

KLASİK YAZILI SINAVLA VE ÇOK SORULU TESTLE ELDE EDİLEN ÖLÇÜMLERİN GÜVENİRLİK VE GEÇERLİĞİ

Yard.Doç.Dr.Satılmış TEKİNDAL

Eğitim Bilimleri Bölümü

Eğitim Fakültesi Ondokuz Mayıs Üniversitesi

ÖZET

Bu araştırmanın amacı; kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklamaların güvenilirlik ve geçerliğini karşılaştırmak ve puanlama anahtarları kullanmanın geçerlik ve güvenilirliğe etkisini ortaya çıkarmaktır. Eldeki çalışma, 1995-1996 yılında, Amasya'daki bazı ilkokulların 5.sınıfları üzerinde yürütülmüştür. Araştırmada kullanılan ölçme araçları, bir çoktan seçmeli test ve bu testin açık uçlu sorulara dönüştürülmesi ile elde edilen kısa cevaplı çok sorulu ve uzun cevaplı az sorulu yazılı yoklamalardır. Eldeki çalışmada, aritmetik ortalama, standart kayma, Pearson Momentler Çarpımı Korelasyon Katsayısı, Z istatistiği ve "t" testi kullanılmıştır. Sonuç olarak; kısa cevaplı çok sorulu yazılı yoklamanın güvenilirlik ve geçerliği açısından yararını ve özellikle anahtarla puanlamanın güvenilirliğe katkısını gösteren sınırlı kanıtlar elde edilebilmiştir, özellikle anahtarla puanlamanın etkisi açıkça ortaya konamamıştır. Puanlama anahtarının rolü, ancak uzun cevaplı az sorulu yazılı yoklamada bir ölçüde ortaya çıkmış olup, bu da beklentiye uygundur.

SUMMARY

The purpose of this study is to compare the reliability and validity of the short-answer test and essay test, also to appear the effects of use scoring key (analytic scoring) on this reliability and validity. This study was based on the fifth grade (12 year old) at some primary schools in Amasya in 1995-1996. The instruments used in the study are the multiple-choose test, short-answer test, and essay test. In this research arithmetical mean, standard deviation, Pearson Product Moment correlation coefficient, Fisher's "z" statistics and "t" test were used. Consequently, compared with the short answer test and essay test there is a limited evidence in the usefulness of the short answer test related to the reliability and validity. It is found that the usefulness of scoring key is limited in short answer test. The effect of scoring key appeared somewhat in essay test. This is favorable for the expectation.

* Bu makale yazarın, Klasik Yazılı Sınavla ve Çok Sorulu Testle Elde Edilen Ölçümlerin Güvenirlik ve

Geçerliği; Pegem Özel Eğitim ve Hizmetleri P.K.115, Kızılay-Ankara, kitabının bir kısmının özetidir.

Yazılı yoklama, okullarımızda çok kullanılan bir sınav türüdür. Bazen açık uçlu sınav diye de anılan yazılı yoklamalar, öğrencilere, soruların yazdırılıp veya yazılı verilip, bunlara belli bir sürede yazılı cevaplar verilmesini gerektirir. Bu sınavın uzun cevaplı az sorulu ve kısa cevaplı çok sorulu olmak üzere iki türü vardır.

Uzun cevaplı az sorulu yazılı yoklamalarda soruların cevapları sınırlandırılmamıştır. Bunun, sınavda az soru sorulmasına, puanlamanın yeteri kadar objektif bir şekilde yapılamamasına ve böylece de sınavın geçerlik ve güvenilirliğinin düşmesine neden olduğu söylenebilir. Kısa cevaplı çok sorulu yazılı yoklamalarda ise, soruların cevapları sınırlandırılarak soru sayısı artırılır ve puanlamanın nispeten objektif yapılması sağlanır, böylece de sınav daha geçerli ve güvenilir hale getirilmiş olur.

Diğer taraftan yazılı yoklamaların puanlanmasına etki eden başka faktörler de vardır. Bu faktörlerin olumsuz etkilerinden kurtulmanın yolları bulunmaya çalışılmış ve bu doğrultuda birçok araştırmalar yapılmıştır (Tekindal 1997-1:121-140). Bu araştırmalar puanlayıcıların sıklıkla birbirine katılmadığını, öğrenci cevaplarına verilen puan miktarının cevaplarla ilgisinin birbirine göre oldukça düşük olduğunu gösterir (Coffman, 1972). Puanlama sürecindeki sübjektiflik maddenin doğasına göre değişir, fakat, sübjektifliğe neden olan etkenler genel olarak geniştir. Bulgular, yazılı yoklamanın iç özelliklerinin (bunlar puanlama ile ilgili değil) ve dış özelliklerinin bir fonksiyonu olarak kalitenin etkilendiğini, bireyler tarafından alınan puanların oldukça kötü olduğunu ve yüzeysel durumun asıl içerikten daha uygun hale getirilmeye çalışıldığını göstermiştir (Townsend ve Diğerleri 1989: 1; Townsend ve diğerleri 1993:1). Yine, araştırmalar, yazılı yoklamaların kalitesine etki eden faktörlerin dört grupta toplanabileceğini gösterir. Bunlar (1) öğrencilerle ilgili olanlar, (2) yazılı yoklamaların kendisi ile ilgili olanlar, (3) içerikle ilgili olanlar ve (4) puanlayıcı ile ilgili olanlardır. Örneğin: Öğrencinin kendisi ile ilgili olarak yapılan bazı araştırmalar Townsend, Kek ve Tuck (1989) ve Townsend, Hick, Thompson, Willon, Tuck ve Moor (1993) tarafından belirtildiğine göre şunlardır; öğrencinin ismi (Mc David ve Harari, 1973), ırk (Fajardo, 1985), cinsiyet (Carter, 1952), fiziksel çekicilik (Landy ve Sigall, 1974) vb. Yazılı yoklamaların kendisi ile ilgili olan araştırmalardan bazıları da, tekrar ve dilbilgisi bakımından doğruluk (Diederich ve diğerleri, 1961; Marshall, 1967) ve el yazısının kalitesi (Markham, 1967)'dir. Puanlar, aynı zamanda, diğer yazılı yoklamaların kalitesinin bir fonksiyonu olarak da değişebilir (Hughes ve

