






<http://www.tayjournal.com>

<https://dergipark.org.tr/tr/pub/tayjournal>

Comparison Of The Scores Emerged By Scoring Rubric And Answer Key

 Yusuf Kasap, PhD Student, Corresponding Author
Hacettepe University, Turkey
yusufkasap5806@hotmail.com
Orcid ID: 0000-0002-5114-1175

 Selahattin Gelbal, Prof. Dr.
Hacettepe University, Turkey
sgelbal@gmail.com
Orcid ID: 0000-0001-5181-7262

 Nuri Doğan, Prof. Dr.
Hacettepe University, Turkey
nuridogan2004@gmail.com
Orcid ID: 0000-0001-6274-2016

Article Type: Research Article
Received Date: 08.03.2021
Accepted Date: 11.06. 2021
Published Date: 28.06.2021

Tr/En: Tr

Plagiarism: This article has been reviewed by at least two referees and scanned via a plagiarism software

Citation: Kasap, Y., Gelbal, S. & Doğan, N. (2021). Comparison of the scores emerged by scoring rubric and answer key. *Türk Akademik Yayınlar Dergisi (TAY Journal)*, 5(1), 95-117.

Abstract

In this research, it is purposed to compare the scoring scale and rubric which is widely used in recent years. The study group of the research, which was designed in the correlational survey model, consists of first-year students in four branches, consisting of 195 students who continue their education in the machinery, metal, motor, electricity and electronics departments of Atatürk Vocational School in Çubuk district of Ankara. 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. According to the results of the research, there is a meaningful difference between the assessment results by scoring rubric and answer key. Also, ANOVA (analysis of variance) test was performed for dependent samples to test whether there was a significant difference between scoring methods in terms of average success. 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.

Keywords: Scoring rubric, answer key, self-assessment, peer assessment

Extended Summary

Introduction

Whether new measurement and evaluation approaches will be effective in enabling individuals to use their different abilities and skills effectively and directing them to the right interests is one of the important questions that must be answered in the field of education today. In addition, the evidence regarding the validity of new measurement and evaluation methods should eliminate doubts about the use of these methods and ensure that these methods are clarified so that accurate decisions can be made about individuals using these methods. The problem status of this research is to determine the relationships between the scoring key used to score the written exam used to evaluate student success in the mathematics course, the grading key, the self-assessment and the scores obtained as a result of peer evaluation. In this context, the aim of the research is to investigate the effects of scoring written exams, which are one of the measurement tools used today, with the scoring key and the grading key on mathematics education. When comparing the effects of these evaluation methods, it is aimed to determine which is suitable in terms of revealing individual differences and revealing the knowledge skills and abilities of individuals. The benefit to teachers is that the new methods are tested to see if the claim of superiority over the traditional one occurs in practice; In whichever way they evaluate the written exam, it will be possible for them to make the right decision in deciding that they will obtain a more aroused evaluation result from mistakes. The benefit to the experts is; since they will obtain aroused measurement results from errors in researches, it can be useful to ensure the reliability of the researches to be carried out. The research is thought to be important in terms of guiding teachers, experts, students. When the researches related to this research are examined;

in a study conducted by Aslanoglu ve Kutlu (2004), the benefit of the graded scoring key in measuring and evaluating school learning was investigated.

A similar study was conducted by Dogan and Parlak (2014). In this research, it is aimed to investigate the compliance levels of the scores obtained using the scoring key and the rubric. Another research was conducted by Dogan and Anadol (2018). In this study, the effect of the experience year on using a rubric on scoring reliability was examined. When the relevant researches were examined, the compliance of the rubric and scoring key and the effect of the rubric on the scorer reliability and the benefit of the measurements were investigated. Since the rubric and scoring key are compared and examined in a multifaceted way in this research, the results obtained from the research will give important ideas about the compared measurement tools because the relevant research is inclusive.

Methods

To investigate the effects of the changing scoring methods used in the research for the changing evaluation approach on students' thoughts on success, performance and course; since it is a research to investigate the effects and relationships of different scoring methods on these situations; relational research has been carried out. The working group of the research consists of first-year students continuing their education in four branches in the machinery, metal, motor, electrical and electronic departments of Ataturk Vocational School in The Cubuk district of Ankara. There are 54 students in the 1st branch, 49 in the 2nd branch, 47 in the 3rd branch and 45 students in the 4th branch. The data collection tools in the research are written polling, scoring key, rubric, teacher evaluation results, self-evaluation results, peer evaluation results, survey. The data of the research was obtained by the mathematics visa exam of 195 students continuing their education in the machinery, metal, motor, electrical and electronic departments of vocational school. ANOVA (variance analysis) was performed in dependent sample repetitive data to test whether there was a significant difference in average success between the two scoring methods in the analysis process of the data. Then, as a result of scoring the written exam results applied to the students with PA and DPA, the coefficient of the relationship between the scores obtained from two different scoring tools was calculated and the results obtained were compared and interpreted. In addition, the internal consistency measure (Cronbach Alpha Reliability) was calculated and the results obtained were compared and interpreted using the item scores obtained as a result of scoring the written exam results applied to the students with PA and DPA.

Results

In this study, scoring key evaluation as a traditional evaluation approach is limited to rating key evaluation, self-evaluation, peer evaluation and survey as a new evaluation approach. The evaluation of the students' achievements with the graded scoring key and the evaluation of the students' achievements with the scoring key were compared. When making this comparison, the academic achievements of the students were taken into account. The reliability of these two methods was compared by looking at consistency, sensitivity and compliance between the scores obtained from the scoring key used and the rubric. In addition, with some survey questions, students were asked to write down their thoughts on self-assessment, peer evaluation and teacher evaluation. According to the findings of the study, there is a significant difference between the

scoring key and the evaluation results obtained from the rated scoring key. However, there was no significant difference between peer evaluation results and self-assessment results. The scores obtained from the binary compared evaluations differ significantly. Another finding of the study is; the correlation obtained from the rubric was slightly higher as a result of calculating the correlations of the scores obtained with the scoring key and the average of the scores obtained from other scoring techniques. As a final finding, when the students know the criteria in the rubric used as a result of the surveys applied to individuals, they make self-evaluation, peer evaluation and take responsibility, as well as they state that they perceive and answer the question more accurately because they know the criteria for answering the questions asked in the exam.

Discussion

The averages of the scores obtained as a result of using different scoring methods for written exam answers differ significantly. In addition, significant relationships were found between DPA and PA scores. On the other hand, the internal consistency criteria obtained from both were high as a result of calculating the alpha coefficients obtained as a result of the application of different scoring tools such as PA and DPA to find the internal consistency criteria alpha coefficient of the test. If judged, the coefficient value obtained from DPA was slightly higher than the coefficient value obtained from pa. [b1] As a result of the questions directed to the students by the questionnaire, the fact that they expressed satisfaction with the scoring process with DPA is an indication of the motivating effect of DPA on students. When the scores given to the students with PA and DPA were compared and examined, it was determined that the students who were below average were closer to the average by taking higher than DPA according to pa, and the scores of the students who were above average were very close to each other with two measurement tools. Accordingly, it was concluded that the evaluation with DPA was in favor of low-achieving students and did not affect the success status of high-achieving students so much. Especially since the intermediate digits are scored in detail in analytical DPA, it can be concluded that students with low success levels are more advantageous. For students who fail to reflect their actual success in multiple choice and short-answer tests, it is stated that evaluation with DPA creates an advantage (Airasian and Russell, 2008).

Recommendations


Choosing DPA as a scoring tool will both increase the sensitivity of scoring and ensure that students are aware of the criteria. By expanding the research process to two semesters, taking written exams from the first semester subjects and taking written exams from the second semester subjects, comparing the results obtained as a result of scoring these written exams with scoring tools will be more effective in comparing the difference between these two scoring tools. In the research process, if the transfer of information to students for the use of the rubric is intensified and if such studies are spread over time, students can be provided with a very good understanding of the criteria. Students' bias can be reduced because their grasp of the criteria allows for unbiased scoring. By reducing sample error by participating students from the same level in different schools, the universe can be better represented and expansion can be made by obtaining more generalized results for comparison of scoring key and rubric.




<http://www.tayjournal.com>

<https://dergipark.org.tr/tr/pub/tayjournal>

Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Karşılaştırılması

 Yusuf Kasap, Doktora Öğrencisi, Sorumlu Yazar
Hacettepe Üniversitesi, Türkiye
yusufkasap5806@hotmail.com
Orcid ID: 0000-0002-5114-1175

 Selahattin Gelbal, Prof. Dr.
Hacettepe Üniversitesi, Türkiye
sgelbal@gmail.com
Orcid ID: 0000-0001-5181-7262

 Nuri Doğan, Prof. Dr.
Hacettepe Üniversitesi, Türkiye
nuridogan2004@gmail.com
Orcid ID: 0000-0001-6274-2016

Makale Türü: Araştırma Makalesi

Geliş Tarihi: 08.03.2021

Kabul Tarihi: 11.06. 2021

Yayınlanma Tarihi: 28.06.2021

Tr/En: Tr

İntihal: Bu makale, en az iki hakem tarafından incelendi ve intihal içermediği teyit edildi.

