

8. SINIF TÜRKÇE DERSİNE İLİŞKİN BAŞARI TESTİ GELİŞTİRME SUCCESS TEST FOR 8TH GRADE TURKISH LESSON DEVELOPMENT

Göknur KALYONCUOĞLU

MEB, kalyoncuoglugoknur@gmail.com, <https://orcid.org/0009-0000-8013-4989>

ÖZET

Öğrenme, öğretme süreçlerinin en önemli öğelerinden ikisi ölçme ve değerlendirmedir. Eğitim sürecinin değerlendirilmesinde sıklıkla başarı testleri kullanılır. Önemli olan ise başarıyı ölçmek için kullanılan bu testlerin niteliğidir. Nitelikli bir başarı testi ancak test geliştirme basamakları takip edilerek hazırlanabilir. Bu çalışmada eğitimde kullanılan başarı testlerinin standardını yükseltmek için madde istatistiği yapılmış sorulardan oluşan bir başarı testi geliştirilmiştir. Bu kapsamda MEB 2019 Türkçe Öğretim Programı'ndan 8. sınıf düzeyinde 17 kritik kazanım belirlenmiş, nitelikli madde hazırlama kurallarına dikkat edilerek çoktan seçmeli madde formuna sahip, her kazanımdan iki adet olmak üzere toplamda 34 adet madde hazırlanmış, bu maddelerden oluşan test kitapçıkları çoğaltılarak Ankara ili Mamak ilçesinde bir okulda 204 sekizinci sınıf öğrencisine uygulanmıştır. Uygulama sonrasında ITEMAN analiz programı kullanılarak test maddeleri analiz edilmiştir. Analiz sonuçlarına göre madde istatistiği çıktıları yorumlanmış, geçerlik ve ayırt edicilik kriterlerine uygun maddeler seçilerek her bir kazanıma ait istenilen özelliklerde 17 madde nihai teste alınmıştır. Bu çalışmada asıl hedef; okullarda uygulanan başarı testlerinin niteliğini artırmak, bu testler yoluyla öğretim süreci hakkında öğretmen ve öğrencilere geri bildirim vermektir. Bu sayede kazanım eksiklikleri belirlenen öğrenciler ve eğitimde büyük rol sahibi olan öğretmenler, süreç içerisindeki eksikliklerini görebilecek ve bu eksiklikleri tamamlama fırsatı bulabileceklerdir.

Anahtar Kelimeler: Türkçe Başarısı, Test Geliştirme, Madde İstatistiği, Kazanım.

178

ABSTRACT

Two of the most important elements of learning and teaching processes are measurement and evaluation. Achievement tests are often used in the evaluation of the educational process. What is important is the quality of these tests used to measure success. A qualified achievement test can only be prepared by following the test development steps. In this study, an achievement test consisting of questions with item statistics was developed to raise the standard of achievement tests used in education. In this context, 17 critical acquisitions at the 8th grade level were determined from the MoNE 2019 Turkish Language Teaching Programme, a total of 34 items with multiple-choice item form, two items from each acquisition, were prepared by paying attention to the rules of preparing qualified items, test booklets consisting of these items were reproduced and applied to 204 eighth grade students in a school in Mamak district of Ankara province. After the application, the test items were analysed using ITEMAN analysis programme. According to the results of the analyses, item statistical outputs were interpreted, and 17 items with the desired characteristics for each outcome were included in the final test by selecting the items suitable for validity and discrimination criteria. The main aim of this study is to increase the quality of achievement tests applied in schools and to give feedback to teachers and students about the teaching process through these tests. In this way, students whose achievement deficiencies are determined and teachers, who have a great role in education, can see their deficiencies in the process.

Keywords: Turkish Success, Test Development, Item Statistics.

1. GİRİŞ

Eğitim, “bireyin davranışlarında kendi yaşantıları yoluyla kasıtlı ve istendik davranış değişikliği meydana getirme süreci” olarak tanımlanır (Demirel, 1999). Eğitim insanlık var olduğundan bu yana hayatımızın içindedir ve zaman içinde kendini şekillendirmiştir. Eğitimin şekillenmesi ve çeşitlenmesi ölçmenin zorunluğunu da beraberinde getirmiştir. Eğitimde ölçme değerlendirme, verdiğimiz eğitimin çıktılarını görmemiz açısından son derece önemlidir. Ölçme için “Geçerli görgül yollarla test edilebilecek kurallar çerçevesinde nesnelere belli özelliklere sahip oluş derecesine göre sayılar veya semboller vermektir.” der Magnusson (Magnusson 1967). Eğitim en önemli parçalarından biri ölçmedir. Ölçme, süreç hakkında hem öğrenene hem öğretene geri bildirim verir. Ölçmenin çeşitli yolları vardır. Doğrudan, dolaylı veya türetilmiş ölçme türlerinden biri kullanılabilir. Eğitimde genellikle dolaylı ölçme kullanılır. Dolaylı ölçmelerde ise ölçek geliştirme en sık karşımıza çıkan yöntemdir. Ölçekler herhangi bir duyuşsal özelliği ölçmek için geliştirilebileceği gibi bilişsel özellikler için de kullanılabilir. Başarı testleri bilişsel özellikleri ölçmek için idealdir. Bir başarı testi hazırlanırken test geliştirme basamakları takip edilmelidir. Bunun için ilk olarak amaç belirlenmelidir. Ardından öğretim programına uygun kazanımlar seçilmeli ve bu kazanımları ölçmesi planlanan maddeler yazılmalıdır. Yazılacak maddelerin türleri testin kullanılacağı dersin içeriğine ya da ölçülmek istenen beceriye göre çeşitlenebilir. Hazırlanan maddeler gerekli düzeltmelerden geçtikten sonra benzer özellikleri taşıyan bir gruba uygulanmalıdır. Sonrasında uygun analiz programlarında incelenen maddelerden istenilen özelliklere sahip olanlar seçilmeli ve nihai teste alınmalıdır.

Bu çalışmada bahsi geçen test geliştirme basamaklarına riayet edilerek hazırlanan 34 adet çoktan seçmeli madde kitapçık haline getirilmiş ve belirlenen örnekleme ön uygulama yapılmıştır. Ardından ITEMAN analiz programında test maddeleri analiz edilmiş ve istenilen özelliklere sahip 17 madde nihai teste alınmıştır. Bir başarı testi nasıl hazırlanır, test hazırlanırken hangi basamakları takip etmek gerekir bu çalışmada bulmak mümkün olacaktır. Çalışma bu açıdan test geliştirmek isteyen eğitimcilere örnek niteliği de taşımaktadır.

Eğitim sürecinin en temel iki ögesi; ölçme ve değerlendirmedir. Eğitimde herhangi bir gözlem sonucunu sayı veya sıfatlarla ifade etme ölçme; ölçme sonuçlarını belli bir ölçüte vurarak ölçülen nitelik hakkında bir yargıya ulaşma çabası ise değerlendirme (Turgut ve Baykul, 2010) şeklinde tanımlanır. Ölçme ve değerlendirme tanımlardan da anlaşılacağı gibi birbiri ile ilişkili olan iki kavramdır; birbirine bağlıdır, birbirini destekler ve birbirinden etkilenir. Başarı testi bağlamında ölçme ve değerlendirme düşünüldüğünde öğretim programından belirlenen kazanımların belli bir eğitim süreci sonunda öğrenciye ulaşip ulaşmadığını görmek ancak doğru bir ölçmeyle mümkündür. Yapılan başarı testlerinde kazanılması beklenen becerilerin gerçekleşip gerçekleşmeme durumuna göre yapılan değerlendirmeler not verme şeklinde öğrenci başarısını direkt etkilediğinden ölçme işleminin doğru yapılması doğru bir değerlendirme için daha da fazla önem arz etmektedir.

Bu açıdan ölçme araçlarının nitelikli olması gerekir. Bu ölçme araçları rastgele sorulardan oluşmamalıdır. Burada da test geliştirme devreye girer. Geliştirilmiş bir testin güvenilirlik, geçerlik ve kullanılabilirlik açısından ideal olması beklenir. Bunlardan özellikle geçerlik ve güvenilirlik oldukça önemlidir. Bu iki özellik birbirine bağlıdır: Bir ölçme aracının geçerli olabilmesi her şeyden önce onun güvenilir ölçme yapabilmesine bağlıdır; bu bakımdan geçerli olan bir ölçme aracının güvenilir olması zorunludur; ancak güvenilir olan bir ölçme aracı aynı zamanda geçerli olmayabilir (Erkuş, 2017). Kullanılabilirlik da hem geçerliği hem güvenilirliği etkileyen bir faktördür. Bir testte süre ile soru sayısı arasında yeterli bir uyum yoksa öğrenciye soruyu cevaplamak için yeterli zaman verilmemişse bu test için; kullanılabilir olmadığı gibi güvenilirliği de düşük bir testtir denilebilir. Ya da kapsam geçerliğine dikkat

edilmeden hazırlanmış bir başarı testinde öğrenci derste işlediği konuların kritik kazanımlarına ait sorularla karşılaşmadığında süreçle ilgili değerlendirme eksik kalacaktır. Bu durum öğrencileri olumsuz etkileyecek öğrencilerin teste olan güvenlerini sarsacaktır. Dolayısıyla doğru hazırlanmış bir test ile bu sorunlar ortadan kalkacaktır.

Bunlardan yola çıkıldığında eğitim kurumlarında yapılan sınavlarda kullanılan soruların çoğu zaman test geliştirme yöntemlerinden bağımsız hazırlanması, hazırlanan bu testlerde geçerlik, güvenilirlik ya da kullanılabilirlik gibi ölçmenin temel prensiplerine dikkat edilmemesi hazırlanan soruların; uygun olmayan koşullarda öğrencilere rastgele uygulanması ve öğrencilere başarı puanı şeklinde geri dönen bu sınavların öğrencide ölçmek istenen becerileri ölçmemesi; sonuç olarak öğrenci ve öğretmene eğitim süreci ile ilgili gerekli dönütleri sağlayamaması okullarda karşılaştığımız belli başlı ölçme problemleridir. Bu sorunlar doğrultusunda 8. sınıf öğrencilerinin MEB 2019 Türkçe Dersi Öğretim Programı'ndan belirtilen kazanımları öğrenme düzeylerinin tespit edilmesi için test geliştirme süreçleri izlenerek standart bir test oluşturmaya ihtiyaç duyulmuştur. Bu sayede öğrenme eksiklikleri belirlenebilecek ve hem öğrencilere hem de öğretmenlere öğrenme süreci ile ilgili dönüt verilmesi mümkün olacaktır.