Keeling, 1984). Puanlayıcılarla ilgili olarak yapılan bazı araştırmalar, erkek veya kadının etkileşim tarzı (Theodory ve Day, 1985), ırk (Erwin ve Calev, 1984), kişisel önyargılar (Acher ve McCarthy, 1988), öğrenci beklentileri (Chase, 1986), pozitif ve negatif ruh hâli (Isen ve Simmonds, 1978)'dir (Townsend ve diğerleri, 1989:1-2; Townsend ve diğerleri, 1993:1-2).

Yazılı yoklamalar analitik, sınıflama, sıralama ve genel izlenimle puanlama yöntemlerinden biri ile puanlanabilir. Bu yöntemlerden analitik puanlama (anahtarla puanlama)'nın dışında kalanlar güvenilirliği düşük puanlama sonuçları verir. Anahtarla puanlamada, sorulara verilebilecek cevaplar önceden hazırlanır ve her cevap en ince ayrıntılarına kadar analiz edilir. Her öğenin puan limitleri belirlenir, cevaplayıcının yazdığı cevapların anahtardaki cevaplara uyma derecesine göre onlara bir puan takdir edilir. Bu yöntemin diğer yöntemlere göre daha güvenilir puanlar vermesi beklenir.

Genel izlenimle puanlamada, puanlayıcı, bir cevap kağıdının tamamını okur ve onun kendisinde bıraktığı genel izlenime göre, onu en iyiden en kötüye (ya da tersi) bir kaç kategorilik puanın birine koyar. Yani hangi puanlama sistemini kullanıyorsa, ona uygun bir puan verir. Bu puanlama şekli, öğrencinin cevabına niçin eksik puan verildiğine dair bir bilgi sağlamaz. Puanlama standartları açıkça belli değildir. Bu standartlar öğretmenden öğretmene ve hatta bir öğretmende değişik zamanlarda farklılıklar gösterir. Çok sakıncalı bir puanlama yöntemi olup, bu yöntemle elde edilen puanların güvenilirliğinin düşük olması beklenir.

Görüldüğü gibi ölçme aracının yapısı ve puanlamanın biçimi, puanlamanın hatasızca yapılıp yapılmamasına etki eder. Kısa cevap gerektiren veya uzun cevap gerektiren yazılı yoklamalarda olduğu gibi, bir cevabın doğru olup olmadığını tayin etmek puanlayıcının takdirine kalırsa, puanlama yöntemi sübjektiftir. Bu durumlarda puanlama güvenilirliği tam olmaz; puanlara, takdir farklılıklarından doğan, puanlama hataları karışır. Bu anahtarla puanlanabilen seçmeli testlerde olduğu gibi, puanlayıcının görevi, seçilen cevabın anahtara uygun olup olmadığını tayin etmekten ibaret ise, puanlama yöntemi objektiftir.

Puanlamanın sübjektif olduğu yazılı yoklamalarda, iyi hazırlanmış bir puanlama anahtarı (analitik anahtar), puanlamanın bir dereceye kadar objektifleşmesine yardım eder. Bu objektifleşmenin kısa cevap gerektiren çok sorulu yazılı yoklamalarda daha çok olması beklenir.

Uzun cevaplı yazılı yoklamalar, az soru sorulmasına imkan verir. Bu da, sınavın geçerliğini düşürücü bir etki yapar. Bu yüzden,

soruların cevapları sınırlandırılarak soru sayısı artırılır. Soru sayısının artırılmasının, sınavın hem geçerliğini hem de güvenilirliğini artırıcı bir etkiye sahip olması beklenir. Soru sayısını artırmanın ve puanlama yöntemlerini değiştirme-nin ölçme sonuçlarına nasıl bir etkide bulunduğu merak konusudur.

İşte, bu çalışmada, yukarıdaki sayfalarda sözü edilen genel izlenimle puanlama mı, yoksa anahtarla puanlama mı daha güvenilir ve geçerli sonuçlar verir sorusu çalışılmıştır.

Problem*

İlkokul beşinci sınıf Sosyal Bilgiler Dersinin bir ünitesi ile ilgili olarak hazırlanmış olan bir çoktan seçmeli test ve bu testin açık uçlu (kısa cevaplı) sorulara dönüştürülmüş şekliyle oluşan kısa cevaplı çok sorulu yazılı yoklama ile yine aynı çoktan seçmeli testin bazı maddelerinin açık uçlu (uzun cevaplı) sorulara dönüştürülmüş şekliyle oluşan uzun cevaplı az sorulu yazılı yoklama arasında geçerlik ve güvenilirlik bakımından ne gibi farklar vardır?