Atıf: Atıf: Kasap, Y., Gelbal, S. & Doğan, N. (2021). Dereceli puanlama anahtarı ve puanlama anahtarından elde edilen puanların karşılaştırılması. *Türk Akademik Yayınlar Dergisi (TAY Journal)*, 5(1), 95-117.

Özet

Bu arařtırmada, puanlama anahtarı ile son yıllarda yaygın olarak kullanılan dereceli puanlama anahtarının karşılařtırılması amaçlanmıřtır. İliřkisel tarama modelinde tasarlanan arařtırmanın çalıřma grubunu, Ankara'nın Çubuk ilçesinde bulunan Atatürk Meslek Yüksekokulu'nun makine, metal, motor, elektrik, elektronik bölümlerinde öğrenimine devam eden 195 öğrencinin oluřturduđu dört řubedeki birinci sınıf öğrencileri oluřturmaktadır. Arařtırma geleneksel deđerlendirme yaklařımı olarak puanlama anahtarıyla deđerlendirme, yeni deđerlendirme yaklařımı olarak dereceli puanlama anahtarıyla deđerlendirme, öz deđerlendirme, akran deđerlendirme ve anketle sınırlandırılmıřtır. Dereceli puanlama anahtarı ve puanlama anahtarıyla öğrencilerin başarılarının deđerlendirilmesi durumu karşılařtırılmıřtır. Bu karşılařtırmada öğrencilerin akademik başarıları dikkate alınmıřtır. Puanlama anahtarı ve dereceli puanlama anahtarından elde edilen puanlar arası tutarlıđa bakılarak, bu iki yöntemin güvenilirlikleri karşılařtırılmıřtır. Ayrıca puanlama yöntemleri arasında ortalama başarı yönünden anlamlı fark olup olmadıđını test etmek için bađımlı örneklemeler için ANOVA (varyans analizi) testi yapılmıřtır. Arařtırmanın bulgularına göre, puanlama anahtarı ve dereceli puanlama anahtarından elde edilen deđerlendirme sonuçları arasında anlamlı fark bulunmuřtur. Buna karşı, akran deđerlendirme ve öz deđerlendirme sonuçları arasında anlamlı fark bulunmamıřtır. Arařtırmanın bulgularından diđer i se; puanlama anahtarı ve dereceli puanlama anahtarıyla elde edilen puanların diđer puanlama tekniklerinden elde edilen puanların ortalamasıyla korelasyonlarının hesaplanması sonucunda dereceli puanlama anahtarından elde edilen korelasyon daha yüksek çıkmıřtır.

Anahtar Kelimeler: Puanlama anahtarı, Rubric, Öz deđerlendirme, Akran deđerlendirme

Giriř

Dünyada 1975–2000 yılları arasında toplumsal, sosyo-kültürel, ekonomik, politik ve teknolojik alanda birçok sarsıcı deđişiklikler olmuřtur. Bu alanlardaki sarsıcı deđişmeler toplumsal hayatı geri dönölmez bir sürece sokmuřtur. Bu sürecin çok boyutlu ve çok geniř çaplı olduđu görölmektedir. Bu deđişimler toplumların hayata bakıř açılarını, deđerlerini, maddi yařamdan beklentilerini köklü bir řekilde etkilemektedir. Toplumsal hayattaki bu deđişimler toplumu oluřturan yapılardaki deđiřimi de zorunlu kılmıřtır. Bu yapılar kendi içinde bu hızlı deđiřimle meydana gelen çağın sorunlarına çözüm bulma adına, evrimleřmeye gitmiřtir. Bu yapılardan biri olan eğitim, bu toplumsal deđiřim dinamiklerini toplum hayatına adapte etme ya da toplumu bu deđiřimlere adapte etme adına en önemli alanlardan biri olarak günümüzde kabul görmektedir. İnsanlıđın yařam standartlarını eğitim vasıtasıyla yükselttiđi herkes tarafından bilinen bir gerçektir. En iřevuruk anlamda “istendik davranıř meydana getirme veya istendik davranıř meydana getirme süreci” olarak ifade edilen eğitim, toplumun önemli deđerlerinin, ahlak kurallarının, bilgi birikimlerinin nesilden nesile aktarılması süreciyle ilgilidir.

Eđitim özetle bireyi istenen bir amaç dođrultusunda kültürlenme sürecidir (Senemođlu, 2004). Eğitim insanın kiřisel ve sosyal alanda başarıya ulařmasında; barıř, demokrasi, hürriyet, eřitlik ve evrensel bütünlük amaçlarına ulařmasında temel araçtır. Diđer yandan eğitim; toplumsal ve ekonomik ilerlemenin önemli unsurlarından biri olarak tüm sektörleri önemli ölçüde etkilemektedir (Arslan ve Eraslan, 2003). İçinde bulunduđumuz zamanın deđiřim sürecinde ihtiyaçlara ve beklentilere yanıt verecek bireyler yetiřtirmek için mümkün olduđunca az hatalı adımlar atılması ve kaynađı insan olan eğitim platformunda mümkün olduđu kadar az hata ile güvenilir sınamaların yapılması gerekliliktir. Bilgi toplumu olgusu yařamımızda

yaygınlaşırken, bireylerden istenen mesleki, teknik, eğitsel bilgi ve beceriler derinden etkilenip değişmiştir (Aşkar, 2004; Şimşek, 1997). Bireyin istendik davranışları kazanması için gerçekleştirilecek eğitim durumlarının bireye bu davranışları kazandırmadaki etki düzeyinin incelenmesi ve değerlendirmesi eğitimin önemli bir ögesini oluşturmaktadır. Bireylerin nitelikli eğitim ürünlerini meydana getirirken becerilerinin ve ürünlerin değerlendirmesinde tercih edilecek değerlendirme yaklaşımlarının etkili olup olmamasının bireyin performansını etkileyeceği herkes tarafından bilinmektedir (Ertürk, 1998). Eğitim, yaygın olarak; insanın kişiliğini geliştirme ve geleceğe yönelik önemli bir yatırım aracı olarak kabul edilmektedir. Bireyin performansını çok yönlü olarak değerlendiren yeni paradigmlar, bireyin kendi yaşantısı vasıtasıyla davranışlarında ortaya çıkan değişiklik sonucunda meydana gelen ürünün niteliğini bu paradigmların anlayışına uygun olarak değerlendirilmesini gerektirir. Böylece ölçmenin çeşitli yaklaşımlarla çok yönlü olarak yapılması sonucu elde edilen verilerin daha kapsamlı olması sağlanacaktır. Bu durum çeşitli tür ve biçimlerde değerlendirme yapmaya imkan sağlarken bireylerin başarı, performans, yetenek, tutum, zeka gibi özelliklerin ölçülmesinde hatayı azaltacağı düşünülmektedir. Bu iddialar güvenilirlik ve geçerlik çerçevesinde incelenecek araştırma kapsamına girmektedir.

Belli bir zaman dilimi sürecinde bir grubun veya topluluğun düşünme şeklini ve davranışlarını belirleyen bir dünya görüşü, bir algı kaynağı, bir izlenceler bütünü, bir bakış açısı ya da modele paradigma denir (Şimşek, 1997). Bazı uygulamaların uygun, bazı uygulamaların da 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). Marby (1999) eğitimde ölçme literatüründe, değerlendirme ile ilgili üç değişik farklı paradigma bulunduğunu belirtmektedir. Bunlar:

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

Psikometrik Paradigma süreçten çok ürünle ilgilidir. Çünkü ürün gözlenebilirdir (Milli Eğitim Bakanlığı [MEB], 2003). Bu tür ölçmenin temel yöntemi, standardize edilmiş çoktan seçmeli testlerdir (Marby, 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:

- Sadece fark etme bilgisini ölçmesi

- Őans başarısı olması
- Ayırt ediciliđi yüksek sorular yazmanın zorluđu
- Dönüt vermenin zorluđu
- Kopya çekmeyi kolaylaştırması

Bu sınavların amacı, çocukları birbirleriyle ve çođunlukla yıllar öncesinden geliştirilmiş bazı sınavları almıő olan çocuklardan oluőan örneklemlerle karşılaőtırmaktır (Conner 1991'den akt. İlhan, 2006). Psikometrik paradigmada öđrenci başarısının anlaşılmasındaki temel strateji karşılaőtırmadır.

