırılır, üretilir. Yapılandırmacılığa göre bilgiyi yapılandırma gereksinimi, bireyin çevresiyle etkileşim sırasında geçirdiği

yaşantılardan anlam çıkarmaya çalışırken ortaya çıkar (Açıkgöz, 2002). Yapılandırmacılıkta bilginin öğrenen tarafından oluşturulan yapı olduğuna ve bu süreçte ön bilgilerin önemli yeri olduğuna inanıldığı için bu yapılar bireye özgüdür. Dolayısıyla bir bireyin kendisi için oluşturduğu yapıları bir başkasına aktarması olanaksızdır. Bir başka deyişle, öğretmen, kendi zihnindeki bilgi, kavram ya da düşünceleri öğrencilerin zihnine aktaramaz. Bunu yapmaya çalışsa bile, öğretmenin anlattıkları öğrenciler tarafından aynen alınmaz. Anlatılanlar öğrenci tarafından yorumlanır ve dönüştürülür. Öğretmenlerin öğretmeye çalıştıklarının, öğrenciler tarafından onların anlattığı şekliyle öğrenilmemesinin nedeni budur. Yapılandırmacılığa göre öğretmenin yapması gereken, öğrenci ile eğitim programı arasında aracılık etmek, öğrencinin bilgiyi yapılandırma sürecini yanlış yönelmeleri önleyerek kolaylaştırmaktır. Çoğu zaman, öğrencilerin yeni bilgileri öğrenebilmeleri için, onların önceki yapılarında değişikliklerin olması gerekebilir. Önceki yapıların değişmesi ise oldukça zordur. Yanlış anlamaların yanlış olduğunu ya da doğrusunu söylemek yetmez. Öğrencilerin kendi yapılarının farkına varması ve yanlışın nedenlerini görmesi sağlanmalı ve sonuçta kanıtlarla öğrenci inandırılmalıdır. Yapılandırmacılığı gelenekselden ayıran en önemli özellik budur. Yapılandırmacılık geleneksel öğretimdeki dayatma eğilimini reddeder (Açıkgöz, 2002).

Yapılandırmacı değerlendirme yaklaşımını benimsemenin en önemli görüşlerinden birisi; öğretmen ve öğrencilerin rollerinin geleneksel değerlendirme uygulamalarından ayrılmasıdır. Yapılandırmacı sınıflarda öğrenciler genellikle hangi ödevleri tamamlayacakları ve değerlendireceğini seçme hakkına sahiptirler. Üstelik öğrenciler sık sık, değerlendirme sürecinde kullanılacak olan ölçütü oluşturmayla ilgili karar verme sürecinde yer alırlar. Ölçütü derecelendirmek için kurallar (açıklama) koyarlar ve akranını ve kendini değerlendirmede yer alırlar.

Kendi kendini değerlendirme, sırasıyla bir sonrakinde hangi ödevi tamamlayacağıyla ve ne öğrenmesi gerektiğiyle ilgili kararlar almasında yol gösterir. Öğretmenler, değerlendirme sürecinde öğrencileri daha aktif bir şekilde yer almaları için cesaretlendirdikleri zaman, daha demokratik bir durumu benimseyerek tüm kontrolünü bırakır (Anderson,1998).

Dolayısıyla öğrencilerin zihinlerinde oluşturdukları bilgiyi ve bu bilgiyi anlamlandırmalarını doğru yorumlamak durumunda olan eğitimciler de artık, değerlendirme yaklaşımlarını bu yeni temellere oturtturarak dönüştürmeleri gerekmektedir. Bireyselleştirilmiş yaklaşım, kişiselleştirilmiş eğitim ve yapılandırmacı öğrenme teorisi ile daha iyi uyum sağlamaktadır (Mabry, 1999).

Bu uyumun nedenini yazar aşağıdaki gibi açıklamıştır:

1. Aynı şeyi farklı şekilde de olsa herkes anlar dolayısıyla testler bireyselleştirilmiş kapsama sahip olmalıdır.
2. Herkesin farklı ilgi ve hedefleri vardır. Dolayısıyla testlerle öğrencinin kendi hedeflerini gerçekleştirmeye hazırlık durumunu ortaya çıkaracak başarılar değerlendirilmelidir.
3. Herkes farklı şekilde cevap verir. Örneğin, bazıları yazılı ifadeye göre sözlü ifadeye daha iyidir. Dolayısıyla, sınavlar öğrencilerin bildiklerini göstermelerinde en iyi fırsat tanıyan yollarda performans göstermelerine izin vermelidir.
4. Bazı öğrenciler diğerlerine göre bazı şeyleri düşünmek için daha fazla zamana ihtiyaç duyar (veya şartlardaki diğer farklılıklara gereksinim duyar). Dolayısıyla, testi alanların bildiklerini gösterebilmeleri için sınavlar farklı şartlara ve sürelerle izin vermelidir.

Bazı öğretmenler tarafından uygulanan bireyselleştirilmiş yaklaşımda kapsam geçerliği kişisel uygunluk ile dengelenir. Testin kapsamı, uygulaması ve kalite standartlarının bireylere uygun hale getirilmesi ile bireyselleştirilmiş değerlendirme, öğrencilerin başarılarının diğerlerinin başarısı veya önceden belirlenen standartlarla karşılaştırılmasından ziyade öğrencilerin ne bildiklerini anlama ve değerlendirme fırsatını maksimize etmektedir. Bu yaklaşımda öğretmenler, öğrencilerinin değerlendirmelerinde çok daha fazla girdi ve bu değerlendirmeler üzerinde çok daha fazla güç sahibi olabilirler. Bazı değerlendirme programlarında, kendi bireysel başarılarına hangi test kapsamının uygun olduğunu, bilgi ve becerilerini gösterebilmelerine en iyi fırsat tanıyan formatların hangileri olduğunu

öğretmenleri ile görüşebilirler. Sıklıkla, neyi en iyi öğrendiğini ve hangi sınav formatını bildiklerini ortaya koymada kendilerine en iyi imkânı tanıdığını bireyin kendisi bilir. Testin nerede ve ne zaman verilmesi gerektiğinin belirlenmesine öğrenciler yardımcı olabilir. Sonuçlar, gelecekteki çalışmalara kılavuzluk etmesi, başarının özet göstergesinin belirlenmesi amacı için kullanılabilir (örneğin, harf veya not ya da her ikisi). Bireyselleştirilmiş değerlendirmede, kişisel veya kendi kendini değerlendirme, tüm anlamalar ve biliş-ötesi bilgi için ve öz bilginin öneminden dolayı hayatidir. Öğretmenlerin ve diğerlerinin değerlendirmelerinden daha önemlidir. Bireyselleştirilmiş yaklaşım dışında, tüm yaklaşımlarda öz değerlendirme çok fazla ihmal edilmektedir. Kişiselleştirilmiş yaklaşım içine tamamen yerleştirilen yeni değerlendirme sistemleri Tablo 1’de sunulan özellikleri gösterme eğilimindedir (Mabry, 1999).

Tablo 1

Yeni Değerlendirme Sistemleri için Bakış Açıları ve Örnek Durumlar

Bakış Açıları	Örnek Durumlar
Öğrenci tarafından seçilen veya öğrenci-öğretmen arasında görüşülen değerlendirme kapsamı	"Biyolojinin bu alanında konuyu anladığımı kanıtlamamın yanı sıra özel ilgi ve çalışmamı göstermek için ne yapabilirim?"
Öğrenci tarafından seçilen veya öğrenci-öğretmen arasında görüşülen değerlendirme zaman ve formatı	"Ne zaman hazır olacağım? Öğrendiklerimi nasıl gösterebilirim?"
Sayısal not, kontrol listesi veya harfler yerine açıklayıcı değerlendirme ve bu değerlendirme sağlamaları	"bu öğrencinin yatak odası için alacağı halının satış fiyatını doğru hesaplaması alan, yüzde, ondalık sistem ile ilgili bilgilerini uygulama durumlarına uygulayabileceğini göstermiştir"
Öğrenme ve diğer bireyselleştirilmiş geri bildirimin düzeltilmesi, iyileştirilmesi veya genişletilmesi	"Öğrenci, alacak ve borçlar ile ilgili federal sistemi anladığını göstermesine rağmen, eyalet ve yerel mercileri söyleyebilmeli ve yerel bölgesinde nasıl oy kullanabileceğini açıklayabilmelidir."
Öğrencinin, öğrenci tarafından belirlenen veya öğrenci öğretmen arasında görüşülerek belirlenen ölçüt ve standartlara göre öz değerlendirmesi	"Dilbilgimin, bu öyküyü birisinin okumasından utanmayacağım düzeye gelişmiş olmasından mutluyum ama daha gerçekçi bir sonuca gelemediğim için üzgünüm"

İnsanlar kendi anladıklarını yapılandıkları için bireyselleştirilmiş değerlendirme sistemi yapılandırımla uyum içindedir. Bireyselleştirilmiş performansların değerlendirilmesi karmaşık ve kötü tanımlanmış bir iştir. Kapsam, format ve ortam ile ilgili esneklik, değerlendirmecilerde olağan olmayan tepkiler gerektirir. Örneğin, değerlendirme makineye yaptırılamaz veya her biri belli bir öğrenci için özelleştirilmiş ve öğrenci için veya kısmen öğrenci tarafından belirlenen bireyselleştirilmiş standartlar ile değerlendirilen birçok performansın karmaşıklığı ile makine baş edemez. Şayet değerlendirme, insanların doğal olarak sergiledikleri değişikliği ayırt etmek ise, insan değerlendirmesi yani öğretmen ve diğerlerinin değerlendirmesi şarttır. Kendi kavrayışlarını oluşturan insanlar için sadece kişiselleştirilmiş değerlendirme yapılandırımla uygundur.

Değerlendirilen ürün ve süreçlerin değişkenliği nedeniyle önceden belirlenen ölçüt ve standartlar, kişiselleştirilmiş değerlendirme için zayıf puanlama kılavuzudur. Söz konusu ölçüt ve standartlar değerlendirmeyi standardize eder ve öğrencinin kendisine özgü başarısından dikkatleri uzaklaştırır. Kişiselleştirilmiş paradigmada, ölçüt ve standartların bireyselleştirilmesi, nitelik hususlarının her bir öğrencinin geçmişi, fırsatları, hedefleri ve benzeri şeyleri dikkate alınarak dengelenmesi çok daha uygundur. Bireyselleştirilmiş değerlendirme işlemleri ve bireyselleştirilmiş standartlar, şu anda tipik olandan tamamen farklı bir değerlendirmeci eğitim türünü zorunlu kılar (Mabry, 1999).

Paradigmaların karşılaştırılması

Değerlendirme amaçları açısından, değerlendirme paradigmatları ve stratejileri arasında bir seçim yapılması avantajlar ve dezavantajlar arasında bir karşılaştırmayı gerektirebilecektir. Eğitim açısından mevcut olan avantajlar bazen uygulanabilirlik açısından dezavantajları içerebilmektedir.

Psikometrik paradigma standartlaştırılmış değerlendirmeyi içerir. Avantajları; yeterli, kamu güvenilirliği, köklü ve kurumsallaşmış teknoloji, standart veya ölçüt

açısından ilgili konumu veya başarıyı verme olarak sıralanabilir. Buna karşılık dezavantajları; olumsuz öğrenci etkisi, olumsuz okul etkisi, müfredat planlama veya düzeltme için sınırlı imkân, öğrencinin ne bildiği ve ne yapabileceği konusunda az ayrıntı vermesidir (Mabry, 1999).

İçeriksel paradigmada gösterilen müfredat hassasiyetinin bir bedeli vardır. İçeriksel paradigma müfredat tabanlı değerlendirmeyi içerir. Avantajları; eğitmenlere bilgi sağlaması, öğretmenleri yetkilendirmesi, öğrencilerin amaçlananı ne kadar iyi öğrendiklerini göstermesidir. Dezavantajları; müfredatı sınırlandırması, ölçümler konusunda eğitilmiş olmayan öğretmenlerin hata yapabilmesi, başka yerlerdeki öğrencilerin karşılaştırmalı olarak ne kadar başarılı olduğunu bildirmemesidir. Müfredat yoğun olarak standartlaştırılmış testlere kıyasla aksine müfredat tabanlı testlerin eğitim açısından faydası daha güçlüdür. Ancak, bilginin sınıf müfredatı, şartlar ve eldeki öğrenciler ile sınırlı olduğu durumlarda genel referans puanları kaybolmaktadır (Mabry,1999).

Bireyselleştirilmiş paradigma faydaları daha fazla artırırken, diğer yandan maliyetleri ağırlaştırmaktadır. Bireyselleştirilmiş paradigma bireyselleştirilmiş değerlendirmedir (Mabry,1999).

Avantajları; duyarlı eğitmenlere bilgi vermesi, öğrenciyi yetkilendirmesi, öğretmenleri yetkilendirmesi, öğrencilerin ne bildiğini ve neler yapabileceğini bildirmesidir. Dezavantajları; Muhakeme ve esneklik üzerine fazla talepler, ölçüm konusunda eğitilmiş olmayan değerlendirmecilerin hata yapması, müfredata veya diğerlerine kıyasla öğrencilerin başarıları konusunda az bilgi vermesidir (Mabry,1999).

Değerlendirme sistemlerini standardize etmek yerine bireyselleştirmek öğrencileri diğer öğrencilerle karşılaştırmadığı gibi sınav olan öğrencileri önceden belirlenmiş ölçütlerle de belirlemez. Normatif değerlendirmenin öğrencileri birbiriyle ya da tek bir standarda uymaya göre karşılaştırmak yerine, bireyselleştirilmiş değerlendirmeler öğrencinin ne öğrenmiş ve ne yapabilir olduğu ile ilgili mümkün olduğu kadar fikir

edinmeyi amaçlar. Diğer yandan bireysel paradigma da kapsam geçerliliği sorunu da ortaya çıkabilir. Öğrencilerin ilgili alana yönelik ne bildiklerini öğrenmek amaçlanmaktadır. Bu amaca ulaşmak için de sınavlar bir ölçme aracı olarak kullanılmaktadır. Öğrencilerin sınavlarda yanıtladıkları sorulara göre genellemeler yapmak için sınavlarında alanı iyi yansıtmaları gerekmektedir.

Yapıcı öğrenim kuramı tarafından dolaylı olarak belirtilen, eğitimin bireylere uyarlanmasına yönelik bireyselleştirmenin sağladığı bariz faydaya rağmen, bireyselleştirilmiş standartların geliştirilmesi ve çok çeşitli performans türlerinin cevaplandırılmasına ilişkin zihinsel kaynak ve talepleri göz korkutucudur. Bireyselleştirilmiş değerlendirme bazı küçük ölçekli, okul tabanlı sistemlerine iyi yerleştirilmiştir. Ancak bireyselleştirmenin bireylerle ilgili aşinalık ve sürdürülen ilgi gerekliliği büyük çaplı uygulamalarda uygulanabilirlikle ilgili sorunlar ortaya çıkarmaktadır. Öte yandan, içeriksel paradigmada olduğu gibi bireyselleştirilmiş paradigmada da birçok zorluklarla karşılaşmaktadır. İnsan tarafından puanlamanın çeşitlilik doğurması, bu çeşitliliğin farklı değerlendirme yöntemleri ve farklı ölçütlerle artması, etkinliklerin kişiye özel hazırlanması ve uygulanması ile daha fazla zaman alıcı etkinlikler olmaları, herkesin öğrenmesi gereken bilgi ve kapsanması gereken öğretim programının ne olması gerektiği, bunların en önemlileridir.

Gerçek şu ki, değerlendirme paradigmalarının eğitim kapsamı ve uygulamaları üzerinde genellikle önemli etkileri olmuştur. Değerlendirme paradigmaları ve bu yaklaşımların politik etkileri öğretme ve öğrenmenin gerçekleştiği mikro düzeyde öğretmen ve öğrenciler tarafından her zaman hissedilecektir ve hissedilmeye devam edecektir (Mabry, 1999).

Ölçme Araçlarında Aranılan Başlıca Nitelikler

Ölçme işinde kullanılacak araçların belli niteliklere sahip olması beklenir. Bu niteliklerden birisi olan güvenilirlik, herhangi bir ölçme araç veya yönteminin ne derece tutarlı ölçüm yapabildiğidir. Başka bir deyişle güvenilirlik, ölçme sonuçlarının

hatalardan arınlık derecesidir. Ölçme ve değerlendirme hangi amaç için yapılırsa yapılsın elde edilen ölçümlerin hatasız ya da az hatalı olması beklenir. Ancak, en duyarlı araçlarla yapılan ölçümlerde bile bir miktar hata vardır. Bu hatalar ölçme sonuçlarına çeşitli yollardan karışmaktadır. Ölçme hataları, ölçmede kullanılan araçtan, ölçme yönteminden, ölçmeyi yapan kişiden, ölçmenin yapıldığı ortamdan ya da üzerinde ölçme yapılan bireyden kaynaklanabilir. Ölçme sonuçlarında sabit, sistematik ve tesadüfi hatalara rastlanabilir (Turgut, 1993, s.27–28).

Ölçme aracının başka bir niteliği ise geçerliliğidir. Geçerlik, bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellik ile karıştırmadan, doğru olarak ölçebilme derecesidir (Tekin, 1996). Örneğin; bir ilköğretim öğrencisinin Türkçe dersindeki başarısının ölçülmesi için bir sınav yapılacaksa, bu sınavın Türkçe dersindeki başarıyı tanımlayan tüm değişkenleri ölçmesi, bu değişkenler dışındaki değişkenleri ölçmemesi istenir. Sınav bu nitelikteyse, verdiği puanlar geçerlidir. Bir ölçme aracı olarak testin güvenilirliğini etkileyen bütün etkenler doğrudan ya da dolaylı olarak testin geçerliğini de etkiler. Bir testin ya da testten alınan puanların geçerliliği yukarıda açıklanmış olan sabit, sistematik ve tesadüfi hataların tamamından etkilenir. Ölçme sonuçlarının geçerliğini artırmak için aşağıdaki hususlar hatırdan tutulmalıdır (Tekin, 1996; Turgut, 1997):

1. Bir ölçme aracı kapsadığı soruların geçerliği oranında geçerli olabilir. Bu nedenle, her bir sorunun o araçla ölçülmek istenen bilgi, beceri ve yeteneklerden en az bir tanesini ölçmesi sağlanmalıdır.
2. Her soru, o soruyla ölçülmek istenen bilgi ve becerilere sahip öğrencilerin tereddütsüz olarak doğru cevaplandırabileceği; bu bilgi ve becerilere sahip olmayanların doğru cevaplandıramayacağı nitelikte olmalıdır.
3. Test, belli bir sürede kazandırılması hedeflenen bütün konuları, bilgi, beceri ve yetenekleri temsil etmelidir.
4. Bunların yanı sıra, güvenilirliği artırmaya yönelik tüm önlemler geçerliği de artırır.

Bir ölçme aracında güvenilirlik ve geçerliğin yanı sıra bulunması gereken bir başka nitelik ise kullanılabilirliktir. Kullanılabilirlik, bir ölçme aracının veya yönteminin, geliştirilme, çoğaltılma, uygulanma ve puanlama yönünden kolaylık, objektiflik ve ekonomiklik özelliklerine sahip olması anlamına gelir (Tekin, 1996).