Alt Problemler

1. Güvenirlikle ilgili olarak

a)Çoktan seçmeli testin, iki yarı yöntemi ve Kuder-Richardson yöntemi (KR-20) kullanılarak hesaplanan güvenilirlik tahminleri nedir?

b)Farklı puanlayıcıların aynı cevaplayıcı grubuna uygulanan kısa cevaplı çok sorulu yazılı yoklama sorularının cevaplarına puanlama anahtarı kullanmadan verdikleri puanlar arasındaki ilişkinin derecesi (puanlayıcı güvenilirliği) nedir? Bu ilişkiler arasında fark var mıdır?

c)Farklı puanlayıcıların, aynı cevaplayıcı grubuna uygulanan kısa cevaplı çok sorulu yazılı yoklama sorularının cevaplarına, puanlama anahtarı kullanarak verdikleri puanlar arasındaki ilişkinin derecesi (puanlayıcı güvenilirliği) nedir? Bu ilişkiler arasında fark var mıdır?

d)Aynı puanlayıcının, aynı cevaplayıcı grubuna uygulanan kısa cevaplı çok sorulu yazılı yoklama sorularına verilen cevapları, puanlama anahtarı kullanılarak puanlaması halinde, bu iki puanlama şekliyle elde edilen puanlar arasındaki ilişkinin derecesi nedir?

* Bu kesimin sonunda ele alınması gereken sayılılar, sınırlamalar, tanımlar, araştırmanın

önemi ve ilgili araştırmalar kısımlarının, yer darlığı nedeni ile üzerinde durulamamıştır.

e)Farklı puanlayıcıların aynı cevaplayıcı grubuna uygulanan uzun cevaplı az sorulu yazılı yoklama sorularının cevaplarına puanlama anahtarı kullanmadan verdikleri puanlar arasındaki ilişkinin derecesi (puanlayıcı güvenilirliği) nedir? Bu ilişkiler arasında fark var mıdır?

f)Farklı puanlayıcıların, aynı cevaplayıcı grubuna uygulanan uzun cevaplı az sorulu yazılı yoklama sorularının cevaplarına, puanlama anahtarı kullanarak verdikleri puanlar arasındaki ilişkinin derecesi (puanlayıcı güvenilirliği) nedir? Bu ilişkiler arasında fark var mıdır?

g)Aynı puanlayıcının, aynı cevaplayıcı grubuna uygulanan uzun cevaplı az sorulu yazılı yoklama sorularına verilen cevapları, puanlama anahtarı kullanmadan ve anahtar kullanarak puanlaması halinde, bu iki puanlama şekliyle elde edilen puanlar arasındaki ilişkinin derecesi nedir?

2. Geçerlikle ilgili olarak

a)Aynı cevaplayıcı grubuna uygulanan çoktan seçmeli testten elde edilen puanlarla bu testteki soruların açık uçlu sorulara dönüştürülmesi ile oluşan kısa cevaplı çok sorulu yazılı yoklamadan puanlama anahtarı kullanılmadan ve puanlama anahtarı kullanılarak elde edilen puanlar arasındaki ilişkiler nedir? Bu ilişkiler arasında fark var mıdır?

b)Aynı cevaplayıcı grubuna uygulanan çoktan seçmeli testten elde edilen puanlarla bu testteki bazı soruların açık uçlu sorulara dönüştürülmesi ile oluşan uzun cevaplı az sorulu yazılı yoklamadan puanlama anahtarı kullanılmadan ve puanlama anahtarı kullanılarak elde edilen puanlar arasındaki ilişkiler nedir? Bu ilişkiler arasında fark var mıdır?

c)Aynı cevaplayıcı grubuna uygulanan kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklama sorularının cevaplarından puanlama anahtarı kullanılmadan ve puanlama anahtarı kullanılarak elde edilen puanlar ile, aynı öğrencilerin sosyal bilgiler dersi birinci dönem "kanaat notları" arasındaki ilişkiler nedir? Bu ilişkiler arasında fark var mıdır?

YÖNTEM

Örneklem

Bu çalışmada, esas olarak, iki değişik ölçme yaklaşımı güvenilirlik ve geçerlik bakımından karşılaştırılmakta olduğundan, sadece bir karşılaştırma araştırmasına imkân verecek verilerin elde edilmesini sağlayacak cevaplayıcı (öğrenci) ve puanlayıcı gruplarından yararlanılmıştır. Bu sebeple bir evren ve bu evreni temsil edecek nitelikte bir örneklem düşünülmemiştir. Bu çalışma, çoktan seçmeli test ile bu testteki

maddelerin (soruların) açık uçlu sorulara dönüştürülmüş şekli olan kısa cevaplı çok sorulu ve uzun cevaplı az sorulu yazılı yoklamaların karşılaştırılması için verilerin toplanmasında işbirliği yapılabileceği düşünülen Amasya Merkez ilkokullarından Atatürk İlkokulu ve Yeşilirmak İlköğretim Okulu beşinci sınıflarındaki şubelerden, dört şubedeki öğrenciler üzerinde yürütülmüştür. Atatürk İlkokulu'nda A, B, C, D ve E olmak üzere beş; Yeşilirmak İlköğretim Okulu'nda A ve B olmak üzere iki beşinci sınıf şubesi vardır. Okulların işleyişinden kaynaklanan sınırlamalar da göz önünde tutularak, Yeşilirmak İlköğretim Okulu'ndan bir şube, Atatürk İlkokulu'ndan da üç şube random olarak seçilmiştir.