Psikometrik paradigmalarla standart hale getirilmiş testlerde karşılaőtılan problemler, öđretmenleri daha kapsamlı, geniő ölçekli deđerlendirmelere ve performans deđerlendirmelere götürmüőtür. Psikometrik paradigma, öđrencilerin yaőantılarından bađımsız, tek tip, standart içeriđi yansıtmamasına karşı, içeriksel paradigma öđrencilerin yaőadığı gerçek deneyimleri yansıtmayı amaçlamıőtır. Bu temel fark içeriksel deđerlendirmenin sadece sıralama ve sınıflama yapmaktan öteye geçip, öđrencilerin gerçekte neler öđrendiklerini ortaya koymaktadır. Puanlayıcı açıısından puanlama zorluđu yaratırken, puanlama bakımından zaman alıcı ve kullanıőlılık yönünden ekonomik deđildir. Uygulamada standart testlere göre daha az soru sorulur bu da konunun kapsamını sınırlar ve kapsam geçerliđiyle ilgili sorunlar meydana getirebilir (Burstein, 1991).

Öđrenci başarısının deđerlendirilmesine yönelik bireyselleőtirilmiő yaklaşım yapılandırmacı öđrenme teorisi ve bireyselleőtirilmiő eğitime daha iyi uyar. Literatürde yapılandırmacılık temel alınarak bireysel paradigmayla ilgili açıklamalar özellikle Wiggins (1993) tarafından yapılmıőtır. Yapılandırmacı öđrenme teorilerinin bireysel paradigmayla çok fazla uyum sađladığı ifade edilir. Bu yapılandırmacı bakıő açısına göre eğitim öđrencilerin bilgiyi nasıl elde edeceđini öđrenmelerine yardımcı olmaktadır. Bu yaklaőtıma göre eğitimin odak noktası öđrenmenin yer aldıđı kültür ve konuları göz önünde bulundurmadır (Anderson, 1998).

Yapılandırmacılık bilme kuramı olarak bilgiyi yapılandırma süreci, bu süreci etkileyen faktörlerle ilgili birçok açıklamayı yapısında bulundurduđundan öđretme ve öđrenme süreçleri üzerinde önemli etkileri bulunmaktadır. Yapılandırmacılıđa göre bilgi, iletişim araçları ve duyularla dıő dünyadan edilgin olarak alınan bir őey deđildir. Aksine; bilgi, öđrenen tarafından yapılandırılır. Yapılandırmacılıđa göre bilgiyi yapılandırma ihtiyacı, bireyin çevresiyle etkileőim halinde iken gördüklerinden anlam çıkarmaya çalıőırken meydana gelir (Açıkgoz, 2002).

Bireyselleőtirilmiő paradigmaya dayalı olarak yapılan performans deđerlendirme okullarda öđrencilerin, hatırlama (bilgi) düzeyinde kazandıkları davranıőlar performansın oluőması ve geliőmesi için fayda sađlarken bu davranıőların yalnız başına kullanılması performansın alt düzeyi olarak ifade edilmektedir. Bu anlamda performans, öđrencilerin okulda öđrendiklerini geliőtirmek ve pekiőtirmek amacıyla sınıf içinde ya da sınıf dıőında bir etkinlik yaptıkları veya bir ürün meydana getirdikleri zaman oluőmaktadır. Baőka bir deyiőle genellikle üst düzey zihinsel beceri gerektiren bir görev icra edilirken ortaya konan çaba sonucunda ortaya çıkan ürün performans olarak ifade edilebilir (Kutlu, Dođan ve Karakaya, 2008).

Eđitim alanında her geen gn nemi artan yeni deęerlendirme yaklařımlarından biri performans odaklı deęerlendirmelerdir. đrenci performansının geliřimini saęlamak amacıyla yapılan st dzey zihinsel beceri gerektiren etkinlikler performans odaklı deęerlendirmeler olarak ifade edilmektedir. Performans odaklı deęerlendirmede đrencilerin basit, yalın, alt dzey dřnme gerektiren grevleri deęil, st dzey dřnme becerisi gerektiren karmařık yapıdaki bazı grevleri icra etmeleri istenmektedir. Genellikle bu karmařık ve st dzey beceri gerektiren grevler bir rnn ortaya koyulmasını veya verilen bir problem durumuna uygun zm yolları belirlemeyi iermektedir. Buradaki en temel ama đrencinin problem zme, st dzey ve yaratıcı dřnme, eleřtirel dřnme, karar verme, empati kurma gibi becerilerini ortaya ıkarmak, geliřtirmek, bu becerilerin ne dzeyde ortaya konulduęunu ve kullanılabildięini tespit etmektir (Kutlu, Doęan ve Karakaya, 2008).

Performans deęerlendirme ayrıca đrencilerin gerek yařam problemlerinin zmnde đrendikleri bilgileri kullanmalarını gerektirir. Bylece đrenciler đrendiklerini gerek yařam problemlerinin zmnde uygulamıř olurlar (Airasian, 1994). Bu duruma uygun deęerlendirme yapabilmek iin en ok kullanılan puanlama yntemlerinden biri dereceli puanlama anahtarıdır. Popham (1997) dereceli puanlama anahtarının, đrencilerin, kompozisyon yazma becerilerinin, sunum yapma becerilerinin, bilimsel proje hazırlama becerilerinin deęerlendirilmesinde bir eřit puanlama kriteri olduęunu ifade etmektedir. Popham dereceli puanlama anahtarını, yapılan alıřma iin becerinin en nemli gzlenebilir yanları olan ltleri listeleyen ve alıřmada nelerin hangi sırada ve nasıl yapılacaęını aıklayan bir puanlama aracı olarak ifade etmektedir. Popham'a (1997) gre dereceli puanlama anahtarı; deęerlendirme ltleri, lt aıklamaları ve puanlama stratejisi olmak zere  blmden oluřur.

lt aıklamaları: đrencilerin performanslarını meydana getiren her bir lte ynelik vermeleri gereken cevaplarındaki niteliksel farklılıkları aıklamayı ifade eder. Bylece đrencinin performansını puanlamak iin verilen lt aıklamaları đrencinin ne yapması gerektięini aıklar (Popham, 1997).

Deęerlendirme ltleri: đrencilerin ltlere ynelik ne dzeyde performans sergilediklerinin belirlenmesini saęlar. Performans ltleri lme aracında mevcuttur. Her performans lt iin tanımlama yapılır ve bunlar deęerlendirme ltleriyle bir tabloda birleřtirilerek dereceli puanlama anahtarı (rubric) oluřturur (Popham, 1997).

Puanlama stratejisi: Puanlama btnsel (holistic) ya da analitik (analytical) olarak yapılabilir. Bunlardan hangisinin kullanılacaęı deęerlendirmenin amacına baęlıdır (Popham, 1997).

Bazı durumlarda yapılan bir deęerlendirmeyi baęımsız ltlere ayrıřtırmak mmkn olamamakta, performansın farklı dzeylerinin ortaya ıkarılması iin belirlenmiř ltler arasında bir farklılık bulunmamaktadır. Byle durumlarda btnsel dereceli puanlama anahtarı kullanılmalıdır (Brookhart, 1999). Analitik dereceli puanlama anahtarı ise, llen bir beceri boyutu alt ltlere ayrıřtırılabildięi durumlarda ve daha ayrıntılı puanlama yapmanın gerekli olduęu durumlarda kullanılmaktadır (Haladyna, 1997).

Analitik bir dereceli puanlama anahtarı, sreklilik arz eden nitelikteki bir alıřmaya verebilecek muhtemel puanlardan oluřan bir lęi ierir. Yksek sayılar en iyi performanslara

verilir. Puanlamada 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 dereceli puanlama anahtarında 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).

Yapılandırmacı öğrenme teorisine dayanan değerlendirme yaklaşımı kapsamında Öz değerlendirme ve Akran değerlendirme de bulunmaktadır. Öz değerlendirme belli bir beceriye yönelik öğrencinin kendisinin ortaya koyduğu performansı değerlendirmesidir. Öz değerlendirme öğrencilerin kendilerini tanımalarına, güçlü ve zayıf yönlerini keşfetmelerine imkân sağlar. Bu değerlendirme öğrencilere kendi performanslarını bağımsız olarak değerlendirme fırsatı vermenin yanı sıra kendi gelişimlerini izlemeye de olanak sağlar. Ayrıca öğrencilerin etkinlikler sürecinde sonraki adımda ne yapacaklarını bilmelerine, öğrenme ve değerlendirme sürecine aktif olarak katılarak kendilerine güven duymalarına olanak sağlar. Öğrenciler öz değerlendirme yaparken öğretmen onların kendilerini objektif olarak ifade etmeleri konusunda bilgilendirir. Öğrencilere yapacakları değerlendirmenin, iyi ve zayıf yönlerini görmelerine ve böylece eksikliklerini gidermelerine yardımcı olacağı söylenir. Bu değerlendirmenin dezavantajlı yönleri de vardır. Genellikle öğrencilerin kendi performanslarını değerlendirirken yanlış davranacakları unutulmamalıdır. Başlangıçta öz değerlendirme, öğrencilerin tecrübesizliğinden dolayı yanlılığa neden olabilir. Yine de öğrenciler tecrübe kazandıkça yapacakları değerlendirmeler daha objektif olacaktır (Bahar, Bıçak, Nartgün, Durmuş, 2008).

