

SEÇME GEREKTİREN, KISA CEVAPLI VE DOĞRU-YANLIŞ TESTLERİNİN MADDE VE TEST ÖZELLİKLERİNİN KARŞILAŞTIRILMASI

Erol KARACA*

ÖZET. - Bu araştırmada seçme gerektiren, kısa cevaplı ve doğru-yanlış test maddelerinin aynı davranışları ölçme yönünden etkililiklerinin ortaya konulabilmesi için madde güçlükleri, ayıricılık gücü indeksleri, test varyansları ve KR-20 güvenilirlikleri arasında anlamlı farklılıkların olup olmadığına bakılmıştır. Bu farklılıkların ortaya konulabilmesi için, seçme gerektiren, kısa cevaplı ve doğru-yanlış tipi maddelerden oluşturulan üç test hazırlanmış ve Eskişehir Mustafa Kemal İlköğretim Okulu yedinci ve sekizinci sınıflarında okuyan 74 öğrenciye üç oturumda uygulanmıştır. Uygulama sonunda elde edilen verilerin işlenmesiyle ilgili olarak her üç teste ait maddelerin güçlük indekslerinin karşılaştırılmasında varyans analizinden yararlanılmış, farklılıkları test etmede ise Scheffé testi kullanılmıştır. Her üç test için madde ayıricılık gücü indeksleri, madde güvenilirlik indeksleri hesaplanıp gruplandırılmıştır. Test güvenilirlikleri KR-20 güvenilirlik katsayısı ile, testlerin varyansları ise F testi ile hesaplanmıştır. Verilerin analizi sonunda; kısa cevaplı teste ait ortalama madde güçlüğü, doğru-yanlış ve seçme gerektiren teste ait ortalama madde güçlüklerinden daha küçük olduğu; doğru-yanlış testinin seçme gerektiren ve kısa cevaplı testten daha çok ayıricı madde içerdiği; kısa cevaplı testin seçme gerektiren ve doğru-yanlış testten daha çok güvenilir olduğu görülmüştür.

Anahtar Sözcükler: Seçme gerektiren test, kısa cevaplı test, doğru-yanlış testi, test maddelerin etkinliği

ACOMPARİSON OF TEST PARTICULARITIES AND TEST ITEMS OF MULTIPLE CHOICE, SHORT ANSWER AND TRUE-FALSE TESTS

ABSTRACT. - In this research, it was looked for if there significant differences in terms of the difficulty indexes discrimination indexes, test variances and reliabilities calculated with KR-20, among the multiple choice, short answer and true-false test items to point out the effectiveness of these tests in terms of measuring same modes of behavior. In order to find out these differences, three tests including the items multiple choice, short answer and of true-false, were prepared. These tests were applied to 74 students studying in 7th and 8th grades in Mustafa Kemal Primary School in Eskişehir. In order to provide the data bases on the processing, the following analyses were made at the end of the application. While comparing the difficulty indexes of the items of all the three tests; the analysis of variance was used, scheffé tests were used to make pairwies comparations. Items discrimination indexes and KR-20 reliability indexes were computed and grouped for each type of tests. The variances of the test were tested with the F test. In results of the research, it was found that the average difficulty belonging to the short answer items was lower than that of the true-false and the multiple choice items;the true-false test items had more items than the short answer and the multiple choice test in terms of discrimination indexes; the short answer test had more reliable items than the true-false and the multiple choice test.

Key Words: Multiple choice test, Short answer test, True-false test, Effectiveness of test items.

* Öğretim Görevlisi Anadolu Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü
Tel: 0-222-3350580/3532 Elektronik Posta: ekaraca@anadolu.edu.tr

GİRİŞ

Bireyin toplum içinde yaşamını sağlıklı bir şekilde sürdürmesi ve insanlığa yararlı olabilmesi için istenilir davranışlarla donanık olması gerekir. İstenilir davranışları geliştirmek, kusurlu davranışları düzeltmek, istenmeyen davranışları silmenin etkili yolu eğitimidir. Planlı davranışların geliştirilmesi amacıyla eğitimin yapıldığı yer okuldur. Eğitime tabi tutulan kişilerde eğitim süreci sonucu olarak planlanan davranış değişikliklerinin meydana gelmesi beklenir. Eğitim işleminin tamamlanmasıyla bu davranışlar eğitilenlere kazandırılmış olabilir. İstenilen davranışların hiç oluşmaması da mümkündür. Bu durumun kontrolü ölçme ve değerlendirmeyi gerektirir.