Tablo 1

Araştırma Kapsamına Alınan Cevaplayıcıların Okullara ve Şubelere Göre Dağılımı

Okul/Şube	Test Uygulanan Öğrenciler		Değerlendirme Dışı Kalan Öğrenciler		Değerlendirmeye Alınan Öğrenciler	
	F	%	f	%	f	%
Yeşilirmak İlköğretim Okulu;						
B Şubesi:	36	25	13	20	23	30
Toplam:	36	25	13	20	23	30
Atatürk İlkokulu;						
A Şubesi:	38	27	10	15	28	37
C Şubesi:	36	25	18	27	18	24
E Şubesi:	32	23	25	38	7	09
Toplam:	106	75	53	80	53	70
Genel Toplam:	142	100	66	100	76	100

Bu çalışmada yararlanan asıl örneklem grubunun okullara ve şubelere göre dağılımı Tablo 1'de gösterilmiştir. Anılan tabloda görüldüğü gibi, araştırma kapsamına giren cevaplayıcı sayısı Yeşilirmak İlköğretim Okulu'nda 36, Atatürk İlkokulu'nda 106 olmak üzere toplam 142'dir. Bu cevaplayıcıların katılım yüzdeleri şöyledir. Yeşilirmak İlköğretim Okulu için yüzde 25, Atatürk İlkokulu için ise yüzde 75'dir. Araştırma kapsamına giren öğrencilerin bir kısmı değerlendirme dışı kalmıştır. Bunun nedeni, öğrencilerin, uygulama yapıldığı günler-deki devamsızlığıdır. Değerlendirmeye alınan öğrencilerin yüzde 30'u Yeşilirmak İlköğretim Okulu'nda, yüzde 70'i de Atatürk İlkokulu'nda kayıtlıdır.

Verilerin Elde Edilmesi

Bu araştırmanın verileri, çoktan seçmeli testin puanları, kısa cevaplı çok sorulu yazılı yoklamalardan ve uzun cevaplı az sorulu yoklamalardan elde edilen cevaplara puanlama anahtarı kullanmadan ve puanlama anahtarı kullanarak verilen puanlar ile öğrencilere, dönem sonunda sosyal bilgiler dersinden verilen dönem sonu “kanaat notları”dır.

Sınav kapsamı olarak, ilkokul beşinci sınıf Sosyal Bilgiler dersinin “Dünyamız, Yurdumuz ve Komşularımız” ünitesi seçilmiştir. Materyal olarak ilkokul beşinci sınıf Sosyal Bilgiler kitabı kullanılmıştır. Bu ünite ile ilgili hedef-konu öğeleri öncelikle belirlenmiştir.

Belirlenen hedef davranışları içerecek şekilde, yeteri kadar dört seçenekli çoktan seçmeli test maddesi hazırlanmış ve bu test beşinci sınıfları okutan tecrübeli dört sınıf öğretmeni ve bir coğrafya öğretmeni ile iki ölçme değerlendirme

uzmanına ayrı ayrı verilerek, onların birbirinden bağımsız olarak eleştirileri alınmıştır. Eleştiriler ışığında düzeltmeler yapılmıştır. Sorular, asıl örneklemin dışında kalan beş beşinci sınıf öğrencisine bireysel olarak ayrı ayrı cevaplatılmış, cevaplandırma sırasında ortaya çıkan güçlük ve yetersizlikler giderilmiştir.

Testin ön denemesi, 1995 yılı Kasım ayında Amasya merkez Gazi İlkokulu, Yavuz Selim İlköğretim Okulu ve Kılıçarslan İkokulu’ndaki 150 öğrenci üzerinde yapılmıştır.

Geliştirilen testin asıl örneklemin dışındaki beşinci sınıflar üzerinde ön denemesi yapıldıktan sonra, %27’lik alt ve üst gruplarla madde analizi yapılarak ve ölçme uzmanları ile sınıf öğretmenlerinin ayrı ayrı görüşleri alınarak nihai test elde edilmiştir.

Geliştirilen bu çoktan seçmeli testin maddeleri kısa cevaplı çok sorulu yazılı yoklama maddelerine dönüştürülmüştür. Ayrıca, bu çoktan seçmeli test maddelerinden seçilen dokuz soru da, beş soruluk uzun cevaplı az sorulu yazılı yoklama sorularına dönüştürülmüştür. Dönüştür-meler sonucunda ortaya çıkan kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklamalar zaman darlığı ve puanlayıcı bulma zorluğu gibi nedenlerle öğrencilere yeniden yazılı olarak uygulanmamış; ancak, üç Sosyal Bilgiler beşinci sınıf dersi okutan öğretmen ile iki Ölçme ve Değerlendirme uzmanının birbirinden bağımsız olarak görüşleri alınarak uyum sağlanmış ve uzun cevaplı az sorulu yazılı yoklama soruları ile kısa cevaplı çok sorulu yazılı yoklama soruları asıl örneklemin dışında kalan 9 öğrenciye sözlü ve birbirinden bağımsız olarak cevaplatılıp, gerekli düzeltmeler yapılmıştır. Bu düzeltmeler yapıldıktan sonra bu iki tür yazılı yoklama indekslerinin normal seviyeye ulaşacağı varsayılmıştır. Puanlama 100 üzerinden yapılmıştır. Bu ölçme araçları, araştırmacının denetim ve gözetiminde ikişer hafta ara ile uygulanmıştır.

Puanlayıcılar

Kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklamalar puanlama anahtarı kullanılmadan ve puanlama anahtarı kullanılarak 14 puanlayıcıya puanlatılmıştır. Puanlayıcılar, ölçme konusuna yabancı olmasın diye, beşinci sınıfı okutan öğretmenlerin gönüllüleri arasından seçilmiştir.

Sosyal Bilgiler Dersi Dönem Sonu Notları

Bu notlar, araştırma kapsamına alınan cevaplayıcı (öğrenci)’ların 1995-1996 öğretim yılı, güz sömestrisi sonunda Sosyal Bilgiler dersinden aldıkları notlardır. Bu notlar ilgili okulların yönetiminden alınmış ve geçerlik ölçütü olarak aynen kullanılmıştır.

Verileri Çözümleme Yöntemleri

Bu çalışmadaki veriler, alt problemlere cevap verecek biçimde aşağıdaki istatistiksel yöntemler kullanılarak çözümlenmiştir.

Güvenirlikle ilgili olarak,

1.Çoktan seçmeli testin güvenirliliği için KR-20 ve iki yarı yöntemi kullanılmıştır.

