

**MADDE İSTATİSTİKLERİ ve MADDE TEPKİ SÜRESİ
ARASINDAKİ İLİŞKİNİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

FATMA ALTUNER

**MERSİN ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**MERSİN
AĞUSTOS - 2019**

**MADDE İSTATİSTİKLERİ ve MADDE TEPKİ SÜRESİ
ARASINDAKİ İLİŞKİNİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

FATMA ALTUNER

**MERSİN ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**




**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**Danışman
Doç. Dr. Önder SÜNBÜL**

**MERSİN
AĞUSTOS - 2019**

ONAY

Fatma ALTUNER tarafından Doç. Dr. Önder SÜNBUİL danışmanlığında hazırlanan "Madde İstatistikleri ve Madde Tepki Süresi Arasındaki İlişkinin İncelenmesi" başlıklı bu çalışma aşağıda imzaları bulunan jüri üyeleri tarafından oy birliği ile Yüksek Lisans tezi olarak kabul edilmiştir.

Görevi	Unvanı, Adı ve Soyadı	İmza
Başkan	Prof. Dr. Adnan KAN	
Üye	Doç. Dr. Önder SÜNBUİL	
Üye	Dr. Öğr. Üyesi Seçil ÖMÜR SÜNBUİL	

Yukarıdaki Jüri kararı Eğitim Bilimleri Enstitüsü Yönetim Kurulu'nun 49/2019 tarih ve 35/02 sayılı kararıyla onaylanmıştır.

Prof. Dr. Adnan KAN
Eğitim Bilimleri Enstitüsü Müdürü



Bu tezde kullanılan özgün bilgiler, şekil, tablo ve fotoğraflardan kaynak göstermeden alıntı yapmak 5846 sayılı Fikir ve Sanat Eserleri Kanunu hükümlerine tabidir.

ETİK BEYAN

Mersin Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliğinde belirtilen kurallara uygun olarak hazırladığım bu tez çalışmada,

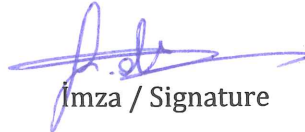
- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlâk kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak kullandığımı,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Bu tezin herhangi bir bölümünü Mersin Üniversitesi veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı,
- Tezin tüm telif haklarını Mersin Üniversitesi'ne devrettiğimi beyan ederim.

ETHIC DECLARATION

This thesis is prepared in accordance with the rules specified in Mersin University Graduate Education Regulation and I declare to comply with the following conditions:

- I have obtained all the information and the documents of the thesis in accordance with the academic rules.
- I presented all the visual, auditory and written informations and results in accordance with scientific ethics.
- I refer in accordance with the norms of scientific works about the case of exploitation of others' works.
- I used all of the referred works as the references.
- I did not do any tampering in the used data.
- I did not present any part of this thesis as an another thesis at Mersin University or another university.
- I transfer all copyrights of this thesis to the Mersin University.

02 Ağustos 2019/02 August 2019


İmza / Signature

Fatma ALTUNER

ÖZET

MADDE İSTATİSTİKLERİ ve MADDE TEPKİ SÜRESİ ARASINDAKİ İLİŞKİNİN İNCELENMESİ

Bu araştırmada, bilgisayar ortamında uygulanan bir İngilizce kelime bilgisi testinin, madde istatistikleri ve madde tepki süresi arasındaki ilişkiyi farklı yetenek düzeyleri (alt, orta, üst) göre incelenmesi amaçlanmıştır.

İlişkisel türde bir çalışma olan bu çalışma araştırma grubu Mersin Üniversitesi Eğitim Fakültesi'nde öğrenim gören 212 öğrenci oluşturmaktadır. Araştırma kapsamında öğrencilere TAO sınav platformunda çoktan seçmeli İngilizce kelime bilgisi testi uygulanmıştır. Teste ilişkin madde istatistikleri hesaplanmış, öğrenciler yetenek düzeylerine göre ayrılmıştır. TAO sınav platformundan elde edilen madde tepki süreleri ile madde istatistikleri arasındaki ilişki için korelasyon katsayısı hesaplanmıştır. Elde edilen veriler R-3.2.0 programı kullanılarak analiz edilmiştir.

Araştırma sonucunda, tepki süresinin maddelerin güçlük düzeyine göre değişiklik gösterdiği görülmüştür. Ayrıca üst yetenek düzeyine sahip öğrenciler kolay maddeleri daha kısa sürede yanıtlarken, zor maddelere daha fazla zaman harcamışlardır. Madde ayırt edicilik indeksinin öğrencilerin yetenek düzeylerine göre tepki süresi üzerinde etkisi olmadığı sonucuna ulaşılmıştır.

Böylece bu çalışmada, madde güçlük indeksinin tepki süresini etkilediği, madde ayırt edicilik indeksinin tepki süresi üzerinde anlamlı olarak etkisi olmadığı sonucuna ulaşılmıştır.