Akran değerlendirme belli bir beceriye yönelik öğrencinin arkadaşını değerlendirmesidir. Öğrencilerin, arkadaşlarının performansını değerlendirirken eleştirel ve yaratıcı düşünme becerileri de gelişecektir. Diğer yandan akran değerlendirme, öğretmene öğrencilerin performanslarına yönelik ilerleme düzeyleri hakkında dönüt vermektedir. Akran değerlendirmede, öğrencilerin taraflı davranışlarını önlemek için ölçütlerin ve ölçüt açıklamalarının öğrencilere bildirilmesi faydalı olacaktır (Bahar, Bıçak, Nartgün, Durmuş, 2008).

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, hangi durumda hangi değerlendirme yaklaşımının kullanılacağı belirlenir. 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. Ölçme işleminde kullanılacak araçların belli özelliklere sahip olması istenir. Bu

özelliklerden birisi güvenilirliktir. Güvenirlik ölçme aracı ile benzer şartlarda art arda yapılan ölçmelerin tutarlı sonuçlar vermesi olarak ifade edilir (Crocker ve Algina, 1986). Başka bir ifadeyle güvenilirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesidir (Turgut ve Baykul, 2011). Diğer yandan ölçme aracının başka bir özelliği ise geçerliğidir. Geçerlik, ölçme aracının ölçmeyi amaçladığı özelliği, amacın dışında başka herhangi bir özellikle karıştırmadan, doğru olarak ölçebilme derecesidir (Tekin, 1996).

Bireylerin farklı yeteneklerini, becerilerini etkili bir şekilde kullanmalarını sağlamada ve onları doğru ilgi alanlarına yönlendirmede, yeni ölçme ve değerlendirme yaklaşımlarının etkili olup olmayacağı, günümüzde eğitim alanında cevaplanması gereken önemli sorulardan birisidir. Ayrıca yeni ölçme ve değerlendirme yöntemlerinin geçerlikleriyle ilgili kanıtlar, bu yöntemleri kullanmak konusunda şüpheleri ortadan kaldırmalı ve bu yöntemleri kullanarak bireyler hakkında tam doğru kararlar verilebilmesi için bu yöntemlerin açıklığa kavuşturulmasını sağlamalıdır. Böylece geleneksel ve yeni değerlendirme yaklaşımlarını nesnellikleri, güvenilirlikleri ve geçerlikleri yönünden karşılaştırmak mümkün olacaktır. Bu problemler için ortaya konulacak doğru çözümler, çağdaş donanımlı bireyler yetiştirmede daha doğru bir yaklaşım belirlemeye yardımcı olacaktır. Bu araştırmanın problem durumu matematik dersinde öğrenci başarısının değerlendirilmesinde kullanılan yazılı sınavın puanlanmasında kullanılan puanlama anahtarı, dereceli puanlama anahtarı, öz değerlendirme ve akran değerlendirme sonucunda elde edilen puanlar arasındaki ilişkilerin belirlenmesine yöneliktir. Bu bağlamda araştırmanın amacı; puanlama araçlarıyla yapılan farklı puanlama yöntemlerinden elde edilen iç tutarlık ölçütünün incelenmesi, puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirmeden elde edilen puanların ortalamalarının karşılaştırılarak incelenmesi ve puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirme sonuçları arasındaki ilişkinin incelenmesine yöneliktir.

Üst düzey düşünme becerilerinin değerlendirilmesi amacına hizmet eden dereceli puanlama anahtarları son yıllarda oldukça popüler araçlar haline gelmekle birlikte dereceli puanlama anahtarının kullanımı ile ilgili alanyazında çok sayıda araştırma yapılmıştır (Akkuş ve Duatepe Paksu, 2006; Anadol ve Doğan, 2018; Baysal ve Gürbüz, 2019; Danışmaz ve Adıbatmaz, 2020; Demir ve Yıldırım, 2019; Özcan ve Taşar, 2018; Sönmez, 2019; Yetiş, 2017). Bu araştırmayla ilgili alanyazın incelendiğinde; Aslanoğlu (2002) tarafından yapılan bir araştırmada dereceli puanlama anahtarının öğrencilerin öğretim sürecinde öğrendikleri bilgilerin ölçülmesinde sağladığı katkı araştırılmıştır. Yapılan araştırmaya benzer bir araştırma Doğan ve Parlak (2014) tarafından yapılmıştır. Bu araştırmada puanlama anahtarı ve dereceli puanlama anahtarı kullanılarak elde edilen puanların uyum düzeylerinin araştırılması amaçlanmıştır. Diğer bir araştırma ise Doğan ve Anadol (2018) tarafından yapılmıştır. Bu araştırmada ise dereceli puanlama anahtarı kullanmaya ilişkin deneyim yılının puanlayıcı güvenilirliğine etkisi incelenmiştir.

Yapılan ilgili araştırmalar incelendiğinde dereceli puanlama anahtarı ve puanlama anahtarının uyumu ve dereceli puanlama anahtarının puanlayıcı güvenilirliğine etkisi ve yapılan ölçmelere sağladığı fayda araştırılmıştır. Bu araştırmada ise dereceli puanlama anahtarı ve puanlama anahtarı çok yönlü olarak karşılaştırılıp incelendiği için ilgili araştırmaları kapsamının da ötesinde detaylı nitelikte olması nedeniyle araştırmadan elde edilen sonuçlar karşılaştırılan ölçme araçları hakkında önemli fikirler verecektir.

Bireylerin farklı yeteneklerini, becerilerini etkili bir şekilde kullanmalarını sađlamada ve onları dođru ilgi alanlarına yönlendirmede, yeni ölçme ve deđerlendirme yaklaşımlarının etkili olup olmayacağı, günümüzde eğitim alanında cevaplanması gereken önemli sorulardan birisidir. Ayrıca yeni ölçme ve deđerlendirme yöntemlerinin geçerlikleriyle ilgili kanıtlar, bu yöntemleri kullanmak konusunda şüpheleri ortadan kaldırmalı ve bu yöntemleri kullanarak bireyler hakkında tam dođru kararlar verilebilmesi için bu yöntemlerin açıklığı kavuşturulmasını sađlamalıdır. Böylece geleneksel ve yeni deđerlendirme yaklaşımlarını nesnellikleri, güvenilirlikleri ve geçerlikleri yönünden karşılaştırmak mümkün olacaktır. Bu problemler için ortaya konulacak dođru çözümler, çağdaş donanımlı bireyler yetiştirmede daha dođru bir yaklaşım belirlemeye yardımcı olacaktır. Bu araştırmanın problem durumu matematik dersinde öğrenci başarısının deđerlendirilmesinde kullanılan yazılı sınavın puanlanmasında kullanılan puanlama anahtarı, dereceli puanlama anahtarı, öz deđerlendirme ve akran deđerlendirme sonucunda elde edilen puanlar arasındaki ilişkilerin belirlenmesine yöneliktir. Bu bağlamda araştırmanın amacı, günümüzde kullanılmakta olan ölçme araçlarından olan yazılı sınavların puanlama anahtarı ve dereceli puanlama anahtarıyla puanlanmasının matematik dersindeki başarıyı belirlemedeki etkilerinin araştırılmasıdır. Bu deđerlendirme yöntemlerinin etkileri karşılaştırılırken hangisinin bireysel farkları ortaya çıkarma ve bireylerin bilgi beceri ve yeteneklerini ortaya koyma yönünden uygun olduğunun saptanması yapılabilir. Buna ek olarak, dereceli puanlama anahtarı ve puanlama anahtarıyla yapılan deđerlendirme sonucunda puanlama güvenilirliklerinin karşılaştırılması da mümkün olacaktır. Öte yandan, puanlama araçlarıyla yapılan farklı puanlama yöntemlerinden elde edilen iç tutarlık ölçütünün incelenmesi, puanlama anahtarı, dereceli puanlama anahtarı, akran, öz deđerlendirmeden elde edilen puanların ortalamalarının karşılaştırılarak incelenmesi ve bu yöntemlerden elde edilen puanlar arasındaki ilişkinin incelenmesi de mümkün olacaktır.

Yöntem

Bu bölümde araştırmanın türü, araştırmanın yapılacağı çalışma grubu, veri toplama araçları, verilerin toplanması ve verilerin analizi üzerine durulmuştur.

Araştırmanın Türü

Deđişen deđerlendirme yaklaşımına yönelik kullanılan deđişen 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 tarama araştırması yapılmıştır (Büyüköztürk ve ark., 2018).