1.1. Araştırmanın Amacı

Eğitimde ve psikolojide insan niteliklerinin ölçülmesinde kullanılan ölçme araçları belli amaçlar doğrultusunda uygulanmaktadır. Derslerde kullanılan testlerin başlıca iki amacı vardır. Bunlardan biri öğrencilerin dersin ünitelerindeki öğrenme eksiklerini ve bu eksiklere yol açmış olabilecek güçlükleri belirlemek suretiyle bunların başka bir üniteye geçmeden önce tamamlanmasına imkân hazırlamaktır. Bunlardan ikincisi de öğrencilerin, dersin tümünün veya birkaç ünitelik bölümlerinin sonunda, hedefler doğrultusundaki öğrenme düzeylerini belirlemektir (Özçelik, 2010). Bu çalışmada da amaç, Türkçe dersinde 8. sınıf öğrencilerin öğretim programına göre belirlenen ünitelerdeki öğrenme durumlarını değerlendirmek, kazanım eksiklerini tespit etmek ve test geliştirme süreçleri izlenerek standart bir test hazırlamaktır. Bu sayede eğitimcilere test geliştirme süreçleri ile ilgili yol gösterici olmak ve yapılan çalışma ile benzer çalışmalara örnek teşkil etmektir.

2. LİTERATÜR

Eğitim, bilgiyi aktarma ve bilgiye ulaşma süreci olarak tanımlanırken gelişimsel döngüsünü ölçme ve değerlendirme ile tamamlar. Eğitimde doğru ölçme, eğitim sürecinin revizyonu için bize veriler sağlar. Ölçme sürecinde ise başarı testlerinin yeri büyüktür. Başarı düzeylerinin tespit edilmesinde her bir öğretmenin kendi ölçütlerine göre yapacağı değerlendirmeler farklı olacağından standartlaştırılmış testler, daha gerçekçi sonuçlar vermesi açısından önemlidir. Standart başarı testleri test geliştirme yöntemleri uygulanarak hazırlandığında ve doğru bir şekilde alanında uzman test uygulayıcıları tarafından ortam ve öğrenci özellikleri dikkate alınarak uygulandığında güvenilirliği yüksek bir ölçme gerçekleşecektir.

Eğitimde öğrenmenin ne derece gerçekleştiğini ölçen farklı türde birçok başarı testi geliştirilmiştir. Yurt dışı kaynaklı Wechsler Bireysel Başarı Testi ve Kaufman Eğitimsel Başarı Testi gibi testler öğrencinin ilerlemesini takip etme, öğrenci hatalarını analiz etme ve bu hatalara göre bireysel eğitim planı oluşturma hatta öğretim yeniden planlama, süreci değerlendirme, bilimsel araştırma yapma ya da öğrenci yerleştirme gibi amaçlarla kullanılabilir (Lichtenberger ve Smith, 2005).

Wechsler Sözel Olmayan Yetenek Testi(WNV) sözel ifadelerle gerek duyulmadan ya da sözel ifadeleri en aza indirerek ölçmeye tasarlanmıştır. Bu test 4-21 yaş aralığına uygulanabilmektedir. Belli alt testleri mevcuttur. Kaufman Kısa Zekâ Testi ise daha geniş bir

yaş aralığına uygundur ancak ülkemizde 4-45 yaş aralığı için norm çalışmaları devam etmektedir. Bu testler üst düzey becerilerin ölçülmesi için hazırlanmış geçerliği, güvenilirliği yüksek testlerdir. Bunların dışında Woodcock-Johnson III başarı testleri ise öğrenme yetersizliği ve özel öğrenme ihtiyacı olan öğrencilere yönelik geliştirilmiş sözel ifade ve temel matematik becerilerini ölçen bir başarı testidir.

Ülkemizde ise geliştirilen standart başarı testlerinden genel yetenek veya genel zekâyı ölçmeye yönelik iki çalışma alanda dikkat çekmektedir. Çalışmalardan ilkinde sekizinci sınıf öğrencilerine eğitsel rehberlik çalışmalarında faydalı olması planlanan Akademik Yetenek Testidir (Sanlı Richards, 2004). Ölçme aracının geliştirilmesinde Crocker ve Algina (1986)'da önerilen yaklaşıma göre önce uygulanacak bir yetenek testiyle ölçülecek özelliklerin belirlenmesi, ardından buna uygun test maddeleri geliştirildiği belirtilmiştir. Bahsi geçen testin Ankara ili için norm çalışması tamamlanmıştır. Bu iki çalışma dışında diğer bir çalışma da Sak (2015), Anadolu-Sak Zekâ Ölçeği olarak adlandırılan ilk yerli zekâ ölçeğidir. Geliştirilen bu ölçeğin ülke normlarının belirlenmesi çalışması devam etmektedir. Sak zekâ ölçeği genel zekâyı ve genel zekâyı oluşturan ana bileşenleri tarafsız olarak ölçen bireysel bir zekâ ölçeği şeklindedir ve yedi alt testi mevcuttur (Sak, 2015). Testin 4-12 yaş aralığındaki çocuklara uygulama süresi 35-45 dakika olabileceği belirtilmiştir. Testin güvenilirlik ve geçerlikle ilgili pilot uygulamaları tamamlanmıştır.

Öğretim programları kapsamında okullarda öğretmenler tarafından ders içeriklerine ait öğrenmelerin gerçekleşme durumunu ölçmek için uygulanabilecek başarı testleri alanında birçok çalışma yapılmıştır. Yurt dışında lise öğrencileri üzerine başarı testi geliştirme çalışması yapan Gucken (1986) ders kitaplarını incelemiş ve kritik olduğunu düşündüğü 30 kazanım belirlemiştir. Belirlenen 30 kazanımdan toplamda 242 çoktan seçmeli madde hazırlanmış ve uygunluk konusunda başka eğitimcilerden uzman görüşü almıştır. Sonuç olarak test uygulandıktan sonra maddeler seçilmiş ve güvenilirlik katsayıları hesaplanmıştır. İstatistik sonuçlarına göre istene özellikleri taşıyan maddeler nihai teste alınmıştır.

Yurt içinde yapılan öğretim programlarına göre hazırlanmış başarı testleri de incelenmiş bu başarı testlerinde de benzer tekniklerin kullanıldığı görülmüştür. 7. sınıf düzeyinde hazırlanan bir başarı testi çalışmasında Millî Eğitim Bakanlığı Öğretim Programı'nda yer alan kazanımlara uygun hazırlanan çoktan seçmeli sorulardan oluşan pilot test 202 kişilik bir örnekleme uygulanmış ve uygulama sonucunda madde istatistikleri hesaplanarak nihai test oluşturulmuştur (Sarıçetin, 2021).

Yurt içi ve yurt dışında yapılan çalışmalar değerlendirildiğinde bir başarı testi geliştirmek için öncelikle iyi bir planlamanın yapılması gerektiğini görülmektedir. Planlamada öncelikle testi hazırlama amacı ve araştırma modeli belirlenmeli ardından evren ve örneklem grubu seçilmelidir. Başarı testleri için hazırlanacak maddelerinin hazırlanmasında öğretim programından yararlanılmalı ve ders içeriklerine ait kazanımlar belirtke tablosunda gösterilmelidir. Ön uygulama yapıldıktan sonra belli analiz programları ile test maddeleri analiz edilmeli ve istenen kriterlere uygun geçerli ve güvenilir maddeler nihai teste alınarak başarı testi standartlaştırılmalıdır.

3. YÖNTEM

3.1. Araştırmanın Modeli

Çalışma nicel araştırma yoluyla gerçekleştirilmiştir. Nicel araştırma modellerinden tarama niteliği taşımaktadır. Tarama araştırmalarının amacı genellikle araştırma konusu ile ilgili var olan durumun fotoğrafını çekerek bir betimleme yapmaktır. Çok sayıda elemandan oluşan bir evrende, evren hakkında genel bir yargıya varmak amacı ile evrenin tümü ya da ondan alınacak bir grup, örnek ya da örneklem üzerinde yapılan tarama düzenlemeleridir.

3.2. Evren ve Örneklem

Evren Ankara ilinde Millî Eğitim Bakanlığına bağlı ortaokullarda eğitim gören tüm 8. sınıf öğrencileriyken Ankara ili Mamak ilçesindeki bir ortaokulda 8. sınıfta eğitim gören 204 öğrenci örnekleme oluşturmaktadır.

3.3. Veri Toplama Araçları

Veri toplama aracı olarak başarı testi kullanılmıştır. Hazırlanan 34 adet çoktan seçmeli madde, kitapçıklar hâlinde öğrencilere sınıf ortamında uygulanmıştır.

3.4. Verilerin Toplanması

Türkçe dersi 8. sınıf düzeyinde belirlenen kazanımlara göre hazırlanan 34 adet çoktan seçmeli madde kitapçık hâlinde çoğaltılarak bir ortaokulda toplam 204 öğrenciye yüz yüze kâğıt kalem testi şeklinde uygulanmış ve veriler tek seferde toplanmıştır.

3.5. Verilerin Analizi

Verilerin analizinde madde istatistikleri ve seçenek analizine uygun bir program olan ITEMAN analiz programı kullanılmıştır. Program çıktılarına göre elde edilen madde istatistikleri yorumlanmış kriterlere uygun olmayan maddeler testten çıkarılmış ayırt ediciliği yüksek, zorluk düzeyi istenen seviyede olan maddeler nihaî teste alınmıştır.

4. BULGULAR VE YORUM

4.1. Test İstatistikleri

Merkeze yığılma ve dağılma ölçüleri aşağıda özetlenmektedir:

Merkeze Yığılma Ölçüleri:

Verinin düzenlenmesiyle ilgili yapılanlar veri kümesinin tamamı hakkında bilgi vermezler. Bu bakımdan veri kümesinin tamamını ifade edecek, onu temsil edecek ve özetleyecek bilgiye ihtiyaç vardır. Bu ihtiyacın bir kısmını merkeze yığılma ölçüleri karşılar. Merkeze yığılma ölçüleri, bir veri kümesinin dağılımında, kümenin etrafında toplandığı veya yığılma eğilimi gösterdiği kısaca kümeyi özetleyen değerlerdir (Turgut & Baykul, 2010). Merkeze yığılma ölçüleri, veri kümesinin bütünü temsil ederler. Testin analizinde, merkeze yığılma ölçülerinden aritmetik ortalama incelenmiştir.

Aritmetik Ortalama:

Aritmetik ortalama merkeze yığılma ölçülerinden en çok kullanılanıdır (Turgut & Baykul, 2010). Aritmetik ortalama gruptaki ölçülerin toplamının ölçü sayısına yani gruptaki kişi sayısına bölümüdür (Özçelik, 2010). Aritmetik ortalama puanların ağırlık merkezidir. Aritmetik ortalamanın özellikleri:

- 1) Aritmetik ortalama bir kümedeki puanların ağırlık merkezidir.
- 2) Bir kümedeki puanların kendi ortalamalarından farklarının toplamı sifıra eşittir.
- 3) Aritmetik ortalama özellikle veri sayısının az olduğu hallerde, uçlardaki kopmalardan etkilenir.
- 4) Bir kümedeki puanlara sabit bir "a" sayısı eklenirse, ölçümlerin aritmetik ortalaması "a" sayısı kadar artar.