Okul öğrenmelerini ölçme nedenleri çeşitlidir: Bunlar; öğrencilerin bir derse ne derecede hazır olduklarını belirleme, dersten öğretilmesi planlanmış olan davranışların daha önceden öğrenilmiş olup olmadığını saptama, diğer bir deyişle, bir dersin aday hedefleri arasından bu dersin belli bir gruba öğretilişi ile ilgili gerçek hedefleri seçme, her bir öğrenme ünitesi sonunda, bu öğrenme ünitesinde öğretilmesi planlanmış olan davranışlardan hangilerinin tam olarak öğrenilmiş, hangilerinin tam olarak öğrenilmemiş ve nihayet öğrenilmemiş olanların da muhtemelen neden öğrenilmemiş olduğunu meydana çıkarma, kısaca söylemek gerekirse, ünitedeki öğrenmeleri izleme ve dersin belli dönemlerinin sonunda, bu dönemde etkileşimlerin ürünü olarak meydana gelen, dersin hedefleri ile tutarlı öğrenme düzeylerini belirleme, kısacası, dersteki erişiyi ortaya koyma (Özçelik, 1992-2) olarak özetlenebilir.

Öğrencilere bazen programın tekrarlatıldığı veya bir kısım öğrencilerin sistemden çıkarıldığı olur. Bütün bu işlemler öğrenci hakkında verilen bazı kararların sonucunda yapılır. Öğrenciler hakkında bu çeşit kararlar çoğu zaman öğrenci başarısına ilişkin bir takım ölçmelere ve değer yargılarına dayanır. Ne yazık ki başarıya dayanan kararlar her zaman isabetli olamaz. İsbetli bir karar verilebilmesi için karara dayanak olacak kadar yeterli bilgi toplanması, toplanan bilgilerde önemli hataların bulunmaması ve bilgilerin doğru yorumlanması gerekir (Turgut, 1988).

Bu kararların doğru alınabilmesi için, özellikleri bilinen, öğrenci başarısını ortaya koymadaki etkililiği önceden denenmiş ölçme araçlarına ihtiyaç duyulur.

Her karar bir değer yargısı içerir. Değer yargılarına gözlem veya ölçme sonuçlarının bir ölçüte vurulması ile ulaşılır. Bir değer yargısının doğruluğu ise, hem dayandığı gözlem veya ölçmelerin doğruluğuna hem de seçilen değerlendirme ölçütünün uygunluğuna bağlıdır.

Sınıftaki öğretimin iyi yürütülebilmesi için, öğrenci başarısının sık sık ölçülüp değerlendirilmesi gerekir. Eğitimde birçok değişken ölçülebilir. Ölçülen büyüklükler çoğu halde öğrencilerin bir takım nitelikleridir. Bu nedenle, ölçme işlemi her zaman ölçülecek büyüklüğün tanımlanmasıyla başlar. Ölçme yukarıda da belirtildiği gibi, araç gerektirir. Ölçme araçları, ölçülen büyüklüğün gözlenip sayılarla ifadesini kolaylaştırır. Her araç bir bakıma gözlemin daha duyarlı yapılmasını sağlar. Eğitimde kullanılan ölçme araçları, ölçülecek büyüklüğün tabiatına ve aracın kullanılacağı gruba göre çeşitlilik gösterir. Ölçme yapacak kimse, değerlendirmenin amacına uygun bir araç seçme; yoksa böyle bir aracı bizzat yapma ihtiyacı duyar (Turgut, 1988).

Eğitimde başarıyı etkileyen faktörler arasında bilişsel olanlar önemli bir yer tutar. Bloom'a (1956) göre, okuldaki öğrenmeler arasındaki değişkenliğin yaklaşık yarısı bilişsel faktörlerden gelir. Bilişsel faktörler denilince yetenekten söz etmek gerekir.

Yetenek, zekânın öğrenmelerden sonraki hali hazır durumu olarak ifade edilebilir. Sözel yetenek de bireyin dili kullanmada, sözel ilişkileri görmede, okuduğunu anlamada belli bir andaki durumunu ifade eden bir kavramdır.

Thurstone Çoklu Faktörler Teorisi'nde belirtilen yeteneklere ilişkin yedi temel faktör belirlenmiştir. Bu faktörler, test geliştirme açısından ve yetenek testlerinde kullanılacak soruların niteliği açısından önemli katkılar getirmiştir. Araştırmada incelenen sözel faktör, yeteneklerle ilgili olarak yapılan faktör analizi tekniklerinin hemen hepsinde ortaya çıkan bir faktördür. Sözel yeteneğin alt faktörlerinden olan sözcükleri anlama, sözcüklerin anlamlarını ve yazılı materyalleri anlama yeteneğini ifade eder. Bir çok yetenek testlerinde kelime hazinesi, cümle tamamlama, cümle düzenleme, sözcüklerin eş anlamını, zıt anlamını bulma ve türlü şekilleri yer alır. Kelime akıcılığı faktörü de sözel faktörlerden bir diğeridir. Bunun diğer faktörlerden farkı, yazılı materyali anlamakla pek o kadar ilgisi olmayışıdır. Kelime akıcılığı faktöründe bir cinsten olan çok sayıda kelimeleri süratle hatırlayabilme önemlidir. Thurstone, kelimeleri anlama ile kelime bolluğu faktörleri arasında korelasyon bulmuştur. Ancak, her iki faktör birbirinden farklı zihinsel faktörlerdir. Sözcükleri anlamak, o sözcüklerin zorluk ve kolaylığı ile yakından ilgilidir (Özgüven, 1994).