Çalışma Grubu

Araştırmanın çalışma grubunu, Ankara'nın Çubuk ilçesinde bulunan Atatürk Meslek Yüksekokulu'nun makine, metal, motor, elektrik, elektronik bölümlerinde öğrenimine devam eden öğrencilerin oluşturduğu dört şubedeki birinci sınıf öğrencileri oluşturmaktadır. Şube 1'de 54, 2. şubede 49, 3. şubede 47 ve 4. şubede 45 öğrenci bulunmaktadır. Araştırma farklı puanlama yöntemlerinin öğrencilerin başarılarını, 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

Yazılı Yoklama

Yazılı sınavlar problem çözme, fikirleri analiz edip organize etme, özgün fikirler üretme, bilgileri yeni durumlara uyarlama, görüşleri değerlendirme vb. gibi davranışların ölçülmesi amacıyla en uygun sınav türü olarak kabul edilir (Doğan, 2007). Yazılı yoklamalarla puanlanma yaparken objektif bir puanlama yöntemi kullanılmıyorsa bazı cevaplara gerektiğinden fazla, bazılarına ise gerektiğinden az puan verileceğinden yapılan ölçmenin geçerliği düşecektir (Turgut ve Baykul, 2011). Bu araştırmada dört şubenin tüm öğrencileri aynı zamanda ve aynı sorularla yazılı yoklama sınavı almışlardır. Ayrıca, yazılı yoklamanın hazırlanması sürecinde belirtke tablosundan yararlanılmış ve üç alan uzmanından kapsam geçerliğine ilişkin olumlu görüş alınmıştır.

Puanlama Anahtarı

Puanlama anahtarı (PA) ile puanlama metodu yazılı yoklamaların, sakıncalarından birçoğunu önler. Çok iyi hazırlanmış bir anahtar, puanlayıcının işini kolaylaştırır, puanlama işlemini çabuklaştırır ve puanları oldukça objektif hale getirir (Turgut ve Baykul, 2011). Bu araştırmada hazırlanan değerlendirme sorularının puanlaması için bir PA hazırlanmıştır. PA hazırlanırken öncelikle matematik vize sınavında sorulan soruların her biri için uzman desteğiyle çözüm basamakları belirlenmiş ve her sorunun çözüm basamaklarına puan atanmıştır. Sonrada, puanlama anahtarının geçerliği ile ilgili 6 uzman görüşü alınmıştır. Böylece, her soru için ölçme kuralı belirlenmiş ve öğrencilerin yazılı kağıtları bu PA'ya göre puanlanmıştır. Öğrencinin PA'dan alacağı en düşük puan 0, en yüksek puan 100'dür. Puanlama anahtarında her sorudaki puanlama basamaklarına verilen puanların eşit olmamasından dolayı α güvenilirlik katsayısını puanlama anahtarıyla hesaplamanın uygun olmamasına rağmen dereceli puanlama anahtarıyla karşılaştırmaya gitmek için puanlama anahtarından elde edilen puanlarla yapılan istatistiksel işlemler sonucunda Cronbach α güvenilirlik katsayısı 0,756 bulunmuştur.

Dereceli Puanlama Anahtarı

Bu araştırma için dereceli puanlama anahtarı (DPA) 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 DPA'ya göre puanlanmıştır. DPA'nın ölçek geçerliği ile ilgili 6 uzman görüşü alınmıştır. Dereceli puanlama anahtarından her soru için elde edilen puanlarla yapılan hesaplamalar sonucunda elde edilen Cronbach α güvenilirlik katsayısı 0,778 bulunmuştur.

Öğretmen Değerlendirme Sonuçları

Meslek yüksekokulunda görev yapan Öğretmen toplamda dört şubenin dersine girmektedir. Bu Öğretmen hem PA hemde DPA ile tüm şubelerdeki öğrencilerin yazılı sınav sonucunu puanlamıştır.

Öz Deđerlendirme Sonuları

Öz deđerlendirme (ÖD) sonuları öđrencilerin DPA ile kendi yazılı sınav sonucunu puanlayarak elde ettikleri puanlardır. DPA'da bulunan ölçüt açıklamalarının puanlama kolaylıđı ve objektiflik sađlamasından dolayı şubelerdeki tüm öđrenciler kendi yazılı kâđıtlarını puanlamada sadece DPA'yı kullanmıřtır.

Akran Deđerlendirme Sonuları

Akran deđerlendirme (AD) sonuları öđrencilerin DPA ile arkadaşlarının yazılı sınav sonucunu puanlayarak elde ettikleri puanlardır. Öđrenciler Akran deđerlendirme yaparken sadece DPA'yı kullanmıřtır. Akran deđerlendirme yaparken öncelikle her bir şubede tarafsız biçimde ikiyeşerli gruplar belirlenmiř ve herbir gruptaki öđrencilerin birbirinin kâđıdını puanlaması sađlanmıřtır.

Anket

Anket hazırlandıktan sonra anketin içeriđiyle ilgili uzman görüşü alınmıřtır. Farklı puanlama yaklaşımlarının uygulandıđı tüm şubelerdeki öđrencilere bu farklı puanlama yaklaşımları sorulmuřtur. Bunun için açık uçlu sorulardan oluřan bir anket hazırlanmıřtır. Öđrenciler cevaplarını yazılı olarak vermiřlerdir.

Verilerin Toplanması

Arařtırmanın verileri Meslek Yüksekokulundaki makine, metal, motor, elektrik, elektronik bölümlerinde öğrenimine devam eden öđrencilerin matematik vize sınavıyla elde edilmiřtir. Veriler üç aşamada toplanmıřtır. Verileri toplamanın ilk aşamasında aynı şartlar altında matematik vize sınavı yazılı yoklaması yapılmıř dört şubenin de katıldıđı bu sınava toplam 195 kiři katılmıř ve bu sınav sonunda elde edilen yazılı yoklama kâđıtları toplanmıřtır. Toplama işleminden sonra bu kâđıtlar Öđretmen tarafından önce PA sonra da DPA ile puanlanması sonucunda elde edilen puanlar analiz için kaydedilmiřtir. Verileri toplamanın ikinci aşamasında yapılan yazılı sınavdan bir hafta sonra öđrencilere yazılı kâđıtları getirilmiř ve yazılı yoklamada sorulan sorular tahtada çözülmüřtür. Sorular çözüldükten sonra öđrencilere DPA ile puanlama yöntemine yönelik yarım saatlik bir seminer verilmiřtir. Seminerin içeriđi; DPA'da bulunan her ölçütü öđrencilerin öğrenmesini sađlama yoluyla öz deđerlendirme ve akran deđerlendirme yapmalarına imkan sađlamaya yöneliktir. Seminer sonrası öđrencilere sınav kâđıtları ve DPA dađıtılmıř ve bu kâđıtlara her öđrenciye adları ve soyadlarının yazılması söylenmiřtir. Öđrenciler ad ve soyadlarını yazdıktan sonra öz deđerlendirmeye başlamıřlardır. Öz deđerlendirmeden sonra öđrencilere yazılı kâđıtlarını arkadaşlarıyla deđiřtirmeleri söylenmiř ve tekrar DPA dađıtılmıřtır. Bu sefer öđrenciler kâđıtlara adlarını ve soyadlarını yazmamaları söylendikten sonra öđrenciler akran deđerlendirmeye başlamıřlardır. Burada akran deđerlendirme alıřmasına toplam 195 öđrenci katıldıđından akran deđerlendirme için, 195 tek sayı olduđundan eşleřtirme oluřurmada sorun yařanmıřtır. Bu sebepten dolayı, bir öđrencinin yazılı kâđıdı öđrencilerden biri tarafından puanlanmıřtır. Bu durumda öđrencilerden bir tanesi iki yazılı kâđıdını puanlamıřtır. Böylece, öz deđerlendirme ve akran deđerlendirmeden elde edilen puanlar analiz için kaydedilmiřtir. Verileri toplamanın üçüncü aşamasında öđrencilere açık uçlu soruları kapsayan anket uygulanmıřtır. Anket uygulaması dönem sonu ulařılan 56 öđrenci üzerinde gerekleřtirilmiřtir. Öđrenciler

ankette bulunan açık uçlu sorulara yazılı yanıt vermiştir. Elde edilen sonuçlar okunarak, yorumlanarak kaydedilmiştir.

Tablo 1. Analizlere alınan öğrenci sayıları

Şubeler	PA	DPA	ÖD	AD	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

Verilerin Analizi

Araştırmanın problemlerine çözüm bulmak amacıyla kullanılan teknikler farklılık göstermektedir. Problemlerin niteliğine göre, çeşitli istatistiksel teknikler ve bazı nitel yöntemler kullanılmıştır. Bu teknikler aşağıda açıklanmıştır. Öncelikle öğrencilere uygulanan yazılı sınav sonuçlarının PA ve DPA ile Öğretmen tarafından puanlanması ve öğrencilerin DPA ile öz değerlendirme ve akran değerlendirme yapımları sonucunda elde edilen başarı puanları arasında fark olup olmadığı test edilmiştir.