Bireyselleştirilmiş Paradigmaya Dayalı Olarak Yapılan Performans Değerlendirme

Okullarda öğrencilerin, hatırlama (bilgi) düzeyinde kazandıkları davranışlar performansın oluşmasına ve gelişmesine katkı sağlarken bu davranışların tek başına kullanılması performansın minimum düzeyi (alt düzey) olarak nitelendirilebilir. Bu bağlamda performans, öğrencilerin okul öğrenmelerini geliştirmek ve pekiştirmek amacıyla sınıf içinde ya da sınıf dışında bir etkinlik yaptıkları veya bir ürün ortaya koydukları zaman oluşmaktadır. Başka bir deyişle genellikle üst düzey zihinsel süreç gerektiren bir görev ya da etkinlik yerine getirilirken ortaya konan çaba ve sonucunda ortaya çıkan ürün performans olarak adlandırılabilir (Kutlu, 2008).

Eğitim çevresinde yaygınlaşan yeni değerlendirme yaklaşımlarından biri performansa dayalı durum belirlemedir. Öğrenci başarısının gelişimini sağlamak amacıyla yapılan üst düzey zihinsel çaba gerektiren çalışmalar performansa dayalı durum belirleme (performance based assessment) olarak adlandırılmaktadır. Performansa dayalı durum belirlemede öğrencilerin basit, yalın, alt zihinsel düşünme gerektiren görevleri değil, daha üst düzey düşünme gerektiren karmaşık yapıdaki bazı görevleri yerine getirmeleri istenmektedir. Genellikle bu karmaşık görevler bir ürünün ortaya çıkarılmasını veya verilen bir problem durumuna çözüm önerileri getirilmesini içermektedir. Buradaki temel amaç öğrencinin yaratıcılık, problem çözme, eleştirel düşünme, karar verme, empati kurma gibi yeteneklerini ortaya çıkarmak, geliştirmek, bu yeteneklerin ne düzeyde geliştiğini ve kullanılabildiğini belirlemektir (Kutlu, 2008).

Bunun yanında, performans değerlendirme öğrencilerin gerçek yaşam problemlerine akademik bilgilerini uygulayabilme ve bunu problem üzerinde gösterebilmeleri ile ilgilenir. Öğrencilerin öğrendiklerini gerçek durumlarda göstermelerini sağlar (Airasian, 1994).

Performans deęerlendirmesi öęretmenlerin öęrenmeyi iyileřtirmek ve deęerlendirmek için kullandıkları uygulamaları desteklemektedir (MEB, 2003).

Bu duruma uygun deęerlendirme yapabilmek için en çok kullanılan puanlama yöntemlerinden biri dereceli puanlama anahtarıdır.

Dereceli Puanlama Anahtarı (Rubric)

Dereceli puanlama anahtarı, gözlemlerimizi puanla tanımlanmış kategorilerden (ölçüt ya da ölçütler) uygun düşen boyuta kaydetmemizi sağlayan bir deęerlendirme aracıdır (Haladyna, 1997). Wiggins'e göre dereceli puanlama anahtarları, öęrencilerin çalışmalarını deęerlendirmede puanlama kılavuzlarıdır. Ve aşağıdaki sorulara cevap verirler:

Performans hangi ölçütlerle deęerlendirilir?

Performans başarısını deęerlendirmek için neyi, nerede aramalıyız?

Performansın nitelięindeki sıra neye benzer?

Hangi puanın tarafsız, güvenilir ve geçerli bir biçimde verildięini nasıl belirleriz ve bu puan ne anlama gelir?

Nitelięin farklı seviyeleri nasıl tanımlanır ve bunlar birbirlerinden nasıl

Farklılık gösterir?

Popham (1997) dereceli puanlama anahtarının, öęrencilerin, yazılı kompozisyonlarının, sözel sunumlarının veya bilimsel projelerinin deęerlendirilmesinde bir çeřit puanlama ölçüsü olduęunu belirtmektedir. Popham dereceli puanlama anahtarını, her bir çalışma için ölçütleri (ölçülecek boyutları) listeleyen ve çalışmada nelerin yapılacaęını gösteren bir puanlama aracı olarak görmektedir. Popham'a göre dereceli puanlama anahtarı; deęerlendirme ölçütleri, ölçüt tanımlamaları ve bir puanlama stratejisi olmak üzere üç bölümden oluşur:

Ölçüt tanımlamaları: Öęrencilerin deęerlendirilmek istenen yanıtlarındaki niteliksel farklılıkları tanımlama yolunu ifade eder. Böylece öęrencinin başarısını

puanlamak için ölçütler öğrencinin ne yapması gerektiğini ortaya koyar.

Değerlendirme Ölçütleri: Öğrencilerin ölçütlere göre ne kadar iyi performans gösterdiklerini açıklamaya yardımcı olur. Performans standartları ölçme aracında mevcuttur.

Örneğin, yazma etkinlikleri için Vermont'un (1991) hazırladığı analitik değerlendirme rehberinde, belirlenen ölçütler ve standartlar Tablo 2'de verilmiştir.

Tablo 2
Vermont'un Analitik Değerlendirme Rehberinde Belirlenen
Performans Düzeyleri ve Ölçütler

ölçütler/performans düzeyleri	Amaç	Organizasyon	Detaylar	Ses/Tını	Grammer/kullanımı
Çok fazla					
Sıklıkla					
Bazen					
Nadiren					

Her performans ölçütünün her standart için tanımı yapılır ve bunlar bir tabloda birleştirilerek dereceli puanlama anahtarı (rubric) oluşturulur.

Puanlama stratejisi: Puanlama bütünsel ya da analitik (analytical) biçiminde olabilir. Bunlardan hangisinin kullanılacağı değerlendirmenin amacına bağlıdır (Popham,1997). Bazı durumlarda yapılan bir değerlendirmeyi bağımsız etkenlere (ölçüt) ayrıştırmak mümkün olamamakta, performansın farklı düzeylerinin ortaya çıkarılması için belirlenmiş ölçütler arasında bir ayrışma bulunmamaktadır. Böyle durumlarda bütünsel (bütünsel) dereceli puanlama anahtarı kullanılmalıdır (Brookhart, 1999). Analitik dereceli puanlama anahtarı ise, ölçülen bir yetenek boyutunun öğelere ayrıştırılabildiğinde ve daha ayrıntılı puanlama yapmak istendiğinde kullanılmaktadır (Haladyna, 1997). Analitik bir dereceli puanlama anahtarı, süreklilik arz eden nitelikteki bir çalışmaya verebilecek muhtemel puanlardan oluşan bir ölçeği içerir. Yüksek sayılar en iyi performanslara verilir. 6 dereceli bir rubrik anahtarında 4,5,6 gibi puanlar yüksek performans düzeylerine atanırken; 0,1 gibi puanlar ise en düşük performans düzeylerine atanır.

Dereceli puanlama anahtarında her performans düzeyinde tanımlamalar yapılarak ölçeğin daha güvenilir ve yansız puanlanması sağlanır. Tanımlayıcılar ölçütler içerdiği ve genellikle standartlara dayandığı için kaliteli bir dereceli puanlama anahtarı performansların güvenilir ve geçerli bir ayrımını sağlar (Wiggins, 1993).

Moskal'a göre (2003) Analitik dereceli puanlama anahtarı performansı değişik bölümlere ayırıp, her bölüm için ayrı bir ölçek kullanarak değerlendirir.

Bütünsel dereceleme ölçekleri ise, daha geniş bir süreci değerlendirmek için tek bir ölçek kullanılır. Performansı ya da sorun durumunu oluşturan her bir bölüm bir bütün olarak değerlendirilir(Moskal, 2003).

Dereceli puanlama anahtarı Hazırlama Aşamaları (Goodrich, 2001):

1. Performansı belirlemede kullanılacak ölçütlerin listelenmesi: Öncelikle farklı performans görevlerine uygun olarak kullanılacak dereceli puanlama anahtarı için ölçütler belirlenmelidir. Bu ölçütler öğrencinin yerine getirmesi istenen performanslar göz önüne alınarak belirlenebilir. Bu noktada, ölçütün performansın önemli bir yönüne odaklandığından emin olunmalıdır. Öğrenci çalışmasının farklı örneklerini ayırt etmenin pek çok yolu vardır. Değerlendirmek istediğimiz öğrenme hedefleri ile ilişki içinde gerekli olan ölçütleri kullanmayı isteriz. Bu ölçütlerin bazıları performansın kapsadığı her olası yol için uygulanabilir olmayabilir. Bunun için en önemli olanları tanımlamaya gereksinim duyarız.
2. Puanlama stratejisi olarak hangi dereceli puanlama anahtarı türünün kullanılacağına karar verilmesi: Amaca uygun olarak analitik ya da bütünsel dereceli puanlama anahtarından biri seçilmelidir.
3. Performans düzeylerinin belirlenmesi ve düzey tanımlarının yapılması: Bu aşamada, en iyi ve en kötü performans düzeyleri (dereceleri) belirlenmeli ve buna göre her ölçüt derecelendirilmelidir. Bu derecelendirme, diğerlerinin değerlendirilen alandaki performansı nasıl puanladıklarının ve tanımladıklarının örneklerini oluşturmada yararlıdır. Bu basamakta etkiliyi etkili olmayandan ayıran çalışmaların özelliklerini

belirlemek önemlidir.

4. Uzman görüşünün alınması: Dereceli puanlama anahtarı hazırlandıktan sonra ilgili alan öğretmenlerinden, ölçme-değerlendirme uzmanlarından görüşler alınmalıdır.

Öz Değerlendirme

Belli bir konuda bireyin kendi kendisini değerlendirmesidir. Öz değerlendirme öğrencilerin kendilerini keşfetmelerine, güçlü ve zayıf yönlerini tanımlarına yardımcı olur. Kendi performans ve gelişimlerini bağımsız olarak ve gerçekçi şekilde değerlendirmeye yönlendirir. Öğrencilerin bir sonraki adımı tanımlayabilmelerine, kendilerine olan güvenlerini artırmalarına, öğrenme ve değerlendirme sürecine aktif olarak katılmalarına olanak sağlar. Öğretmen değerlendirme sürecinde öğrencileri dürüst ve açık olarak kendilerini ifade etmeleri konusunda cesaretlendirir. Öğrencilere bu değerlendirmenin, kendi gelişimlerini izlemek ve eksikliklerini gidermek amacıyla yapıldığı açıklanır. Bu tür değerlendirmenin olumsuz yönleri de vardır. Genellikle kendi performanslarını değerlendirirken yanlılığın varlığı göz ardı edilmemelidir. Başlangıçta kendini değerlendirme, öğrencilerin deneyimsizliği nedeniyle yanılığa neden olabilir. Yine de öğrenciler daha fazla deneyim kazandıkça aldıkları kararlar daha doğru olacaktır. Ölçütleri belirlenmiş, yapılandırılmış öz değerlendirme formları ile öz değerlendirmeler yapılabileceği gibi Teknoloji ve Tasarım dersinde proje geliştirme sürecinde tutulacak “günlükler” yardımı ile de öz değerlendirme yapılabilir. Ancak yapılandırılmış formlar yardımı ile değerlendirme yapmak genellikle ekonomik ve kullanışlıdır (Tekindal, 2006).

Akran Değerlendirme

Öğrencilerin, arkadaşlarının hazırladığı ödev, araştırma, proje, rapor vb. çalışmalarını değerlendirmesidir. Öğrenciler, arkadaşlarının çalışmalarındaki yeterlik düzeylerini değerlendirirken kendilerinin eleştirel düşünme becerilerini de geliştirirler. Akran değerlendirme, öğretmene öğrencilerin gelişim ve yeterlik düzeyleri hakkında geri

bildirim sağlar. Akran değerlendirmede, öğrencilerin yanlış davranışlarını önlemek için ölçütlerin öğrencilere verilmesi yararlı olur (Tekindal, 2006).

Performans Dereceleme ölçekleri geliştirmede güvenilirlik sorunları

Dereceli puanlama anahtarının puanlayıcı ve puanlama güvenilirliğinin sağlanması önemlidir. Dereceli puanlama anahtarı, iyi tanımlanmış puan kategorileri ile değerlendirme tamamlandığında ya da puanlayıcının dikkatsizliği durumunda uygun puanlamayı sürdürmeyi desteklemelidir. Aşağıdaki sorular verilen bir dereceli puanlama anahtarının anlaşılabilirliğini değerlendirmek için kullanılabilir.

1. Puanlama ölçütleri iyi tanımlanmış mı?
2. Puanlama ölçütleri arasında kesin farklılıklar var mı?
3. İki bağımsız puanlayıcı puanlama sonucunda aynı puanları elde ediyor mu?

Eğer bu soruların herhangi birine yanıt hayırsa belirsiz puanlama kategorilerinin düzeltilmesi gerekir.

Dereceli puanlama anahtarının daha iyi açıklanmasının yöntemlerinden biri derecelenmiş yanıt anahtarının kullanımıdır. Derecelenmiş yanıt anahtarı dereceli puanlama anahtarının ayrıntılarını gösteren puanlanmış yanıtların bir bütünüdür. Bir puanlama aracı, puan düzeyleri arasındaki farkları açıklamak için puanlama süreci boyunca derecelenmiş yanıt anahtarını belirtir.

Puanlama kategorilerini açıklamak için yapılan her çabadan sonra, diğer öğretmenlerden yanıtların bir grubunu değerlendirmek için derecelenmiş yanıt anahtarı ve dereceli puanlama anahtarı kullanmaları istenebilir. Öğretmenler tarafından verilen puanlar arasındaki bazı farklılıklar ileri derecede açıklama gerektiren dereceli puanlama anahtarının bileşenlerini ima edecektir. Yorumdaki bazı farklılıkların tartışılması ve görüşülmesi gereken dereceli puanlama anahtarı için uygun düzeltmelerin yapılması gerekir. Bu görüşme sürecinin çok fazla zaman tüketebilmesine rağmen, bu süreç güvenilirliği artırabilir (Yancey, 1999).

Başka bir güvenilirlik sorunu, sorumlu öğrenci kitlesine verilen dereceli puanlama anahtarının uygunluğudur. Öğrencilerin bir grubunun performanslarını uygunca ölçen bir dereceli puanlama anahtarı, farklı bir öğrenci grubunun performansını

uygunca ölçmeyebilir. Mesela, bir konuya bir görev iyi bir şekilde oturtulursa öğrencilerin topluluğu konuya aşına olabilir. Ve diğer öğrenci topluluğu konuya aşına olmayabilir. Verilen konuya aşına olmayan öğrenciler konunun bilgisinin yetersizliğine bağlı olarak düşük puan alabilirler. Bu öğrenciler bilinen bir konuda aynı materyalle kapsanmış farklı bir görevi tamamlayacak olurlarsa onların puanları yükselebilir. Puanlar ve performanstaki çeşitlilik değerlendirmenin amacıyla ilişkisiz olduğunda puanlar güvenilir değildir.

Bazen puanlama sürecinde, öğretmenler dereceli puanlama anahtarında ifade edilmeyen kapalı ölçütü fark ederler. Her zaman bu durum olursa öğrencilerin yanıtları doğru vermelerinin, ölçütleri iyi bilmeleriyle sağlandığı dikkate alınarak dereceli puanlama anahtarının önceden öğrencilerle paylaşılması gerekmektedir. Dereceli puanlama anahtarı değerlendirme öncesi öğrencilerle paylaşılırsa, ifade edilmeyen ölçüt için öğrencilerin sorumlu tutulması gerekmez. Kapalı ölçütü tanılamak; gelecek değerlendirmeler için dereceli puanlama anahtarını mükemmelleştirmede öğretmene yardım edebilir (Moskal, 2000).

Performans Dereceleme ölçekleri geliştirmede geçerlilik sorunları

Bir değerlendirme aracı ya da bir görevin geliştirilmesi veya seçiminden önce, değerlendirme sonuçlarının geçerli yorumuyla ilgili endişelerin başlaması gerekir. İyi düzenlenmiş bir dereceli puanlama anahtarı, yetersiz derecede düzenlenmiş bir değerlendirme aracını düzeltmez. Kanıtlanmış geçerliliğin değerlendirmenin amacına bağlı olmasından dolayı, öğretmenlerin, sorumlu öğrencilerin neleri öğrenmesinin gerekliliği ile ilgili beklentileri ve bu yetenekleri nasıl icra edeceklerini net bir şekilde ifade etmeleri gerekir.

Değerlendirme aracının, değerlendirmede amaçlanan hedeflere uygun olup olmadığına yönelik kanıtlar sağlayabilmesi için, değerlendirme ile ilgili hedeflerin numaralandırılarak bir liste haline getirilmesi ve bu amaca uygun yazılacak soruların yanına uygun amaçların sıralanması tavsiye edilmektedir (Hanny, 2000). Böylece, değerlendirmeye uygun olmayan hedefler açık bir şekilde ortaya

çıkacaktır. Değerlendirme aracını bu yönüyle ele almak, dereceli puanlama anahtarının uygunluğunu değerlendirmede de kullanılabilir. Sonraki adımda, ulaşılmak istenen her bir hedefle uyum sağlayacak puanlama ölçütleri oluşturulabilir. Hedeflerden biri, puanlama kategorilerinde yansıtılmamışsa, dereceli puanlama anahtarının verilen hedefi incelemek için gerekli olan kanıt sağlayamadığı ifade edilebilir.

Eğer puanlama ölçütlerinden herhangi biri hedeflerle ilişkili değilse, değerlendirmenin ölçülmek istenen yapıyla ya da içerikle uygunluğu ile ilgili, dolayısıyla dereceli puanlama anahtarı ile ilgili soru işaretleri oluşmaktadır. Bu soru işaretlerini gidermek için kullanılacak süreç Tablo 3'deki gibi özetlenebilir (Moskal ve Leydens, 2000):

Tablo 3

Dereceli Puanlama anahtarı Geliştirmede İzlenecek Adımlar

1. Adım: Değerlendirmenin amacını ve hedeflerini belirleme	2. Adım: Her bir hedef için puanlama ölçütü oluşturma	3. Adım: Aşağıdaki sorulara cevap aranması: 1.Puanlama ölçütleri ile her bir hedef ölçülebiliyor mu? 2.Hedeflerle ilgisiz olan herhangi puanlama ölçütü var mı?
---	---	--

Değerlendirmenin amacı ve hedeflerini yansıtacak Tablo 3'ün üçüncü adımındaki soruların cevapları sayesinde geçerlikle ilgili hangi tür kanıt veya kanıtların üzerinde durulması gerektiği ile ilgili bilgiler sağlanabilir. Bunlar yapı, içerik ya da ölçütlerle ilgili kanıtlar olabilir. Değerlendirme aracının amacı, bireyin belirli bir konudaki bilgisi ile ilgili kanıt ortaya çıkarmaksa, o halde içerikle ilgili kanıtlar ele alınmalıdır. Değerlendirme aracı, bireyin içsel süreçlerini, akıl yürütmeyi, problem çözmeyi veya diğer süreçleri ölçmek için tasarlanmışsa, yapı geçerliliği ile ilgili

kanıtlara bakmak gerekmektedir. Değerlendirme aracının amacı, bireyin performansının okul dışında veya farklı bir durumda nasıl olacağı ile ilgili ise, ölçüt dayanaklı kanıtlara önem verilmelidir. Değerlendirmenin amacına göre puanlama ölçeğinin uygun olup olmadığını ortaya koymak için Tablo 4'te yazılı olan soru listesi yararlı olabilir (Moskal ve Leydens, 2000).