Anahtar kelimeler: Madde Tepki Süresi, Madde Güçlük İndeksi, Madde Ayırt Ediciliği, Yetenek Düzeyleri

Danışman: Doç. Dr. Önder SÜNBÜL, Mersin Üniversitesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Mersin.

ABSTRACT

INVESTIGATION OF THE RELATIONSHIP BETWEEN ITEM STATISTICS AND ITEM RESPONSE TIME

In this study, it is aimed to investigate the relationship between item statistics and item response time according to different ability levels (lower, middle, upper).

The universe of this study, which is a relational type, consists of 212 students studying at Mersin University Faculty of Education. Within the scope of the research, a multiple choice English vocabulary test was applied to students on TAO exam platform. Item statistics of the test were calculated and students were divided according to their ability levels. The correlation coefficient was calculated for the relationship between item response times and item statistics obtained from TAO exam platform. Obtained data were analysed using R-3.2.0 program.

As a result of the study, it was seen that the response time varies according to the difficulty level of the items. In addition, students with top talent level responded to easier items in a shorter period of time and spent more time on difficult items. It was concluded that item discrimination index had no effect on response time according to students' ability levels.

Thus, in this study, it was concluded that item difficulty index had a significant effect on response time and item discrimination index had no significant effect on response time.

Keywords: Item Response Time, Item Difficulty Index, Item Discrimination, Ability Levels.

Advisor: Assoc. Professor Önder SÜNBÜL, Mersin University, Mersin University, Department of Measurement and Evaluation in Education, Mersin.

TEŐEKKÜR

Yüksek lisans öğrenimim boyunca bana destek olan, yol gösteren tezimin her aşamasında yardımını esirgemeyen, emek harcayan çok değerli hocam ve danışmanım Doç. Dr. Önder SÜNBUİL' e teşekkür ederim.

Tez sürecinde bilgilerini benimle paylaşan, görüş ve eleştirileriyle çalışmama katkı sunan desteğini ve sevgisini esirgemeyen çok değerli hocam Dr. Öğretim Üyesi Seçil ÖMÜR SÜNBUİL' e teşekkür ederim.

Tez savunma jürimde bulunan değerli zamanını ve emeğini harcıyarak çalışmama katkı sunan Prof. Dr. Adnan KAN' a teşekkürlerimi sunarım.

Her zaman yanımda olan, bana varlığıyla cesaret ve güven veren canım annem ve babama, biricik kardeşim Emine ALTUNER' e ve sürekli yanımda olup sevgisi ve sabrıyla beni destekleyen Arda ALP' e teşekkürü borç bilirim.



İÇİNDEKİLER

	Sayfa
İÇ KAPAK	
ONAY	
ETİK BEYAN	
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
TABLolar DİZİNİ	v
ŞEKİLLER DİZİNİ	vi
KISALTMALAR ve SİMGELER	vii
1. GİRİŞ	1
1.1. Bir Ölçüte Göre Yapılan Değerlendirme Türleri	2
1.2. Yapılış Amacına Göre Değerlendirme	2
1.3. Madde ve Test İstatistikleri	4
1.3.1. Madde Güçlük İndeksi	4
1.3.2. Madde Ayırt Edicilik Gücü İndeksi	5
1.4. Test İstatistikleri	6
1.4.1. Testin Ortalaması	6
1.4.2. Testin Ortalama Güçlüğü	7
1.4.3. Test Varyansı ve Standart Sapması	7
1.4.4. Testin Güvenirliđi	7
1.4.5. Testin Standart Hatası	8
1.5. Ölçme Aracı Uygulama Şekilleri	9
1.5.1. Kâğıt Kalem Testleri	9
1.5.2. Bilgisayar Destekli Testler (BDT)	9
1.5.3. Bilgisayar Ortamında Bireye Uyarlanmış Test	11
1.6. Tepki Süresi	12
1.7. Araştırmanın Amacı	13
1.8. Araştırmanın Önemi	14
1.9. Problem	14
1.9.1. Problemler	14
1.10. Araştırmanın Sayılıtları	15
1.11. Araştırmanın Kapsam ve Sınırlılıkları	15
2. İLGİLİ ARAŞTIRMALAR	16
3. YÖNTEM	18
3.1. Yöntem	18
3.2. Araştırmanın Modeli	18
3.3. Araştırma Grubu	18
3.4. Veri Toplama Araçları	18
3.5. Uygulama	21
3.6. Verilerin Analizi	22
4. BULGULAR ve YORUM	23
5. TARTIŞMA, SONUÇ ve ÖNERİLER	33
5.1. Sonuç ve Tartışma	33
5.2. Öneriler	333
KAYNAKLAR	355
EKLER	387
ÖZGEÇMİŞ	40