Sözel yeteneğin ölçülmesinde seçme gerektiren testler büyük ölçüde kullanılmaktadır. Seçme gerektiren testlerde, maddelerin ve testlerin özellikleri sayısal olarak belirtilebilir. Ayrıca testin özelliklerini maddenin özelliklerine bağlayan fonksiyonlar yardımıyla testin özellikleri bir dereceye kadar kontrol altında tutulabilir. Seçme gerektiren testlerin bu özelliği önemli bir avantaj olarak görülebilir (Baykul, 1989).

Seçme gerektiren test maddeleri cevaplayıcıya madde kökünde sorulan sorunun cevabı olan ve çeldirici adı verilen, cevap diye verilen ifadelerden oluşturulur. Böylece cevaplayıcıya sorunun cevabı da verilmiş olmaktadır. Bu husus bazı psikometristler tarafından böyle maddelerin köklü davranışları ölçmeyeceği iddiasını getirmektedir. Diğer bir görüş de bu iddiaya karşı gelmekte ve sorular iyi yazılırsa çok üst düzey davranışların ölçülebileceğini savunmaktadır (Turgut, 1988). Şüphesiz çoktan seçmeli maddelere getirilen diğer bir eleştiri de doğru cevabın ölçülen özelliğe sahip olmayanlar tarafından şansa bulunabileceği hususudur. Belirtilen bu hususlardan birincisi maddelerin, dolayısıyla ölçme aracının geçerliği, ikincisi de güvenilirliği ile ilgilidir.

Doğru-yanlış testleri özel bir seçme gerektiren test olduğu için, seçme gerektiren testler için getirilen eleştiriler doğru-yanlış testleri için de geçerlidir. Ancak Turgut'a göre (1988), doğru-yanlış testleriyle üst düzeydeki davranışlar da ölçülebilir. Kısa cevaplı testlerin cevaplayıcının, cevabı kendisinin yazması sebebiyle davranışları daha köklü bir şekilde ölçtüğü ifade edilmektedir (Turgut, 1988). Davranışların ölçülüp ölçülmemesi, ölçme aracının daha çok geçerliği ile ilgilidir. Geçerlik ile bir ölçme aracının veya yönteminin, onunla ölçülmek istenen değişkeni ölçüp ölçmediği; ölçebiliyorsa, onu başka değişkenlerden ne derece arınık olarak ölçtüğü ifade edilir (Turgut, 1988).

Bir ölçme aracının diğer önemli özelliği güvenilirliğidir. Güvenirlik ise, herhangi araç ve yöntemin ölçtüğü değişkeni ne derece duyarlıkla ölçebildiği; başka bir deyişle ölçme sonuçlarının tesadüfi hatalardan ne derece arınık olduğudur (Turgut, 1988).

Bir ölçme aracının diğer özellikleri arasında, test puanlarının özellikleri de gelir. Bu istatistikler, testin aritmetik ortalaması ve standart kaymasıdır.

Bir testle elde edilecek ölçme sonuçlarının kullanış amaçlarına hizmet edebilmesi, bu sonuçların yeterince hatasız ve geçerli olmasına, bu da testin amaçlara uygun niteliklerde hazırlanmasına bağlıdır. Bir testin nitelikleri ise geniş ölçüde, onu oluşturan maddelerin niteliklerine bağlıdır (Turgut, 1979).

Test özellikleri, test ve madde özellikleri yardımıyla madde özelliklerinin önceden bilinmesi halinde kestirilebilir. Ayrıca bu ilişkiler yardımıyla istenen niteliklerde ölçme araçları da elde edilebilir.

Bu araştırmada seçme gerektiren, kısa cevaplı ve doğru-yanlış testlerinin madde ve test özellikleri arasında ne gibi farklılıkların olduğu incelenmiştir. Araştırma sonuçlarının test geliştiricilerine, alınacak kararlara ve veri toplamada kullanılan araçların özelliklerinin bilinmesine katkı sağlayacağı umulmaktadır.