Tüm 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 bağımlı örneklem tekrarlı verilerde ANOVA (varyans analizi) yapılmıştır. Daha sonra öğrencilere uygulanan yazılı sınav sonuçlarının PA ve DPA ile puanlanması sonucunda iki farklı puanlama aracından elde edilen puanlar arası ilişki katsayısını SPSS paket programından faydalanarak hesaplanmış ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır. Ayrıca öğrencilere uygulanan yazılı sınav sonuçlarının PA ve DPA ile puanlanması sonucunda yazılı yoklamada bulunan madeler arası tutarlığı incelemek için iç tutarlık ölçüsü (Cronbach Alfa Güvenirliği) excell paket programında alfa güvenirlilik katsayısının formülü kullanılarak hesaplanmış ve elde edilen sonuçlar karşılaştırılıp yorumlanmıştır.

Araştırmanın Etik İzinleri

Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik Kurul İzin Bilgileri:

Araştırma 2019 öncesinde toplanan verileri içerdiği için etik kurul izin belgesine ihtiyaç bulunmamaktadır.

Bulgular

PA, DPA, akran ve öz değerlendirmeden elde edilen puanların karşılaştırılması amacıyla öncelikle betimsel istatistiklerin incelenmesi gerçekleştirilmiştir. Yapılan değerlendirmelerden elde edilen puanların betimsel istatistikleri Tablo 2’de özet olarak verilmiştir.

Tablo 2. Puanlama yöntemleriyle elde edilen puanlara ilişkin betimsel istatistikler

Puanlama yöntemleri	N	Aritmetik ortalama	Standart sapma	Standart hata
PA	195	35,46	19,59	1,40
DPA	195	40,95	18,62	1,33
ÖD	195	62,28	21,82	1,56
AD	195	60,22	21,68	1,55

Standart sapma ve standart hata yönünden inceleme yapıldığında en yüksek standart sapma ve standart hata öğrencilerin DPA ile kendi sınav kâğıtlarını puanlamaları sonucu elde edilen puanlarda meydana gelmiştir. En düşük standart sapma ve standart hata, Öğretmenin sınav kâğıtlarını DPA ile puanlanması sonucu elde edilen puanlarda meydana gelmiştir. Öğretmenin PA ile elde ettiği puanların standart sapma ve standart hatası DPA ile elde edilen puanların standart sapma ve standart hatasından büyük çıkmıştır. Öğrencilerin kendi sınav kâğıdını DPA ile puanlaması sonucu elde edilen puanların standart sapma ve standart hatası, öğrencilerin arkadaşlarının sınav kâğıdını değerlendirmesi sonucunda elde edilen puanların standart sapma ve standart hatasından daha büyük çıkmıştır. Ayrıca Öğretmenin yaptığı puanlamadan elde edilen standart sapma ve standart hata değerleri, öğrencilerin yaptıkları puanlamanın standart sapma ve standart hata değerlerinden küçük çıkmıştır.

Farklı puanlama yöntemlerinden elde edilen puanların ortalamaları arası farkların anlamlılığını daha detaylı test etmek için bağımlı örneklem için tekrarlı verilerde varyans analizi yapılmıştır. Varyans analizi sonuçları Tablo 3'te verilmiştir.

Tablo 3. Uygulanan puanlama yöntemlerinin karşılaştırılması

Varyansı kaynağı	Kareler toplamı	Sd	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			

Tablo 3 incelendiğinde varyans analizi sonucu göstermektedir ki, farklı puanlama yöntemleri arasında anlamlı bir fark bulunmaktadır. Puanlama tekniklerine ilişkin ortalamalar arasındaki farklılığın hangi puanlama yöntemleri arasındaki farktan kaynaklandığını belirlemek için ikili karşılaştırmalar Sheeffe tekniğinden yararlanılarak yapılmış ve sonuçlar Tablo 4'te verilmiştir.

Tablo 4. Uygulanan puanlama yöntemleri için ikili karşılaştırmalar

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

** p < 0,01

İkili karşılaştırma sonuçlarına göre en büyük farklılık, Öğretmenin PA ile puanlama sonucunda elde ettiği puanların ortalaması ile öğrencilerin kendi sınav kâğıtlarını DPA ile puanlamaları sonucunda elde edilen puanların ortalaması arasında olmuştur. En küçük fark ise öğrencilerin kendi sınav kâğıtlarını DPA ile puanlamaları sonucunda elde ettikleri puanların ortalaması ile öğrencilerin akranlarının sınav kâğıtlarını DPA ile puanlamaları sonucunda elde ettikleri puanların ortalaması arasında olmuştur. Sonuç olarak ($\alpha=0,01$) düzeyinde yalnızca öğrencilerin kendi sınav kâğıtlarını DPA ile puanlamaları sonucunda elde ettikleri puanların ortalaması ile akranların sınav kâğıtlarını DPA ile puanlamaları sonucunda elde ettikleri puanların ortalaması arasında anlamlı bir fark gözlenmemiştir. Öğretmenin PA'dan elde ettiği puanların ortalaması ile DPA'dan elde ettiği puanların ortalaması arasında da anlamlı fark olduğu gözlenmektedir. Ulaşılan bu sonuçlara göre, Öğretmenin DPA'dan elde ettiği puanların ortalaması, PA'dan elde ettiği puanların ortalamasından büyüktür. Bunun sonucunda, DPA puanlama yaparken daha çok adım içerdiği için daha fazla detaya puan vermeyi mümkün kılarak duyarlı davranmış ve kısmi bilgiye puan verme olanağı yaratmıştır. Diğer yandan PA daha az puanlama adımı içermesinden dolayı DPA'ya göre duyarsız ve daha yüzeysel puanlama yapmaya neden olmuş olabilir. Öğrencilerin DPA ile kendi sınav kâğıtlarını puanlamaları sonucunda elde ettikleri puanlarının ortalaması, arkadaşlarının sınav kâğıtlarını DPA ile 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 cömert davrandıkları söylenebilir. Puanlama anahtarı, dereceli puanlama anahtarı, akran, öz değerlendirme sonuçları arasındaki ilişkiler Tablo 5'te verilmiştir.

Tablo 5. Puanlama yöntemleri arasındaki korelasyonlar

Pearson Korelasyon	PA	DPA	ÖD	AD
PA	1.00			
DPA	0.98	1.00		
ÖD	0,742**	0,723**	1.00	
AD	0,516**	0,485*	0,598**	1.00

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

Tablo 5 incelendiğinde farklı puanlama teknikleri ile elde edilen puanlar arasında manidar korelasyonların bulunduğu görülmektedir. PA'dan elde edilen puanlarla DPA ve öz değerlendirme puanları arasında, DPA 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. PA'dan elde edilen puanlarla akran değerlendirme puanları ve DPA'dan elde edilen puanlarla öz değerlendirme puanları arasında $\alpha = 0,05$ düzeyinde önemli ilişkiler olduğu bulunmuştur. PA'dan elde edilen puanlar ile DPA'dan elde edilen puanlar arasındaki korelasyon pozitif çok yüksek çıkmıştır. Buna dayanarak öğretmenin PA ve DPA ile objektif puanlama yaptığı söylenebilir. PA puanları ve öğrencilerin öz değerlendirme yapımları sonucunda elde edilen puanlar arasındaki korelasyon pozitif yüksek ve PA puanları ile öğrencilerin akran değerlendirme yapımları sonucunda elde edilen puanlar arasındaki korelasyonun pozitif orta düzeyde olduğu söylenebilir. DPA'dan elde edilen puanlar ile öğrencilerin öz değerlendirme yapımları sonucunda elde edilen puanlar arasındaki korelasyon pozitif yüksek, DPA'dan elde edilen puanlar ile öğrencilerin akran değerlendirme yapımları

sonucunda elde edilen puanlar arasındaki korelasyon ise pozitif orta düzeyde bulunmuştur. Son olarak öğrencilerin öz değerlendirme yapımları sonucunda elde edilen puanlar ile akran değerlendirme yapımları sonucunda elde edilen puanlar arasındaki korelasyon pozitif orta düzeydedir. Ayrıca, bu iki puanlama aracının benzer puanlama basamakları içermesi de ilişkinin yüksekliğinde bir etken olarak görülebilir.