Turgut & Baykul (2010), tepe değer, ortanca ve ortalamayı şu şekilde karşılaştırmaktadır:

Aritmetik ortalama, tanımından dolayı, matematiksel işlemler için daha uygundur. Örneğin ortanca bir sıra sayısına dayalı olduğundan toplama işlemine uygun değildir. Bu nedenle ölçümlerle ileri analizler yapılacaksa aritmetik ortalamanın kullanılması daha uygun olur.

Aritmetik ortalamanın ölçümlerin ağırlık merkezi olması, onun ölçme sonuçları arasındaki-özellikle uçlardaki- kopmalardan fazlaca etkilenmesine yol açar. Ortanca sıralanmış ölçümlerde tam ortaya düştüğünden kopmalardan etkilenmez; dolayısıyla ölçümlerin yığıldıkları noktayı daha iyi temsil eder. Bu durumda kopmaların bulunduğu veri kümelerinde, daha ileri analizler yapılmıyacaksa merkeze yığılma ölçüsü olarak ortancanın kullanılması uygun olur.

Sadece frekansa dayalı, merkeze yığılma ölçüleri arasında ve en az bilgi taşıyanı tepe değerdir. Bir veri kümesinde bazen tepe değer bulunmayabilir veya birden fazla tepe değer bulunabilir. Bu nedenlerle tepe değer kullanılması alanı çok sınırlıdır. Merkeze yığılma ölçüleri kullanışlılıklarına göre sıraya konulacak olursa ilk olarak aritmetik ortalama, daha sonra ortanca ve en sonda tepe değer gelecektir.

Merkezi Dağılım Ölçüleri

Merkeze yığılma ölçüleri bir kümedeki puanların yığıldıkları değeri belirtir, bu durum puanların kümesinin bir özelliğidir, fakat kümeyi tam olarak betimleyemez (Turgut & Baykul, 2010). Puanların birbirinden ne kadar ayrıldıkları da önemlidir. Bu bölümde merkezi dağılım ölçülerinden varyans ve standart sapma incelenmiştir.

Varyans ve Standart Sapma:

Çok kullanılan bir dağılım ölçüsü varyanstır. Bunun karekökü olan standart sapmadır. Bir veri kümesinde ölçme sonuçlarının aritmetik ortalamadan farklarının karelerinin aritmetik ortalamasına varyans denir. Varyansın karekökü ise standart sapmayı verecektir. Standart sapma, dağılımın ölçüsüdür. Standart sapma veya varyans büyüdükçe dağılım yayvanlaşır, küçüldükçe sivrileşir (Turgut & Baykul, 2010).

Standart sapma, bir ölçüler kümesindeki değerlerin aritmetik ortalamadan ne derecede uzaklara yayılmakta olduklarını puan biriminde gösterebilen bir özelliktir.

Normal Dağılım:

Normal olasılık dağılımı bir sürekli dağılımdır.

Standart normal dağılım özelliklerini Turgut ve Baykul (2010) şu şekilde açıklamaktadırlar:

- 1.Aritmetik ortalaması $\bar{X}=0$ ve Standart sapması $S=1$ 'dir.
- 2.Standart normal dağılımın eğrisi $\bar{X}= 0$ noktasında yatay eksene göre simetrik.
- 3.Standart normal dağılımın eğrisi ile yatay eksen arasında kalan alan "1"e eşittir.
- 4.Standart normal dağılımın yatay eksenindeki değerleri (Standart puanlar) (z) $(-\infty, +\infty)$ aralığında değerler alır. Ancak, z'nin $(-3,+3)$ aralığındaki değerler için eğrinin altında kalan alan, toplam alanın 0,99'undan fazladır.

Çarpıklık (Kayışlılık) ve Basıklık (Sivrilik) Katsayısı :

Bir dağılımda ortalama ve ortanca ayrı ayrı noktalar üzerinde ise dağılım çarpıktır. Çarpıklık katsayısının sıfırdan küçük olması çarpıklığın sola (negatif), büyük olması ise sağa (pozitif) doğru olduğunu, sıfıra eşitse dağılımın ortalamaya göre simetrik olduğunu gösterir (Arıcı, 1993). Arıcı (1993), çarpıklığın ve basıklığın aşağıdaki gibi hesaplandığını ifade etmektedir:

Çarpıklık katsayısı iki formülle hesaplanabilir:

$$\text{Çarpıklık katsayısı: } \frac{3(X - \text{Ortanca})}{S}$$

$$\text{Çarpıklık katsayısı: } \frac{\sum (X - \bar{X})^3 / n}{S^3}$$

Basıklık katsayısı, dağılımın genişliği yorumlanmak istendiğinde kullanılır. Basıklık katsayısının sıfırdan küçük olması dağılımın basık, büyük olması sivri, sıfıra eşit olması ise dağılımın standart normal dağılıma uygun olduğunu gösterir. Basıklık katsayısı aşağıda verilen formül ile hesaplanabilir (Büyüköztürk & Çokluk & Köklü, 2010):

$$\text{Basıklık Katsayısı: } \frac{\sum (X - \bar{X})^4 / n}{S^4} - 3$$

Çarpıklık ve basıklık katsayısının sıfır olduğu bir dağılımın standart normal dağılıma uygun olduğu söylenebilir. Ancak her iki katsayının da sıfır olduğu durumlara uygulamada pek rastlanmaz. Sıfıra yakın değerlere ulaşıldığında dağılımın normal dağılıma uygun olduğu kabul edilir ve buna göre yorumlama yapıp uygun istatistikler seçilebilir (Büyüköztürk & Çokluk & Köklü, 2010).

Ölçmenin Standart Hatası:

Güvenirliği saptamanın bir diğer yolu da, ölçmenin standart hatasını belirlemektir. Ölçmenin standart hatası, tek uygulamaya dayalı testlerin güvenilirliğini belirlemede kullanılabilir. Hatanın ölçüsü, ölçmenin standart hatasıdır (Erkuş, 2003). Grubun heterojenliğinden güvenilirlik katsayısı etkilenir fakat ölçme hatası etkilenmez. Bu yüzden ölçmenin standart hatası bireysel puanları yorumlamada daha uygundur (Anastasi, 1988). Ancak testten teste, birim farklılıklarından dolayı doğrudan karşılaştırmalarda güvenilirlik katsayısı daha kullanışlı bir ölçüdür.

Ölçmenin standart hatası; bir bireyin sonsuz sayıda tekrarlı ölçümlerinde gözlenen ve gerçek puanı arasındaki uygunsuzluğun standart sapması olarak görülebilir. Ölçmenin standart hatası, gerçek puan yerine gözlenen puan kullanıldığında yapılan ölçme hatasıdır.

Ölçmenin standart hatasını hesaplarken aşağıdaki formül kullanılır:

$$\text{ÖSH} = s \sqrt{1 - r}$$

Formülde:

S = Testin standart sapması

r = Teste ait güvenilirlik katsayısını göstermektedir.

Formül kullanılarak teste ait ölçmenin standart hatası bulunmuştur. Formül aşağıdaki gibi uygulandığında:

$$\text{ÖSH} = 20.33 \sqrt{1 - (.91)} = 6.99 \text{ bulunmuştur.}$$

Bu testten puanı 76 olan öğrencinin gerçek puanı ± 1 Sh ile (% 68'lik bir alanda) 69 ile 83 puan arasındadır.

Görüldüğü gibi, bireysel puanlar, ham puanlara dayanarak çok kolay yorumlanabilmektedir. Cronbach (1971)'e göre ölçmenin standart hatası yardımıyla bireyin farklı alanlardaki görelî durumları bulunabilir.

Ölçme ve değerlendirmede herhangi bir özelliğe ilişkin bir ölçme sonucunda elde edilen ham puanların kullanılabilirliği oldukça sınırlıdır (Kan, 2006). Bu sorunu giderebilmek için farklı dağılımlardan elde edilen ölçme sonuçları ortalaması ve standart sapması aynı olan bir dağılımda ifade edilen puanlara dönüştürülür. Böyle bir dönüştürme; ölçme sonuçlarının merkez, eğilim ve değişim ölçülerinin daha kolay anlaşılmasına ve yorumlanmasına ayrıca farklı dağılımlara sahip ya da aynı dağılımdaki ölçme sonuçlarının birbirleriyle karşılaştırılmasına olanak sağlar.

Bunlardan en çok kullanılanı ve bilineni birim normal dağılım içerisinde ifade edilebilen Z ve T puanlarıdır (Kan, 2006).

Z ve T puanları aşağıdaki gibi bulunabilmektedir:

$$Z \text{ puanı} = (\text{Ham Puan} - \text{Ortalama Puan}) / \text{Standart Sapma}$$

T puanı; en basit haliyle açıklayacak olunursa ortalaması 50 standart sapması olan dağılıma ait puanlardır (Baykul, 2000).

T puanı hesaplaması şu şekilde yapılabilmektedir:

$$T \text{ puanı} = (10 \cdot Z) + 50$$

Yukarda ÖSH'yi hesaplamak için 76 puanı örnek alınmıştır. Yine ham puanı 76 olan öğrencinin Z ve T puanlarını hesaplamak için:

$$Z \text{ puanı} = (76-48) / \text{Standart Sapma}$$

$$Z \text{ puanı} = 1,4$$

$$T \text{ puanı} = (10 \cdot 1,4) + 50$$

$$T \text{ puanı} = 14 + 50 = 64 \text{ bulunmuştur.}$$

T puanları yukarıda da görüldüğü gibi ondalık ve negatif değerlerden arınıktır. Yani ondalık ve negatif değerler almaz. Bu sebeple anlaşılması, yorumlanması ve üzerinde işlem yapılması daha kolaydır.

Son olarak test istatistikleriyle ilgili olarak testin ortalama güçlüğü bulunmak istenebilir. Testin ortalama güçlüğü, test puanlarının ortalamasının madde sayısına bölünmesiyle elde edilebilir.

$$\text{Testin Ortalama Güçlüğü} (\bar{p}) = (\bar{X} / K)$$

$$\text{Testin Ortalama Güçlüğü} (\bar{p}) = 0,6$$

Testin ortalama güçlüğüne bakıldığında test maddelerinin genel olarak kolaya yakın olduğu söylenebilir. Yani test maddeleri, grubun % 60'ı tarafından doğru olarak cevaplanmıştır denebilir.

4.2. Madde İstatistikleri

Madde analizi, madde istatistiklerinin hesaplanması, doğrudan teste konulabilecek maddelerin belirlenmesi ve bu maddeler üzerinde yapılabilecek düzeltme çalışmalarının ne doğrultuda olacağını saptanması, teste konulması mümkün olmayan maddelerin ayıklanması amaçlarıyla yapılır.

Madde istatistiklerinde bakılması gereken temel veriler; madde ayırıcılık gücü, madde güvenilirliği ile madde güçlük katsayısı, madde standart sapması, madde basıklık ve çarpıklık katsayılarıdır. Bu bilgiler doğrultusunda maddelerin, madde ayırıcılık gücü, madde güçlüğü, madde basıklık katsayısı, madde çarpıklık katsayısı, madde güvenilirliği hesaplanmıştır. Ayrıca grup % 27'lik alt ve üst gruplara ayrılarak, her bir maddenin alt ve üst gruplarda cevaplanma

yüzdeleri ve frekanslarına bakılmıştır. Ayrıca, çift serili korelasyon katsayısı da hesaplanmıştır. Buna başvurulmasının nedeni sürekli değişken olan madde puanlarının, yapay olarak iki kategorili (1), (0) hâle getirilmesidir.