Tablo 4
Geçerlilik Türlerine Uygun Kanıt Toplamak İçin Sorulabilecek Sorular

içerik	Yapı	ölçüt
1)Değerlendirme ölçütleri asıl değerlendirmenin içeriğine yabancı bir içeriğe işaret ediyor mu?	1)İstenen yapının tüm önemli yönleri puanlama ölçütleri ile değerlendirilebiliyor mu?	1)Puanlama ölçütü, benzer performanslarda veya sonraki performanslarda başarıyla ilgili öneriler yapacak farklılığı ne kadar yansıtabiliyor?
2) Değerlendirme ölçütü ölçülmesi amaçlanan içeriğin her yönünü yansıtıyor mu?	2) ilgilenilen yapıyla ilişkisiz olan bir değerlendirme ölçütü var mıdır?	2) Değerlendirme aracı kullanılarak değerlendirilebilecek, daha ilerdeki veya benzer performansların önemli elemanları nelerdir?
3) Problemin, içeriğinde bulunan ve puanlama ölçeği ile değerlendirilmesi gereken ama değerlendirilemeyen bir yönü var mıdır?		3) Puanlama ölçütleri, gelecekteki veya daha ilerdeki performansların önemli elemanlarını ne kadar ölçebilmektedir?
		4) Daha ilerdeki veya benzer performansların, puanlama ölçütlerinde yansıtılmayan önemli yönleri var mıdır?

Çoğu değerlendirme birden çok amaca hizmet eder. Örneğin bir problem hem öğrencinin kesirlerle ilgili bilgisini, hem de problemi çözerken kullandığı akıl yürütme süreçlerini ölçebilir. Bu tür durumlarda, değerlendirmelerin geçerliliği için birden çok kanıt çeşidi ele alınmalıdır (Moskal ve Leydens, 2000).

Sonusal geerlik ynnden ele alındığında bu tr kanıtlar, deęerlendirme sonularının kullanımı veya sonularının incelenmesi ile ilgili kanıtlardır. rneęin, bir ęretmen erkek ve kız ęrencilerin performanslarının deęerlendirilmesinde kullanmıř olduęu dereceli puanlama anahtarının erkek ęrencilerin daha dřk sonular elde etmesine yol amıř olduęunu fark etmektedir. Byle bir sonucun yorumu, erkek ęrencilerin arařtırılan alanda kızlar kadar yeterli olmadığı biiminde olabilir. Fakat tespit edilen bu farklılıęın deęerlendirmenin amacı ile iliřkisiz bir deęiřkenden kaynaklandığı da ifade edilebilir. Bylece, problemin tamamlanması, zgn amalara uygun olmayan bir ierik bilgisi ya da yapı gerektiriyor olabilir. Sonusal geerlik, deęerlendirmenin rnlerini incelemek ve bu rnleri deęerlendirme sonularından olası alternatif yorumlar ıkarmak iin kullanılır (Moskal ve Leydens, 2000).

İeriksel Paradigmaya Dayalı Olarak Yapılan Performans Deęerlendirme

İeriksel paradigmanın temel farklılıęı, bir deęerlendirme kapsamının, ęrencilerin yařantısına bakılmaksızın her ęrenci iin standardize edilmiř aynı kapsam deęil ęrencilerin fiilen grdkleri mfredatı yansıtmasıdır. İeriksel deęerlendirmeler, sadece sıralamak ve semek iin deęil ęrencilerin fiilen ęrendiklerinin ortaya ıkarılması iin tasarlanır. Standardize testler, genelde uzaktan izleyicilerin (rneęin; yasa koyucular, dięer politika belirleyiciler, kamu) tarafından istenen alt sınır karřılařtırmalarını verirken mfredata dayalı deęerlendirmelerden elde edilen geri bildirim, mfredatı bozmaktan ziyade destekleyen, ana izleyicilerle (rneęin, ęrenci, ęretmen, veli) mfredat geliřtirme ve uyarlanmasına yarar saęlar. İeriksel paradigma ęretmen hkmlerine deęer verir ve ęrencilere karmařık bilgi ve becerilerini gsterme ve yansıtmaya fırsatı sunar. Eęitim fonları ve dięer kaynakların tahsisatındaki sorunlu eřiřsizlik dnemine, ieriksel deęerlendirme ęrencilerin eřiř olmayan ęrenme fırsatlarının dikkate alınması olanaęını saęlamaktadır. Yani, eęitim kaynakları adil olarak daęıtılmasa bile, deęerlendirme bu adaletsizlięi, mevcut ęrenme imknlarından zaten mahrum bırakılan ęrencileri cezalandırmak iin dřk puan kullanarak pekiřtirmek durumunda deęildir.

İçeriksel paradigmada da sorunlar mevcuttur. Yapıcı cevap sorularının hazırlanması ve puanlama zaman alır bu nedenle pahalıdır. Uygulanması da zaman aldığından, bir test oturumunda nispeten bir kaç soru sorulabilir. Bu da konu kapsamını veya test edilen alanı sınırlamakta ve kapsam geçerliği konusunda kuşku uyandırmaktadır. Ayrıca, eğitimin uygun amaçları veya kapsamı konusunda evrensel bir mutabakat olmamasına rağmen, bazıları, değerlendirmenin genel konu veya becerilerden ziyade birçok müfredat ile ilgili olduğunda ihmal edilebilen şeyler olduğunu düşünmektedir.

Puanlaması sonuç itibarı ile öznel ve makine değil de insan tarafından puanlama sıkıntılı değişkenlikle sonuçlanabilir. Zira, değerlendirmecinin birisi bir öğrenciyi geçer puan verirken diğeri vermez ya da bir değerlendirmeci değerlendirme sırasında yorgun olabilir veya diğeri ihlaller içinde olabilir ki bu da öğrenci performansı niteliği ile ilgisiz puan değişikliklerine sebep olabilir. Tutarsız puanlar, yeni değerlendirme sistemleri ile ilgili zorluk ve kuşkuyla beraberinde getiren güvenilirlik sorunlarıdır (Marby, 1999).

Öğrencilerin performanslarını değerlendirmek için en çok kullanılan araçlardan birisi yazılı yoklamadır (özellikle açık uçlu sorular).

Yazılı yoklamalar

Yazılı yoklamalar yanıtlayıcının sorulan soruların yanıtlarını düşünüp hatırlayarak ve hatırladığı yanıtı organize edip yazılı olarak sunduğu sınav türüdür. Kolay hazırlanmasından dolayı öğretmenler tarafından çok kullanılır. Hem öğrencilerin hemde öğretmenlerin en alışık olduğu sınav türüdür. İlköğretimden başlayarak her eğitim kademesinde kullanılmaktadır. Nesnel sınav türlerinin hazırlanmasındaki güçlükler ve uzmanlaşma gerekliliği yazılı sınavın çok kullanılmasının diğeri bir nedeni olarak görülebilir. Ayrıca, diğeri birçok sınav türünün öğrencilerin üst düzey davranışlarını ölçemeyeceği düşüncesi, nesnel testlerin önem ve avantajlarının görülememesi ve okullardaki teknik donanım, araç-gereç ve materyal eksikliği de öğretmenlerin yazılı sınav kullanmayı tercih etmelerindeki sebepler arasında görülmektedir. Yazılı sınavlar problem çözme, problemleri organize etme, yeni ve orijinal fikirler üretme, bilgileri yeni durumlarda işe koşma, görüşleri değerlendirme,

fikirleri analiz etme vb. gibi davranışların ölçülmesi için en uygun sınav türü olarak kabul edilir (Doğan, 2007).

Yazılı Yoklamaların Puanlanması

Yazılı puanlamada kullanılabilecek birkaç yöntem vardır.

Genel İzlenimle Puanlama: Bir cevap kâğıdı baştan sona okunur ve edinilen genel izlenime göre puanlanır.

Sınıflama: Ele alınan bir cevap kâğıdı tamamen okunduktan sonra ‘iyi- orta- zayıf’ diye adlandırılan üç sınıftan birine konur. Böylece tüm kâğıtlar teker teker okunduktan sonra sınıflandırılmış olur. Tüm kâğıtlar tekrar okunarak kendi sınıflar içerisinde alt sınıflara ayrılabilir. Böylece her üç sınıf içerisinde de kâğıtlar sınıflarına göre sıralanmış olur. Bir sonraki işlem her üç sınıfa verilecek puan aralığını saptamak ve kâğıtları puanlamaktır.

Sıralama: Okunacak kâğıt sayısı az olduğu hallerde kâğıtlar birbirleriyle kıyaslanarak okunur ve kâğıtlar en iyiden en zayıfa göre bir sıraya konabilir.

Anahtarla Puanlama: Yazılı soruların cevapları kısa ve önemli fikirler halinde listelenir. Böylece liste hazırlandıktan sonra her cevaba ve cevapların alt bölümlerine puan takdir edilir. Bu tür puanlamaya analitik puanlama da denir.

Anahtarın hazırlanmasında dikkat edilecek esaslar:

1. Cevaplara verilecek puanın yuvarlak bir sayı olması her soruya ve soruların alt bölümlerine verilecek ağırlıkların kolayca hesaplanmasına yardım eder. Toplam puanın büyüklüğü, en küçük ayrıntıya verilecek puan 1 olmak üzere seçilmelidir.
2. Bir sınavdaki soruların bir kısmı diğerlerine oranla daha fazla veya daha önemli bilgileri yokluyorsa bu sorular daha fazla ağırlık alması puanların geçerliğini artırır.
3. Puanlama anahtarı kesinlikle sınavdan önce hazırlanmalıdır ve cevap anahtarı hazırlanırken cevaplar çok yönlü düşünülmelidir. Cevapların çeşitli olacağı biliniyorsa cevap anahtarı birkaç sınav kâğıdı okunduktan sonra da hazırlanabilir.

4. Puanlama Anahtarı mümkünse birden fazla puanlayıcı tarafından okunup anahtarlanmalıdır (Turgut, 1993, s.60–61).

Yazılı yoklamalarda güvenilirlik

Eğitim sisteminde ölçme önemli bir yer tutar. Ölçtüğümüz nesnenin yada bireyin özelliği hakkında bize bilgi verir. Bundan dolayı ölçmenin hatasız (güvenilir) olması gerekmektedir. Çünkü hatasız ya da en aza indirgenmiş hatalı ölçmeler bizim doğru sonuca doğru yaklaşmamızı sağlar. Bunun için eğitimde yapılacak ölçmelerin mutlaka güvenilir olması gerekmektedir. Hatalar bireyden, test aracından veya çevre koşullarından kaynaklanabilir. Buradan yola çıkarak güvenilirlik ölçme sonuçlarının tesadüfi hatalardan arınırlık derecesini ifade eden bir kavramdır (Turgut, 1993,s.31).

Puanlamada güvenilirliği etkileyen faktörler şöyle özetlenebilir:

- Yazılı yoklamalarda elde edilecek puanların güvenilirliği genellikle soru sayısına bağlı olarak artar. Kısıtlı soru sayısı güvenilirliği sınırlayan bir etkidir.
- Sorunun açık olmayışı ve belirsizliği de güvenilirliği etkileyen faktörlerden biridir. Soruda ne istenildiğini anlamayan bir cevaplayıcı istenen cevabı değil kendi anladığı cevabı verecektir. Böyle bir cevabı puanlamak güçleşir.
- Yazılı yoklamalarda soru güçlüğü istatistiksel yöntemlerle ayarlanamadığından bazen soru güçlüğü sınavın amacına göre ayarlanamayabilir. Bu tür soruların güvenilirliği bütün diğer etkenler sabit kalmak şartıyla ancak her soru, sınıfın %50'si tarafından doğru cevaplanabildiği takdirde en büyük değerine ulaşır (Turgut, 1997).

Yazılı yoklamalarda geçerlilik

Ölçme araçlarının güvenilirliği kadar geçerli olması da önemlidir. Sınavda sorulacak soruların derste öğretilmiş konuları öğrenmeden de cevaplandırılabilir nitelikte olması sınav sorularının derste işlenen konuların tümünü içine almaması veya dengeli olarak kapsamaması, kopya yapılmış olması gibi etkenler sınavdan elde

edilecek puanların geçerliğini düşürür. Sınavların sınıf ortamında işlenen konuları kapsıyorken işlenmemiş konuları kapsamaması gerekmektedir. Kapsam geçerliğinin yanı sıra sorulan sorular kopya çekmeyi zorlaştırmalı, ipuçları vermemeli veya tahmin ile cevaplandırılabilir biçimde olmamalıdır. Ayrıca her yıl uygulanan sınav soruları üzerinde değişiklik yapılmalıdır. Bir sorunun yalnız konusu değil soruluş tarzı da geçerliği etkiler. Cevaplayıcılar iyice anlamadıkları sorulara kendi yorumlarına göre cevap verirler. Sorular kesinlikle anlam bakımından açık olmalı ve verilecek cevap sınırlanmalıdır. Puanlama yanlılığı da geçerliği etkileyen diğer bir faktördür. Yazılı yoklamaların puanlanması sırasında objektif bir puanlama yöntemi kullanılmıyorsa bazı cevaplara gerektiğinden fazla bazılarına ise gerektiğinden az puan verilebilir bu da geçerliği düşürecektir (Turgut, 1997).

Eğer eğitimdeki amacımız anlama, problem çözme ve bilgiyi yeni durumlarda kullanma yeteneğini geliştirme ise öğrenenin bilgiyi etkin bir biçimde oluşturduğu yapılandırmacılık gibi öğrenci merkezli yaklaşımlar tercih edilmelidir. Yapılandırmacı öğrenme kuramı eğitimde çok baskın olan geleneksel ve nesnelci paradigmaya karşıdır. Yapılandırmacılara göre bilgi duyu organları ile çevreden pasif bir biçimde alınmaz, öğrenen tarafından etkin bir biçimde yapılandırılır. Öğreneni öğrenme sürecinin merkezine alan yapılandırmacı öğrenme kuramı, davranışçı yaklaşımın etkisindeki geleneksel eğitim programının tüm öğelerinde önemli değişiklikler yaratmıştır. Yapılandırmacı ortamın bilişsel öğrenme ürünlerini kazandırılmasında geleneksel ortamdan daha etkili olduğu araştırmalarla kanıtlanmıştır (Koç ve Demirel, 2003).

Turgut (1997) güvenilirliğin ölçme sonuçlarının hatalardan arınlık derecesi olduğunu ifade etmektedir. Bu nedenle güvenilirlik kolayca kanıtlanabilir. Ancak bu sınavın amacına hizmet ettiğinin kesin kanıtı değildir. Daha önemli özellik olan geçerlilik, güvenilirlik kadar kolay kanıtlanamaz. Test geliştirme uzmanları çok yüksek düzeyde güvenilirlik elde etmek için uğraşırlar, çünkü bir testin güvenilir puanlar ortaya koyması, öncelikle güvenilirliğin gösterilmesi geçerlilik iddiasını kanıtlamak için bir destek olarak alınabildiği için istenen bir niteliktir (Mabry, 1995; Moss, 1994).

Toplumdaki deęiřen beklentilere ve çağın beklentilerine cevap verecek bilgi ve beceri yönünden donanımlı bireyler yetiřtirmenin önemi, deęiřen eęitim ve ölçme deęerlendirme yaklařımları ile vurgulanmaktadır. Fakat, geleneksel yaklařımları nesnellikleri, güvenirlikleri ve geçerliliklerinin yüksek oluřlarından dolayı göz ardı edilmemeleri gerekmektedir.

Bireylerin farklı yeteneklerini, becerilerini etkili bir řekilde kullanmalarını saęlamada ve onları doęru ilgi alanlarına yönlendirmede, yeni ölçme ve deęerlendirme yaklařımlarının etkili olup olmayacaęı, günümüzde eęitim alanında cevaplanması gereken önemli sorulardan birisidir. Ayrıca yeni ölçme ve deęerlendirme yöntemlerinin geçerlilikleriyle ilgili kanıtlar, bu yöntemleri kullanmak konusunda řüpheleri ortadan kaldırmak ve bu yöntemleri kullanarak bireyler hakkında tam doęru kararlar verilebilmesi için bu yöntemlerin açıklıęa kavuřturulması gerekmektedir. Böylece geleneksel ve yeni deęerlendirme yaklařımlarını nesnellikleri, güvenirlikleri ve geçerlilikleri yönünden karřılařtırmak mümkün olacaktır.

Bu problemler için ortaya konulacak doęru çözümler, çağdař donanımlı bireyler yetiřtirmede daha doęru bir yaklařım belirlemeye yardımcı olacaktır.

Problem Cümlesi

Matematikte öęrenci bařarısının deęerlendirilmesinde kullanılan yazılı sınavın puanlanmasında kullanılan puanlama anahtarı, dereceli puanlama anahtarı, öz deęerlendirme ve akran deęerlendirme sonucunda elde edilen puanlar arasındaki iliřkiler nasıldır?

Alt Problemler

1.Alt Problem: Puanlayıcının puanlama anahtarı ve dereceli puanlama anahtarından elde edilen bařarı puanlarıyla, öęrencilerin öz deęerlendirme ve akran deęerlendirme yapmaları sonucunda elde edilen puanların ortalamaları arasında anlamlı fark var mıdır?

2.Alt Problem: Puanlama anahtarı ve dereceli puanlama anahtarı kullanılarak elde edilen puanlar arasında nasıl bir ilişki vardır?

3.Alt Problem: Puanlama anahtarı ve dereceli puanlama anahtarından elde edilen puanların iç tutarlılık ölçüleri (Crombach Alfa Güvenirliği) arasında farklılık var mıdır?

4.Alt Problem: Puanlama anahtarı ve dereceli puanlama anahtarıyla elde edilen puanlar için hesaplanan puanlayıcı güvenilirlik değerleri arasında farklılık var mıdır?

5.Alt Problem: Puanlayıcının puanlama anahtarını ve dereceli puanlama anahtarının kullanmasıyla elde edilen puanların, öğrencilerin öz değerlendirme ve akran değerlendirme yapmaları sonucunda elde edilen puanların, puanların ortalamasıyla korelasyonları arasında farklılık var mıdır?

6.Alt Problem: Öğrencilerin başarılarını değerlendirmek için kullanılan yazılı sınav kâğıtlarının, puanlama anahtarı ile puanlayıcı tarafından puanlanmasının yanı sıra öğrencilerin yazılı sınav kâğıtlarının puanlayıcı, öğrencinin kendisi, öğrencinin arkadaşı tarafından dereceli puanlama anahtarıyla puanlamasıyla ilgili olarak öğrenciler neler düşünmektedir?

Sayıtlar

Öğrencilerin başarılarını değerlendirmek için kullanılan yazılı sınavın, puanlanmasında kullanılan puanlama anahtarı ve dereceli puanlama anahtarının uygulanması sonucunda elde edilen sonuçları karşılaştırmak için bir yarıyılık süre yeterlidir.

Sınırlılıklar

1. Araştırma Meslek Yüksekokulunda 1.sınıfta öğrencilere verilen Genel Matematik (Mat-101) dersi ile sınırlandırılmıştır.
2. Bir yarıyıl (dönem) ile sınırlandırılmıştır.

Tanımlar ve Kısaltmalar

Değerlendirme paradigmalarının değerlendirme yaklaşımları anlamında kullanıldığı bu araştırmada ele alınan değerlendirme yöntemleri aşağıdaki gibi tanımlanabilir.