2.Farklı puanlayıcıların puanlama anahtarı kullanarak ve bu anahtarı kullanmadan verdikleri puanların ilişkisinin incelenmesinde Pearson Momentler çarpımı korelasyon katsayısını hesaplama yoluna gidilmiştir.

Geçerlikle ilgili olarak

1.Aynı öğrenci grubuna uygulanan çoktan seçmeli testten alınan puanlar ile bu testin dönüştürülmesi ile oluşan kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklama sorularına verilen cevaplara anahtar kullanarak ve anahtar kullanmadan verilen puanlar arasındaki ilişkinin incelenmesi için Pearson Momentler çarpımı korelasyon katsayısını hesaplama yoluna gidilmiştir. Korelasyon katsayılarının ve iki korelasyon katsayısı arasındaki farkın mânidarlığını test etmek için, r değerleri Fisher’in z istatistiğine çevrilerek “t” testi kullanılmıştır.

2.Kısa cevaplı çok sorulu yazılı yoklama sorularının cevaplarına cevap anahtarı kullanarak ve kullanmadan verilen puanlar arasındaki ilişkiyi ve uzun cevaplı az sorulu yazılı yoklama sorularının cevaplarına cevap anahtarı kullanarak ve kullanmadan verilen puanlar arasındaki ilişkiyi belirlemek için Pearson Momentler

çarpımı korelasyon katsayısını hesaplama yoluna gidilmiştir. Korelasyon katsayılarının mânidarlığını test etmek için, r değeri Fisher'in z istatistiğine çevrilerek "t" testi uygulanmıştır.

3.Çoktan seçmeli test soruları cevaplarına verilen puanlar, kısa cevaplı çok sorulu yazılı yoklama soruları cevaplarına cevap anahtarları kullanarak ve kullanmadan verilen puanlar ve uzun cevaplı az sorulu yazılı yoklama soruları cevaplarına cevap anahtarlı ve anahtarsız verilen puanlar ile Sosyal Bilgiler dersi dönem sonu notları arasındaki ilişkinin bulunmasında Pearson Momentler çarpımı korelasyon tekniği kullanılmıştır. Korelasyon katsayılarının ve iki korelasyon katsayısı arasındaki farkın mânidarlığının test edilmesinde, r değeri Fisher'in z istatistiğine çevrilerek "t" testi kullanılmıştır.

4.Pearson Momentler çarpımı korelasyon katsayısının sıfırdan mânidar derecede farklı olup olmadığını test etmek için "t" testi uygulanmıştır.

$$t = r \sqrt{\frac{N-2}{1-r^2}}$$

r:Test edilecek korelasyon katsayısı
N:Ölçme sayısı

İki korelasyon katsayısı arasındaki farkın mânidarlığını test etmek için şu formül kullanılır.

$$z = \frac{Z_{r_1} - Z_{r_2}}{\sqrt{1/(N-3) + 1/(N-3)}}$$

5.Kısa cevaplı çok sorulu ve uzun cevaplı az sorulu yazılı yoklamaların geçerliklerinin mânidarlığı Turgut (1971:64)'ta verilen aşağıdaki formülle hesaplanmıştır. Bu yol testler arasındaki korelasyonları da hesaba kattığından daha doğrudur.

$$t = (r_{1y} - r_{2y}) \sqrt{\frac{N-3}{2(1-r_{12})(1-R^2)}}$$

r_{1y}=İlgili yazılı yoklama puanlarının ölçüt puanlar ile korelasyonu

r_{2y}=Diğer ilgili yazılı yoklama puanlarının aynı ölçüt puanlar ile korelasyonu

r₁₂=Farkları test edilecek iki testin puanları arasındaki korelasyon

İki testin ölçüt puanları ile, o iki testin birleşik korelasyonu

$$R^2 = b_1 r_{1y} + b_2 r_{2y}$$

formülü ile belirlenir.

Burada b ağırlıkları da,

$$b_1 + b_2 r_{12} = r_{1y}$$

$$b_1 r_{12} + b_2 = r_{2y}$$

denklemlerinden hesaplanır.

BULGULAR VE YORUM*

Bu bölümde, ilkokul beşinci sınıf Sosyal Bilgiler dersinden kapsam olarak seçilen "Dünyamız, Yurdumuz ve Komşularımız Ünitesi"ne yönelik olarak geliştirilen bir çoktan seçmeli test ve bu testin açık uçlu sorulara dönüştürülmesi ile elde edilen kısa cevaplı çok sorulu yazılı yoklama ve uzun cevaplı az sorulu yazılı yoklamaların nitelikleri üzerinde durulmuştur. Aşağıda araştırmanın problemlerine ilişkin bulgular sırasıyla açıklanmış ve bunların yorumları yapılmıştır.

Çoktan Seçmeli Test, Kısa Cevaplı Çok Sorulu Yazılı Yoklama ve Uzun Cevaplı Az Sorulu Yazılı Yoklamaların Güvenirlik ve Geçerlikleri**

Aşağıda bu üç ölçme aracının güvenirlik ve geçerlikleri üzerinde durulmuştur.

Güvenirlik

Çoktan seçmeli testin KR-20 güvenirliliği 0.95, iki yarı yöntemi ile bulunan güvenirliliği de 0.78 bulunmuştur.

Ondört puanlayıcının, puanlama anahtarları kullanmadan kısa cevaplı çok sorulu yazılı yoklamaya verdikleri puanlar arasındaki korelasyonların en büyüğü 0.98, en küçüğü 0.68 olup, ortanca korelasyon 0.84'dür. Farklı

* Yer darlığı nedeni ile, ölçme araçlarına ilişkin ortalama ve standart kayma gibi betimsel istatistikler üzerinde durulamamıştır.