Madde Güçlük İndeksi:

Bir maddeye doğru cevap verenler sayısının testi alanların sayısına oranı bir maddenin güçlüğüdür (Turgut & Baykul, 2010). Madde güçlüğü “P” sembolüyle gösterilir ve aşağıdaki gibi hesaplanır:

$$P = \frac{n(D)}{N}$$

P: Madde güçlüğü

n(D): Maddeyi doğru cevaplayanlar sayısı

N: Testi alan öğrenci sayısı olarak ifade edilmektedir.

Madde güçlük katsayısı “0” ile “1” arasında değerler alır; “1”e yaklaştıkça madde kolaylaşır, “0”a yaklaştıkça madde zorlaşır.

Madde güçlük indeks aralıkları ile ilgili farklı görüşler olsa da genel olarak;

- 0.29 ve altı zor madde
- 0.30-0.49 orta güçlükte madde
- 0.50-0.69 kolay madde
- 0.69 ve üzeri çok kolay madde, şeklinde belirlenmiştir.

Madde güçlük indeksi orta güçlükte olan maddeler ideal maddelerdir. Testin özelliğine göre zor ya da kolay maddeler de belirli sayıda teste alınabilir. Madde güçlük indeksi hem bir ortalama hem bir yüzde hem de bir olasılık belirtir. Örneğin, bir maddenin güçlük indeksi 0,46 ise bu katsayı, testi alan grupta bu maddeye ait puanların ortalamasıdır. Bu anlamıyla 0,46 bu madde yönünden sınıfın ortalama başarısını belirtir ve sınıfın başarısının temsilcisidir. Diğer taraftan 0,46 sınıfta maddeyi doğru cevaplayanların tüm öğrencilere oranını gösterir. Bu anlamıyla da 0,46 sınıfın 0,46’sı tarafından doğru cevaplandırıldığını belirtir. Bu katsayı aynı zamanda, testi alan sınıftan rastgele seçilen bir öğrencinin bu maddeyi doğru cevaplama olasılığının 0,46 olduğunu ifade eder.

Üst ve alt gruptan hiç kimsenin doğru cevaplayamadığı bir maddenin güçlüğü 0; üst ve alt gruptan bütün öğrencilerin doğru cevaplandığı bir maddenin güçlük derecesi ise +1 olur (Tekin, 2000).

Madde Ayırt Edicilik İndeksi:

Madde ayırt edicilik indeksi, maddenin ilgili davranışa sahip olanla olmayana ne ölçüde ayırdığını gösterir. Madde ayırt ediciliğinin diğer bir karşılığı madde geçerliliğidir. Ayırma kavramı maddenin ölçülmek istenen davranışı ölçüp ölçmediğini ifade ettiği için bu katsayıya madde geçerliği de denir (Turgut & Baykul, 2010). Bir maddenin ayırt edicilik gücü r_j sembolüyle gösterilir ve aşağıdaki formül yardımıyla hesaplanır.

$$r_j = \frac{\bar{X}_{jD} - \bar{X}_X}{S_X} \cdot \frac{\sqrt{p}}{\sqrt{1-p}}$$

r_j : Madde ayırt edicilik indeksi

\bar{X}_{jD} : Maddeyi doğru cevaplayanların test puanlarının ortalaması

\bar{X}_X : Testin ortalaması

S_x : Testin standart sapması

p: Madde güçlüğü

Madde ayırt edicilik indeksi bir korelasyon katsayısıdır. Diğer korelasyon katsayıları gibi (-1) ile (+1) aralığında değerler alır. Bu katsayının (+) işaretli değerleri maddenin ölçtüğü niteliğin testin bütünüyle ölçülen nitelikle uyumlu olduğunu; (-) işaretli değerleri, maddenin ölçtüğü değişkenin testin bütünüyle ölçülen nitelikle ters yönde ilişki içinde bulunan bir nitelik ölçtüğünü; 0 (sıfır) ve civarındaki değerleri de maddenin ölçtüğü niteliğin testin bütünüyle ölçülen nitelik arasında ilişki bulunmadığını belirtir.

Madde ayırt edicilik indeksi hesaplanırken çift serili korelasyon ve alt ve üst % 27'lik gruplara göre "d" değeri hesaplanmıştır.

Çift Serili Korelasyon Katsayısı:

Çift serili korelasyon katsayısı sürekli iki değişkenden birinin aslında normal dağılıma sahipken yapay olarak iki kategorili hâle getirilmesi durumunda kullanılır. Bu korelasyona eğitimde ve psikolojideki test geliştirme çalışmalarında sık sık başvurulur. Test geliştirmede, madde adı verilen sorular puanlanırken her doğru cevaba 1 ve yanlış cevaplama, boş bırakma, birden çok cevap işaretleme hâlinde de 0 verilir. Madde puanları ile testin bütününden alınan puanlar arasındaki korelasyon ayırma gücü denilen maddenin önemli bir özelliğini belirtir (Turgut & Baykul, 2010).

Bu testte test puanları sürekli madde puanları 1 ve 0 olmak üzere iki kategorilidir. Dolayısıyla madde puanları aslında sürekli olduğu hâlde yapay olarak iki kategorili hâle getirilmiş kabul edilir.

Çift serili korelasyon katsayısı aşağıdaki şekilde hesaplanır:

$$r_{bis} = \frac{\mu_{iD} - \mu_X}{S^2(X)} \frac{p_j}{y_j}$$

μ_{iD} : O maddeyi doğru cevaplayanların başarı puanı ortalaması

μ_X : Tüm test puanlarının ortalaması

p_j : İlgili maddenin güçlük indeksi

$S^2(X)$: Tüm test puanlarının varyansı

y_j : Normal dağılım eğrisi altında kalan p_j karşılık gelen ordinat yüksekliğidir.

Test puanları dağılımı çift modlu ya da fazla çarpık ise bu katsayı -1'den küçük veya +1'den büyük değerler alabilir (Baykul, 2000). Bunun yanı sıra % 27'lik alt ve üst gruplar arasındaki korelasyon hesaplama yoluyla maddenin ayırt edicilik indeksi (d=discriminant) bulunabilir. Bunun formülü aşağıda verilmiştir:

$$d = \frac{D_{üst} - D_{alt}}{N}$$

d: Madde ayırt edicilik indeksi

$D_{üst}$: Üst grupta o maddeye doğru cevap veren öğrenci sayısı

D_{alt} : Alt grupta o maddeye doğru cevap veren öğrenci sayısı

N: Üst veya alt gruptaki öğrenci sayısı

Burada beklenen üst gruptaki bireylerin hepsinin ilgili maddeye doğru cevap vermesi ve alt gruptan hiçkimsenin ilgili maddeye doğru cevap vermemesidir. Bu durumda “d” değeri, 1’e eşit olacaktır. Test geliştirmede, maddelerin testin bütünüyle ölçülen nitelikte uyumlu olması arzu edilir. Bu nedenle ayırıcılık gücü katsayısı (-) işaretli olanlarla 0 ve 0 civarında olan maddeler teste konulmaz. Teste konulacak maddeler için aşağıdaki değerler genel ölçüt olarak kabul edilir (Ebel, 1965). Madde ayırıcılık gücü;

- 0,19 ve daha küçük olan maddeler test konulmaz.
- 0,20-0,29 arasında olan maddeler genel olarak düzeltilmeli ve geliştirilmelidir.
- 0,30-0,39 arasında olan maddeler iyi maddelerdir ancak geliştirilebilir.
- 0,40 ve üzeri maddeler oldukça iyi maddelerdir, teste direkt konulabilir.

Basıklık katsayısı normal dağılımlarda 0’a eşittir. Bu katsayı pozitif değerler aldığında maddelerin normalden daha sivri, negatif değerler aldığında ise maddelerin normalden daha basık dağılım gösterdiği şeklinde yorumlanır.

Çeldiricilerin Beklenen Frekansı:

Toplam cevaplayıcı sayısından o maddeyi doğru cevaplayan ve boş bırakanların sayısı çıkarılarak elde kalan sayı, kalan çeldirici sayısına bölünürse çeldiricilerin beklenen frekans değeri elde edilir. Çünkü bir cevaplayıcının sorulan maddeyi biliyorsa cevaplama, bilmiyorsa boş bırakması varsayılmıştır. Bunun sonucunda çeldiricilerin frekanslarının birbirine yakın olması beklenir. Frekansı, beklenen frekanstan büyük olan çeldiricilerin öğrencilere çok cazip geldiği; frekansı beklenen frekanstan küçük olan çeldiricilerin ise zayıf kaldıkları söylenir (Baykul, 2000).

Bu açıklamalar kapsamında, aşağıda verilen bölümde her bir madde için elde edilen madde istatistikleri yorumlanarak nihai testi oluşturacak maddeler belirlenmiştir. Pilot uygulamada ardışık şekilde birbirini izleyen iki madde aynı kazanımı ölçmektedir. Bu nedenle yorumlar yapılırken birbirini izleyen iki madde bir arada yorumlanmıştır. Madde istatistikleri daha iyi çalışan maddenin nihai testi oluşturmak üzere seçilmesine buna göre karar verilmiştir. Bu şekilde ölçülmesi amaçlanan kazanımlardan en az bir soru nihai testte yer alması hedeflenmektedir. Sonrasında test istatistiklerine ilişkin bulgular yorumlanmıştır. Yazılımın verdiği analiz çıktısı Ek-3’te verilmiştir.

Madde İstatistiklerine İlişkin Bulgular

Testte kullanılan 34 adet maddeye ilişkin bulgular tablo ve yorumlarıyla birlikte aşağıda verilmiştir. Yorumlar yapılırken her bir kazanıma ait iki madde peş peşe verilmiş ve bu iki maddeden uygun kriterlere sahip olan madde nihai teste alınmıştır. Nihai teste alınması uygun görülen maddelerle ilgili kararlar her kazanıma ait ikinci maddenin son cümlesinde italik verilmiştir.

Madde Analiz Tablolarında Kullanılan Kavramların Karşılığı:

Item Statistics: Madde istatistiği

Seg.No.: Madde numarası

Prop. Correct: Güçlük indeksi

Disc. Index: Ayırt edicilik indeksi

Prop. Total: Seçeneği işaretleyen bireylerin oranı

Endorsing Low: Alt grup

Endorsing High: Üst grup

Key: Doğru seçenek

Other: Boş bırakanlar

Tablo 1. 1.Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
1	1-1	.94	.14	.31	A	.03	.06	.00	-.21
					B	.94	.86	.99	.31 *
					C	.02	.04	.00	-.16
					D	.01	.03	.00	-.15
					Other	.00	.00	.00	-.05

Birinci maddenin madde güçlüğü .94, ayırt edicilik indeksi ise .14 çıkmıştır. Bu durumda madde çok kolay ve ayırt ediciliği düşüktür. Ayırt edicilik değeri olması gerektiği değerin altında çıkmıştır. Maddenin ideal bir madde olduğu söylenemez.