Dereceli puanlama anahtarı (Analitik rubric): Performansın ortaya çıkarılmasında etkili olan tüm sürece odaklanan puanlama aracıdır.

Puanlama Anahtarı: Süreçten çok sonucu değerlendirmek için kullanılan puanlama aracıdır.

Bu araştırma kapsamına alınanlar yazılı sınav, dereceli puanlama anahtarı (Analitik rubric), ve puanlama anahtarlarıdır.

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı, günümüzde kullanılmakta olan ölçme araçlarından olan yazılı sınavların puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanmasının matematik eğitimi üzerindeki etkilerinin araştırılmasıdır. Bu değerlendirme yöntemlerinin etkileri karşılaştırılırken hangisinin bireysel farkları ortaya çıkarma ve bireylerin bilgi beceri ve yeteneklerini ortaya koyma yönünden uygun olduğunun saptanması yapılabilir. Öte yandan, dereceli puanlama anahtarı ve puanlama anahtarıyla yapılan değerlendirme sonucunda puanlama ve puanlayıcı güvenilirliklerinin karşılaştırılması da mümkün olacaktır.

Günümüzde öğrenci başarısını ölçmek ve değerlendirmek sadece puanlama anahtarıyla öğrencilerin sınıflarını geçip kaldıkları hakkındaki kararları veren geleneksel anlayış dışına taşmak durumundadır. Ölçme ve değerlendirme uygulamaları, okul programlarında yer alan davranışların öğrenilememesine ya da eksik öğrenilmesine neden olan ve dolayısıyla öğrenci başarısının artmasına engel olan sistemdeki bu olumsuzlukları saptamada kullanılan bir araç ve bu olumsuzluklara dayanarak doğru kararlar almaya yardımcı olan bir alan olarak görülmelidir. Bu nedenle, öğrencilerin okul hedeflerinde yer alan davranışları ne derece kazandıklarını ortaya koyan ölçme ve değerlendirme etkinlikleri yanında,

öğrencilerin kazandıkları bu davranışları ne derece kullandıklarını gösterecek ölçme ve değerlendirme etkinliklerine gereksinim vardır. Her iki durumda da, öğrencilerin öğretim sürecindeki durumlarının ortaya konması söz konusu olmaktadır. Bu ise, eğitim süreci içinde durum belirleme (assessment) kavramının önemini ortaya çıkarmıştır (Kutlu, 2002).

Durum belirleme, eğitimin iyileştirilmesini amaçlayan ve süreklilik gösteren bir işlemdir. Durum belirleme, beklentileri daha açık ve net bir biçimde ortaya koymakla birlikte eğitim kalitesi için uygun ölçütler ve yüksek standartlar öne sürmektedir. Ayrıca gösterilen performansın bu beklenti ve standartlarla ne derece uyumlu olduğunu sistematik olarak analiz edip yorumlamakta ve elde edilen sonuçların performansı artırma, performansa ilişkin bir açıklamada bulunma ve belgeye dayandırmada kullanılmasını sağlamaktadır (Stiggins, 1994).

Eğitim ortamında görülen bu eksiklikler ölçme ve değerlendirme amacıyla kullanılması gereken yeni yaklaşımları gündeme getirmiştir. Bu yaklaşımlara "performansa dayalı durum belirleme yaklaşımları" adı verilmektedir (Aslanoğlu, 2003). Bilginin çok çabuk ve sürekli olarak değişmesi ve öğrencilerin, bilgi toplumunun sürekli değişen ihtiyaçlarına karşılık verebilen bireyler olarak yetişmesi için, öğrencilerin bireysel ihtiyaçlarını göz önünde bulunduran ve onların kendilerini gerçekleştirmelerini sağlayan eğitim paradigmalarının uygulanması gerekmektedir. Yazılı sınavların klasik yöntemlerle değerlendirilmesinin dezavantajlarından birisi, klasik yöntemlerin süreci değerlendirmede yetersiz kalmasıdır. Öğrencilerin her biri bir soruyu çözerken farklı yollar izlemektedirler. Dolayısıyla, standart bir puanlama anahtarıyla bu farklı çözümlerin sürecini inceleyerek puanlama yapıldığında anahtara uyan çözüme puan verilirken, uymayan çözüme puan verilmez. Bu sonuç klasik yöntemle değerlendirmenin süreç değerlendirmede yetersizliğinin belirtisidir. Diğer bir dezavantaj ise, her sınav için bir puanlama anahtarının hazırlanması ve her soru için verilecek puanı ölçme kuralına göre doğru puanlamanın zor olmasıdır. Bazen, çok önemli bir çözüm basamağı önemsenmeyerek düşük bir puan atanırken bazen de basit bir çözüm basamağına yüksek puan atanması söz konusu olabilir. Bunun sonucunda çözüm basamaklarına ölçme kuralına göre yanlış puanlar verilebilir.

Diğer yandan yazılı sınavların dereceli puanlama anahtarıyla değerlendirilmesinin avantajlarından birisi süreç değerlendirmenin mümkün olmasıdır. Öğrenciler soruyu çözerken farklı yollar izleseler de her ölçüt için performans düzeyine göre puan verilir. Bu da öğrencilerin kendilerine özgün farklı çözümlerini doğru bir şekilde ölçmeyi sağlayabilir. Diğer bir avantaj ise, birden çok sayıda yapılan sınav ve birden çok öğretmenin puanlama yapması için aynı dereceli puanlama anahtarının kullanılması mümkündür. Böylece, yeni yöntemlerin geleneksel olana üstünlük iddiasının uygulamada gerçekleşip gerçekleşmediğinin test edilmesinin öğretmenlere sağlayacağı fayda; yazılı sınavı hangi yöntemle değerlendirirlerse hatalardan daha fazla arınık bir değerlendirme sonucu elde edeceklerinin kararını vermede doğru karar vermeleri mümkün olacaktır. Uzmanlara sağlayacağı fayda ise; araştırmalarda hatalardan arınık ölçme sonuçları elde edeceklerinden yapılacak araştırmaların güvenilirliğinin sağlanması noktasında faydalı olabilecektir. Araştırmanın öğretmenlere, uzmanlara, öğrencilere yönlendirici olması yönünden önemli olduğu düşünülmektedir.

İlgili Araştırmalar

Araştırma günümüzde kullanılmakta olan değerlendirme yöntemlerinden yazılı sınavların puanlama anahtarı ve dereceli puanlama anahtarıyla değerlendirilmesi, öz değerlendirme ve akran değerlendirmenin matematik dersinin değerlendirilmesi üzerindeki etkilerinin karşılaştırılmasıyla ilgili bir araştırmadır. Bu araştırmayla ilgili olduğu düşünülen bazı araştırmalar aşağıda sıralanmıştır:

Ankara Özel Tevfik Fikret Okulları'nda dereceli puanlama anahtarı uygulamalarının yaygınlaştırılmasına temel oluşturması amacıyla Nisan 2002 tarihinde başlayan ve iki aylık bir süreyi kapsayan bir araştırma yapılmıştır. Bu araştırma okulun 6/C sınıfındaki 24 öğrenci üzerinde yürütülmüştür. Öğrencilerden Fen bilgisi dersi kapsamında kendilerine önerilen 3 performans görevinden birini seçmeleri ve bu konuyla ilgili çalışmaları sınıf ortamında sunmaları istenmiştir. Öğrencilere performans göreviyle birlikte sunularının hangi ölçütlere göre puanlanacağını gösteren dereceli puanlama anahtarı verilmiştir. Öğrenciler sunularını yaparken bilgisayar, projeksiyon, tepegöz, teyp vb. teknolojik araçları kullanmışlardır. Buna göre araştırmanın bulguları ışığında dereceli

puanlama anahtarının okul öğrenmelerinin ölçülmesinde ve değerlendirilmesinde önemli görev üstleneceği ve okul öğrenmelerine katkı getireceği düşünülmüştür. Ayrıca öğrencilerin mevcut bilgi ve becerilerini, belirlenen ölçütler açısından değerlendirebilmesi yanında, bizzat öğrencinin kendisini değerlendirme sürecine katması açısından da önemli olduğu görülmüştür (Aslanoglu, 2002).

Anderson (1998), tarafından yapılan araştırmada geleneksel değerlendirmeden alternatif değerlendirmeye geçişte meydana gelen değişiklikleri araştırmıştır. Anderson'un yaptığı araştırmanın bulgularına göre geleneksel değerlendirme işlemlerinden alternatif değerlendirmeye doğru ilerlemek büyük bir teorik gelişmedir ve değişim zaman alır. Öğretmenler için yavaş ilerlemek ve öğrencilerin alternatif değerlendirme uygulamalarının arkasındaki hedef ve istekleri konuşmak önemlidir. Alternatif değerlendirmeye yönelen öğretmenler öğrencileri dinlemeli, onlarla görüşmeli ve onlardan dönüt (eleştiri) istemelidir. Bu öğretmenlerin gelenekselden alternatif değerlendirmeye geçiş boyunca ne çalıştıkları ve çalışmadıklarıyla ilgili görüş kazanmalarını sağlar. Alternatif değerlendirme paradigması çoğunlukla nesnel değerlendirmeye katkı sağlar. Öğretim, öğretme ve değerlendirmeyi birbirine bağlar. Öğrenci öğrenimini ilerleten bir öğretim aracıdır ve öğrencilerin işlerini değerlendirirken adil, düşünceli ve yaratıcı olmalarında öğretmenlere yardımcı olan önemli ve güçlü bir araçtır.

Meyer ve arkadaşlarının Manhattan'da bir ilköğretim okulunda yaptıkları araştırmada bu okulda okuyan alt gelir seviyesindeki ailelerin çocuklarını, onların en iyi çalışmalarının koleksiyonu ve sözlü sınavlarla değerlendirme ile ilgili çalışmalar yapmışlardır. Bulguları, bu alternatif tekniklerde başarılı olan öğrencilerin, üniversitelere kabul edildiklerini ve üniversitelerde başarılı olma eğiliminde olduklarını göstermektedir (Akt: Mathews, 2004).

Poyraz (2006), tarafından Celal Bayar Ü. Eğitim Fakültesi, ilköğretim fen bilgisi eğitimi bölümünde yapılan araştırmada ilköğretim fen bilgisi dersi öğretiminde işbirlikli öğrenme yönteminin kullanıldığı eğitim ortamlarında başarıyı ölçmede çoktan seçmeli testlerin, doğru-yanlış soru cümleleriyle yapılandırılmış ve kısa cevaplı testlere göre etkisi araştırılmıştır. Yapılan denemeler sonucunda, fen bilgisi

dersi öğretiminde işbirlikli öğrenme sonucu öğrenci başarısını ölçmede çoktan seçmeli testler ile doğru-yanlış soru cümleleriyle yapılandırılmış testlerin aynı oranda etkili olduğu, kısa cevaplı testlerin ise çoktan seçmeli testlere göre daha az başarı kaydettiği saptanmıştır. Bu durum çoktan seçmeli testler ile doğru-yanlış soru cümleleriyle yapılandırılmış testlerin başarıyı ölçmede aynı oranda etkili olduğunu, fen bilgisi öğretiminde çoktan seçmeli testlerin ve doğru-yanlış soru cümleleriyle yapılandırılmış testlerin birbirinin yerine kullanılabileceğini göstermektedir. İşbirlikli öğrenme yöntemi kullanılarak işlenen fen bilgisi dersi öğretiminde başarıyı ölçmede; çoktan seçmeli testlerin kısa cevaplı testlere göre başarıyı ölçmede daha etkili olduğunu, fen bilgisi öğretiminde işbirlikli öğrenme yönteminin uygulandığı eğitim ortamlarında çoktan seçmeli test kullanılmasının kısa cevaplı test kullanılmasına göre daha fazla başarı kaydettiğini göstermektedir. Araştırma sonuçlarına göre genel anlamda; çoktan seçmeli test tekniğinin başarıyı ölçmede kısa cevaplı test tekniğine göre daha etkili olduğu, doğru-yanlış soru cümleleriyle yapılandırılmış test tekniği ile aynı oranda etkili olduğu saptanmıştır. Bu durumun nedeni kısa cevaplı testlerin öğrencide bir tek doğru cevap yerine, doğru cevaba benzer birden fazla yanlış cevabı çağrıştırmasından; doğru-yanlış soru cümleleriyle yapılandırılmış testlerin ise fen kavramlarını ve fen bilgisi dersine ait ana temaları yoklamada etkili olmasından kaynaklanmaktadır.

Yıldız ve Uyanık (2004), tarafından yapılan araştırmada Türkiye de okullarda uygulanan ölçme ve değerlendirmeye yönelik çalışmaların önemli ölçüde yetersiz olması nedeniyle araştırmada günümüz değerlendirme sisteminin eksik yönleri gösterilerek bu noktada yeni bir yaklaşım tanıtılmaya çalışılmıştır. Adının “ Alternatif Model “ olarak ifade edildiği ve geleneksel değerlendirmeden farklı olan bu model, klasik yöntemdeki öğretmen merkezli eğitimi hedeflemektedir. Bu yaklaşımda önerilen ve gündeme getirilen fikirlerin, eğitim sisteminin yeniden şekillendirilmesinde bir kilometre taşı olacağı düşünülmektedir. Araştırmaya göre matematik eğitiminde ve ölçme ve değerlendirme yöntemlerinde her zaman eski yöntemlerin iyileştirilmesi ve gerek duyulması durumunda yeni yöntemlerin oluşturulması gerekmektedir.

Bunun için;

1. Matematik eğitiminde öğrenciye kazandırılmak istenen hedef ve davranışlarla neyin, nasıl değerlendirileceği konusundaki çalışmaların matematiğin doğasına aykırı olmadan yeniden ele alınması gerekir.
2. Açık sınav yöntemi tercih edilmelidir. Bu tür bir değerlendirmenin eğitim ve öğretime zarar verici bir yanı yoktur. Aksine geleneksel değerlendirme sisteminin getirdiği korku, öfke ve kaygı gibi olumsuzluklar bu yöntem ile ortadan kaldırılmış olur.
3. Matematik eğitimindeki değerlendirme sonuçlarına mutlak doğrular olarak bakmak ve çok çabuk sonuç almayı beklememek gerekir. Burada öğrencinin matematik öğrenirken bulunduğu psikolojik durum, bireysel farklılık ve faktörler dikkate alınmalıdır.
4. Ölçme, objektif ve öğrencinin özgüvenini kazanmaya yönelik olmalıdır. Çünkü öğrencinin özgüveni, “zor ders” olan matematik dersinin eğitiminde önemli bir faktör olmaktadır.
5. Ölçme ve değerlendirme süreci mutlaka süreklilik göstermelidir.
6. Eğitim sisteminin yapısı yeniden ele alınarak ders içeriklikleri günün koşullarına göre düzenlenmelidir. Çünkü ancak bu yolla sistemdeki yanlışlıklar ve onların verdiği zararlar ortadan kaldırılarak eğitim sürecine yeni bir dinamizm kazandırılmış olur.

Geleneksel öğretim yöntemi ile öğrenci katılımlı öğretim yöntemi arasında öğrenci başarısına etkisi açısından bir farklılığın bulunup bulunmadığını kontrol gruplu ön-test son-test desenini kullanarak araştıran Teker (1990), öğrenci katılımlı öğretimin öğrenci başarısını geleneksel öğretime göre daha fazla etkilediğini ortaya çıkarmıştır. Araştırmacı, öğrenci katılımlı öğretim yönteminin uygulanmasında karşılaştığı problemlere de değinerek yeni öğretim yöntemi konusunda daha çok çeviri ve araştırma yapılmasını önermektedir.

Taşlı (1997) tarafından, "Coğrafya öğretiminde öğrenci-merkezli aktif yöntemlerin öğretmen-merkezli pasif yöntemlere karşı bir üstünlüğü var mıdır?" sorusuna yanıt bulmak amacıyla Demirci Eğitim Fakültesi Sınıf Öğretmenliği bölümünden 78 öğrenci ile kontrol gruplu ön-test son-test desenini kullanarak deneysel bir çalışma yapılmıştır. Araştırmada deney grubundaki öğrencilerin gruplara ayrılarak coğrafya

dersindeki konuları paylaşmış ve araştırmıştır. Ayrıca, deney grubuna öğrenci kılavuzları verilmiş ve öğrencilerin dersteki etkinliklerini bu kılavuzlara uyarak gerçekleştirmiştir. Ön-test ve son-test puanları arasında anlamlı bir fark olup olmadığını öğrenmek için uygulanan t-testi sonucunda anlamlı bir fark olmadığı belirlenmiştir.

Eğitimi Araştırma ve Geliştirme Merkezi (The Center for Educational Research and Innovation - CERİ) tarafından sekiz ülkedeki araştırmacılarla sınıf gözlemleri ve öğrencilerle görüşme yöntemlerinin uygulandığı bir çalışma yapılmıştır. Araştırma kapsamında İngiltere, Avustralya, Birleşmiş Milletler, Almanya, Danimarka, Finlandiya, Hollanda ve İspanya'da toplam 74 sınıftaki öğretmen ve öğrenciler gözlenmiştir. Bu çalışma, öğrencilerin hedef belirleme sürecine katılımının etkin öğrenmenin geliştirildiği sınıflarda bile nadiren yapıldığını göstermektedir. Öğrencilerle yapılan görüşmeler öğrencilerin çoğunun kendi bilgi ve beceri düzeylerini birbirlerine sorarak ya da arkadaşlarıyla karşılaştırarak bulmak için değişik yaklaşımları uyguladıklarını ortaya koymaktadır. Bu araştırmanın bulgularına göre, sınıflarda etkin öğrenmeyi geliştirmenin öğretmenlerin rolü üzerinde açık bir etkisi vardır. Öğretmenler öğrencilere daha çok sorumluluk vermişlerdir. Öğretmenler hedefler, yöntemler ve öğrenme denetimini öğrencilerle araştıran daha demokratik öğretmen rolünü onaylamışlardır. Ayrıca, öğrencilerin değerlendirmeye ekip olarak katıldıkları işbirlikçi değerlendirme süreçleri öğrencinin bireysel öğrenme hedeflerini belirlemesine yardımcı olabilir. Araştırmanın sonucunda; öğrenciler soruların yanıtlarını özgürce bulabilmeli, öğrencilerin merak, ilgi, biliş ötesi becerileri dikkate alınmalı ve okullar öğrenmeyi öğrenme becerisini geliştirmeli önerilerine ulaşılmıştır (Stern ve Huber 1997).

McCombs ve Whisler (1997) üç lisede yaşanan öğrenci-merkezli öğretime doğru değişim sürecini araştırmıştır. Bu okullar 400 öğrencisi bulunan bir kırsal kesim okulu olan Nederland Lisesi, 1200 öğrenciye hizmet veren bir kenar mahalle okulu olan Mundelein Lisesi ve 2500 öğrenciye sahip bir şehir okulu olan Westbury Lisesidir. Araştırmacıların araştırmak için bu okulları seçmesinin nedenleri bu okullarda akademik başarısızlık, okuldan atılma ve okul dışında faaliyetlerle

ilgilenme olaylarının gözlenmesi ve öğretmenlerin öğrenci merkezli ilkelerden yoksun bir eğitim almış olmasıdır. Araştırma kapsamında bu üç lisede öğrenci-merkezli ilkelerin uygulanmasına öncülük eden öğretmenlerin birbirleri ile konuşmaları ve değişim öykülerini anlatmaları istendi. Araştırmanın verileri bu karşılıklı konuşmalardan elde edildi. Bu okullardaki öğretmenlerin değişim sürecindeki ortak yaşantılarından biri eğitim programını hazırlarken öğrencilerin görüşlerini almaya gereksinme duyduklarını fark etmeleridir. Ayrıca, öğretmenler öğrenci merkezli uygulamalara yönelmenin her birey ve tüm okulun tüm öğrenciler için en iyiyi istemesi ve tüm öğrencilerin başarmayı istemesiyle başladığını belirtti. Araştırmanın sonunda öğretmen eğitimi programlarının öğrenci-merkezli öğretim ilkelerini öğretmesi gerektiğine karar verilmiştir.