YÖNTEM

Araştırmanın Türü

Bu araştırmada seçme gerektiren, kısa cevaplı ve doğru-yanlış maddelerin güçlük indeksleri, ayırıcılık gücü indeksleri, güvenilirlik ölçüleri, bu tip maddelerden oluşturulan testlerin aritmetik ortalamaları, ortalama güçlükleri, varyansları, testin KR-20 güvenilirliklerine göre ortaya çıkan farklılıklar incelendiğinden araştırma temel nitelikli bir araştırma sayılabilir.

Veri Toplama Araçları

Araştırma verilerinin toplanmasında araştırmacı tarafından geliştirilen “Kelime İlişkileri” ve “Okuduğunu Anlayabilme” yeteneklerini ölçmeye yönelik seçme gerektiren, kısa cevaplı ve doğru-yanlış tipi maddelerden oluşturulan üç test hazırlanmıştır.

Testlerin hazırlanmasında, önce ilköğretim okullarındaki Türkçe dersi öğretim programının yedinci ve sekizinci sınıflarına ait “okuduğunu anlayabilme ve kelime ilişkilerine yönelik hedefler ve davranışları” değiştirilerek alındı. Daha sonra ilgili hedefe ulaşıldığına kanıt olabilecek nitelikte kritik ve gözlenip ölçülebilir olup olmadığı, davranışların hedefi tam olarak yansıtıp yansıtmadığı konularında program geliştirme ve ölçme-değerlendirme uzmanlarının görüşlerine başvuruldu. Uzmanların görüş ve eleştirileri dikkate alınarak her bir hedef ile ilgili davranışlar listesi hazırlandı.

Davranışları yoklayacak şekilde seçme gerektiren, kısa cevaplı ve doğru-yanlış tipi maddelerden oluşan 83'er maddelik üç ön deneme testinin bu maddeleri çeşitli kaynaklardan seçilen parçalar (cümle, paragraf, metin), çoğu kez istenilen nitelikte ve sayıda soru çıkarılmasına elverecek biçimde düzenlenerek yeniden yazıldı.

Önce seçme gerektiren tipteki maddeler yazıldı; bu maddeler kısa cevaplı ve doğru yanlış tipteki maddelere dönüştürüldü. Seçme gerektiren maddeler dört seçenekli hale getirildi. Doğru-yanlış tipteki maddelerin başında öğrencilerin, cevaplarını “D” veya “Y” simgeleri ile belirtecekleri boşluklar mevcuttur. Kısa cevaplı maddelerde ise öğrencilerin cevaplarını yazabilecekleri, noktalarla belirtilen, boşluklar bulunmaktadır.

En sonunda, denenmek üzere seçme gerektiren yetenek testi, kısa cevap gerektiren yetenek testi ve doğru-yanlış yetenek testi olmak üzere, yukarıda belirtilen sayıda madde içeren üç ön deneme testi oluşturuldu.

Ön Deneme Uygulaması: Hazırlanan ön deneme testlerinin yönergelerine cevaplama süreleriyle ilgili herhangi bir zamanlama konulmadan, testler

Mustafa Kemal İlköğretim Okulu yedinci ve sekizinci sınıf öğrencilerinden 24, 25 ve 26 mevcudluk üç şubeden toplam 74 öğrenciye uygulandı.

Uygulamanın başlangıcında her grupta sınıfların Türkçe öğretmenleri ve araştırmacı tarafından motivasyonun sağlanması amacıyla, bu testlerde alınan puanların sınıf geçmelerine etkisi olmasa dahi, öğrencilerin bu alandaki yeteneklerini görmeleri bakımından önemli olduğu, alacakları puanların kendilerini tanımlarında yardımcı olacağı öğrencilere açıklandı.

Öğrenciler birbirlerinden yararlanamayacak şekilde oturtuldu ve uygulama sınıftaki öğrencilere testlerin verilmesi ile başladı. Cevaplama işlemini tamamlayan öğrencilerin cevaplama süreleri cevap kağıtlarına yazıldı. Bu sürelerden yararlanılarak testlerin cevaplandırılmasının ne kadar bir zaman dilimi gerektirdiği saptandı.

Ön deneme uygulaması sonunda cevaplama sürelerinin seçme gerektiren test için 45-55 dakika; kısa cevaplı test için 55-65 dakika; doğru-yanlış testi için 45-60 dakika arasında değiştiği gözlemlendi. Bundan, bu üç testteki maddelerin her biri için, yaklaşık, 55 dakikalık bir cevaplama zamanı verilmesi gerektiği saptandı.

Ön deneme grubunun cevap kağıtları, her üç test için ayrı ayrı puanlanarak, her bir öğrencinin her bir testten aldığı puanlar bulundu. Puanlama, her üç testin maddeleri için doğru cevaplara "1", yanlış cevaplanmış maddelerle boş bırakılanlara "0" puan verilerek yapıldı. Kısa cevaplı testin maddelerinde kısmi puanlama yapılmadı. Bir maddeye verilen cevap, anahtara tamamen uygunsa "1", aksi halde "0" ile puanlandı.