PA ile öz değerlendirme uygulaması sonucu elde edilen puanlar arası korelasyonun yüksek çıkması sonucu ise öğretmenin puanlama aracıyla puanlamayı objektif biçimde yapmasıyla birlikte, öğrencilerin de verilen seminer sonucunda DPA ile puanlamayı öğrenmesi ve ayrıca, DPA'nın öğrencilerin kendilerini değerlendirirken etkili bir puanlama aracının göstermesi gereken özelliği göstermesi nedeniyle korelasyonun yüksek çıkmasına neden olduğu sonucuna varılmıştır. PA ile öğrencilerin akran değerlendirme yapması sonucunda elde edilen puanlar arası ilişkinin orta düzeyde çıkmasının nedeninin bazı öğrencilerin arkadaşlarının kâğıtlarını değerlendirirken gelişigüzel puanladıklarından kaynaklı olduğu düşünülmektedir. Böylece, yapılan gelişigüzel puanlamanın korelasyon üzerinde olumsuz etkide bulunabileceğinin sonucuna varılabilir. PA ile öz değerlendirme puanları arasında elde edilen korelasyona çok yakın korelasyon DPA ile öz değerlendirme puanları arasında çıkmıştır. Yani puanlayıcının DPA ile çok iyi puanlama yapması ve öğrencinin de verilen seminer sonucunda DPA ile puanlamayı öğrenmesi bunun yanında, DPA'nın etkili bir puanlama araç özelliği göstermesi nedeniyle puanlar arasındaki korelasyonun yüksek çıktığı sonucuna varılabilir. PA ile akran değerlendirme puanları arasında elde edilen korelasyona çok yakın korelasyon, DPA ile akran değerlendirme puanları arasında çıkmıştır. Yani Öğretmenin DPA'yı çok iyi kullanması karşısında öğrenciler DPA ile 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 DPA ile puanlarken kendi kâğıtlarında yaptıkları puanlama hatasından daha fazla hata yapsalar da öğrencilerin kendi kâğıtlarını ve arkadaşlarını kâğıtlarını puanlarken benzer davranış gösterdiklerinden dolayı korelasyon katsayısının ortanın üstünde çıktığı sonucuna varılabilir. Yukarıda varılan sonuçlara daha genel ilave yapılacak olursa öğretmen PA'yı ve DPA'yı kullanmayı iyi bildiği ve puanlama yaparken çok az miktarda puanlama hatası yaptığının sonucuna varılabilir. Öğrenciler ise DPA'yı Öğretmen 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 puanların iç tutarlık anlamında güvenilirlikleri Tablo 6'da verilmiştir.

Tablo 6. Puanlama yöntemlerinden elde edilen Cronbach Alfa (α) güvenilirlikleri

Puanlama yöntemleri	Soru sayısı	N	α
PA	10	195	0,756
DPA	10	195	0,778
ÖD	10	195	0,665
AD	10	195	0,667

Tablo 6 incelendiğinde PA'dan elde edilen iç tutarlık ölçütü olan güvenilirlik katsayısı 0,756 bulunmuş olduğundan PA'dan elde edilen ve sınav sorularının iç tutarlığını veren güvenilirlik katsayısının kabul edilebilir düzeyde olduğu söylenebilir. DPA'dan elde edilen iç tutarlık ölçütü olan güvenilirlik katsayısı 0,778 çıktığından sınav sorularının iç tutarlığının DPA ile hesaplanması sonucunda PA'da bulunan katsayıdan göreceli olarak büyük ve kabul edilebilir düzeydedir. Varılan bu sonuçlara göre DPA ile elde edilen iç tutarlık güvenirliliği ile PA ile elde edilen iç tutarlık güvenirliliğinin karşılaştırılması sonucunda DPA'dan elde edilen iç tutarlık güvenirliliği daha yüksek çıkmıştır. Bunun nedeni DPA'da her soru için daha fazla puanlama basamağın kullanılmasının duyarlılığı artırması olabilir.

Güvenirlilik katsayıları arasındaki farkların istatistiksel olarak anlamlı olup olmadığı Fisher'in Z istatistiği ile test edilmiştir.

Tablo 7. Puanlama yöntemlerinden elde edilen Cronbach Alfa (α) güvenirlilik farklarının incelenmesi

N:196	α	PA:0,756	DPA:0,778	AD:0,667
	ÖD:0,665	Zr: 1.818(P<0.05)	Zr:2.343(P<0.05)	Zr:0.035(P>0.05)
	AD:0,667	Zr: 1.783(P<0.05)	Zr:2.308(P<0.05)	
	DPA:0,778	Zr: 0.53 (P>0.05)		

Tablo 7'de analiz sonuçları incelendiğinde, puanlama işlemlerinden elde edilen alfa güvenirlilik katsayıları arasında farkın manidar olup olmadığını test etmek için Fisher'in z testinden yararlanılmıştır. Elde edilen sonuçlara göre PA ile ÖD, PA ile AD, DPA ile ÖD, DPA ile AD güvenirlilikleri arasında anlamlı fark bulunmakta iken DPA ile PA, ÖD ile AD güvenirlilikleri arasında anlamlı fark bulunmamıştır. Buna göre Öğretmenin yaptığı puanlamanın güvenirliliği ile öğrencilerin yaptığı puanlamanın güvenirliliği arasında anlamlı fark olduğu söylenebilir.

Sonuç ve Tartışma

Bu araştırmada yaygın kullanılan ölçme yaklaşımlarından olan yazılı sınavların puanlamasında yararlanılabilecek PA, DPA, öz değerlendirme ve akran değerlendirme ile elde edilen puanların karşılaştırılması amaçlanmıştır. Ulaşılan sonuçlar araştırma problemine bağlı olarak maddeler halinde verilmiştir.

Yazılı sınav cevapları için farklı puanlama yöntemlerinin kullanılması sonucunda elde edilen puanların ortalamaları anlamlı düzeyde farklılaşmaktadır. Öğrencilerin kendi kendilerini ve akranlarını puanlamaları sonucunda elde edilen ortalama puanlar, öğretmenin DPA kullanarak verdiği ortalama puandan manidar düzeyde yüksek çıkmıştır. Öğrenciler kendi kendilerine ve akranlarına öğretmenlerinin verdiği kadar puan vermişlerdir. Ayrıca, öğrencilerin kendilerine yanlı davranmalarından dolayı kendilerine verdiği puanların ortalamaları görece olarak akranlarına verdiği puanların ortalamasından daha yüksek çıkmıştır.

Farklı puanlama yaklaşımlarından elde edilen puanlar arasında anlamlı ilişkiler bulunmuştur. Özellikle öğretmenin puanlama araçlarını iyi kullandığı ve tutarlı puanlama yaptığı PA ve DPA ile verdiği puanlar arasındaki ilişkinin yüksek çıkmasıyla doğrulanabilir. Fakat öğrencilerin akranlarını ve kendilerini değerlendirmeleri sonucu verdikleri puanlar arasındaki ilişki Öğretmenin PA ve DPA ile teknikleriyle verdiği puanlar arasındaki ilişkiden düşük çıkmıştır. Bu sonuç ise öğrencilerin DPA'yı öğretmen kadar iyi kullanamadığı veya yansız puanlama yapmada güçlük çektiği şeklinde yorumlanabilir.

Yazılı sınavın iç tutarlık ölçütü alfa katsayısını bulmak için PA ve DPA gibi farklı puanlama araçlarının uygulanması sonucunda elde edilen alfa katsayılarının hesaplanması sonucunda her ikisinden elde edilen iç tutarlık ölçütleri yüksek çıkmıştır. Elde edilen güvenilirlik düzeyleri arasında anlamlı farklılık çıkmamıştır. Bu sonuç, Parlak ve Dođan (2014) tarafından PA ve DPA yaptıkları güvenilirlik çalışmasında DPA ve PA ile yapılan puanlamadan elde edilen puanları uyum düzeyinin yüksek olması bulgusu ile tutarlıdır.

Öğrencilere anketle yönlendirilen soruların neticesinde, onların DPA ile yapılan puanlama işleminden memnun olduklarını ifade etmeleri DPA'nın öğrenciler üzerindeki motive edici etkisinin bir göstergesidir. Öğrencilerin büyük çoğunluğu kendi kendilerinin yazılı kâğıtlarını puanlarken ve arkadaşlarının yazılı kâğıtlarını puanlarken çok heyecanlandıklarını ve kendilerinin puanlamada hak sahibi olmalarının çok demokratik olduğunu ifade etmişlerdir. Bu bulgulara dayanarak öz ve akran değerlendirmeden yararlanmanın hem öğrencilerin birbiriyle olan etkileşimlerini artıracığı hemde ölçütlerden haberdar olmalarını sağlayarak öğrenme düzeylerini artıracığı söylenebilir.