Öğrencilerin öğretmen ve derse ilişkin puanlarını geleneksel, derse dayalı öğretim ile öğrenci-merkezli öğretim yöntemlerinin nasıl etkilediğini bulmayı amaçlayan bir çalışmada 241 kişilik bir proje grubu ve 217 kişilik bir karşılaştırma grubu yer almıştır. Her iki grup da 10'ar alt gruba ayrıldı. Muhasebe dersinde bu yarıyıl boyunca proje grubunun öğretiminde işbirlikçi öğrenme, ikili, grup ve sınıf tartışmaları ve projeye dayalı öğrenmeyi içeren öğrenci merkezli teknikler kullanıldı. Karşılaştırma grubundaki öğrenciler ise daha önce kullanılan geleneksel, derse dayalı yöntemler ile öğrenim gördü. İki gruptaki öğrencilere de öğretmen ve dersi değerlendirmeleri için 12 nitelikten oluşan beşli Likert ölçeği türünde bir değerlendirme formu verildi. Bu formu kullanarak öğretmenlerini ve dersi değerlendirmeleri istendi. Araştırmada elde edilen verilerin sonuçlarına göre, proje ile karşılaştırma grubu öğretmenleri arasında hazırlık, sunu, konuşma ve yazma, ilginçlik, sorular, isteklilik ve puanlama nitelikleri açısından 0.01 düzeyinde anlamlı bir fark bulunmuştur. İki grup arasında anlamlı fark çıkmayan nitelikler ise etkili olarak kullanılan ders süresi, ödev, test etme ve politikalar (p<.05). Araştırmada öğrencilerin öğrenci-merkezli bir ortamda öğrenmeyi geleneksel derse dayalı bir sınıfta öğrenmeye yeğlediği de bulunmuştur (Keller, Chris ve Thompson 1999).

Hollanda'nın iki büyük kentindeki dört okulda gerçekleştirilen dört nitel çalışmayı içeren bir araştırmada öğretmen ile öğrencilerin öğrenmeye ilişkin görüşleri,

öğretmenin görüşleri ile davranışları, öğrencilerin görüşleri ile davranışları ve öğretmen ile öğrencilerin davranışları arasında karşılaştırmalar yapılmıştır. Araştırmaya katılan öğrenciler Almanca öğrenmek için okula devam eden toplam 23 yetişkin kadındır. Araştırmaya katılan öğretmen sayısı ise dördütdür. Her çalışmada birden çok veri toplama yöntemi kullanılmıştır. Bunlar öğretmen ile görüşme, öğrencilerle (tesadüfen seçilen beş kişi) görüşmeler ve iki sınıf gözlemidir. yapılan çözümlemeler sonucunda bilginin kullanımı ya da uygulanmasının bütün öğretmenler için önemli bir öğrenme modeli olduğu görülmüştür. Ayrıca ideal öğrenciyi öğrencilerin öğretmenlere göre daha edilgen ve tereddütlü sözcüklerle tanımladığı görülmüştür. Öğrenciler öğretmenlerin öğrenme sürecinde merkezi bir rol oynaması gerektiğini düşündüklerini bildirmişlerdir. Bununla birlikte öğretmenlerin öğrencilerden daha etkin bir tutum beklediği fakat onları derste çok edilgen gördüğü anlaşılmıştır. Araştırmacılar öğrencilerin bu görüşünün kültürel özgeçmişlerinden kaynaklanabileceğini belirtmektedir. (Lunenberg ve Volman 1999).

Illinois'te Tech Prep Projesine katılan yönetici, öğretmen, danışman ve işçilerle yapılan dört çalışmanın sonucunda öğretmenlerin gerçek dünyada uygulama yapmaları, öğrencilerin bireysel ya da işbirliği içinde çalışma becerilerini geliştiren öğrenci-merkezli öğretim yaklaşımını kullanmayı öğrenmeleri ve bu yaklaşımın örnek olunarak öğretilmesi gerektiği bulguları elde edilmiştir (Roegge, Wentling ve Bragg 1999).

Sorunla doğrudan ilişkili araştırmanın yapılmaması ve dolaylı yapılan araştırmalarında konuyu detaylı olarak açıklamadaki yetersizlikleri farklı puanlama araçlarının işlevleri üzerine araştırma yapmanın gerekliliğini ortaya koymaktadır. Dereceli puanlama anahtarı ve puanlama anahtarının karşılaştırılması bu işlevin temel boyutlarından birini temsil etmektedir.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, araştırmanın yapılacağı grup, verilerin elde edilmesi ve verilerin analizi üzerine durulmuştur.

Araştırmanın Türü

Değişen değerlendirme yaklaşımına yönelik kullanılan değişen puanlama yöntemlerinin, öğrencilerin başarı, performans ve derse yönelik düşünceleri üzerindeki etkilerini araştırmak üzere; farklı puanlama yöntemlerinin bu durumlar üzerindeki etkilerinin ve ilişkilerinin araştırılmasına yönelik bir araştırma olduğundan ilişkisel araştırma yapılmıştır.

Çalışma Grubu

Araştırmanın çalışma grubunu, Ankara'nın Çubuk ilçesinde bulunan Atatürk Meslek Yüksekokulu'nun makine bölümünün dört şubesindeki 1. sınıf öğrencileri oluşturmaktadır.

1. şubede 54, 2. şubede 49, 3. şubede 47 ve 4. şubede 45 öğrenci bulunmaktadır.

Araştırma farklı puanlama yöntemlerinin öğrencilerin başarı, performanslarını belirlemeye dayalı bir araştırma olduğundan araştırmaya uygun çalışma grubu olarak 195 kişi seçilmiştir. Araştırmacı dört şubenin öğrencilerine yönelik ortak bir vize sınavı uygulamış ve elde edilen sonuçlar araştırmacıya teslim edilmiştir. Buna göre toplam 195 yazılı sınav sonucu araştırmacıda mevcut olmuştur.

Veri Toplama Araçları

Bu araştırmada veri toplama araçları aşağıda listelenmiştir.

- Yazılı yoklama (Ek: 1)

Dört şubenin tüm öğrencileri aynı zamanda ve aynı sorularla yazılı yoklama sınavı almışlardır.

- Puanlama Anahtarı (Ek: 2)

Hazırlanan değerlendirme sorularının puanlaması için bir puanlama anahtarı hazırlanmıştır. Puanlama anahtarı hazırlanırken öncelikle matematik vize sınavında sorulan soruların her biri için uzman desteğiyle çözüm basamakları belirlenmiş ve her sorunun çözüm basamaklarına puan atanmıştır. Böylece, her soru için ölçme kuralı belirlenmiş ve öğrencilerin yazılı kağıtları bu puanlama anahtarına göre puanlanmıştır. Öğrencinin puanlama anahtarından alacağı en düşük puan 0 ve alacağı en yüksek puan 100'dür. Puanlama anahtarında her sorudaki puanlama basamaklarına verilen puanların eşit olmamasından dolayı α güvenirlik katsayısını puanlama anahtarıyla hesaplamanın uygun olmamasına rağmen dereceli puanlama anahtarıyla karşılaştırmaya gitmek için puanlama anahtarından elde edilen puanlarla yapılan istatistiksel işlemler sonucunda Cronbach α güvenirlik katsayısı 0,756 bulunmuştur. Ayrıca puanlama anahtarının geçerliliği ile ilgili 6 uzman görüşü alınmıştır.

- Dereceli puanlama anahtarı (Ek: 3)

Hazırlanan değerlendirme sorularının puanlaması için bir dereceli puanlama anahtarı hazırlanmıştır. Dereceli puanlama anahtarı hazırlanırken daha önce matematik dersinin değerlendirildiği dereceleme ölçeklerinden faydalanarak ve uzman desteği alınarak her sorunun çözümü için 4 tane ölçüt yazılmıştır. Hazırlanan her ölçütün icra edilme derecesi için (1-4) arası 4 ayrı performans düzeyi belirlenmiştir. Böylece, her soru için elde edilen ölçütlerle ölçme kuralı belirlenmiş ve öğrencilerin yazılı kağıtları dereceli puanlama anahtarına göre puanlanmıştır. Öğrencinin 1 sorudan alacağı en yüksek puan 4 ölçütün her birinden en yüksek 4 puan aldığı düşünülürse $4 \times 4 = 16$ puan olacaktır. Böylece, yazılı sınavda 10 soru bulunduğundan öğrencinin alacağı en yüksek toplam puan $10 \times 16 = 160$ puan ve en düşük puan ise 0 olacaktır. Puanlanan her öğrencinin notu (0–100) arası not sistemine dönüşüm formülüyle dönüştürülmüştür.

X: Öğrencinin dereceli puanlama anahtarından aldığı puan

T: Öğrencinin (0-100) arası not sistemine dönüştürülmüş puanı

$T = \frac{100 * X}{160}$,dereceli puanlama anahtarından elde edilen notu dönüştüren dönüşüm formülü

Bu formülden elde edilen puanlar; puanlama ve puanlayıcı güvenirliğinin hesaplanmasında, diğer puanlama yöntemleriyle korelasyonların hesaplanmasında, dereceli puanlama anahtarından elde edilen puanlar ortalamasıyla diğer puanlama yöntemlerinden elde edilen puanların ortalaması karşılaştırılırken kullanılmıştır.

Alfa güvenirliği hesaplanırken her soru için puanlar dönüştürülmemiştir. Dereceli puanlama anahtarından her soru için elde edilen puanlarla yapılan hesaplamalar sonucunda elde edilen Cronbach α güvenirlik katsayısı 0,778 bulunmuştur. Ayrıca, ölçek geçerliliği ile ilgili 6 uzman görüşü alınmıştır.

- Öğretmen değerlendirme sonuçları

Öğretmenlerin puanlama anahtarı ve dereceli puanlama anahtarıyla öğrencilerin yazılı sınav sonucunu puanlamasından elde ettikleri puanlar.

- Öz değerlendirme sonuçları

Öğrencilerin dereceli puanlama anahtarıyla kendi yazılı sınav sonucunu kendi kendilerinin puanlayarak elde ettikleri puanlar.

- Akran değerlendirme sonuçları

Öğrencilerin dereceli puanlama anahtarıyla arkadaşının yazılı sınav sonucunu puanlayarak elde ettikleri puanlar.

- Anket (Ek: 4)

Farklı puanlama yaklaşımlarının uygulandığı tüm şubelerdeki öğrencilere bu farklı puanlama yaklaşımları sorulmuştur. Bunun için açık uçlu sorulardan oluşan bir anket hazırlanmıştır. Öğrenciler cevaplarını yazılı olarak vermişlerdir.

Verilerin Toplanması

Araştırmanın verileri Meslek Yüksekokulu makine bölümü matematik vize sınavıyla elde edilmiştir. Veriler üç aşamada toplanmıştır.

Verileri toplamanın ilk aşamasında aynı şartlar altında matematik vize sınavı yazılı yoklaması yapılmış dört şubenin de katıldığı bu sınava toplam 195 kişi katılmış ve bu sınav sonunda elde edilen yazılı yoklama kâğıtları toplanmıştır. Toplama işleminden sonra bu kâğıtlar araştırmacı, puanlayıcılar(öğretmenler) tarafından önce puanlama anahtarı sonra da dereceli puanlama anahtarıyla puanlanması sonucunda elde edilen puanlar analiz için kaydedilmiştir.

Verileri toplamanın ikinci aşamasında yapılan yazılı sınavdan bir hafta sonra öğrencilere yazılı kâğıtları getirilmiş ve yazılı yoklamada sorulan sorular tahtada çözülmüştür. Sorular çözüldükten sonra öğrencilere dereceli puanlama anahtarıyla puanlama yöntemine yönelik yarım saatlik bir seminer verilmiştir. Seminerin içeriği; dereceli puanlama anahtarında bulunan her ölçütü öğrencilerin öğrenmesini sağlamaya ve onların gerek ölçütlerin bilincinde sınava hazırlanmaları gerekse öz değerlendirme ve akran değerlendirme yapmalarını sağlamaya yöneliktir. Seminer sonrası öğrencilere sınav kâğıtları ve dereceli puanlama anahtarı dağıtılmış ve bu kâğıtlara her öğrenciye adları ve soyadlarının yazılması söylenmiştir. Öğrenciler ad ve soyadlarını yazdıktan sonra öz değerlendirmeye başlamışlardır. Öz değerlendirmeden sonra öğrencilere yazılı kâğıtlarını arkadaşlarıyla değiştirmeleri söylenmiş ve tekrar yeni dereceleme ölçek kâğıtları dağıtılmıştır. Bu sefer öğrenciler kâğıtlara adlarını ve soyadlarını yazmamaları söylendikten sonra öğrenciler akran değerlendirmeye başlamışlardır. Burada akran değerlendirme çalışmasına toplam 195 öğrenci katıldığından akran değerlendirme için, 195 tek sayı olduğundan eşleştirme oluşturmada sorun yaşanmıştır. Bu sebepten dolayı, bir öğrencinin yazılı

kâğıdı öğrencilerden biri tarafından puanlanmıştır. Bu durumda öğrencilerden bir tanesi iki yazılı kâğıdını puanlamıştır. Böylece, öz değerlendirme ve akran değerlendirmeden elde edilen puanlar analiz için kaydedilmiştir.

Verileri toplamanın üçüncü aşamasında öğrencilere açık uçlu soruları kapsayan anket uygulanmıştır. Anket uygulaması dönem sonu ulaşılan 56 öğrenci üzerinde gerçekleştirilmiştir. Öğrenciler ankette bulunan açık uçlu sorulara yazılı yanıt vermiştir. Elde edilen sonuçlar okunarak, yorumlanarak kaydedilmiştir.

Tablo 5

Çalışma grubuna alınan şubeler ve bu şubelerde uygulanan değerlendirmelere ve ankete katılan öğrenci sayıları

	Puanlama anahtarı	Dereceli puanlama anahtarı	Öz değerlendirme	Akran değerlendirme	Anket
1. Şube	54	54	54	54	12
2. Şube	49	49	49	49	15
3. Şube	47	47	47	47	18
4. Şube	45	45	45	45	11

Meslek Yüksekokulunda birinci sınıflardan çalışma grubuna alınan 4 şube vize sınavına girdikten sonra elde edilen yazılı kâğıtları dört farklı puanlama tekniği ile puanlanmıştır. Bununla birlikte, dönem sonunda öğrencilere anket uygulaması yapılmıştır.

Tabloda bu değerlendirme süreçlerine katılan 195 öğrenci sayısını ve ankete katılan 56 öğrenci sayısını vermektedir. Değerlendirme süreçlerine yeterince öğrenci katılmasına rağmen ankete yeterince öğrenci katılamamıştır. Bunun nedeni, Okulda okuyan öğrencilerin iş hayatlarını okulla birlikte devam ettirmeleridir. Ancak sınavlarda öğrencileri bir araya toplamak mümkün olmuştur. Anketin dönem sonu yapılmasından dolayı yeterince öğrenciye ulaşamamıştır.

Verilerin Çözümlemesi

Araştırmanın alt problemlerine çözüm bulmak amacıyla kullanılan teknikler farklılık göstermektedir. Alt problemlerin niteliğine göre, çeşitli istatistiksel teknikler ve bazı nitel yöntemler kullanılmıştır. Bu teknikler aşağıda alt problemlerin sırasına göre açıklanmıştır.

Birinci alt problemde öğrencilere uygulanan yazılı sınav sonuçlarının puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanması sonucunda elde edilen başarı puanları arasında fark olup olmadığı test edilmiştir. Her iki puanlama için uygulama aynı evrenden olduğundan, örneklemelerin evren varyanslarının aynı olacağı varsayımının yanı sıra örnekleme alınan grubun örneklem hacmi yönünden yeterince geniş olması nedeniyle iki puanlama yöntemi arasında ortalama başarı yönünden anlamlı fark olup olmadığını test etmek için ANOVA(Varyans analizi) yapılmıştır.

İkinci alt problemde öğrencilere uygulanan yazılı sınav sonuçlarının puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanması sonucunda iki farklı puanlama aracından elde edilen madde puanları SPSS paket programından faydalanarak korelasyon katsayısı hesaplanmış ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır.

Üçüncü alt problemde öğrencilere uygulanan yazılı sınav sonuçlarının puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanması sonucunda elde edilen madde puanları kullanılarak iç tutarlılık ölçüsü (Crombach Alfa Güvenirliği) excell paket programında alfa güvenirlik katsayısının formülü kullanılarak hesaplanmış ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır.

Dördüncü alt problemde öğrencilere uygulanan yazılı sınav sonuçlarının puanlama anahtarıyla puanlanmasının yanı sıra, dereceli puanlama anahtarıyla puanlaması yapılırken iki farklı puanlayıcı hem puanlama anahtarıyla hemde dereceli puanlama anahtarıyla puanlama yapmıştır. İki farklı puanlayıcı tarafından puanlama anahtarıyla yapılan puanlama ve iki farklı puanlayıcı tarafından dereceli puanlama anahtarıyla

yapılan puanlama sonuçları SPSS paket programından faydalanılarak korelasyon katsayıları bulunup ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır.

Beşinci alt problemde puanlayıcının puanlama anahtarını ve dereceli puanlama anahtarını kullanmasıyla elde edilen puanlarla birlikte öğrencilerin öz değerlendirme ve akran değerlendirme yapımları sonucunda puanlar elde edilmiştir. Her puanlama yönteminden elde edilen puanların diğer üç puanlama yönteminden elde edilen puanların ortalamasıyla arasındaki korelasyonu SPSS paket programından faydalanılarak bulunmuş ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır.

Altıncı alt problemde ise nitel veri toplama yöntemi uygulanmıştır. Nitel araştırma yöntemleri, nicel verilere derinlik, ayrıntı ve anlam kazandırmak amacıyla çok etkili bir biçimde kullanılabilir(Şimşek, 2004;Yıldırım, 2004). Bu amaçla, öncelikle açık uçlu sorulardan oluşan bir anket hazırlanıp öğrenci grubuna diğer uygulamalardan sonra uygulanmış ve uygulama sonunda, belirli anket maddelerinde belirli yığılma olup olmadığına bakmak için frekans tablosu oluşturulmuş ve tablo yardımıyla anket sonuçları yorumlanmıştır.

BÖLÜM III

BULGULAR VE YORUMLAR

Puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirme elde edilen puanların ortalamalarının karşılaştırılarak incelenmesi

1.Alt Problem: Puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirme ve akran değerlendirme kullanılarak elde edilen puanların ortalamaları arasında anlamlı fark var mıdır?