Her üç testin ön deneme madde analizi yapılarak her bir maddedeki cevapların seçeneklere dağılışı çıkarıldı. Her bir maddenin güçlük derecesi (\bar{p}), ayırt etme güçleri (r_j) ve güvenilirlikleri (r) hesaplandı. Seçme gerektiren maddelerin güçlük derecelerinin (.11; .97), ayırt etme güçlerinin (-.13; .57) ve madde güvenilirliklerinin (-.06; .25); kısa cevaplı maddelerin güçlüklerinin (.14;.95), ayırt etme güçlerinin (-.02; .55) ve madde güvenilirliklerinin (-.01; .27); doğru-yanlış maddelerin güçlüklerinin (.18; .97), ayırt etme güçlerinin (-.06; .50) ve madde güvenilirliklerinin (-.03; .23) aralıklarında değiştiği gözlemlendi. KR-20 güvenilirlikleri ise seçme gerektiren test için .84, kısa cevaplı test için .89 ve doğru-yanlış testi için .82 bulunmuştur.

Esas Uygulama: Son denemede kullanılan maddeler, ön denemede kullanılan maddelerden seçildiğinden, son denemede kullanılan maddelerin yazım şekli, ön denemede kullanılan maddelerin yazım şekli ile aynıdır. Seçme gerektiren maddelerden oluşturulacak nihai test için ön deneme sonuçlarından

yararlanılmıştır. Buna göre, ayırcılık gücü düşük olan maddeler teste alınmamıştır. Çeldiricileri iyi işlemeyen maddelerin çeldiricileri düzeltilmiş, yeniden yazılması gerekli görülenler değiştirilmiş ve bazı maddeler daha açık ifade edilmiştir. Böylece nihai testler 60 maddeden oluşturulmuştur.

Hazırlanışları yukarıda açıklanan ölçme araçları, bilginin toplandığı öğrenci grubunun dışında aynı seviyedeki başka şubelerden random olarak oluşturulan 24 kişilik bir gruba uygulanmıştır. Uygulama, araştırmacının gözlemi altında, ders saatleri içinde, ders sorumlusu öğretmenler tarafından yapılmıştır.

Öğrenciler testleri üç oturumda cevaplamışlardır. Üç test çeşidinden birini alan üç sınıf, birer hafta arayla ikinci ve üçüncü testi almışlardır. Uygulamanın başlangıcında ön denemede olduğu gibi her grupta, sınıfların Türkçe öğretmenleri ve araştırmacı tarafından, motivasyonunun sağlanması amacıyla, bu testlerde alınan puanların sınıf geçmelerine etkisi olmasa dahi öğrencilerin bu alandaki yeteneklerini görmeleri bakımından önemli olduğu, alacakları puanların kendilerini tanımalarında yardımcı olacağı öğrencilere açıklandı. Uygulamada öğrencilere seçme gerektiren testin cevaplandırılması için 50; kısa cevaplı testin cevaplandırılması için 60; doğru-yanlış testin cevaplandırılması için de 55 dakikalık süre verildi. Uygulama sırasında bir aksaklık gözlenmemiştir.

Verilen İşlenmesi

Öğrencilerden alınan cevap kağıtları, cevapsız bırakılan ya da birden çok cevap işaretlenen madde bulunup bulunmadığı bakımından iyice gözden geçirilmiştir. Kontrol sonunda hatalı kağıda rastlanmamıştır. Daha sonra her bir cevap kağıdı, önceden hazırlanmış olan “cevap anahtarı” kullanılarak, elle, araştırmacı tarafından puanlanmıştır.

Seçme gerektiren, kısa cevaplı ve doğru-yanlış testlerinin madde ve test özellikleri arasında ne gibi farklılıkların olduğunun incelenmesi için, her üç testin maddelerinin güçlük dereceleri, ayırt etme güçleri, madde güvenilirlikleri ve KR-20 güvenilirlikleri hesaplanmıştır. Buna göre, seçme gerektiren maddelerin güçlük derecelerinin (.04; .96), ayırt etme güçlerinin (-.19; .68) ve madde güvenilirliklerinin (-.08; .34); kısa cevaplı maddelerin güçlüklerinin (.13; .96), ayırt etme güçlerinin (-.41; .60) ve madde güvenilirliklerinin (-.18; .30); doğru-yanlış maddelerin güçlüklerinin (.17; .96), ayırt etme güçlerinin (-.39; .65) ve madde güvenilirliklerinin (-.15; .32) aralıklarında değiştiği görülmektedir. KR-20 güvenilirlikleri ise seçme gerektiren test için .79, kısa cevaplı test için .86 ve doğru-yanlış testi için .84'dür.