** Yine aynı nedenle bu indekslerle ilgili bulgular özetlenmiştir. Ayrıntılı bilgi için, ilk sayfada verilen, yazara ait kaynağa başvurulabilir.

puanlayıcıların aynı yazılı yoklamaya verdikleri puanları arasındaki korelasyonlar ikişerli olarak test edilmiş ve bazıları arasındaki farkların mânidar olduğu görülmüştür. 14 puanlayıcıya ait güvenilirlik katsayısı 0.98'dir.

Puanlayıcıların, puanlama anahtarı kullanılarak kısa cevaplı çok sorulu yazılı yoklamaya verdikleri puanlar arasındaki korelasyonlar 0.98 ile 0.60 arasında değişmekte olup, ortanca korelasyon 0.85'dir. Puanlayıcıların aynı yazılı yoklamaya benzer şekilde verdikleri puanlar arasındaki korelasyonlar ikişerli olarak karşılaştırılmış ve bazıları arasındaki farkların mânidar olduğu görülmüştür. Puanlayıcıların tamamına ait güvenilirlik 0.98'e yükselmiştir.

Adı geçen puanlayıcıların kısa cevaplı çok sorulu yazılı yoklamaya puanlama anahtarı kullanmadan ve kullanılarak verdikleri puanların ortalamaları arasındaki korelasyon 0.92 bulunmuştur.

Aynı puanlayıcıların puanlama anahtarı kullanmadan uzun cevaplı az sorulu yazılı yoklamaya verdikleri puanlar arasındaki korelasyonlar 0.98 ile 0.29 arasında değişmektedir. Ortanca korelasyon da 0.67'dir. Bu yazılı yoklama ile ilgili bazı puanlayıcıların puanları arasındaki korelasyonlar birbirinden mânidar bir şekilde farklı çıkmıştır. 14 puanlayıcının birlikte güvenilirliği ise 0.96'dır.

Ondört puanlayıcının puanlama anahtarı kullanmadan kısa cevaplı çok sorulu yazılı yoklamaya verdikleri puanlar arasındaki korelasyonların ortancası, aynı puanlayıcıların uzun cevaplı az sorulu yazılı yoklamaya aynı şekilde verdikleri puanlar arasındaki korelasyonların ortancasından daha yüksektir.

Söz konusu puanlayıcıların puanlama anahtarı kullanarak uzun cevaplı az sorulu yazılı yoklamaya verdikleri puanlar arasındaki korelasyonlar 0.99 ile 0.72 arasında değişmekte olup, ortanca korelasyon 0.82'dir. Anılan yazılı yoklamaya, ilgili puanlayıcıların verdikleri puanlar arasındaki korelasyonların ikili karşılaştırma farklarının bazıının mânidar olduğu görülmüştür. Puanlayıcıların birlikte güvenilirliği de 0.98'dir.

Aynı puanlayıcıların aynı cevaplayıcı grubuna uygulanan uzun cevaplı az sorulu yazılı yoklamaya puanlama anahtarı kullanarak ve kullanmadan verdikleri puanların ortalamaları arasındaki korelasyon 0.83 bulunmuştur.

Ondört puanlayıcının puanlama anahtarı kullanarak ve kullanmadan kısa cevaplı çok sorulu yazılı yoklamaya verdikleri puanların ortalamaları arasındaki korelasyon, aynı puanlayıcıların puanlama anahtarı kullanarak ve kullanmadan uzun cevaplı az sorulu yazılı yoklamaya verdikleri puanların ortalamaları arasındaki korelasyondan daha yüksek bulunmuştur.

Geçerlik

Bilinen puanlayıcıların puanlama anahtarı kullanmadan kısa cevaplı çok sorulu yazılı yoklamaya verdikleri puanların ortalamaları ile aynı cevaplayıcı grubuna uygulanan çoktan seçmeli test puanları arasındaki korelasyon katsayısı 0.76'dır. Adı geçen puanlayıcıların aynı yazılı yoklamaya puanlama anahtarı kullanarak verdikleri puanların ortalamaları ile aynı çoktan seçmeli test puanları arasındaki korelasyon ise 0.72 bulunmuştur.

Ondört puanlayıcının puanlama anahtarı kullanmadan uzun cevaplı az sorulu yazılı yoklamaya verdikleri puanların ortalamaları ile, aynı cevaplayıcı grubuna uygulanan çoktan seçmeli test puanları arasındaki korelasyon katsayısı 0.74'dür. Söz konusu puanlayıcıların aynı yazılı yoklamaya puanlama anahtarı kullanarak verdikleri puanların ortalamaları ile, aynı çoktan seçmeli test puanları arasındaki korelasyon ise 0.66'dır.

Çoktan seçmeli test ölçütüne göre, kısa cevaplı çok sorulu yazılı yoklamaya puanlama anahtarı kullanılarak verilen puanların geçerliği, uzun cevaplı az sorulu yazılı yoklamaya aynı şekilde verilen puanların geçerliğinden fark kanıtı 0.05 düzeyinde mânidar olmasa da daha yüksektir.

Kısa cevaplı çok sorulu yazılı yoklamaya puanlama anahtarı kullanmadan ondört puanlayıcının verdiği puanların ortalamaları ile, cevaplayıcıların tamamı (N=76)'nın dönem sonu kanaat notları arasındaki korelasyon 0.23'dür. Aynı yazılı yoklamaya puanlama anahtarı kullanarak aynı puanlayıcıların verdikleri puanların ortalamaları ile, aynı kanaat notları arasındaki korelasyon 0.21 bulunmuştur.