Tablo 2. 2. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
2	1-2	.87	.25	.36	A	.04	.10	.01	-.21
					B	.04	.08	.01	-.19
					C	.04	.08	.01	-.18
					D	.87	.73	.98	.36 *
					Other	.01	.00	.00	-.06

189

İkinci maddenin madde güçlüğü .87, ayırt edicilik indeksi ise .25 çıkmıştır. Madde çok kolay ve ayırt ediciliği düşüktür. Ayırt edicilik değeri 1. maddeye göre daha yüksek ve “madde revize edilip teste konulabilir” kategorisinde yer almaktadır. Seçenek dağılımlarına bakıldığında hem alt hem de üst grubun çeldiricilere dağılımlarının birbirine yakın oranlarda olduğu görülmektedir. Aynı kazanımı ölçen bu iki madde istatistikleri açısından değerlendirildiğinde *ikinci maddenin nihai teste alınması daha uygundur.*

Tablo 3. 3. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
3	1-3	.76	.44	.45	A	.09	.17	.02	-.23
					B	.06	.12	.00	-.23
					C	.76	.51	.95	.45 *
					D	.07	.13	.02	-.21
					Other	.03	.00	.00	-.13

Üçüncü maddenin madde güçlüğü .76, ayırt edicilik indeksi ise .44 çıkmıştır. Bu durumda madde kolay ve ayırt ediciliği yüksektir. Seçenek dağılımlarına bakıldığında hem alt hem de üst grubun çeldiricilere dağılımlarının birbirine yakın oranlarda dağıldığı görülmektedir. Üçüncü maddenin gerek ayırt edicilik gerekse çeldirici dağılımı açısından uygun şekilde çalıştığı görülmektedir.

Tablo 4. 4. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
4	1-4	.56	.58	.49	A	.13	.19	.04	-.19
					B	.15	.27	.02	-.29
					C	.56	.29	.88	.49 *
					D	.15	.21	.06	-.18
					Other	.02	.00	.00	-.09

Dördüncü maddenin madde gücü .56, ayırt edicilik indeksi ise .58 çıkmıştır. Madde, orta güçlük sınırını aşmış kolaya yakındır ve maddenin ayırt ediciliği yüksektir. Ayırt edicilik değeri üçüncü maddeye göre daha yüksektir. Seçenek dağılımlarına bakıldığında yine hem alt hem de üst grubun çeldiricilere dağılımlarının birbirine yakın oranlarda olduğu görülmektedir. Aynı kazanımı ölçen üçüncü ve dördüncü maddeler, madde istatistikleri açısından değerlendirildiğinde *dördüncü maddenin nihai teste alınması daha uygun görünmektedir.*

Tablo 5. 5. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
5	1-5	.69	.40	.38	A	.14	.22	.07	-.19
					B	.08	.15	.02	-.21
					C	.08	.14	.03	-.18
					D	.69	.48	.88	.38 *
					Other	.01	.00	.00	-.06

Beşinci maddenin madde gücü .69, ayırt edicilik indeksi ise .40 çıkmıştır. Bu durumda madde kolay ve ayırt ediciliği yüksektir. Seçenek dağılımlarına bakıldığında A çeldiricisinin B ve C çeldiricilere göre daha fazla çalıştığı ancak bu durumun ayırt ediciliğe olumsuz bir etkisi olacak düzeyde gerçekleşmediği görülmektedir.

Tablo 6. 6. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
6	1-6	.43	.34	.30	A	.20	.29	.12	-.17
					B	.43	.25	.59	.30 *
					C	.09	.18	.02	-.24
					D	.26	.25	.27	.01
					Other	.02	.00	.00	-.09

Altıncı maddenin madde gücünün .43, ayırt edicilik indeksinin ise .34 çıktığı görülmektedir. Madde orta güçlükte, ayırt ediciliği ise kabul edilebilir düzeydedir. Seçenek dağılımlarına bakıldığında D çeldiricisinin özellikte üst grupta iyi çalıştığı görülmektedir. Bu durum bir çeldiricide olması istenen bir durum değildir ve ayırt ediciliği de bir miktar olumsuz etkilediği yorumu yapılabilir. Yukarıdaki değerlendirmelerden sonra aynı kazanımı ölçen beş ve altıncı maddeler karşılaştırıldığında *altıncı maddenin nihai teste alınması daha uygundur.*

Tablo 7. 7. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key	
7	1-7	.58	.59	.50	A	.58	.30	.89	.50	*
					B	.13	.24	.02	-.27	
					C	.11	.19	.04	-.21	
					D	.14	.21	.04	-.20	
					Other	.04	.00	.00	-.11	

Yedinci maddenin madde güçlüğü .58, ayırt edicilik indeksinin ise .59 çıktığı görülmektedir. Madde kolaya yakın ve ayırt ediciliği de iyi düzeydedir. Çeldiricilerin de alt grupta seçeneklere giden öğrenci sayısına bakıldığında istenen düzeyde çalıştığı görülmektedir. Bu bakımdan madde ideal bir maddedir denilebilir.

Tablo 8. 8. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key	
8	1-8	.49	.64	.53	A	.49	.22	.87	.53	*
					B	.12	.20	.03	-.21	
					C	.09	.14	.02	-.18	
					D	.26	.39	.07	-.29	
					Other	.03	.00	.00	-.09	

Sekizinci maddenin madde güçlüğü .49, madde ayırt edicilik indeksinin ise .64 olduğu görülmektedir. Maddenin bu verilere göre oldukça ayırt edici olduğu söylenebilir. Madde güçlüğü bakımından ise orta güçlükte ideal bir maddedir. Yedinci ve sekizinci madde istatistikleri karşılaştırıldığında iki madde arasında ciddi bir fark olmamasına rağmen doğru seçeneğe alt gruptan gitme oranı fazla olduğu için *sekizinci maddenin nihai teste alınması daha uygundur.*

Tablo 9. 9. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key	
9	1-9	.37	.52	.45	A	.21	.33	.07	-.26	
					B	.37	.17	.68	.45	*
					C	.22	.30	.10	-.19	
					D	.15	.16	.11	-.07	
					Other	.05	.00	.00	-.03	

Dokuzuncu maddenin madde güçlük indeksi .39 yani orta güçlükte, ayırt edicilik indeksi ise .52 olarak bulunmuştur. Madde ayırt edicilik verisine göre maddenin bilenle bilmeyeni ayırt ettiği söylenebilir. Seçenekler incelendiğinde çeldiricilere alt gruptan gitme oranının dengeli olduğu dolayısıyla çeldiricilerin iyi çalıştığı görülmektedir.

Tablo 10. 10. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
10	1-10	.50	.50	.42	A	.23	.28	.13	-.16
					B	.14	.22	.05	-.20
					C	.50	.29	.78	.42 *
					D	.09	.16	.03	-.20
					Other	.04	.00	.00	-.09

Onuncu maddenin madde güçlük indeksi .50 yani orta güçlük sınırında kolayla yakın, ayırt edicilik indeksi ise .50 yani iyi ayırt edici bir madde olarak bulunmuştur. Seçenekler incelendiğinde çeldiricilere alt gruptan gitme oranının dengeli olduğu dolayısıyla çeldiricilerin iyi çalıştığı görülmektedir. Bu şartlarda dokuz ve onuncu maddelerin ikisi de istenilen standartlarda olmasına karşın üst grubun çeldiricilere gitme durumu değerlendirildiğinde *onuncu maddenin nihai teste alınması daha uygundur denilebilir.*

Tablo 11. 11. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
11	1-11	.75	.40	.41	A	.04	.08	.00	-.20
					B	.14	.21	.06	-.19
					C	.75	.52	.92	.41 *
					D	.07	.16	.01	-.25
					Other	.01	.00	.00	-.09

On birinci maddenin madde güçlüğü .75, ayırt edicilik indeksi ise .40 çıkmıştır. Bu durumda madde için çok kolay ve ayırt ediciliği yüksektir denilebilir. Seçenek dağılımlarına bakıldığında hem alt hem de üst grubun çeldiricilere dağılımlarının birbirine yakın oranlarda olduğu görülmektedir. Maddenin iyi bir madde olduğu söylenebilir.

Tablo 12. 12. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
12	1-12	.45	.43	.38	A	.45	.25	.68	.38 *
					B	.18	.25	.10	-.17
					C	.26	.36	.14	-.21
					D	.10	.12	.08	-.08
					Other	.01	.00	.00	-.07

On ikinci maddenin madde güçlük indeksi .45, madde ayırt edicilik indeksi ise .43 olarak bulunmuştur. Bu istatistiklere göre madde için orta güçlükte ve ayırt edici bir madde olduğu söylenebilir. Seçenekler incelendiğinde çeldiricilere alt gruptan gitme durumunun dengeli olduğu ancak C çeldiricisinin diğerlerine göre daha iyi çalıştığı söylenebilir. On birinci madde ile karşılaştırıldığında on ikinci maddenin orta güçlükte olması ve bu maddenin çeldiricileri daha iyi çalışması sebebiyle *on ikinci maddenin nihai teste alınması daha uygundur.*

Tablo 13. 13. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
13	1-13	.58	.39	.35	A	.22	.25	.17	-.10
					B	.58	.39	.77	.35 *
					C	.07	.14	.02	-.22
					D	.11	.19	.04	-.21
					Other	.02	.00	.00	-.07

On üçüncü maddenin madde güçlüğü .58, ayırt edicilik indeksi ise .39 çıkmıştır. Bu durumda madde için kolay ve ayırt edici bir madde diyebiliriz. Seçenek dağılımlarına bakıldığında A çeldiricisine diğer çeldiricilere göre üst grup daha fazla gitmiştir bu durum ayırt ediciliği bir miktar etkilemiştir diyebiliriz.

Tablo 14. 14. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
14	1-14	.26	.49	.47	A	.52	.51	.37	-.11
					B	.26	.09	.58	.47 *
					C	.14	.25	.03	-.27
					D	.06	.12	.01	-.19
					Other	.02	.00	.00	-.08

On dördüncü maddenin madde güçlüğü .26, ayırt edicilik indeksinin ise .49 çıktığı görülmektedir. Madde için zor ve ayırt ediciliği yüksek bir madde olduğu söylenebilir. Seçenekler incelendiğinde A çeldiricisinin üst grubu etkileyecek şekilde çalıştığı görülmektedir. Maddenin ayırt ediciliği iyi olmasına rağmen on üçüncü madde ile karşılaştırıldığında on üçüncü maddenin güçlük düzeyi daha ideal olduğu için *nihai teste on üçüncü maddenin alınması daha uygundur.*

Tablo 15. 15. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
15	1-15	.63	.39	.36	A	.12	.19	.05	-.19
					B	.12	.16	.06	-.14
					C	.63	.43	.82	.36 *
					D	.12	.20	.06	-.18
					Other	.01	.00	.00	-.09

On beşinci maddenin madde güçlük indeksi .63, ayırt edicilik indeksi ise .39 olarak bulunmuştur. Bu durumda maddenin kolay ve ayırt edici olduğu söylenebilir. Seçenekler incelendiğinde çeldiricilere, alt grubun gitme durumunun dengeli olduğu dolayısıyla çeldiricilerin iyi çalıştığı görülmektedir. Alt gruptakilerden doğru seçeneğe giden sayısının fazla olmasının ayırt ediciliği bir miktar etkilediği söylenebilir. Buna karşın madde ideal bir maddedir ve teste alınabilir.