BULGULAR VE YORUM

Bu arařtırmada seme gerektiren, kısa cevaplı ve doęru-yanlıř testlerinin madde ve test özellikleri arasında fark olup olmadığına iliřkin elde edilen bulgular beř bařlık altında incelenerek yorumlanmıřtır.

Seme Gerektiren, Kısa Cevaplı ve Doęru-Yanlıř Testlerinin Aynı Davranıřları Yoklayan Maddelerinin Güçlük İndekslerinin Karřılařtırılması

Seme gerektiren, kısa cevaplı ve doęru-yanlıř testlerinin aynı davranıřları yoklayan maddelerinin güçlük indeksleri arasında fark olup olmadığını saptamak üzere söz konusu testlerin aynı davranıřları yoklayan maddelerinin güçlük indeksleri hesaplanmıř ve ortalama güçlükler Tablo 1’de verilmiřtir.

Tablo 1 Seme Gerektiren, Kısa Cevaplı ve Doęru-Yanlıř Testlerinin Aynı Davranıřları Yoklayan Maddelerinin Güçlük İndekslerine Ait İstatistikler

TESTLER	N	K	Σp_j	\bar{p}_j	$\bar{z}p$
Seme Gerektiren Test	24	60	42.38	.71	58.75
Kısa Cevaplı Test	24	60	35.04	.58	50.34
Doęru-Yanlıř Test	24	60	41.38	.69	57.42

Tablo 1’deki bilgilere göre en düşük ortalama kısa cevaplı maddelere, en yüksek ortalama da seme gerektiren maddelere aittir. Her üç test maddelerinin güçlük indeksleri z’ye dönüřtürülerek normalleřtirilmiř, ortalamalar elde edilmiř ve daha sonra bu ortalamalar arasındaki farkların manidar olup olmadığına tek yönlü varyans analiziyle bakılarak Tablo 2’deki sonuçlar elde edilmiřtir.

Tablo 2 Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Tipindeki Maddelerin Güçlük İndekslerine Ait Varyans Analizi Tablosu

Değişken Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F
Gruplar Arası	2450.23	2	1225.11	6.06**
Hata	35790.54	177	202.21	
Toplam	38240.77	179		

**p<.01

Tablo 2'deki varyans analizi sonuçlarına göre ortalama güçlükler arasında manidar farklılıklar vardır. Bu manidar farkların hangi gruplardan kaynaklandığını anlamak için, varyans analizi yapılan madde güçlük indekslerinin ortalamaları Scheffe testi ile Tablo 3'te karşılaştırılmıştır.

Tablo 3 Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin Ortalama Madde Güçlüklerinin Karşılaştırılması

TESTLER	Seçme Gerektiren (58.75)	Kısa Cevaplı (50.34)	Doğru-Yanlış (57.42)
Seçme Gerektiren (58.75)	-		
Kısa Cevaplı (50.34)	8.41**	-	
Doğru-Yanlış (57.42)	1.33	-7.08**	-

**p<.01

Tablo 3'teki bilgilerden kısa cevaplı maddelere ait ortalama madde güçlüğü, doğru-yanlış ve seçme gerektiren maddelere ait ortalama madde güçlüklerinden daha küçük olduğu .01 manidarlık düzeyinde söylenebilir.

Madde güçlük indeksi 1'e yaklaştığı oranda maddenin öğrencilere kolay geldiği söylenebilir. Buna göre her üç madde tipi de öğrencilere çok zor gelmemiştir. Bununla birlikte söz konusu bulgulara dayalı olarak, hem ortalama yönünden hem de madde bazında, kısa cevaplı maddelerin doğru-yanlış ve seçme gerektiren maddelerden cevaplayıcılara daha zor geldiği, ortalama güçlük yönünden doğru-yanlış maddelerle seçme gerektiren maddeler arasındaki farkın manidar olmadığı söylenebilir. Bu bulgu, Tripp (1982) ve Lancaster (1987) tarafından yapılan çalışmaların bulgularıyla bir paralellik göstermemektedir. Bu konuda daha başka çalışmalara ihtiyaç vardır.

Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin Aynı Davranışları Yoklayan Maddelerinin Ayırıcılık Gücü İndekslerinin Karşılaştırılması

Seçme gerektiren, kısa cevaplı ve doğru-yanlış testlerinin aynı davranışları yoklayan maddelerinin ayırıcılık gücü indeksleri arasında fark olup olmadığını saptamak üzere söz konusu testlerin aynı davranışları yoklayan maddelerinin ayırıcılık gücü indeksleri hesaplanmış ve bu değerler gruplandırılarak Tablo 4'te verilmiştir.