Adı geçen puanlayıcıların puanlama anahtarı kullanmadan uzun cevaplı az sorulu yazılı yoklamaya verdikleri puanların ortalamaları ile, cevaplayıcıların tamamının dönem sonu kanaat notları arasındaki korelasyon 0.12'dir. Aynı yazılı yoklamaya adı geçen puanlayıcılar tarafından puanlama anahtarı kullanılarak verilen puanların ortalamaları ile, ilgili kanaat notları arasındaki korelasyon da 0.11 bulunmuştur.

Her iki yazılı yoklamanın da kanaat notları ile uyumlarının düşük olmasının nedeni, dört farklı öğretmenin verdiği Sosyal Bilgiler dersi dönem sonu notlarının birlikte kullanılması olabilir.

Ondört puanlayıcının puanlama anahtarı kullanarak ve puanlama anahtarı kullanmadan kısa cevaplı çok sorulu ve uzun cevaplı az sorulu yazılı yoklamalara verdikleri puanların her birinin ortalamaları ile, örneklem grubunun farklı kısımlarını oluşturan N=28, N=23, N=18 ve N=7 cevaplayıcı gruplarının her birinin kanaat notları arasındaki korelasyonlar (fark kanıtı olmasa da) bir ölçüt grubundan diğerine değişme göstermektedir.

Çoktan seçmeli test ile sosyal bilgiler dersi dönem sonu kanaat notları arasındaki uyuma da bakılmıştır. Bu uyum korelasyonları N=28 için 0.58, N=23 için 0.72, N=18 için 0.79, N=7 için 0.90 ve grubun tamamı N=76 için ise 0.72 bulunmuştur. Bu katsayılar göre, notlar ile çoktan seçmeli test puanları arasında yüksek ilişkiler vardır denilebilir.

Özet olarak, kısa cevaplı çok sorulu yazılı yoklamanın anahtarla puanlanmasının güvenilirlik ve geçerliğe yararı açıkça ortaya konulamamıştır. Anahtarla puanlamanın güvenilirliğe katkısını gösteren sınırlı kanıt elde

edilmiştir. Uzun cevaplı az sorulu yazılı yoklamada puanlama anahtarı kullanmanın güvenilirlikle ilgili yararı bir ölçüde görülmüş, fakat geçerliğine ilişkin yararı açıkça bulunamamıştır.

TARTIŞMA

Bu araştırmada, bir çoktan seçmeli test maddelerinin açık uçlu sorulara dönüştürülmesi ile oluşturulan uzun cevaplı az sorulu ve kısa cevaplı çok sorulu yazılı yoklamaların güvenilirlik ve geçerlikleri çalışılmıştır.

Kısa cevaplı çok sorulu yazılı yoklamalar, 14 puanlayıcıya puanlama anahtarı ve anahtarsız olarak yaptırılmıştır. Puanlama anahtarı kullanmadan verilen puanlar arasındaki korelasyonların ortancası 0.84 olup, 14 puanlayıcının tamamına ait güvenilirlik de 0.98'dir. Anılan yazılı yoklamaya puanlama anahtarı kullanılarak verilen puanlar arasındaki korelasyonların ortancası 0.85, 14 puanlayıcının tamamına ait güvenilirlik ise 0.98'dir. Adı geçen yazılı yoklamaya, ilgili puanlayıcıların puanlama anahtarı kullanarak verdikleri puanlar ile, puanlama anahtarı kullanmadan verdikleri puanlar arasındaki korelasyonların ortancası 0.82'dir. Bu sonuçlar, puanlayıcıların anılan yazılı yoklamayı güvenilir bir şekilde puanladığı ve puanlama anahtarı kullanmadan, puanlama anahtarı kullanıyormuşçasına dikkatli davrandıkları şeklinde yorumlanabilir.

Uzun cevaplı az sorulu yazılı yoklamalar da aynı puanlayıcılar tarafından puanlama anahtarı kullanılarak ve kullanılmadan puanlanmıştır. Puanlama anahtarı kullanılmadan verilen puanlar arasındaki korelasyonların ortancası 0.67, 14 puanlayıcının tamamına ait güvenilirlik de 0.96 bulunmuştur. Adı geçen yazılı yoklamaya puanlama anahtarı kullanılarak verilen puanlar arasındaki korelasyonların ortancası 0.82, puanlayıcıların tamamına ait güvenilirlik ise 0.98'dir. İlgili yazılı yoklamaya puanlama anahtarı kullanılarak ve kullanılmadan verilen puanlar arasındaki korelasyonların ortancası 0.74'dür. Bu sonuçlara göre, puanlayıcıların puanlama anahtarı kullanarak daha dikkatli puanlama yaptıkları, puanlama anahtarı kullanmadan ise biraz daha dikkatsiz davrandıkları söylenebilir.

Yazılı yoklamaların, çoktan seçmeli test ve dönem soru kanaat notları ölçütlerine göre geçerlikleri hesaplanmıştır. Kısa cevaplı çok sorulu yazılı yoklamaya puanlama anahtarı kullanılmadan verilen puanların ortalamaları ile çoktan seçmeli test puanları arasındaki korelasyon 0.76, anılan yazılı yoklamaya puanlama anahtarı kullanılarak verilen puanların ortalamaları ile aynı çoktan seçmeli test puanları arasındaki korelasyon ise 0.72'dir. Bu korelasyonlar arasındaki fark mânidar değildir. Bu durumda, puanlama anahtarı kullanmanın anılan yazılı yoklamaların geçerliğine önemli bir katkı getirdiğini gösteren mânidar bir kanıt yoktur.

Uzun cevaplı az sorulu yazılı yoklamaya puanlama anahtarı kullanılmadan ve kullanılarak verilen puanların ortalamaları ile, çoktan seçmeli test puanları arasındaki korelasyonlar sırasıyla 0.74 ve 0.66'dır. Bu korelasyon katsayıları arasındaki fark 0.05 düzeyinde mânidar değildir. Bu durumda, ilgili yazılı yoklamayı puanlama anahtarı kullanarak puanlamanın sınavın geçerliğine bir katkı getirdiğini gösteren mânidar bir kanıt yoktur.