Tablo 16. 16. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
16	1-16	.58	.38	.34	A	.28	.35	.19	-.15
					B	.06	.11	.02	-.17
					C	.07	.13	.02	-.20
					D	.58	.39	.77	.34 *
					Other	.01	.00	.00	-.07

On altıncı maddenin madde güçlüğü .58, ayırt edicilik indeksinin ise .38 çıktığı görülmektedir. Madde için kolay ve ayırt ediciliği yüksek bir madde denilebilir. Seçenekler incelendiğinde A çeldiricisine alt ve üst grupta da gidilme oranı fazladır. Üst grupta .19, alt grup değeri .39'dur. Bu durum seçeneğin iyi çalıştığını gösterir. Ancak istenilen durum, üst gruptan çeldiricilere giden sayılarının düşük olmasıdır. Bu yüzden aynı kazanımı ölçen iki madde düşünüldüğünde *on beşinci maddenin nihai teste alınması daha uygundur.*

Tablo 17. 17. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
17	1-17	.53	.60	.49	A	.12	.18	.03	-.18
					B	.11	.20	.02	-.24
					C	.53	.27	.87	.49 *
					D	.18	.25	.07	-.20
					Other	.06	.00	.00	-.14

On yedinci maddenin madde güçlük indeksi .53, ayırt edicilik indeksi ise .60 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni çok iyi ayırt ettiği söylenebilir. Seçenekler incelendiğinde çeldiricilere alt gruptan giden kişi sayılarının dengeli olduğu dolayısıyla çeldiricilerin iyi çalıştığı görülmektedir.

Tablo 18. 18. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser. Key
18	1-18	.39	.62	.53	A	.11	.16	.06	-.15
					B	.39	.13	.75	.53 *
					C	.45	.60	.18	-.33
					D	.04	.08	.01	-.17
					Other	.02	.00	.00	-.10

On sekizinci maddenin madde güçlük indeksi .39, madde ayırt edicilik indeksi ise .62 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni çok iyi ayırt ettiği söylenebilir. Seçenekler incelendiğinde çeldiricilere giden kişi sayılarının dengeli olduğu, C çeldiricisinde bu farkın bir miktar açıldığı özellikle alt gruptan bu çeldiriciye giden kişi sayısının fazla olduğu görülmektedir. C çeldiricisi için daha güçlü bir çeldiricidir denilebilir. Aynı kazanımı

ölçen iki madde karşılaştırıldığında *on yedinci maddenin nihai teste alınmasının daha uygun olduğu söylenebilir.*

Tablo 19. 19. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
19	1-19	.66	.39	.37	A	.66	.44	.83	.37	*
					B	.13	.21	.05	-.21	
					C	.07	.12	.03	-.17	
					D	.13	.19	.09	-.14	
					Other	.02	.00	.00	-.11	

On dokuzuncu maddenin madde güçlük indeksi .66, ayırt edicilik indeksi ise .39 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni çok iyi ayırt ettiği söylenebilir. Seçenekler incelendiğinde çeldiricilere alt grubun gitme durumunun dengeli olduğu dolayısıyla çeldiricilerin iyi çalıştığı görülmektedir.

Tablo 20. 20. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
20	1-20	.34	.43	.39	A	.12	.19	.05	-.18	
					B	.32	.32	.26	-.06	
					C	.16	.23	.06	-.19	
					D	.34	.18	.61	.39	*
					Other	.07	.00	.00	-.11	

Yirminci maddenin madde güçlük indeksi .34, madde ayırt edicilik indeksi ise .43 olarak bulunmuştur. Madde güçlük indeksine göre maddenin orta güçlükte olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni çok iyi ayırt ettiği söylenebilir. Seçenekler incelendiğinde çeldiricilere giden kişi sayılarının dengeli olduğu, B çeldiricisinde bu farkın bir miktar açıldığı görülmektedir. C çeldiricisi için diğerlerine nispeten daha güçlü bir çeldiricidir denilebilir. Aynı kazanımı ölçen iki madde karşılaştırıldığında *on dokuzuncu maddenin nihai teste alınmasının daha uygun olduğu söylenebilir.*

Tablo 21. 21. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
21	1-1	.78	.43	.45	A	.13	.26	.03	-.30	
					B	.78	.52	.95	.45	*
					C	.05	.11	.01	-.22	
					D	.04	.09	.00	-.20	
					Other	.01	.00	.00	-.04	

Yirmi birinci maddenin madde güçlük indeksi .78, madde ayırt edicilik indeksi ise .43 olarak bulunmuştur. Madde güçlük indeksine göre maddenin çok kolay olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni iyi ayırt ettiği söylenebilir. Seçenek

analizine bakıldığında D çeldiricisinin üst grupta hiç çalışmadığı, A çeldiricisinin alt grupta iyi çalıştığı söylenebilir.

Tablo 22. 22. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
22	1-2	.61	.68	.58	A	.17	.34	.03	-.35
					B	.08	.15	.02	-.19
					C	.12	.24	.02	-.27
					D	.61	.25	.92	.58 *
					Other	.01	.00	.00	-.08

Yirmi ikinci maddenin madde güçlük indeksi .61, ayırt edicilik indeksi ise .68 olarak bulunmuştur. Madde güçlük indeksine göre maddenin çok kolay olduğu ancak kolay olmasına rağmen madde ayırt edicilik verisine göre maddenin bilenle bilmeyeni çok iyi ayırt ettiği söylenebilir. Seçenekler incelendiğinde A çeldiricisine alt grupta iyi çalıştığı görülmektedir. Bu madde ile yirmi birinci madde karşılaştırıldığında *nihai teste yirmi ikinci maddenin alınmasının daha uygun olduğu söylenebilir.*

Tablo 23. 23. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
23	1-3	.72	.57	.53	A	.10	.22	.02	-.27
					B	.10	.23	.02	-.30
					C	.72	.38	.95	.53 *
					D	.07	.16	.02	-.22
					Other	.01	.00	.00	-.07

Yirmi üçüncü maddenin madde güçlük indeksi .72, ayırt edicilik indeksi ise .57 olarak bulunmuştur. Madde güçlük indeksine göre maddenin çok kolay olduğu, madde ayırt edicilik verisine göre maddenin ayırt ediciliğinin de iyi olduğu görülmektedir. Seçenekler incelendiğinde çeldiricilere alt ve üst grupta gidenler arasında dengeli bir dağılım olduğu söylenebilir.

Tablo 24. 24. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
24	1-4	.59	.68	.57	A	.59	.22	.90	.57 *
					B	.17	.32	.04	-.31
					C	.11	.19	.04	-.21
					D	.12	.25	.02	-.29
					Other	.01	.00	.00	-.06

Yirmi dördüncü maddenin madde güçlük indeksi .59, ayırt edicilik indeksi ise .68 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre maddenin ayırt ediciliğinin oldukça iyi olduğu görülmektedir. Seçenekler incelendiğinde çeldiricilere alt ve üst grupta gidenler arasında dengeli bir dağılım

olduğu söylenebilir. Madde yirmi üçüncü madde ile karşılaştırıldığında ayırt edicilik gücü daha fazla olduğu için *yirmi dördüncü maddenin nihai teste alınması daha uygundur.*

Tablo 25. 25. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
25	1-5	.58	.58	.50	A	.21	.36	.08	-.29
					B	.58	.27	.85	.50 *
					C	.14	.24	.05	-.24
					D	.06	.11	.02	-.17
					Other	.01	.00	.00	-.06

Yirmi beşinci maddenin madde güçlük indeksi .58, ayırt edicilik indeksi ise .58 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin oldukça iyi olduğu görülmektedir. Seçenekler incelendiğinde çeldiricilere alt ve üst grupta gidenler arasında dengeli bir dağılım olduğu söylenebilir. Yalnız alt grupta A çeldiricisinin diğerlerine göre daha iyi çalıştığı görülmektedir.

Tablo 26. 26. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
26	1-6	.27	.19	.23	A	.26	.25	.25	-.02
					B	.20	.24	.15	-.10
					C	.18	.24	.11	-.15
					D	.27	.19	.37	.23 *
					Other	.10	.00	.00	.01

Yirmi altıncı maddenin madde güçlük indeksi .27, ayırt edicilik indeksi ise .19 olarak bulunmuştur. Madde güçlük indeksine göre maddenin zor olduğu, madde ayırt edicilik indeksine göre maddenin ayırt ediciliğinin çok zayıf olduğu görülmektedir. Ayırt edicilik kriterlerine uygun olmayan bu maddenin teste tutulmaması gerekir. Dolayısıyla aynı kazanımı ölçen iki madde dikkate alındığında *yirmi beşinci maddenin nihai teste alınması uygundur.*

Tablo 27. 27. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
27	1-7	.31	.41	.42	A	.31	.13	.54	.42 *
					B	.46	.52	.34	-.18
					C	.06	.11	.03	-.16
					D	.16	.23	.09	-.18
					Other	.01	.00	.00	-.04

Yirmi yedinci maddenin madde güçlük indeksi .31, ayırt edicilik indeksi ise .41 olarak bulunmuştur. Madde güçlük indeksine göre maddenin orta güçlükte olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin iyi olduğu söylenebilir. Seçenekler

incelendiğinde B çeldiricisine alt ve üst grupta giden sayısının fazla olduğu söylenebilir. Bu seçeneğin biraz zayıflatılması önerilir.

Tablo 28. 28. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
28	1-8	.39	.41	.38	A	.39	.19	.60	.38	*
					B	.05	.10	.01	-.18	
					C	.04	.08	.01	-.16	
					D	.51	.61	.37	-.22	
					Other	.01	.00	.00	-.06	

Yirmi sekizinci maddenin madde güçlük indeksi .39, ayırt edicilik indeksi ise .41 olarak bulunmuştur. Madde güçlük indeksine göre maddenin orta güçlükte olduğu, madde ayırt edicilik indeksine göre ise maddenin ayırt ediciliğinin iyi düzeyde olduğu söylenebilir. Seçenekler incelendiğinde D çeldiricisine alt ve üst grupta giden sayısının fazla olduğu görülmektedir. Bu durum ayırt ediciliği etkilediğinden bu seçeneğin biraz zayıflatılması önerilir. Aynı kazanımı ölçen iki madde düşünüldüğünde madde güçlük indeksi bakımından test için istenilen düzeyde olan *yirmi sekizinci maddenin nihai teste alınması daha uygundur.*

Tablo 29. 29. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
29	1-9	.27	.29	.31	A	.07	.13	.03	-.17	
					B	.40	.49	.24	-.23	
					C	.25	.23	.29	.05	
					D	.27	.14	.42	.31	*
					Other	.01	.00	.00	-.03	

Yirmi dokuzuncu maddenin madde güçlük indeksi .27, ayırt edicilik indeksi ise .29 olarak bulunmuştur. Madde güçlük indeksine göre maddenin zor olduğu, madde ayırt edicilik indeksine göre ise maddenin ayırt ediciliğinin biraz zayıf olduğu, madde üzerinde bir miktar çalışılması gerektiği söylenebilir. Seçenekler incelendiğinde B ve C çeldiricisine alt ve üst grupta giden sayısının fazla olduğu söylenebilir. Bu seçeneklerden zayıflatılması önerilir.