Aynı yazılı yoklamanın uygulandığı öğrencilerin puanları hesaplanırken iki farklı puanlama aracı uygulanmış bu puanlama araçları puanlama anahtarı ve dereceli puanlama anahtarıdır. Puanlama anahtarı puanlayıcı tarafından puanlamada kullanılmış, dereceli puanlama anahtarı ise puanlayıcı, öğrencinin kendisi ve öğrencinin arkadaşı tarafından puanlanmıştır. Toplam dört puanlama yönteminden elde edilen puanlar için ayrı ayrı betimsel istatistikler hesaplanmıştır. Sonuçlar Tablo 6'daki gibidir.

Tablo 6

Puanlama anahtarı ve dereceli puanlama anahtarından elde edilen sonuçların karşılaştırılması ve bu araçların kullanımı sonucu elde edilen betimsel istatistikler

Puanlama Yöntemleri	N	$\bar{X}$	S_s	S_H	Alt sınır	Üst sınır
Puanlama Anahtarı (P.A)	195	35,46	19,5	1,40	32,69	38,229
Dereceli puanlama anahtarı (D.Ö)	195	40,95	18,6	1,33	38,32	43,581
Öz Değerlendirme (Ö.D)	195	62,28	21,8	1,56	59,20	65,364
Akran Değerlendirme (A.D)	195	60,22	21,6	1,55	57,15	63,284

Tabloda verilen betimsel istatistikler dört farklı değerlendirme yönteminden elde edilmiştir. Bunlardan ortalamaların güven sınırları daha detaylı incelenirse; %95 güven düzeyinde ortalamaların güven aralığı yönünden inceleme yapıldığında alt ve üst sınırlar arasındaki en büyük fark standart hatası en büyük olan; öğrencilerin kendi sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde edilen puanların ortalamasının güven aralığında çıkmıştır. En küçük fark ise standart hatası en küçük olan; puanlayıcının dereceli puanlama anahtarıyla puanlaması sonucunda elde edilen puanların ortalamasının güven aralığında çıkmıştır. Ayrıca öğrencilerin yaptıkları puanlamadaki standart hataların büyük çıkmasından dolayı öğrencilerin yaptıkları puanlamadaki puanların ortalamalarının güven aralığı, puanlayıcının yaptığı puanlamadaki puanların ortalamalarının güven aralığından büyük çıkmıştır.

Varılan bu sonuca göre puanlama yöntemlerinden elde edilen puanların ortalamalar arası farklarının anlamlılığını daha detaylı test etmek için bağımlı örneklem için tekrarlı verilerde kullanılan tek faktörlü varyans analizi yapılmıştır.

Tablo 7

Farklı puanlama yöntemleriyle puanlama sonucunda elde edilen puanların ortalamaları arasında anlamlı bir farklılık olup oluşmadığını belirleme

Varyansın kaynağı	Kareler toplamı	S_d	Kareler ortalaması	F	P
Puanlama yöntemleri	98914,863	3	32971,621	76,773	0,000
Hata	333268,788	776	429,470		
Toplam	432183,651	779			

*p<0,05

Varyans analizi göstermektedir ki, incelenen dört farklı puanlama yöntemlerinden elde edilen puanların ortalamaları arasında anlamlı bir fark olduğu gözlenmiştir. Bu farklılığın hangi puanlama yöntemlerinde meydana geldiğini incelemek için Tablo 8'de ikili karşılaştırmalar yapılmıştır.

Tablo 8
Uygulanan Farklı Puanlama Yöntemleri İçin İkili Karşılaştırmalar

(I) Faktör1 Faktör1	(j)	Ortalama Fark	p
P.anahtarı	D.Ö	-5,490(*)	0,000
Değ	Öz	-26,821(*)	0,000
A.Değ		-24,760(*)	0,000
D.ölçeği	Öz	-21,331(*)	0,000
değ		-19,269(*)	0,000
A.değ			
Öz değ		2,062	0,141
A.değ			

İkili karşılaştırma sonuçlarına göre en büyük farklılık, puanlayıcının puanlama anahtarıyla puanlama sonucunda elde ettiği puanların ortalaması ile öğrencilerin kendi sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde edilen puanların ortalaması arasında olmuştur. En küçük fark ise öğrencilerin kendi sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde ettikleri puanların ortalaması ile öğrencilerin akranlarının sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde ettikleri puanların ortalaması arasında olmuştur. Sonuç olarak ($\alpha=0,05$) anlam düzeyinde yalnızca öğrencilerin kendi sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde ettikleri puanların ortalaması ile öğrencilerin akranlarının sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde ettikleri puanların ortalaması arasında anlamlı bir fark gözlenmemiştir. Karşılaştırma daha detaylı olarak ele alınırsa; puanlayıcının puanlama anahtarından elde ettiği puanların ortalaması ile dereceli puanlama anahtarından elde ettiği puanların ortalaması arasında da anlamlı fark olduğu gözlenmektedir. Ulaşılan bu sonuçlara göre, puanlayıcının dereceli puanlama anahtarından elde ettiği puanların ortalaması, puanlama anahtarında elde ettiği puanların ortalamasından büyüktür. Bunun sonucunda, dereceli puanlama anahtarı puanlama yaparken daha çok adım içermesinden dolayı daha fazla detaya puan vermeyi mümkün kılmakta duyarlı davranmış ve kısmi bilgiye puan verme olanağı

yaratmıştır. Diğer yandan puanlama anahtarı daha az puanlama adımı içermesinden dolayı dereceli puanlama anahtarına göre duyarsız ve daha yüzeysel puanlama yapmaya neden olmaktadır. Öğrencilerin dereceli puanlama anahtarıyla kendi sınav kâğıtlarını puanlamaları sonucunda elde ettikleri puanlarının ortalaması, arkadaşlarının sınav kâğıtlarını dereceli puanlama anahtarıyla puanlamaları sonucunda elde ettikleri puanların ortalamasından görece olarak fazla çıkmıştır. Fakat ortalamalar arasında anlamlı fark çıkmamıştır ($p > 0,05$). Bunun sonucunda, öğrencilerin kendi kâğıtlarını veya arkadaşlarının kâğıtlarını puanlarken biraz daha fazla puan verdikleri söylenebilir.

Puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirme sonuçları arasındaki ilişkinin incelenmesi

2. Alt Problem: Puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirme ve akran değerlendirme kullanılarak elde edilen puanlar arasında nasıl bir ilişki vardır?

Toplam dört puanlama yönteminden elde edilen puanlar arasındaki korelasyonlara bakılarak, farklı puanlama yöntemleri arasında ilişki araştırılmış, bunun için dört puanlama yönteminin sonuçları arasındaki korelasyonlar hesaplanmıştır. Sonuçlar Tablo 9’da verilmiştir.

Tablo 9

Gruba Uygulanan Yazılı Sınav Sonucunda Kullanılan Puanlama Araçlarıyla Yapılan Puanlama Anahtarıyla Değerlendirme, Dereceli puanlama anahtarıyla Değerlendirme, Öz Değerlendirme, Akran Değerlendirme Yöntemleri Arasındaki Korelasyonlar

Pearson Korelasyon	N	PA	DÖ	ÖD	AD
Puanlama anahtarı (PA)	195	1.00			
Dereceli puanlama anahtarı (DÖ)	195	0,98**	1.00		
Öz değerlendirme (ÖD)	195	0,74**	0,72**	1.00	
Akran değerlendirme (AD)	195	0,51**	0,48*	0,59**	1.00

*. $P < 0,05$ ** . $P < 0,01$

Tablo 9 incelendiğinde farklı puanlama teknikleri ile elde edilen puanlar arasında manidar korelasyonlar bulunduğu görülmektedir. Puanlama anahtarından elde edilen puanlarla dereceli puanlama anahtarı ve öz değerlendirme puanları arasında, dereceli puanlama anahtarı ile elde edilen puanlarla öz değerlendirme puanları arasında ve öz değerlendirmeden elde edilen puanlarla akran değerlendirme puanları arasında $\alpha = 0,01$ düzeyinde önemli ilişkiler olduğu saptanmıştır. Puanlama anahtarından elde edilen puanlarla akran değerlendirme puanları ve dereceli puanlama anahtarından elde edilen puanlarla öz değerlendirme puanları arasında $\alpha = 0,05$ düzeyinde önemli ilişkiler olduğu bulunmuştur. Puanlama anahtarından elde edilen puanlar ile dereceli puanlama anahtarından elde edilen puanlar arasındaki korelasyonun çok yüksek, puanlama anahtarı puanları ve öğrencilerin dereceli puanlama anahtarı ile kendi kendilerini öz değerlendirme yapmaları sonucunda elde edilen puanlar arasındaki korelasyon yüksek ve puanlama anahtarı puanları ile öğrencilerin arkadaşlarını akran değerlendirme yapmaları sonucunda elde edilen puanlar arasındaki korelasyonun orta düzeyde olduğu söylenebilir. Dereceli puanlama anahtarından elde edilen puanlar ile öğrencilerin dereceli puanlama anahtarı ile kendi kendilerini öz değerlendirme yapmaları sonucunda elde edilen puanlar arasındaki korelasyonun yüksek, dereceli puanlama anahtarından elde edilen puanlar ile öğrencilerin arkadaşlarını akran değerlendirme yapmaları sonucunda elde edilen puanlar arasındaki korelasyonun ise orta düzeyde olduğu yorumu yapılabilir. Son olarak öğrencilerin dereceli puanlama anahtarı ile kendi kendilerini öz değerlendirme yapmaları sonucunda elde edilen puanlar ile öğrencilerin arkadaşlarını akran değerlendirme yapmaları sonucunda elde edilen puanlar arasındaki korelasyon orta düzeydedir.

Varılan bu sonuçlara göre puanlama anahtarı ile dereceli puanlama anahtarından elde edilen puanlar arası korelasyonun çok yüksek çıkmıştır. Bunun sonucunda, puanlama anahtarını da dereceli puanlama anahtarını da aynı puanlayıcının puanlaması ve bu puanlayıcının bu iki puanlama aracını çok iyi kullanması çok büyük etken olarak gözlenmektedir. Ayrıca, bu iki puanlama aracının benzer puanlama basamakları içermesi de ilişkinin yüksekliğinde bir etken olarak görülebilir. Puanlama anahtarı ile

öz değerlendirme uygulaması sonucu elde edilen puanlar arası korelasyon yüksek çıkması sonucu ise puanlayıcının puanlama aracıyla puanlamayı çok iyi yapmasıyla birlikte, öğrencilerin de verilen seminer sonucunda dereceli puanlama anahtarıyla puanlamayı öğrenmesi ve iki puanlama aracının benzer puanlama basamakları içermesi ve ayrıca, dereceli puanlama anahtarının öğrencilerin kendilerini değerlendirirken etkili bir puanlama aracının göstermesi gereken özelliği göstermesi nedeniyle korelasyonun yüksek çıkmasına neden olduğu sonucuna varılmıştır.

Puanlama anahtarı ile öğrencilerin arkadaşlarının sınav kâğıtlarını dereceli puanlama anahtarıyla akran değerlendirme yapması sonucunda elde edilen puanları orta düzeyde çıkmasının nedeninin bazı öğrencilerin arkadaşlarının kâğıtlarını değerlendirirken gelişigüzel puanladıklarından kaynaklı olduğu düşünülmektedir. Bu yargıya destekleyici fikir olarak öğrencilerin öz değerlendirmeleriyle puanlama anahtarı puanları arasındaki korelasyonun yüksek çıkması alınabilir. Öğrenciler kendi kâğıtlarını puanlarken daha özverili davrandıkları söylenebilir. Böylece, yapılan gelişigüzel puanlamanın korelasyon üzerinde olumsuz etkide bulunabileceğinin sonucuna varılabilir.

Puanlama anahtarı ile öz değerlendirme puanları arasında elde edilen korelasyona çok yakın korelasyon dereceli puanlama anahtarı ile öz değerlendirme puanları arasında çıkmıştır. Çıkan bu sonuç, puanlama anahtarı ve öz değerlendirmede olduğu gibidir. Yani puanlayıcının dereceli puanlama anahtarıyla çok iyi puanlama yapması ve öğrencinin de verilen seminer sonucunda dereceli puanlama anahtarıyla puanlamayı öğrenmesi bunun yanında, dereceli puanlama anahtarının etkili bir puanlama araç özelliği göstermesi nedeniyle puanlar arasındaki korelasyonun yüksek çıktığı sonucuna varılabilir.

Puanlama anahtarı ile akran değerlendirme puanları arasına elde edilen korelasyona çok yakın korelasyon dereceli puanlama anahtarı ile akran değerlendirme puanları arasında çıkmıştır. Çıkan bu sonuç puanlama anahtarı ve akran değerlendirmede olduğu gibidir. Yani puanlayıcının dereceli puanlama anahtarını çok iyi kullanması karşısında öğrenciler dereceli puanlama anahtarıyla arkadaşlarının sınav kâğıtlarını

puanlarken dikkatsiz davranmalarından kaynaklı olarak korelasyonun düşük çıktığı sonucuna varılabilir.

Öz değerlendirme ile akran değerlendirme puanları arasındaki korelasyon ortanın üstünde çıkmıştır. Bunun sonucunda öğrencilerin arkadaşlarının kâğıtlarını dereceli puanlama anahtarıyla puanlarken kendi kâğıtlarında yaptıkları puanlama hatasından daha fazla hata yapsalar da öğrencilerin kendi kâğıtlarını ve arkadaşlarının kâğıtlarını puanlarken benzer davranış gösterdiklerinden dolayı korelasyon katsayısının ortanın üstünde çıktığı sonucuna varılabilir.

Yukarıda varılan sonuçlara daha genel ilave yapılacak olursa puanlayıcı puanlama anahtarını ve dereceli puanlama anahtarını kullanmayı iyi bildiği ve puanlama yaparken çok az miktarda puanlama hatası yaptığının sonucuna varılabilir. Öğrenciler ise dereceli puanlama anahtarını puanlayıcı kadar iyi kullanmamakla birlikte kendi kâğıtlarını puanlarken az miktarda puanlama hatası, arkadaşlarının kâğıdını puanlarken çok miktarda puanlama hatası yaptıklarının sonucuna varılabilir.

Kullanılan puanlama araçlarıyla yapılan farklı puanlama yöntemlerinden elde edilen iç tutarlılık ölçütünün incelenmesi

3. Alt Problem: Puanlama anahtarı ve dereceli puanlama anahtarından elde edilen puanların iç tutarlılık ölçüleri (Crombach Alfa Güvenirliği) ne düzeydedir.

Yazılı sınavda toplam 10 soru sorulmuş, her maddeye verilen puanlar dereceli puanlama anahtarı ve puanlama anahtarıyla kaydedilmiştir. Kaydedilen puanlar Excel paket programında (Cronbach Alfa (α) güvenirligi) iç tutarlılık formülüyle hesaplanmıştır. Sonuçlar Tablo 10'da verilmiştir.

Tablo 10

Gruba uygulanan yazılı sınavın puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanması sonucunda elde edilen Cronbach Alfa (α) güvenirliklerinin incelenmesi

	Soru sayısı	N	α
Puanlama	10	195	0,756

anahtarı			
Dereceli puanlama anahtarı	10	195	0,778

Tablo 10 incelendiğinde puanlama anahtarından elde edilen iç tutarlılık ölçüsü olan güvenirlik katsayısı 0,756 bulunduğundan dolayı puanlama anahtarından elde edilen ve sınav sorularının iç tutarlılığını veren güvenirlik katsayısı kabul edilebilir düzeydedir. Dereceli puanlama anahtarından elde edilen iç tutarlılık ölçüsü olan güvenirlik katsayısı 0,778 çıktığından sınav sorularının iç tutarlılığının dereceli puanlama anahtarıyla hesaplanması sonucunda puanlama anahtarında bulunan katsayıdan göreceli olarak büyük ve kabul edilebilir düzeydedir.

Tablo 10'da bulunan alfa değerlerine göre, dereceli puanlama anahtarıyla elde edilen iç tutarlılık güvenirliği ile puanlama anahtarıyla elde edilen iç tutarlılık güvenirliğinin karşılaştırılması sonucunda dereceli puanlama anahtarından elde edilen iç tutarlılık güvenirliği az miktarda yüksek çıkmıştır. Dereceli puanlama anahtarında her sorudaki ölçütler için daha fazla puanlama basamağı varken puanlama anahtarında mevcut olan puanlama basamak sayısı daha az olması nedeniyle, dereceli puanlama anahtarında birimin sayısal büyüklüğü puanlama anahtarındaki birimin sayısal büyüklüğünden daha küçük değer almıştır. Ya da dereceli puanlama anahtarı puanlayıcıyı daha iyi yönlendirmiş olabilir. Sonuç olarak, birimlerin küçük olmasının güvenirliği olumlu yönde etkilemesinden dolayı ve puanlayıcının puanlarken doğru yönlendirilmesinden dolayı dereceli puanlama anahtarından elde edilen iç tutarlılık güvenirliği puanlama anahtarından elde edilen iç tutarlılık güvenirliğinden biraz yüksek çıktığının sonucuna varılmıştır.

İki farklı puanlayıcının yazılı sınavı ve dereceli puanlama anahtarını puanlaması sonucu elde edilen puanlayıcı güvenirliklerinin incelenmesi

4.Alt Problem: Puanlama anahtarı ve dereceli puanlama anahtarıyla elde edilen puanlar için hesaplanan puanlama ve puanlayıcı güvenirlik değerleri ne düzeydedir?

Hem puanlama anahtarı hemde dereceli puanlama anahtarı iki ayrı puanlayıcı tarafından kullanılmıştır. Puanlayıcılar iki puanlama aracını da kullanmayı iyi

bilmektedirler. Puanlayıcı güvenirliği hesaplanırken önce, puanlama anahtarını birinci puanlayıcı puanladıktan sonra ikinci puanlayıcı puanlamış sonrada, dereceli puanlama anahtarını birinci puanlayıcı puanladıktan sonra, ikinci puanlayıcı puanlamıştır. Bu iki puanlama aracının iki puanlayıcı tarafından kullanılması sonucunda her sınav kağıdı için elde edilen puanlardan iki puanlayıcı arasındaki pearson korelasyon katsayısı hesaplanmıştır. Sonuçlar Tablo11’de verilmiştir.

Tablo 11

Yazılı sınavı puanlama anahtarı ve dereceli puanlama anahtarıyla iki ayrı puanlayıcının iki puanlama aracını da ayrı ayrı kullanmasıyla elde edilen puanlayıcı güvenirliklerinin incelenmesi

Pearson Korelasyon	N	I. puanlayıcı PA	II. puanlayıcı PA	I. puanlayıcı DÖ	II. puanlayıcı DÖ
I. Puanlayıcı PA	195	1			
II. Puanlayıcı PA	195	0,995 **	1		
I. Puanlayıcı DÖ	195	0,918 **	0,922 **	1	
II. Puanlayıcı DÖ	195	0,928 **	0,929 **	0,997 **	1

** P<0,01

Tablo 11 incelendiğinde puanlama anahtarının uygulanması sonucunda I. Puanlayıcının puanlamasıyla II. puanlayıcının puanlaması arasındaki puanlayıcı güvenirliği 0,995 bulunmuştur. Dereceli puanlama anahtarının uygulanması sonucunda I. Puanlayıcının puanlamasıyla II. puanlayıcının puanlaması arasındaki puanlayıcı güvenirliği 0,997 bulunmuştur. Sonuçlar betimsel olarak karşılaştırıldığında her iki puanlayıcı güvenirliğinin aynı olduğu söylenebilir. Öte yandan her iki puanlayıcının iki farklı puanlama aracıyla yaptığı puanlamalardan elde edilen puanlar arasındaki korelasyonun puanlama güvenirliği olarak ele alındığı durumda I. puanlayıcının puanlama güvenirliği 0,918 ve II. Puanlayıcının puanlama

güvenirliği ise 0,929 olarak bulunmuştur. Sonuçlar betimsel olarak karşılaştırıldığında her iki puanlayıcının da puanlama güvenirliğinin aynı olduğu söylenebilir. Öte yandan, puanlayıcı güvenirlikleri puanlama güvenirliğinden görece daha yüksektir. Buradan ölçme aracı değiştiğinde puanlayıcının puanlama mantığı değişirken, aynı ölçme aracını farklı puanlayıcılar benzer şekilde kullanabildiği sonucuna varılabilir. Sonuç olarak, her iki puanlama aracının kullanılması sonucu elde edilen puanlama ve puanlayıcı güvenirliklerinin yüksek olduğu söylenebilir.

Puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirme yöntemlerinden elde edilen puanların, puanlar ortalamasıyla korelasyonlarının incelenmesi

5.Alt Problem: Puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirme ve akran değerlendirme sonucunda elde edilen puanların diğer puanların ortalamasıyla korelasyonları ne düzeydedir?

Her teknik için elde edilen puanlarla diğer üç puanlama yönteminden elde edilen puanların ortalamaları arasındaki korelasyonlar hesaplanmıştır. Bu korelasyonlar; bir puanlama yöntemiyle diğer puanlama yöntemlerinin benzerliği olarak kabul edilmiştir. Doğrusal ilişkinin engellenmesi için korelasyonu hesaplanan puanların ortalamadan arındırılması sonucunda düzeltmeler yapılarak daha gerçekçi sonuçlar elde edilmiştir. Elde edilen korelasyonlar her bir puanlama yönteminin diğerleriyle benzerliğini vermektedir. Sonuçlar Tablo 12'deki gibidir:

Tablo 12

Farklı puanlama yöntemlerinin ortalamalarla korelasyonunun incelenmesi

Korelasyon katsayısı	Soru Sayısı	N	r
Puanlama anahtarı	10	195	0,835
Dereceli puanlama anahtarı	10	195	0,812
Öz değerlendirme	10	195	0,788
Akran	10	195	0,578

değerlendirme			
---------------	--	--	--

Tablo 12 incelendiğinde, puanlama anahtarının diğer puanların ortalamasıyla korelasyonu 0,835, dereceli puanlama anahtarının diğer puanların ortalamasıyla korelasyonu 0,812, öz değerlendirme puanlarından elde edilen korelasyon katsayısı 0,788, akran değerlendirme puanlarından elde edilen korelasyon katsayısı 0,578 olarak bulunmuştur. Sonuçlar karşılaştırıldığında ilk üç puanlama yöntemlerinin ortalama puanlarla korelasyonu yüksek çıkarken, dördüncü puanlama yönteminin korelasyonu daha küçük çıkmıştır. İlk üçü kendi arasında karşılaştırıldığında en yüksek korelasyon puanlama anahtarı, sonra dereceli puanlama anahtarı ve en son olarak öz değerlendirmeye yapılan puanlama yöntemi gelmektedir.

Bulunan bu sonuçlara göre puanlar dağılımının benzerliği konusunda ilk iki yöntemin diğer puanlarla benzerliği yüksek ve birbirine yakın çıkmıştır. Bunun sebebi, puanlayıcının puanlama yaparken kullanılan ölçütleri iyi bilmesidir. Öz değerlendirmede ise ortalama ile korelasyonu, ilk iki puanlama yöntemine göre göreceli olarak düşüktü olsa yeterince yüksektir. Bunun nedeni öğrenciler puanlayıcı kadar olmasa da ölçütlerin bilincinde davranmışlardır. Sonuncu puanlama yönteminde korelasyonun diğerlerine göre düşük düzeyde çıkmasının nedeni, öğrenciler akranlarının kâğıtlarını puanladıkları için her ne kadar ölçütlerin bilincinde olsalar da, öz değerlendirmede olduğu kadar özverili davranmamaları neticesinde ortalama ile korelasyonu ortanın biraz üstüne çıkmıştır. Sonuç olarak, ölçütlere uygun puanlama yapılması durumunda puanların güvenilirliğinin daha yüksek çıkacağı sonucuna varılabilir.

6. Alt Problem: Öğrencilerin başarılarını değerlendirmek için kullanılan yazılı sınav kâğıtlarının, puanlama anahtarı ile puanlayıcı tarafından puanlanmasının yanı sıra öğrencilerin yazılı sınav kâğıtlarının puanlayıcı, öğrencinin kendisi, öğrencinin arkadaşı tarafından dereceli puanlama anahtarıyla puanlamasıyla ilgili olarak öğrenciler neler düşünmektedir?

Bu alt problemde, araştırmada kullanılan farklı değerlendirme teknikleri ile tanışan öğrencilerin bu uygulamalara ait görüşleri anket yoluyla belirlenmeye çalışılmıştır.

Araştırma süresince öğrencilerin, yalnız puanlama anahtarıyla puanlanmasıyla uygulanan değerlendirme tekniği ile dereceli puanlama anahtarı kullanılarak puanlayıcı, öğrencinin kendisi ve öğrencinin arkadaşının puanlamasıyla uygulanan değerlendirme tekniklerinin kullanıldığı 2006–2007 eğitim öğretim yılının 1.döneminin, onların bakış açısından bir karşılaştırmasına ulaşmak amaçlanmaktadır.

Aşağıda sırayla bireylere sorulan sorular verilmiş, ardından bireylerin sorulara verdikleri cevaplar açıklanmıştır.

Soru 1: Yazılı sınav kâğıtlarının puanlanmasında dereceli puanlama anahtarı kullanılmasının sana göre kuvvetli ve zayıf yönleri nelerdi?

Dereceli puanlama anahtarının kuvvetli yönleri; öğrencilerin verdikleri cevaplardan çoğunlukla dereceli puanlama anahtarıyla yapılan puanlama tekniklerinden hoşnut oldukları, zevk aldıkları, net bir şekilde anlaşılmaktadır. Puanlayıcının puanlama yaparken daha detaylı puanlamasından dolayı öğrencilerin soruya verdikleri cevabın her adımından puan almaları onları memnun ettiğini, inandırıcı olduğunu ve puanlama anahtarına göre çok iyi bir puanlama tekniği olduğunu ifade etmektedirler. Öğrencilerin büyük çoğunluğu kendi kendilerinin yazılı kâğıtlarını puanlama ve arkadaşlarının yazılı kâğıtlarını puanlarken çok heyecanlandıklarını ve onların puanlamada hak sahibi olmalarının çok demokratik olduğunu ifade etmektedirler. Zayıf yönler ise bazı arkadaşlarının onların kâğıtlarını puanlarken düşük puan takdir ettiklerinden yakınmaktadırlar.

Soru 2: Yazılı sınav kâğıtlarının puanlanmasında dereceli puanlama anahtarıyla puanlama yapmayı öğrenmenin sana faydası ne oldu?

Öğrencilerin çoğundan alınan yanıt sınava yönelik çalışırken uygun metotla çalışarak, soruyu çözerken verecekleri cevap için izleyecekleri adımı öğrenmiş olmalarından girdikleri sınavdan bekledikleri not konusunda daha isabetli olduklarını ifade etmektedirler. Ayrıca öğrencilerin sınav kâğıtlarının puanlama ölçütlerini ve her ölçütün bilgi basamaklarını bilmelerinin sınav sonrası itirazları ortadan kaldırdığını ve sınav sonuçlarının daha güvenilir olarak aydınlandığını ifade

etmektedirler. Puanlama anahtarıyla yapılan puanlamanın nasıl yapıldığını bilmemelerinin onların sınav sonunda güvensizliklerinden kaynaklı olarak itirazlarına neden olduğunu öğretmenleriyle bazen sorun yaşadıklarını söylemektedirler.

Soru 3: Dereceli puanlama anahtarıyla yapılan değerlendirme senin matematik dönem sonu başarını etkiledi mi?

Öğrencilerin büyük bir çoğunluğuna yönlendirilen bu sorunun sonucunda en fazla tekrar eden ifade; “tüm ölçütleri öğrenmeleri sayesinde derse olan yakınlıklarının arttığı ve çalışmalarına düzen geldiği” ifadesidir. Bununla birlikte soruları yanıtlarken nasıl başlayıp nasıl devam edeceklerinin ve ne yaparlarsa kaç puan alacaklarının bilgisine sahip olmalarından dolayı sınavda soru çözme tekniğini de kazandıklarından başarılarının olumlu yönde etkilendiğini ifade etmektedirler. Sonuç olarak, öğrencilerin ölçütlerden haberdar olmaları onların cevap verirken doğru yönelmelerini sağlamıştır.

Soru 4: Puanlama anahtarıyla yapılan puanlama konusunda neler düşünmekteisin?

Öğrencilerin geneli puanlama anahtarındaki puanlama yöntemini bilmemelerinden dolayı soruya yanıt verirken hangi adımları izlediklerini bilemediklerini ifade etmektedirler. Bunun sonucunda başarılarının olumsuz etkilendiğini ve sınav sonucu beklentilerinin isabetli olmadığını, böylece oluşan itirazlar sonunda öğretmenle tartışmalar yaşandığını söylemektedirler.

Ulaşılan bulgular aşağıdaki gibi özetlenebilir:

Bu bölümde araştırmada kullanılan veriler analiz edilerek alt problemlerin çözümlerinden elde edilen bulgulardan birincisinde: İncelenen dört farklı puanlama yönteminden elde edilen puanların ortalamalarının karşılaştırılması sonucunda, sadece öz değerlendirme ve akran değerlendirme puanlarının ortalaması arasında anlamlı bir fark bulunmamıştır. Diğer yöntemlerin ortalamaları arasında anlamlı fark bulunmuştur.

İkinci alt problemde elde edilen bulgu: Puanlama anahtarı ve dereceli puanlama anahtarı kullanarak yapılan dört farklı puanlama yönteminde elde edilen puanlar arasındaki ilişkinin incelenmesi sonucunda, $\alpha=0,05$ manidarlık düzeyinde tüm puanlama yöntemleri arasındaki ilişki anlamlı çıkmıştır. Diğer yandan, $\alpha=0,01$ manidarlık düzeyinde puanlayıcının dereceli puanlama anahtarıyla değerlendirmesinden elde edilen puanlarla akran değerlendirmeden elde edilen puanlar arasındaki ilişki anlamsız çıkmıştır.

Üçüncü alt problemde elde edilen bulgu: Puanlama anahtarı ve dereceli puanlama anahtarıyla değerlendirme yapılması sonucunda, elde edilen iç tutarlılık ölçüleri (Cronbach Alfa güvenirliği) arasında puanlama yöntemlerine göre farklılık betimsel olarak karşılaştırılmıştır. her iki yöntemden elde edilen güvenirlik katsayıları birbirine çok yakın çıkmakla birlikte, dereceli puanlama anahtarından elde edilen katsayı biraz fazla çıkmıştır.

Dördüncü alt problemin çözümünde elde edilen bulgu: Puanlama anahtarı ve dereceli puanlama anahtarıyla iki farklı puanlayıcının puanlama yapması sonucunda elde edilen puanlayıcı güvenirlik katsayılarının betimsel olarak karşılaştırılması sonucunda, her iki puanlama yönteminden elde edilen puanlayıcı güvenirlik katsayıları yüksek çıkmıştır. Daha hassas karşılaştırma yapıldığında dereceli puanlama anahtarından elde edilen puanlayıcı güvenirliği biraz fazla çıkmıştır.

Beşinci alt problemde elde edilen bulgu: Dört farklı puanlama yönteminde elde edilen puanların diğer puanlar ortalamasıyla korelasyon katsayılarının hesaplanması sonucunda, Puanlama anahtarı ve dereceli puanlama anahtarıyla değerlendirme, öz değerlendirme sonucunda elde edilen puanların diğer puanlar ortalamasıyla korelasyonu yüksek çıkarken akran değerlendirmeden elde edilen puanların diğer puanlar ortalamasıyla korelasyonu orta düzeyin biraz üzerinde çıkmıştır.

Altıncı alt problemde elde edilen bulgu: Dört farklı puanlama yöntemine ilişkin anket sonuçlarının incelenmesi sonucunda, dereceli puanlama anahtarıyla yapılan değerlendirme, öz değerlendirme, akran değerlendirmeyeyle ilgili olarak öğrenci

görüşleri olumluyken puanlayıcının puanlama anahtarıyla yaptığı değerlendirmeyle ilgili görüşleri olumlu olmamıştır.

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu bölümde, yapılan araştırmanın bulgularından yola çıkarak ulaşılan sonuçlara ve sonuçlara dayalı yapılacak önerilere yer verilmiştir.

SONUÇLAR

Bu araştırmanın başlangıcında ülkemizde genellikle kullanılan eğitim ve ölçme değerlendirme yaklaşımlarından yazılı sınavlar ve projelerin puanlamasında sıklıkla kullanılan puanlama anahtarı ile genellikle yeni değerlendirme yaklaşımı olarak vurgulanan ve aynı amaçla kullanılabilecek dereceli puanlama anahtarının karşılaştırılması amaçlanmıştır. Ulaşılan sonuçlar maddeler halinde verilmiştir.

1) Yazılı sınav sonuçlarının farklı puanlama araçlarıyla yapılan farklı puanlama yöntemlerinin kullanılması sonucunda elde edilen ortalama başarı puanları farklılaşmaktadır. Öğrencilerin kendi kendilerini ve akranlarını puanlamaları sonucunda buldukları ortalama başarı puanı, puanlayıcının dereceli puanlama anahtarıyla bulduğu başarı puanından yüksek çıkmıştır. Öğrenciler kendi kendilerine ve akranlarına puanlayıcının verdiğiinden daha fazla puan vermişlerdir. Ayrıca, öğrencinin kendine yanlı davranmasından dolayı kendisine verdiği puan akranına verdiği puandan daha fazla çıkmıştır.

2) Kullanılan farklı puanlama araçlarıyla yapılan farklı puanlama yöntemlerinden elde edilen puanlar arasında anlamlı ilişkiler bulunmuş olup puanlayıcının puanlama araçlarını iyi kullandığının ve tutarlı puanlama yaptığının kanıtı puanlama anahtarıyla ve dereceli puanlama anahtarıyla yaptığı puanlama arasındaki ilişkinin yüksek çıkmasıyla doğrulanabilir. Fakat öğrencilerin akranlarını ve kendilerini değerlendirmeleri sonuçları için elde edilen ilişkinin düşük çıkması sonucunda ise

öğrencilerin dereceli puanlama anahtarını puanlayıcı kadar iyi kullanamadığı ve tutarlı puanlama yapamadığını gösterebilir. Ayrıca, puanlayıcının öğrencilerden daha tutarlı puanlama yaptığı sonucuna varılmıştır.

3) Yazılı sınavın iç tutarlılık ölçütü alfa katsayısını bulmak için puanlama anahtarı ve dereceli puanlama anahtarı gibi farklı puanlama araçlarının uygulanması sonucunda elde edilen alfa katsayılarının hesaplanması sonucunda her ikisinden elde edilen iç tutarlılık ölçütleri yüksek çıkmıştır. Betimsel olarak yargıda bulunulursa dereceli puanlama anahtarından elde edilen katsayı değeri puanlama anahtarından elde edilen katsayı değerinden biraz yüksek çıkmıştır.

4) İki farklı puanlayıcının dereceli puanlama anahtarı ve puanlama anahtarıyla yazılı sınavı puanlamaları sonucunda puanlayıcıların oldukça tutarlı puanlama yaptığı gözlenmiştir.

5) Puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirme, akran değerlendirme sonuçlarının diğer puanların ortalamalarıyla korelasyonlarının hesaplanması sonucunda elde edilen sonuçlar karşılaştırıldığında puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirmeyle yapılan puanlama yöntemlerinin diğer yöntemlerin ortalamasıyla korelasyonu yüksek çıkarken, akran değerlendirmeyle yapılan puanlama yönteminin ortalamayla korelasyonu daha küçük çıkmıştır. İlk üçü kendi arasında karşılaştırılmaları sonucunda en yüksek korelasyon puanlama anahtarı, sonra dereceli puanlama anahtarı ve en son olarak öz değerlendirmeyle yapılan puanlama yöntemi gelmektedir.

6) Öğrencilere anketle yönlendirilen soruların neticesinde, dereceli puanlama anahtarıyla yapılan puanlama tekniklerinden hoşnut oldukları, zevk aldıklarını yazmaları dereceli puanlama anahtarının öğrenciler üzerindeki olumlu etkisinin kanıtıdır. Puanlayıcının puanlama yaparken daha detaylı puanlamasından dolayı öğrencilerin soruya verdikleri cevabın her adımından puan almaları onları memnun ettiğini, inandırıcı olduğunu ve puanlama anahtarına göre çok iyi bir puanlama tekniği olduğunu öğrencilerin çoğunluğu yazarak ifade ettiklerinden dereceli

puanlama anahtarı kullanmak yararlı olabilir. Öğrencilerin büyük çoğunluğu kendi kendilerinin yazılı kâğıtlarını puanlama ve arkadaşlarının yazılı kâğıtlarını puanlarken çok heyecanlandıklarını ve onlarında puanlamada hak sahibi olmalarının çok demokratik olduğunu ifade etmeleri, hem öğrencilerin birbiriyle olan etkileşimlerini artıracak hemde ölçütlerden haberdar olmalarını sağlayacaktır.

7) Sonuç olarak puanlama anahtarı olarak dereceli puanlama anahtarının seçilmesi hem puanlamanın duyarlılığını artıracak hemde öğrencilerin ölçütlerden haberdar olmasını sağlayacaktır. Çünkü öğrencinin puanlama anahtarındaki ölçütleri anlaması, dereceli puanlama anahtarındaki ölçütleri anlamasından daha zordur. Dolayısıyla dereceli puanlama anahtarı gerek duyarlılığı gerekse kullanım kolaylığıyla daha güvenilir, daha belirleyici ve daha kullanışlıdır. Ayrıca puanlama anahtarı sadece bir tane derse yönelik hazırlanırken dereceli puanlama anahtarı birbirine yakın birden çok derse uygulanabilir olması da kullanışlılığının kanıtıdır.

ÖNERİLER

1) Bu araştırmada, araştırma süreci iki döneme genişletilerek, iki dönemlik yazılı sınavların puanlama araçlarıyla puanlanması sonucunda elde edilen sonuçların karşılaştırılması bu iki puanlama aracı arasındaki farkı karşılaştırmada daha etkili olacaktır.

2) Araştırma sürecinde dereceli puanlama anahtarının kullanımına yönelik öğrencilere yapılan bilgi aktarımı yoğunlaştırılır ve yapılan bu tip çalışmalar zamana yayılırsa öğrencilerin ölçütleri çok iyi kavraması sağlanabilir. Öğrencilerin ölçütleri kavraması yansız puanlamayı sağlayarak öğrencilerin yanlılığını azaltabilir.