Örneklemin tamamı (N=76)'nın Sosyal Bilgiler dersi kanaat notları ile kısa cevaplı çok sorulu yazılı yoklamaya puanlama anahtarı kullanılmadan ve kullanılarak verilen puanların ortalamaları arasındaki korelasyonlar sırasıyla 0.23 ve 0.21'dir. Bu korelasyonlardan birincisi 0.05 düzeyinde mânidar, diğeri ise değildir. Adı geçen korelasyonlar arasındaki fark da 0.05 düzeyinde mânidar değildir. Uzun cevaplı az sorulu yazılı yoklamaya puanlama anahtarı kullanılarak ve kullanılmadan verilen puanların ortalamaları ile ilgili notlar arasındaki korelasyonlar sırasıyla 0.12 ve 0.11'dir. Bu korelasyonlar 0.05 düzeyinde mânidar olmayıp, aralarındaki fark da mânidar değildir. Bu geçerlik göstergelerinin düşük çıkmasına, dönem sonu kanaat notlarının objektif verilere dayanmaması, dört ayrı öğretmen tarafından verilmesi ve bu öğretmenlerin farklı kanaatlara sahip olması, yani, öğretmenler arası farkların olması neden olmuş olabilir. Durum böyle olmakla birlikte, bu durumun denenen yazılı yoklamalarda benzer etkiye sahip olacağı düşünüldüğünde, bu verilere bakılarak da puanlama anahtarı kullanarak puanlama yapmanın puanlama geçerliğine önemli bir katkı getirdiğini gösteren mânidar bir kanıt olmadığı söylenebilir.

Örneklemin tamamını oluşturan 76 öğrenciye dört farklı öğretmen tarafından Sosyal Bilgiler dersi dönem sonu notu verilmiştir. Yani N=28, N=23, N=18 ve N=7 kişilik grupların her birine birer öğretmen not vermiştir. Örneklemin oluşturulan bu grupların her birinin dönem sonu notları ile kısa cevaplı çok sorulu ve uzun cevaplı az sorulu yazılı yoklamalara puanlama anahtarı kullanmadan ve puanlama anahtarı kullanılarak verilen puanlar arasındaki korelasyonlar hesaplanmıştır. Bu korelasyonlar arasındaki farkların hiç biri 0.05 düzeyinde mânidar bulunmamıştır. Bu durumda da, her iki yazılı yoklamanın geçerliğine ilişkin olarak anahtar kullanarak puanlamanın bir yararı olduğunu gösteren mânidar kanıt elde edilmemiştir.

Sonuç olarak, kısa cevaplı çok sorulu yazılı yoklamanın güvenilirlik ve geçerliği açısından yararını ve özellikle anahtarla puanlamanın güvenilirliğe katkısını gösteren sınırlı kanıt elde edilebilmiştir; özellikle anahtarla puanlamanın etkisi açıkça ortaya konulmamıştır. Puanlayıcılar anahtarla puanlamada daha iyi yetiştirilerek başka bir araştırmada bu konu üzerinde durulabilir. Puanlama anahtarının rolü, ancak uzun cevaplı az sorulu yazılı yoklamada bir ölçüde ortaya çıkmakta olup, bu da beklentiye uygundur. Eldeki çalışmaya dayalı olarak geliştirilen bazı öneriler şunlardır:

1. Uygulayıcılar için öneriler:

a)Uzun cevaplı az sorulu yazılı yoklama sonuçlarına puanlama anahtarı kullanılarak verilen puanların daha güvenilir olması nedeni ile, yazılı yoklamalar önceden hazırlanan bir puanlama anahtarı kullanılarak puanlanmalıdır.

b)Bu araştırmadaki ondört puanlayıcının tamamına ait güvenilirliğin bir puanlayıcı güvenilirliğinden daha yüksek bulunması nedeniyle, yazılı yoklamalarda puanlayıcı sayısını artırmanın güvenilirliği artırdığı söylenebilir. Bu yüzden, yazılı yoklamaların puanlanmasında birden fazla puanlayıcı çalıştırılıp, onların verdikleri puanların ortalamaları (veya toplamları) alınmalıdır.

2.Araştırmacılar için öneriler:

a)Benzer araştırmalar daha üst seviyelerde (ortaokul, lise vb.) de yapılmalıdır.

b)Bu araştırmada denenmeyen, aynı puanlayıcının bir yazılı yoklama kağıdına değişik zamanlarda verdiği puanların geçerlik ve güvenilirlikleri çalışılmalıdır.

KAYNAKÇA

Coffman, W.E.(1972). "On the Reliability of Ratings Essay Examination". **NCME Measurment In education**. Vol.3.

Tekindal, Satılmış (1997-1). **Klasik Yazılı Sınavla ve Çok Sorulu testle Elde Edilen Ölçümlerin Güvenirlik ve Geçerliği**. Samsun:Cem Ofset.

Towsend, A.R. Michael, Lynley Hicks, Jacquilyn D.M. Thompson, Keri M.Wilten, Bryan F.Tuck and Dennis W. Moore (1993). "Effects of Introduction and Conclusion in Assesment of Student Essays." **Journal of Education Psychology**. Vol.85, No:4.

Towsend, A.R., Michael Lee Yong Kek and Bryan F.Tuck (1989). "The Effect of Mood on The Reliability of Essay Assesment." **British Journal of Educational Psychology**. Vol 59.

Turgut, M. Fuat (1986). **Eğitimde Ölçme ve Değerlendirme Metodları**. (Dördüncü Baskı), Ankara. Saydam Matbaası.