Tablo 30. 30. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key	
30	1-10	.74	.47	.46	A	.08	.17	.02	-.23	
					B	.74	.48	.94	.46	*
					C	.08	.17	.02	-.22	
					D	.08	.17	.02	-.24	
					Other	.01	.00	.00	-.07	

Otuzuncu maddenin madde güçlük indeksi .74, ayırt edicilik indeksi ise .47 olarak bulunmuştur. Madde güçlük indeksine göre maddenin çok kolay olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin oldukça iyi olduğu söylenebilir. Seçenekler

incelendiğinde oldukça dengeli bir dağılım olduğu görülmektedir. Bu durumda aynı kazanımı ölçen iki madde karşılaştırıldığında *otuzuncu maddenin nihai teste alınması daha uygundur.*

Tablo 31. 31. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser.	Key
31	1-11	.24	.41	.45	A	.14	.16	.10	-.09	
					B	.38	.48	.25	-.22	
					C	.21	.26	.14	-.14	
					D	.24	.08	.48	.45	*
					Other	.04	.00	.00	.02	

Otuz birinci maddenin madde güçlük indeksi .24, ayırt edicilik indeksi ise .41 olarak bulunmuştur. Madde güçlük indeksine göre maddenin zor olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin iyi olduğu söylenebilir. Madde zor olsa bile ayırt ediciliğinin iyi olması kullanacağımız testin niteliğine göre tercih sebebi olabilir. Seçenekler incelendiğinde B çeldiricisine giden kişi sayısının fazla olduğu görülmektedir. B çeldiricisinin biraz zayıflatılması önerilir.

Tablo 32. 32. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser.	Key
32	1-12	.48	.59	.50	A	.48	.19	.79	.50	*
					B	.09	.17	.02	-.22	
					C	.27	.36	.14	-.22	
					D	.15	.26	.05	-.25	
					Other	.01	.00	.00	-.04	

Otuz ikinci maddenin madde güçlük indeksi .48, ayırt edicilik indeksi ise .59 olarak bulunmuştur. Madde güçlük indeksine göre maddenin orta güçlükte olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin oldukça iyi olduğu söylenebilir. Seçenekler incelendiğinde C çeldiricisine alt grubun gitme oranının fazla olduğu görülmektedir. Aynı kazanımı ölçen iki madde karşılaştırıldığında *otuz ikinci maddenin nihai teste alınması daha uygundur.*

Tablo 33. 33. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Point Biser.	Key
33	1-13	.30	.26	.28	A	.17	.25	.08	-.20	
					B	.28	.25	.29	.02	
					C	.30	.18	.44	.28	*
					D	.20	.25	.15	-.12	
					Other	.05	.00	.00	-.07	

Otuz üçüncü maddenin madde güçlük indeksi .30, ayırt edicilik indeksi ise .26 olarak bulunmuştur. Madde güçlük indeksine göre maddenin zor olduğu, madde ayırt edicilik verisine göre ise maddenin bilenle bilmeyeni iyi ayırt etmediği, madde ayırt ediciliğinin

geliştirilmesi gereken düzeyde olduğu söylenebilir. Seçeneklerde düzeltmeye gidilmesi önerilir.

Tablo 34. 34. Madde İstatistiği Tablosu

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser. Key
34	1-14	.65	.59	.52	A	.11	.25	.02	-.30
					B	.65	.32	.91	.52 *
					C	.11	.21	.02	-.26
					D	.12	.20	.04	-.21
					Other	.01	.00	.00	-.08

Otuz dördüncü maddenin madde güçlük indeksi .65, madde ayırt edicilik indeksi ise .59 olarak bulunmuştur. Madde güçlük indeksine göre maddenin kolay olduğu, madde ayırt edicilik verisine göre ise maddenin ayırt ediciliğinin oldukça iyi olduğu söylenebilir. Seçenekler incelendiğinde dengeli bir dağılım olduğu görülmektedir. Aynı kazanımı ölçen iki madde karşılaştırıldığında *otuz dördüncü maddenin nihai teste alınması daha uygundur.*

Yukarıda aktarılan bulgulardan sonra nihai teste;

2, 4, 6, 8, 10 12, 13, 15, 17, 19, 22, 24, 26, 28, 30, 32, 34. maddeler alınmıştır.

4.3. Test İstatistiklerine İlişkin Bulgular

Aşağıda uygulanan testin toplam test puanları üzerinden temel betimsel istatistiklere yer verilmiştir:

Tablo 35. Test İstatistikleri

Ortalama	16,35
Standart Sapma	3,83
Varyans	14,67
Çarpıklık	0,11
Basıklık	-0,55
Güvenirlik	0,74

Yukarıdaki istatistikler incelendiğinde, çarpıklık katsayısı sıfırdan büyük olduğu için çarpıklığın sağa (pozitif) olduğunu göstermektedir. Basıklık katsayısının sıfırdan küçük (eksi değer almış) olması dağılımın basık olduğunu göstermektedir. Merkezi dağılım ölçülerinden standart sapma (3,83) ve varyans (14,67) incelendiğinde dağılımın yayvanlaştığı görülmektedir.

Testin güvenilirliğini hesaplamak KR-20 formülü kullanılmıştır. Testteki maddeler 1(doğru) ve 0(yanlış) şeklinde puanlandığında alfa güvenilirlik katsayısı yerine Kuder Richardson formülüne dönüşür. KR-20 ile testin iç güvenirligi hesaplanabilmektedir (Kan, 2006). Kr-20 aşağıdaki formül kullanılarak hesaplanmaktadır:

$$KR.20 = \frac{K}{K-1} \left[\frac{S_x^2 - \sum p_i (1 - p_i)}{S_x^2} \right]$$

Yukardaki formülde K, madde sayısını; S_x^2 , testin varyansını; p_i , maddeyi doğru cevaplayanların yüzdesini göstermektedir. KR-20 hesaplandığında, güvenilirlik katsayısı .74 bulunmuştur. Psikolojik bir test için hesaplanan güvenilirlik katsayısının .70 ve daha yüksek olması test puanlarının güvenilirliği için genel olarak yeterli görülmektedir (Büyüköztürk, 2010). Bu durumda bulunan katsayı, testin güvenilirliğinin oldukça yüksek olduğunu göstermektedir. KR-20 ile bulunan bu yüksek katsayı, testin içindeki maddelerin aynı şeyi ölçüyor olduğunu göstermektedir. Yani test içindeki maddelerin iç tutarlılığı oldukça yüksektir.

İç tutarlılık katsayısı olan KR-20, testin maddeleri arasındaki uyumu, dolayısıyla homojenliği ölçmektedir (Erkuş, 2003). KR-20 katsayısının yüksek olması, testin tek boyutlu olduğu ya da tek bir yapıyı ölçtüğü anlamına gelmektedir ki bu da istenen bir durumdur. Bu test ile ölçülmesi hedeflenen yapı Türkçe başarısıdır. Grubun heterojen olması ve soruların homojen olması bu katsayının yüksek olmasında etkili olmuştur.

5. SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç

Testte alan toplan puanlardan yola çıkılarak test değerlendirildiğinde, puan dağılımının normal dağılıma sahip olduğu ve sağa çarpık olduğu belirlenmiştir. Testte yer alan maddelerin güçlük düzeyleri genel olarak ortalama güçlükte bulunmuştur. Bu durum varyansın maksimum değeri almasını sağlamış ve maddeler, ölçülmek istenen özelliği ölçme bakımından bireyler arasındaki farklılıkları en iyi şekilde ortaya çıkarmıştır.

Maddelerin 5 tanesi zor, 9 tanesi orta güçlükte, 13 tanesi kolay ve 8 tanesinin de çok kolay soru olduğu analizler sonunda belirlenmiştir. Kolay maddeler bizim aldığımız kriterler doğrultusunda orta güçlük sınırına yakındır bu bakımdan orta güçlük sınıfında da değerlendirilebilir çünkü belirlenen güçlük sınırları araştırmacılara göre farklılık gösterebilmektedir. Bu rakamlar bir başarı testi için oldukça idealdir. Bir başarı testinde orta güçlüğü yakın soruların yoğunlukta olması istenir. Uç noktalarda az sayıda soru olması ve testin çoğunluğunun orta güçlükte maddelerden oluşması sınıfın genelini yakalamak adına tavsiye edilen durumdur.

Testteki maddelerin ayırt edicilik indekslerine bakıldığında 5 maddenin ayırt edicilik indekslerinin .30'dan küçük çıktığı saptanmıştır. Bu maddelerden 4'ü nihai testin dışında bırakılmış ve yerine aynı kazanımı ölçen diğer maddeler alınmıştır. Bir madde ise üzerinde çalışılması planlanarak nihai teste alınmıştır. Bu bağlamda nihai testi oluşturan maddeler bilenle bilmeyeni iyi bir şekilde ayırt eden ve beklenen düzeyde çalışan maddelerdir.

Testin iç tutarlılık katsayısını bulmak üzere KR-20 formülü uygulanmıştır. KR-20 iç tutarlılık katsayısının .70'den büyük (.74) çıkması, testin tek boyutlu olduğunu göstermektedir. Bu bakımdan oluşturulacak nihai test ideal bir test niteliği taşımaktadır.

Yukarıda aktarılan bulgulardan sonucunda nihai teste, pilot uygulamada kullanılan her bir kazanımdan iki adet olmak üzere hazırlanan 34 adet maddenin çoğunda her iki madde de nihai teste alınacak standardı sağlamıştır. Bu durumda madde güçlük ve ayırt edicilik indeksi dışında seçeneklere bakılmış ve alt ve üst grubun çeldiricilere gitme durumlarına göre nihai teste alınması daha uygun olan maddenin belirlenmesi için bir tercih yapılmıştır. Sonuç olarak 34 adet maddeden madde numarası 2, 4, 6, 8, 10, 12, 13, 15, 17, 19, 22, 24, 26, 28, 30, 32, 34

olan toplam 17 madde nihai teste alınmıştır. Bu maddeler kazanım özellikleri, bilişsel düzeyleri, güçlük ve ayırt edicilik indeksleri ile birlikte aşağıdaki tabloda verilmiştir.