Öğrencilere PA ve DPA ile verilen puanlar karşılaştırılıp incelendiğinde, ortalamanın altında yer alan öğrencilerin, PA'ya göre DPA'dan daha yüksek alarak ortalamaya yaklaştıkları, ortalamanın üstünde yer alan öğrencilerin ise iki ölçme aracı ile aldıkları puanların birbirine çok yakın olduğu belirlenmiştir. Buna göre, DPA ile değerlendirmenin düşük başarı gösteren öğrencilerin lehine olduğu, yüksek başarı gösteren öğrencilerin başarı durumlarını o kadar etkilemediği sonucuna varılmıştır. Özellikle analitik DPA'da ara basamaklar ayrıntılı puanlandığı için, başarı düzeyi düşük olan öğrencilerin daha avantajlı olduğu sonucuna varılabilir. Çoktan seçmeli ve kısa cevaplı testlerde gerçek başarılarını yansıtamayan öğrenciler için DPA ile değerlendirmenin avantaj yarattığı ifade edilmektedir (Airasian ve Russel, 2008).

Öneriler

Puanlama aracı olarak DPA'nın seçilmesi hem puanlamanın duyarlılığını artıracak hemde öğrencilerin ölçütlerden haberdar olmasını sağlayacaktır. Çünkü öğrencinin PA'daki ölçütleri anlaması, DPA'daki ölçütleri anlamasından daha zordur. Dolayısıyla DPA gerek duyarlılığı gerekse kullanım kolaylığıyla daha güvenilir, daha belirleyici ve daha kullanışlıdır. Ayrıca PA sadece bir tane derse yönelik hazırlanırken DPA birbirine yakın birden çok derse uygulanabilir olması da kullanılabilirliğin kanıtıdır.

Bu araştırmada, araştırma süreci iki döneme genişletilerek, ilk dönem konularından yazılı sınav yapıp ve ikinci dönem konularından yazılı sınav yapılarak, bu yazılı sınavların puanlama araçlarıyla puanlanması sonucunda elde edilen sonuçların karşılaştırılması bu iki puanlama aracı arasındaki farkı karşılaştırmada daha etkili olacaktır.

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 sağlamasından dolayı öğrencilerin yanlışlığı azaltılabilir.

Araştırma, aynı bireylere uygulanan dersin özelliklerine yakın derslere de uygulanırsa dereceli puanlama anahtarının sahip olduğu özellikleri belirlemede daha etkili olacaktır. Uygulanan 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 iyi veri kaynağı olabilir.

Araştırmaya farklı okullardaki aynı düzeyden öğrencilerin katılması ile örneklem hatası azaltılarak, evren daha iyi temsil edilerek, puanlama anahtarı ve dereceli puanlama anahtarının karşılaştırması için daha genellenebilir sonuç elde edilerek genişletme yapılabilir. Böyle bir araştırmada ayrıca genişleyen örneklemin puanlama araçlarından elde edilen puanlayıcı ve alfa güvenirliklerine ve yordama geçerliğine etkisi olup olmadığına bakılabilir.

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

Öğretmenin puanlama yaparken daha detaylı puanlamasından dolayı öğrencilerin soruya verdikleri cevabın her adımından puan almalarının onları memnun ettiği, inandırıcı olduğu ve PA'ya göre çok iyi bir puanlama tekniği olduğu öğrencilerin çoğunluğu tarafından ifade edilmesine dayanarak öğrencilerin yazılı sınavlarını puanlamada DPA kullanmaları önerilebilir.

Kaynakça

- Açıkgöz, Ün. K. (2002). *Aktif öğrenme*. İzmir: Eğitim Dünyası Yayınları.
- Airasian, P. W. (1994). *Classroom Assessment*. Second Edition. New York: McGraw Hill.
- Airasian, P. W. & Russel, M. K. (2008). *Classroom assessment. Concepts applications*. USA: McGraw-Hill.
- Akkuş, O., & Duatepe-Paksu, A. (2006). Construction of a proportional reasoning test and its rubrics. *Eurasian Journal of Educational Research*, 6(25), 1-10.
- Anadol, H. Ö., ve Doğan, C. D. (2018). Dereceli puanlama anahtarlarının güvenilirliğinin farklı deneyim yıllarına sahip puanlayıcıların kullanıldığı durumlarda incelenmesi. *Elementary Education Online*, 17(2).
- 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. https://dhgm.meb.gov.tr/yayimler/dergiler/Milli_Egitim_Dergisi/160/arslan-eraslan.htm adresinden erişim 03.02.2008 tarihinde.
- Aslanoğlu, E. A. ve Kutlu, Ö. (2004). Eğitimde sunu becerilerinin değerlendirilmesinde dereceli puanlama anahtarı uygulamasına ilişkin bir araştırma. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, cilt 36, sayı:1-2.
- Aşkar, P. (2004). *Matematik etkinlikleri bildiriler kitabı*. Ankara: Matematikçiler Derneği Yayınları.
- Bahar, M., Nartgün, Z., Durmuş, S., ve Bıçak, B. (2008). *Geleneksel-Alternatif Ölçme ve Değerlendirme*. Ankara: Pegem Yayınları.
- Baysal, E. A., & Gürbüz, O. C. A. K. (2019). İngilizce Derslerinde Hazırlanan Portfolyoları Değerlendirme: Bir Dereceli Puanlama Anahtarı (Rubrik) Geliştirme Çalışması.
- Brookhart, S. M. (1999). *The art and science of classroom assessment: The missing part of pedagogy*. ASHE-ERIC.
- 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.
- Büyükoztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2018). *Bilimsel Araştırma Yöntemleri*. Ankara: Pegem Yayıncılık.
- Conner, C. (1991). *Assesment and testing in primary School*. Cambridge: The Falmer Press.
- Crocker. L. & Algine, J. (1986) *Introduction to clasical and modern test theory*. New York: Holt. Rinehart and Winston.
- Danışmaz, Z. K., ve Adıbatmaz, F. B. K. (2020). Çocuk Gelişimi Alanında Dereceli Puanlama Anahtarı Hazırlama: Deneysel Bir Uygulama. *Çocuk ve Gelişim Dergisi*, 3(5), 12-28.
- Demir, S. B., ve Yıldırım, Ö. (2019). Yazılı anlatım becerilerinin değerlendirilmesi için dereceli puanlama anahtarı geliştirme çalışması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 47, 457-473.
- Dikli, S. (2003). Traditional and alternative assessments. *The Turkish Online Journal of Educational Technology*, 2(3), 13-19.
- Doğan, N. (2007). "Yazılı Sınavlar". *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Yayınları.
- Doğan, N., ve Parlak, B. (2014). Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Uyum Düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)* 29(2), 189-197 [Nisan 2014]
- Doğan, D., ve Anadolu, H.Ö. (2018). *İlköğretim Online*, 2018; 17(2): s.1066-1076. [Online]: <http://ilkogretim-online.org.tr/doi/10.17051/ilkonline.2018.419355>
- Ertürk, S. (1998). *Eğitimde Program Geliştirme*. Ankara: Meteksan A.Ş.
- Haladyna, T. M. (1997). *Writing test item to evaluate higher order thinking*, USA: Altyn & Bacon.
- Mabry, L. (1999). *A critical guide to altemative assessment*. California: Convin Press INC.
- Millî Eğitim Bakanlığı (MEB), (2003). *Öğrenci merkezli eğitim uygulama modeli*. Ankara: Millî Eğitim Bakanlığı Yayını, Milli Eğitim Basımevi.

- Moskal, B. M. (2000). "Scoring rubrics: What, when and how?" Practical assessment, Research & Evaluation, 7 (3)
- Kutlu, Ö., Doğan, C., D. ve Karakaya, İ. (2008). Öğrenci Başarısının Belirlenmesi. Ankara: PegemA Yayıncılık.
- Özcan, H., ve Taşar, M. F. (2018). Öğretmen adaylarının bilimin doğası anlayışlarının değerlendirilmesine yönelik bir dereceli puanlama anahtarı deseninin geliştirilmesi. *Online Fen Eğitimi Dergisi*, 3(2), 35-46.
- Popham, J. W. (1997). What's wrong and what's right with rubric. *Educational Leadership*, 55, (2), 12
- Senemoğlu, N. (2004). *Kuramdan uygulamaya gelişim, öğrenme ve öğretim*. Ankara: Gazi Kitabevi.
- Sönmez, H. (2019). Bilişsel ve Duyuşsal/Üstbilişsel Temelli Hikâye Değerlendirme Stratejisi ve Dereceli Puanlama Anahtarı. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 20(3), 1079-1119.
- Şimşek, H. (1997). *21. Yüzyılın eşliğinde paradigmlar savaşı: Kaostaki Türkiye*. İstanbul: Sistem yayıncılık.
- Tekin, H. (1996). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları
- Turgut, M. F., ve Baykul, Y. (2011). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Yayınları Onbirinci Baskı.
- Wiggins, G. P. (1993). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.
- Yetiş, V. A. (2017). Performans Değerlendirmelerinde Dereceli Puanlama Anahtarı: Yazı Anlatım Örneği. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 30(2), 683-703.