Tablo 4 Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin Aynı Davranışları Yoklayan Maddelerinin Ayırıcılık Gücü İndekslerine Ait Özet Bilgiler

TESTLER	N	K	.19'dan Küçük	.19-.29	.29'dan Büyük
Seçme Gerektiren Test	24	60	26	10	24
Kısa Cevaplı Test	24	60	23	11	26
Doğru-Yanlış Test	24	60	19	5	36

Tablo 4'te madde ayırıcılık gücü indeksi sayıları verilmiştir. Madde ayırıcılık gücü indeksi .19 ve daha yukarısı olan maddelerin bilen ile bilmeyeni daha iyi ayırdığı kabul edildiğine göre doğru-yanlış testinin daha çok sayıda (36) iyi ayırıcı özellikte maddelere sahip olduğu söylenebilir. Daha sonra kısa cevaplı test (26) ve seçme gerektiren (24) test gelmektedir.

Madde ayırıcılık gücü indeksinin 1'e olabildiğince yaklaşması istenen bir durumdur. Buna göre her üç madde tipinde farklı değerler olmakla birlikte pozitif değerler ağırlıktadır.

Dođru-yanlıř test maddelerinde ayırıcılık gc indeksi yksek olanların sayısı seme gerektiren ve kısa cevaplı madde sayılarından daha yksek olduđundan ve ayırıcılıđı yksek olan maddelerden oluřan bir testin geerliđinin de yksek olacađı sylenebilir. Bu durumda dođru-yanlıř maddelerle kısa cevaplı maddelerin seme gerektiren maddelere gre daha geerli madde oluřturabileceđi ihtiyatla da olsa sylenebilir. Ancak bu konuda daha geniř kapsamlı arařtırmalara gereksinim vardır.

Seme Gerektiren, Kısa Cevaplı ve Dođru-Yanlıř Testlerinin Aynı Davranıřları Yoklayan Maddelerinin Gvenirliklerinin Karřılařtırılması

Seme gerektiren, kısa cevaplı ve dođru-yanlıř testlerinin aynı davranıřları yoklayan maddelerinin gvenirlikleri arasında fark olup olmadıđını saptamak zere sz konusu testlerin aynı davranıřları yoklayan maddelerinin gvenirlikleri hesaplanmış ve buna iliřkin zet bilgiler Tablo 5’de verilmiřtir.

Tablo 5 Seme Gerektiren, Kısa Cevaplı ve Dođru-Yanlıř Testlerinin Aynı Davranıřları Yoklayan Maddelerinin Gvenirliklerine Ait zet Bilgiler

TESTLER	N	K	.15’den Kk	.15-.19	.20 ve Byk
Seme Gerektiren Test	24	60	43	8	9
Kısa Cevaplı Test	24	60	33	15	12
Dođru-Yanlıř Test	24	60	37	12	11

Tablo 5’de madde gvenirlik sayıları verilmiřtir. Madde gvenirlik indeksi, madde ayırıcılık gc indeksi ve madde standart kayması ile dođru orantılı olduđundan gvenirliđi yksek olan maddelerden oluřturulan bir testin gvenirliđi de yksek olur. Buna gre kısa cevaplı testin daha ok sayıda (12) yksek gvenirliđe sahip madde ierdiđi sylenebilir. Daha sonra dođru-yanlıř testi (11) ve seme gerektiren test (9) gelmektedir.

Seme Gerektiren, Kısa Cevaplı ve Dođru-Yanlıř Testlerinin Varyanslarının Karřılařtırılması

Seme gerektiren, kısa cevaplı ve dođru-yanlıř tipi maddelerden oluřturulan testlerin varyansları arasında fark olup olmadıđını saptamak zere sz

konusu test maddelerinin varyansları hesaplanmış ve elde edilen sonuçlar Tablo 6’da verilmiştir.

Tablo 6 Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin Varyansları

TESTLER	N	K	Varyanslar
Seçme Gerektiren Test	24	60	41.09
Kısa Cevaplı Test	24	60	79.57
Doğru-Yanlış Test	24	60	59.90

Tablo 6’da görüldüğü gibi, kısa cevaplı maddelerden oluşturulan testin varyansı en büyük (79.57) ve seçme gerektiren maddelerden oluşturulan testin varyansı da en küçük (41.09) bulunmuştur. Varyanslar arasındaki farkların manidar olup olmadığı F testiyle yoklanmış ve Tablo 6’daki sonuçlar elde edilmiştir.

F testi ile en büyük varyans (79.57) en küçük varyansa (41.09) oranlanmak sureti ile yoklanmıştır. Bu hipotez sonunda manidar bir fark olmadığı gözlenmiştir. Bu bilgilere göre araştırma konusu testlerin varyanslarının homojen olduğu söylenebilir.

Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin KR-20 İle Hesaplanan Güvenirliklerinin Karşılaştırılması

Seçme gerektiren, kısa cevaplı ve doğru-yanlış tipi maddelerden oluşturulan testlerin KR-20 ile hesaplanan güvenirlikleri arasında fark olup olmadığını saptamak üzere söz konusu test maddelerinin KR-20 ile hesaplanan güvenirlikleri bulunmuş, Fisher’in Z İstatistiğine dönüştürülmüştür. Elde edilen sonuçlar Tablo 7’de özetlenmiştir.

Tablo 7 Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin KR-20 İle Hesaplanan Güvenirliklerine Ait Z İstatistiği Tablosu

TESTLER	N	K	KR-20 Güv. Kat.	Fisher Z'si	Z
Seçme Gerektiren Test	24	60	.79	1.07	1.16*
Kısa Cevaplı Test	24	60	.86	1.29	
Doğru-Yanlış Test	24	60	.84	1.22	

*p>.05

Tablo 7'deki bilgilere göre seçme gerektiren testin güvenirliliği (.79), kısa cevaplı testin güvenirliliği (.86) ve doğru-yanlış testinin güvenirliliği (.84) bulunmuştur. Ancak bu üç değer arasındaki fark Fisher'in Z istatistiğine dönüştürülerek karşılaştırılmış manidar bulunmamıştır. Bu durumda bu üç testin KR-20 güvenirlikleri arasında manidar bir farkın bulunmadığı söylenebilir. Bu bulgu, Lancaster (1987) tarafından yapılan çalışmanın bulgularıyla tutarlı değildir. Bu konuda daha ayrıntılı çalışmalara ihtiyaç vardır.

SONUÇ VE ÖNERİLER

Araştırmaya alınan öğrencilere seçme gerektiren, kısa cevaplı ve doğru-yanlış testlerin aynı davranışları yoklayan maddelerin güçlükleri karşılaştırıldığında kısa cevaplı maddelere ait ortalama güçlüğü, doğru-yanlış ve seçme gerektiren maddelere ait ortalama güçlükten daha küçük olduğu görülmüştür.

Bu testlerin madde ayırıcılık gücü indeksleri karşılaştırıldığında, doğru-yanlış testinin daha çok ayırıcı madde içerdiği görülmüştür. Bu durum daha kapsamlı araştırmalarla desteklenirse yeteneği ortaya koymasına bakımdan, doğru-yanlış testlerinin daha ayırıcı olduğu söylenebilir.

Araştırma sonucunda kısa cevaplı testler, doğru-yanlış ve seçme gerektiren testlere göre genel olarak öğrencilere daha zor gelmektedir. Fakat bu üç testin güvenirlikleri arasında manidar bir farkın olmadığı ve seçme gerektiren maddelerden oluşturulan testlerin diğer ikisine göre daha geçerli olduğu izlenimi edinilmiştir. Hazırlanan testler bu haliyle standartlaştırılması ve norm oluşturulması için yeterli olmaktan uzaktır. Yalnızca İlköğretimin yedinci ve sekizinci sınıf öğrencileri üzerinde sınırlı olarak "okuduğunu anlama" ve "kelime ilişkileri" yeteneğine ait maddelerin istatistiklerinin karşılaştırılmasında bir fikir vermektedir. Testlerin daha geniş bir örnekleme

uygulanmasında ve faktör analizi tekniđi ile sözü edilen yeteneklerin yapı geçerliđinin belirlenmesinde fayda vardır.

Seçme gerektiren, kısa cevaplı ve doğru-yanlıř türü maddeler kullanılarak daha farklı yeteneklerin ölçümüne gidilebilir. Ayrıca; yař, cinsiyet vb. gibi durumlar göz önüne alınarak farklı gruplar üzerinde çalıřmalara gidilebilir.

KAYNAKÇA

Baykul, Y. (1989). *Test Geliştirme Ders Notları*. Ankara: Hacettepe Üniversitesi, Çoğaltma.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook 1. Cognitive Domain*. New York: Mckay.

Lancaster, D. M. (1987). "A Comparison of Item Type and Source on Difficulty and Discrimination Ability", *Presented at the Annual Meeting of Midsouth Educational Research Association in Mobile Alabama*, November 13.

Özçelik, D. A. (1992-2). *Ölçme ve Değerlendirme*. Ankara: ÖSYM Yayınları.

Özgüven, İ. E. (1994). *Psikolojik Testler*. Ankara: Yeni Doğu Mat.

Tekin, H. (1977). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Mars Mat.

Tripp, A. M. (1982). "A Comparison of the Effect of Conventional and Complex Multiple-Choice Options on Item Difficulty and Discriminations", *Dissertation Abstracts International*, 43:3886-A.

Turgut, M. F. (1979). *676 Test Geliştirme Teknikleri Ders Notu: 1*. Ankara: Hacettepe Üniversitesi.

Turgut, M. F. (1988). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Saydam Matbaacılık, Altıncı Baskı.