Tablo 36. Nihai Teste Alınan Maddeler

Nihai Teste Alınan Maddeler					
Soru No.	Pilot Testteki Soru Sırası	Kazanım	Bilişsel Düzey	Güçlük (p)	Ayırt Edicilik(r)
1	2	T.8.3.10. Geçiş ve bağlantı ifadelerinin metnin anlamına olan katkısını değerlendirir.	Kavrama	0,87	0,25
2	4	T.8.3.5. Bağlamdan yararlanarak bilmediği kelime ve kelime gruplarının anlamını tahmin eder.	Uygulama	0,56	0,58
3	6	T.8.3.6. Deyim, atasözü ve özdeyişlerin metne katkısını belirler.	Kavrama	0,43	0,34
4	8	T.8.3.7. Metindeki söz sanatlarını tespit eder.	Kavrama	0,49	0,64
5	10	T.8.3.8. Metindeki anlatım bozukluklarını belirler.	Kavrama	0,50	0,50
6	12	T.8.3.11. Metindeki anlatım biçimlerini belirler.	Kavrama	0,45	0,43
7	13	T.8.3.16. Metnin konusunu belirler.	Uygulama	0,58	0,39
8	15	T.8.3.17. Metnin ana fikrini/ana duygusunu belirler.	Uygulama	0,63	0,39
9	17	T.8.3.20. Okuduğu metinlerdeki hikâye unsurlarını belirler.	Kavrama	0,53	0,60
10	19	T.8.3.26. Metin türlerini ayırt eder.	Kavrama	0,66	0,39
11	22	T.8.3.14. Metinle ilgili soruları cevaplar.	Kavrama	0,61	0,68
12	24	T.8.4.16. Yazdıklarını düzenler.	Uygulama	0,59	0,68
13	26	T.8.4.18. Cümlelerin öğelerini ayırt eder.	Analiz	0,58	0,58
14	28	T.8.4.19. Cümle türlerini tanıır.	Uygulama	0,39	0,41
15	30	T.8.4.20. Fiillerin çatı özelliklerinin anlama olan katkısını kavrar.	Uygulama	0,74	0,47
16	32	T.8.3.19. Metnin içeriğine uygun başlık/başlıklar belirler.	Kavrama	0,48	0,59
17	34	T.8.3.9. Filimsilerin cümledeki işlevini kavrar.	Kavrama	0,65	0,59

5.2. Tartışma

Araştırma kapsamında geliştirilen başarı testi standart bir test özelliği taşır. Standart testler daha güvenilir ölçme sağlaması açısından önemlidir. Ancak standart testlerin de eleştirildiği bazı noktalar vardır. Örneğin başarı testleri hazırlandıktan sonra uzun yıllar kullanıldığı için bu testlerin ders ait öğretim programındaki kazanımların sadece bir kısmını içermesi durumunda tüm kazanımların ölçülmesi mümkün olmayacaktır. Ayrıca çoktan seçmeli maddelerle oluşturulan bir test, üst düzey becerileri özellikle Bloom taksonomisine göre son üç bilişsel basamağı (analiz, sentez, değerlendirme) ölçmekte yetersiz kalacağı fikri genel kanıdır. Nihai teste alınan maddeler arasında 11 adet kavrama düzeyinde, 5 tane uygulama ve 1 tane de analiz düzeyinde madde bulunmaktadır. Bu bakımdan oluşturulan başarı testinde de son iki üst beceri alanında madde bulunmamaktadır. Çoktan seçmeli maddelerin bu bakımdan sınırlıkları bu testte de ön plana çıkmıştır denilebilir. Testte her yıl aynı maddelerin tekrara düşme kaygısının önüne geçilmesi için ise test içeriklerinin bir süre

sonra revize edilmesi ve teste yeni pilot maddeler eklenerek testin geliştirilmeye devam etmesi yerinde olacaktır (Ansley, 1997).

Hazırlanan sorular analiz edildikten sonra elde edilen sonuçlara göre madde güçlük ve ayırt edicilik değerleri dikkate alındığında nihai teste alınması planlanan birinci maddenin 0.25 değerinde yani ayırt edicilik kriterlerine göre geliştirilmesi gereken madde aralığında olduğu görülmektedir.(Ebel, 1965) Bir maddenin ayırt ediciliğinin neden düşük çıktığı konusu tartışılan bir konudur. Bunu çeldiricilerin çalışması veya sorunun güçlük indeksinin yüksek olması dolayısıyla alt gruptan da çok sayıda cevaplayıcıyı seçeneğe çekmesi düşünülebilir.

Madde analizi yanında bazı durumlarda çeldirici analizi de yapmak gerekebilir. Çeldirici analizi, güçlük ve ayırt edicilik indekslerinden sonra madde analizinin önemli bir bileşenidir. Thissen v.d. (1989), "Multiple Choice Models: The Distractors are Also Part of the Item" isimli çalışmasında seçeneklerde doğru cevabın ve çeldiricilerin, maddeyi oluşturan temel yapıtaşlarından olduğunu bu bakımdan, madde analizinin de bir parçası olduğunu belirtmişlerdir. Haladyna ve Downing (1993)'e göre ise cevaplayıcıların %5'inden daha azının seçtiği bir çeldirici etkili olmayan bir çeldirici olarak kabul edilir. Fakat bu yorum araştırmacıyı her zaman doğru sonuca götürmeyebilir. Bir maddede alt grup, çeldiriciler yerine doğru seçeneğe daha çok gitmişse bu durum maddenin kolay olmasından da kaynaklanabilir. Nihai teste alınan birinci madde ayırt ediciliğinin düşük olması yanında maddenin seçenek durumu gözden geçirildiğinde çeldiricilere gitme oranlarında bir sapma görülmediğinden maddenin kolay bir madde olduğu söylenebilir.

5.3. Öneriler

Pilot uygulama sonucunda elde edilen verilere göre genel itibari ile nihai teste alınabilecek uygun maddeler elde edilmiştir. Bu maddeler çoğunlukla orta güçlükte, kısmen kolay ve ayırt ediciliği yüksek maddelerdir. Bunun yanında pilot testteki bazı maddelerin ayırt ediciliklerinin düşük olduğu tespit edilmiştir. Testteki maddelerden 5 tanesinin ayırt edicilik indekslerine .30'dan küçük çıktığı saptanmıştır. Madde ayırt ediciliği için ölçüt olarak aşağıdaki gösterge kabul edilirse (Ebel, 1965):

- 0,19 ve daha küçük olan maddeler test konulmaz.
- 0,20-0,29 arasında olan maddeler genel olarak düzeltilmeli ve geliştirilmelidir.

Pilot test sonucunda 1, 2, 26, 33 ve 29. maddelerin testte alınması için geliştirilmesi gerekmektedir. Bu maddelerin seçenek analizlerine bakıldığında bazı seçeneklerin çeldirici özelliğinin hem alt hem üst grubu etkilediği görülmüştür. Bu maddelerden 1, 26, 33, 29 numaralı maddeler nihai teste alınmamış, 2 numaralı madde düzeltme yapılarak alınmıştır. Kazanımı ölçen başka madde kalmadığında ayırt edicilikleri düşük olan maddelerde redaksiyona gidilerek bu maddeler de kullanılacak hâle getirilebilir. Çeldiricilerin düzeltilmesi halinde maddelerin ayırt edicilik düzeyleri de artacaktır. Bu maddelerden nihai teste alınmayan 29. maddenin hem ayırt ediciliği düşük hem zor bir madde olduğu görülmektedir. Bu maddenin de seçenekleri üzerinde çalışılması gerekir.

Özetle başarı testi için her kazanımdan iki adet olmak üzere ön uygulamaya alınan toplam 34 adet sorunun 29 tanesi nihai testte kullanılmaya uygun sorular olarak sonuçlanmıştır. Bu beklenenin üzerinde bir rakamdır. Bu sonuç öncelikle iyi bir test planı yapmaktan geçer. Test planı yapıldıktan sonra nitelikli madde yazma kriterlerine uygun maddeler yazılmalı veya seçilmelidir. Pilot testteki maddelerin ön uygulama sonuçları nihai testin belirleyicisi olmuştur. Bu pilot uygulama neticesinde nihai testte kullanılmak üzere istenilen kriterlere uygun soru kalabilmesi için test geliştirme basamaklarının mutlaka takip edilmesi gerekir.

KAYNAKÇA

- Anastasi, A. (1988). *Psychological testing*. Macmillan: New York.
- Ansley, T. (1997). The Role of Standardized Achievement Tests in Grades K12. In G.D. Phye (Ed.), *Handbook of Classroom Assessment. Learning, Achievement and Adjustment*. California: Academic Press.
- Baykul Y. (2000). *Eğitimde ve psikolojide ölçme, klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2010). *Sosyal bilimler için istatistik (5.Baskı)*. Ankara: Pegem Akademi
- Cronbach, L.J. (1971). Test validation. R. L. Thorndike. "Educational Measurement Wahington D.C.
- David, Magnusson, (1967). *Test teorisi*. Avusturalya: Addison-Wesley Yayıncılık.
- Demirel, Ö. (2010). *Eğitim sözlüğü*. Ankara: Pegem Akademi.
- Ebel, R. L. (1965). *Measuring Educational Achievement*. Prentice Hall, Inc.: New Jersey
- Erkuş, A., Sünbül, Ö., Sünbül, S. Ö., Yormaz, S., & Aşiret, S. (2017). *Psikolojide ölçme ve ölçek geliştirme-II ölçme araçlarının psikometrik nitelikleri ve ölçme kuramları*. Ankara: Pegem Akademi.
- Gucken, J. E. (1986). *The construction, validation, and use of a diagnostic test in percent for high school students*. Temple University, Ann Arbor.
- Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice item? *Educational and Psychological Measurement*, 53(4), 999-1010
- Kan, A. (2006). *Eğitimde ölçme ve değerlendirme*. Ankara: Ertem Matbaası.
- Lichtenberger, E. O., & Smith, D. R. (2005). *Essentials of WIAT II and KTEA II Assessment Essentials of Psychological Assessment*. New Jersey: John Willey & Sons.
- Özçelik, D. A. (2010). *Okullarda ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Sak, U. (2015). Development of a new intelligence test. Paper presented at The Cutting-edge Research on Talent Development Conference, Nuremberg, Germany.
- Sanlı Richards, N. (2004). *İlköğretim sekizinci sınıf düzeyi için akademik yetenek testi geliştirmeye ilişkin bir çalışma*. Doktora tezi, Ankara Üniversitesi, Ankara.
- Sarıçetin, Z. (2021). *7. sınıf ışığın madde ile etkileşimi ünitesine yönelik başarı testi geliştirme çalışması*. Yayınlanmamış yüksek lisans tezi, Mersin Üniveristesi, Mersin.
- Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları.
- Thissen, D., Steinberg, L., & Fitzpatrick, A.R. (1989). Multiple-Choice Models: The Distractors Are Also Part of the Item. *Journal of Educational Measurement*, 26(2), 161-176.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- MEB Ölçme, Değerlendirme Sınav Hizmetleri Genel Müdürlüğü <https://odsgm.meb.gov.tr/>