3) Dereceli puanlama anahtarının farklı derslerde aynı sonucu verip vermeyeceği, farklılıklar varsa, bu farklılıkları yaratan kaynakların neler olduğunun araştırılması dereceli puanlama anahtarının etkililiği ile ilgili yararlı bir araştırma olabilir. Böyle bir araştırmada kullanılan yeni yöntemlerin, puanlama aşamasının derslere göre

farklılaştığı ve benzerlik gösterdiği yönlerin ortaya konabilmesi, bu yöntemlerin işleyişi ile ilgili veri kaynağı olabilir.

4) Araştırmaya farklı okullardaki aynı düzeyden öğrencilerin katılması ile çalışma grubu genişletilirse, puanlama anahtarı ve dereceli puanlama anahtarının karşılaştırması için daha güvenilir ve genellenebilir sonuçlar elde edilebilir. Böyle bir araştırmada ayrıca genişleyen çalışma grubunun, puanlama araçlarından elde edilen puanlayıcı ve puanlama güvenilirliği, alfa güvenilirliklerine ve diğer puanlar ortalamasıyla korelasyonuna etkisi olup olmadığına bakılabilir.

5) Dereceli puanlama anahtarı çeşitlendirilerek ve standart puanlama anahtarı ölçüt kabul edilerek, yazılı sınavın standart hale getirilmiş puanlama anahtarı ile puanlanması ve çeşitlenen dereceli puanlama anahtarı kullanarak puanlama yapılması sonucunda elde edilen puanlar karşılaştırılabilir. Buradan elde edilecek bulgular, dereceli puanlama anahtarının geçerliliği ile ilgili kanıtlar sağlamada kullanılabilir.

KAYNAKÇA

- AÇIKGÖZ, Ün. K. (2002). *Aktif Öğrenme*. İzmir: Eğitim Dünyası Yayınları.
- AIRSAN, P. W. (1994). *Classroom Assessment*. Second Edition. New York: McGraw Hill.
- ANDERSON, R. S. (1998). *New Directions for Teaching and Learning Changing The Way, We Grade Student Performance: Classroom Assesment and New Learning Paradigm*, Editors: Bruce W. Speck.
- ARSLAN, M. Metin ve ERASLAN, Levent. (2003). Yeni Eğitim Paradigması ve Türk Eğitim Sisteminde Dönüşüm Gerekliliği *Milli Eğitim Dergisi*, sayı: 160, Ankara. www.sosbil.aku.edu.tr/makale/c9s1m6.pdf adresinden erişim 03.02.2008 tarihinde.
- ASLANOĞLU, E. A. (2002). Dereceli Puanlama Anahtarı uygulamasına ilişkin bir araştırma. *Ankara ün. Eğitim Bilimleri Fak Dergisi*, sayı:1-2
- AŞKAR, P. (2004). Matematik Etkinlikleri Bildiriler Kitabı. Ankara: Matematikçiler Derneği Yayınları.
- BERLAK, H.(1992). *Toward the Development of a New Science of Educational Testing and Assesment*. (Sayfa 181-206). Albany: State University of New York Press.
- BROOKHART, S. M. (1999). The Art and Science of Classroom Assessment: The Missing Part of Pedagogy. ASHE-ERIC Higher Education Report (Vol. 27,

No.1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.

BURSTEIN, L. (1991). Performance Assessment for Accountability Purposes. Chicago: Taking the Plunge and Assessing the Consequences. Paper Presented at the Annual Meeting of the American Educational Association.

CONNER, C.(1991). *Assesment and Testing in primary School*. Cambridge: The Falmer Press.

DİKLİ, S. (2003). Traditional and Alternative Assessments. *The Turkish Online Journal of Educational Technology – TOJET*. ISSN: 1303-6521 Volume 2, Issue 3, Article 2

DOĞAN, N. (2007). ‘‘Yazılı Sınavlar’’. (*Eğitimde Ölçme ve Değerlendirme*). Editör: Hakan ATILGAN. Ankara: Anı Yayınları.

ERTÜRK, S. (1998). *Eğitimde Program Geliştirme*. Ankara: Meteksan A.Ş.

FRECHTLING, J.A. (1991). *Performance Assessment: Moonstruck or the Real Thing? Educational Measurement: Issues and Practice*, 10(4), 23-25.

GOODRICH, A. H. (2001). The effects of instructional rubrics on learning to write. *Current Issues ID Education*.

HALADYNA, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*, USA: Altyn & Bacon.

HANNY, R. J. (2000). Assessing the SOL in Classrooms. College of William and Mary. <http://www.wm.edu/education/SURN/solass.html> adresinden erişim 01.03.2008 tarihindedir.

- İLHAN, E.Gül Çelebi.(2006) “Yeni Değerlendirme Yaklaşımlarının Matematik Eğitimindeki Yansımalarının İncelenmesi (Uygulamalı Bir Araştırma).”(Yayımlanmış Yüksek Lisans Tezi). Ankara: Hacettepe Üniversitesi.
- KELLER, B.,C.CHRIS & H. THOMPSON.(1999). ‘*Effects of Student Evaluation in Calculus.*’ Educational Research Quarterly, XXIII, 1, (Sayfa 59-73).
- KOÇ, G. ve DEMİREL, M.(2003). “Davranışçılıktan Yapılandırmacılığa Eğitimde Yeni Bir Paradigma.”*Milli Eğitim Dergisi*.sayı:159
www.buldansempozyumu.com/kitap/2.oturum/5.pdf adresinden erişim 15.03.2008 tarihindedir.
- KUTLU, Ö. (2002). Öğretmen Yetiştirme Programlarının Yeni Ölçme ve Değerlendirme Yaklaşımlarını İçerme Düzeyi. Uluslararası Katılımlı 2000’li Yıllarda I.Öğrenme ve Öğretme Sempozyumu. 29-31 Mayıs, Marmara Üniversitesi, Atatürk Eğitim Fakültesi.
- KUTLU, Ö. (2008). *Öğrenci Başarısının Belirlenmesi.* ‘*Performansa ve Portfolyoya Dayalı Durum Belirleme*’ Ankara: Pegem Yayıncılık.
- LEYDENS, J.L., MOSKAL, B.M. (2000). Scoring Rubric Development: Validity and Reliability. Practical Assessment, Research & Evaluation, 7(10).
[http ://P AREonline.net/getvn.asp?v=7&n= 10](http://PAREonline.net/getvn.asp?v=7&n=10) adresinden erişim 01.04.2008 tarihindedir.
- LUNENBERG, M..L. & M.. VOLMAN. (1999). “*Active Learning: Views and Actions of Students and Teachers in Basic Education.*” *Teaching and teacher Education*, 15, 431-445
- MABRY, L. (1995). Naturally Occuring Reliability. San Francisco: Paper Presented at the Annual Meeting of the American Educational Association.

MABRY, L. (1999). *A Critical Guide to Alternative Assessment*. California: Convin Press INC.

MATHEWS. J. (2004). Portfolio Assessment. Education Next. Summer 2004.

McCOMS, B.L. & J.S. WHISLER.(1997). *The Learner-Centered Classroom and School*. San Fransisco, Jossey-Bass Publishers.

MEB (2003). ‘‘Öğrenci Merkezli Eğitim Uygulama Modeli’’, Milli Eğitim Bakanlığı Yayını, Milli Eğitim Basımevi(Ankara şubesi), kordinatör: Oğuz ERBİL.

MOSS, P.A. (1994). *Can There Be Validity Without Reliability?* Educational Researcher, 23(2), 5-12.

MOSKAL, B.M. (2003). *Classroom Performance Assessments and Scoring Rubrics* Part II. ERIC Digest.

www.eric.ed.gov/ericwebporta adresinden erişim 20.03.2008 tarihindedir.

MOSKAL, B. M. (2000). "Scoring rubrics: What, when and how?" *Practical Assessment, Research & Evaluation*, 7 (3) Available Online: <http://ericae.net/pare/getvn.asp?v=7&n=3> adresinden erişim 01.03.2008 tarihindedir.

PRIESTLEY, M. (1982). *Performance assesment in education and training: alternative techniques*, Nation Evaluation Systems, Inc Education Technology Publications Englewood cliffs, New Jersey, Printed in the United States of America

POPHAM, J.W.(1997). What’s wrong and What’s right with rubric. Educational Leadership. 55, (2), 12

- POYRAZ, S. (2006). İlköğretim Fen Bilgisi Öğretiminde İşbirlikli Öğrenme Yönteminin Kullanıldığı Eğitim Ortamlarında Başarıyı Ölçmede Çoktan Seçmeli Testlerin Diğer Testlere Göre Etkileri Cilt:14 No:2 *Kastamonu Eğitim Dergisi* (sayfa 497-502)
- ROEGGE, C.A., T.L. WENTLING & D.D. BRAGG. (1999). *Using Tech Prep Principles to Improve Teacher Education*.
<http://borg.lib.vt.edu/ejournals/JVTE/v13nl/roegge.html> adresinden erişim 10.03.2008 tarihindedir.
- SENEMOĞLU, N. (2004). *Kuramdan Uygulamaya Gelişim, Öğrenme ve Öğretim* sayfa(86)
- STIGGINS, R. J. (1994). *Student-Centered Classroom Assesment*. New york: Macmillan publishing Company.
- STERN, D.& G. HUBER. (1997). *Active Learning for Student and Teachers: Reports from Eight Countries, Germany*.
- ŞİMŞEK, H. (1997). *21. Yüzyılın Eşiğinde Paradigmalar Savaşı: Kaostaki Türkiye*, İstanbul: Sistem yayıncılık.
- ŞİMŞEK, H., YILDIRIM, A. (2004). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*, Ankara: Seçkin Yayınevi.
- TAŞLI, İ. (1997). ‘‘Öğrenci Merkezli Yöntemlerle Coğrafya Öğretimi.’’ (Yayımlanmış Doktora Tezi), İzmir: Dokuz Eylül Üniversitesi, 1997.
- TEKER, D. (1990). ‘‘Öğrenci Merkezli Öğretim (Saynergoji) Yöntemiyle Geleneksel Öğretim Yönteminin Başarısına Etkisi (Uygulamalı Bir Araştırma).’’ (Yayımlanmış Yüksek Lisans Tezi), Eskişehir: Anadolu Üniversitesi.

TEKİN, H.(1996). *Eğitimde Ölçme ve Değerlendirme*, Ankara: Yargı Yayınları

TEKİNDAL, S. (2006). 2006 – 2007 Eğitim – Öğretim yılı Kutoren İlköğretim Okulu Teknoloji ve Tasarım Dersi I. Dönem Zümre Öğretmenleri Toplantı Tutanağıdır.
www.kocaeli.edu.tr

TURGUT, M. F. (1993). *Eğitimde Ölçme ve Değerlendirme Metodları*. Ankara: Saydam Matbası Dokuzuncu baskı.

TURGUT, M. F. (1997). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Yargıcı Matbası.

YANCEY, K.B. (1999). "Looking back as we look forward: Historicizing writing assessment." *College Composition and Communication*, 50, 483-503.

YILDIZ,İ., UYANIK,N. (2004). “Matematik Eğitiminde Ölçme ve Değerlendirme.” Cilt:12, No:1 *Kastamonu Eğitim Dergisi*.

VERMONT DEPARTMENT OF EDUCATION (1991). *Vermont's Assessment Program*. Montpelier: Author.

WIGGINS, G.P. (1993). *Educative assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass.

Ek 1

Mat-101 Kodlu Dersin Yazılı Sınav Soruları

Bu sorular meslek yüksek okulu öğrencilerine sorulacak. Soruların zorluk derecesi düşük olacak.

SORULAR

Soru 1) a,b,c reel sayılardır. $a^3.b^5 > 0$, $a.c^3 < 0$, $b^3.c^4 > 0$ olduğuna göre a,b,c'nin işaretleri sırasıyla ne olur?

Soru 2) a,b,c sıfırdan büyük birer tamsayı ve $3a+2b+c=87$ olduğuna göre b'nin en büyük değeri nedir?

Soru 3) ab iki basamaklı sayısının rakamları yer değiştirilerek ba iki basamaklı sayısı elde ediliyor. ab sayısı ba sayısına dönüşünce 63 azaldığına göre a-b nedir?

Soru 4) a ve b sıfırdan farklı birer rakam ise $\frac{(aaa + bbb)}{a + b}$ oranı kaçtır?

Soru 5) 123

ab

yandaki çarpma işleminde a'nın değeri nedir?

*

...

246

+

....

Soru 6) $\frac{2}{3} - \frac{2}{3} : \frac{2}{3} + \frac{2}{3} * \frac{2}{3}$ işleminin sonucu nedir?

Soru 7) $\frac{0,51}{0,017} + \frac{0,003}{0,00002} + \frac{3,5}{0,07}$ işleminin sonucu nedir?

Soru 8) Boyutları 18, 24 ve 36 metre olan bir depoya küp şeklinde araç malzeme kutuları yerleştirilmek isteniyor. Bu depoya en az kaç kutu yerleştirilebilir?

Soru 9) $0,000355 = X \cdot 10^{-6}$ olduğuna göre X nedir?
 $355 \cdot 10^{-6} = X \cdot 10^{-6}$ $355 = X$

Soru 10) $\frac{2 \cdot 10^{-4} + 3 \cdot 10^{-5}}{5 \cdot 10^{-6}}$ kesrinin değeri nedir?

$$\frac{20 \cdot 10^{-5} + 3 \cdot 10^{-5}}{50 \cdot 10^{-5}} = \frac{5 \cdot 10^{-5}}{50 \cdot 10^{-5}} = \frac{5}{50} = \frac{1}{10}$$

Ek 2

Mat-101 Kodlu Dersin Yazılı Sınavını Puanlama İçin Hazırlanmış Puanlama Anahtarı

SORULAR

Soru 1) a,b,c reel sayılardır. $a^3.b^5 > 0$, $a.c^3 < 0$, $b^3.c^4 > 0$ olduğuna göre a,b,c'nin işaretleri sırasıyla ne olur?

1.adım $c^4 > 0$,

2.adım $c^4 > 0$ olduğundan $b^3 > 0$ olur. böylece b'nin işareti + olur.

3.adım b + olduğundan $a^3 > 0$ olur. böylece a'nın işareti + olur.

4. adım a + olduğundan $c^3 < 0$ olur. Böylece c'nin işareti – olur.

5.adım sonuç +,+,- olur.

Soru 2) a,b,c sıfırdan büyük birer tamsayı ve $3a+2b+c=87$ olduğuna göre b'nin en büyük değeri nedir?

b'nin en büyük olabilmesi için a ve c'nin en küçük olması gerekir.

a'nın katsayısı daha büyük olduğu için $a=1$ ve $c=2$ diyebiliriz.

$3.1+2b+2=87 \rightarrow 2b+5=87 \rightarrow 2b=82 \rightarrow b=41$ olur.

Soru 3) ab iki basamaklı sayısının rakamları yer değiştirilerek ba iki basamaklı sayısı elde ediliyor. ab sayısı ba sayısına dönüşünce 63 azaldığına göre a-b nedir?

$$ab=10a+b \quad ba=10b+a \quad ab - ba = 63$$

$$(10a + b) - (10b + a) = 63 \rightarrow 9(a - b) = 63 \rightarrow a - b = 7$$

Soru 4) a ve b sıfırdan farklı birer rakam ise $\frac{(aaa + bbb)}{a + b}$ oranı kaçtır?

$$aaa = 100a+10a+a \quad bbb=100b+10b+b$$

$$\frac{(100a+10a+a) + (100b+10b+b)}{a + b} = \frac{111(a + b)}{a + b} = 111$$

Soru 5) $\begin{array}{r} 123 \\ ab \\ * \\ \hline \end{array}$

*

...

246

+

....

$123 . a = 246 \rightarrow b = 2$ dir.

yandaki çarpma işleminde a'nın değeri nedir?

Soru 6) $\frac{2}{3} - \frac{2}{3} : \frac{2}{3} + \frac{2}{3} * \frac{2}{3}$ işleminin sonucu nedir?

$$\frac{2}{3} - 1 + \frac{4}{9} = \frac{6}{9} - \frac{9}{9} + \frac{4}{9} = \frac{10}{9} - \frac{9}{9} = \frac{1}{9}$$

Soru 7) $\frac{0,51}{0,017} + \frac{0,003}{0,00002} + \frac{3,5}{0,07}$ işleminin sonucu nedir?

$$\frac{510}{17} + \frac{300}{2} + \frac{350}{7} = 30 + 150 + 50 = 230$$

Soru 8) Boyutları 18, 24 ve 36 metre olan bir depoya küp şeklinde araç malzeme kutuları yerleştirilmek isteniyor. Bu depoya en az kaç kutu yerleştirilebilir?

$$\begin{array}{r|l} 18 & 24 & 36 & 2 \\ 9 & 12 & 18 & 2 \\ 9 & 6 & 9 & 2 \\ 9 & 3 & 9 & 3 \\ 3 & 1 & 3 & 3 \\ 1 & 1 & 1 & 1 \end{array}$$

(18, 24, 36) ebob = 2.3=6

kutu sayısını bulmak için $\frac{18 \cdot 24 \cdot 36}{6 \cdot 6 \cdot 6} = 3 \cdot 4 \cdot 6 = 72$ tane kutu

Soru 9) $0,000355 = X \cdot 10^{-6}$ olduğuna göre X nedir?

$$355 \cdot 10^{-6} = X \cdot 10^{-6} \quad 355 = X$$

Soru 10) $\frac{2 \cdot 10^{-4} + 3 \cdot 10^{-5}}{5 \cdot 10^{-6}}$ kesrinin değeri nedir?

[illegible][illegible]

Ek 4

Anket Soruları

Soru 1: Yazılı sınav kâğıtlarının puanlanmasında dereceli puanlama anahtarı kullanılmasının sana göre iyi ve kötü yönleri nelerdi?

- Puanlama Anahtarıyla mı yoksa dereceli puanlama anahtarıyla puanlama yapılması sana daha uygun düşer? Niçin?
- Dereceli puanlama anahtarıyla kendi kendini değerlendirme fikri sana göre nasıldı?

Soru 2: Yazılı sınav kâğıtlarının puanlanmasında dereceli puanlama anahtarıyla puanlama yapmayı öğrenmenin sana faydası ne oldu?

- Dereceli puanlama anahtarının sana sağladığı olanaklar yeterli oldu mu? Açıklar mısın?

Soru 3: Dereceli puanlama anahtarıyla yapılan değerlendirme senin matematik dönem sonu başarını etkiledi mi?

- Dönem sonu başarını artırmada dereceli puanlama anahtarının en etkili yönleri nelerdir? Açıklar mısın?

Soru 4: Puanlama anahtarıyla yapılan puanlama konusunda neler düşünmekteisin?

- Puanlama anahtarıyla yapılan puanlamanın senin için dezavantajları nedir? Açıklar mısın?

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı: Yusuf KASAP

Doğum Yeri ve Tarihi: Ankara 01.11.1975

Eğitim Durumu

Lisans Öğrenimi : Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü

Yüksek Lisans Öğrenimi: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Eğitim
Bilimleri A.B.D. Eğitimde Ölçme ve Değerlendirme

Bildiği Yabancı Diller : İngilizce

İş Deneyimi

Çalıştığı Kurumlar : Gazi Üniversitesi

: Kosgeb Hacettepe/Ankara

:Tübitak Hacettepe/Ankara

İletişim

E-Posta Adresi:yusufkasap5806@hotmail.com

Tarih:02.06.2008