TABLULAR DİZİNİ

	Sayfa
Tablo 3.1. Deneme Uygulamasında Maddelere İlişkin Madde Ayırt Edicilik İndeksleri Ve Madde Güçlük İndeksleri	19
Tablo 3.2. Deneme Uygulamasına Ait Test İstatistikleri	19
Tablo 3.3. Asıl Kelime Bilgisi Testi Uygulamasındaki Maddelere İlişkin Madde Ayırt Edicilik İndeksleri ve Madde Güçlük İndeksleri	20
Tablo 3.4. Asıl Uygulamaya Ait İstatistikler	21
Tablo 4.1. Alt ve Üst Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları	26
Tablo 4.2. Alt ve Orta Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları	27
Tablo 4.3. Üst ve Orta Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları	27
Tablo 4.4. Alt ve Üst Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları	31
Tablo 4.5. Alt ve Orta Yetenek Grupları için Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları	32
Tablo 4.6. Üst ve Orta Yetenek Grupları için Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları	32

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 4.1. Alt Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları	23
Şekil 4.2. Orta Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları	24
Şekil 4.3. Üst Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları	25
Şekil 4.4. Tüm Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları	26
Şekil 4.5. Alt Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları	28
Şekil 4.6. Orta Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları	29
Şekil 4.7. Üst Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları	30
Şekil 4.8. Tüm Gruptaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları	31

KISALTMALAR ve SİMGELER

Kısaltma/Simge	Tanım
BDT	Bilgisayar Destekli Test
BOBUT	Bilgisayar Ortamında Bireye Uyarlanmış Test
KTK	Klasik Test Teorisi



1. GİRİŞ

Belirli bir amaç için bir arada bulunan birbirleriyle bağlantılı parçalar sistemleri oluşturur. Sistem girdi, süreç, çıktı ve kontrol olmak üzere dört parçadan oluşur ve bunlar birbirleriyle etkileşim halindedir. Eğitim sistemimizin girdileri arasında, istenmeyen bazı insan davranışlarını bir amaç doğrultusunda değiştiren bireylere yeni davranışlar kazandırmak, farklı öğretim yöntemleri geliştirmek, toplumun gelenek ve görenekleri, eğitimle ilgili tüm düzenlemeler, bireylerin nitelikleri, eğitimle ilgili bilgiler, kullanılan materyaller tüzükler, yönetmelikler ve yasalar, eğitim veren öğretmenler yer alır (Baykul,1999). Girdilerin geliştirilmesi ve değiştirilmesi için yapılan etkinliklerin tamamı süreci oluşturur. Süreç sonunda ortaya çıkan yeni ve değişen ürünler ise sistemin çıktılarıdır. Eğitim sisteminin kontrolü ise değerlendirme aracılığıyla gerçekleştirilir. Bireylerin bilgi düzeylerini belirlemek, gelişimlerini takip etmek, baskın yanlarını ön plana çıkarmak gelişimlerine yönelik geri bildirimde bulunmak, öğrenme güçlüklerini belirlemek, öğretimin ve öğretim materyallerinin etkinliğini belirlemek, gelecekteki öğrenme süreçlerini planlamaya veri sağlayabilmek ölçme ve değerlendirme ile mümkündür.

Alanyazın incelendiğinde ölçme ile ilgili birçok tanımlamayla karşılaşmak mümkündür. Magnusson (1967) ölçme kavramını; geçerli görgül yollarla test edilebilecek kurallar çerçevesinde nesnelere belli özelliklere sahip oluş derecelerine göre sayılar veya semboller vermek, Lord ve Novick (1968), davranış alanındaki ilişkileri koruyacak ve onların özelliklerini niteleyecek biçimde, deneysel birimlerin tanımlanmış özelliklerine göre sayılarla ayrılma işlemi, Arıcı (1972), nesnelere ya da bireylere göre sayı ile ifade edilmesi işlemi, Turgut (1983), bir niteliğin gözlenip gözlem sonucunun sayılarla veya başka sembollerle gösterilmesi, Tekin (1993), nesnelere belli bir özelliğe sahip olup olmadığını, sahipse oluş derecesinin gözlenip, gözlem sonuçlarının sembollerle ve özellikle sayı sembolleriyle ifade edilmesi şeklinde tanımlamaktadır. Yapılan tanımlamalardan yola çıkıldığında, ölçme kavramı, özetle, belirli özelliklerin gözlenmesi, gözlemden elde edilen bulguların sayı ve semboller kullanarak ifade edilmesidir. Sonuçları anlamlandırmak için bu sonuçların uygun bir ölçütle karşılaştırılması ve elde edilen sonuçlar ile bir yargı oluşturulması gerekir. Yapılan bu işlem, ölçme ve ölçme de kullanılan ölçme araçlarının önemli bir parçası olan değerlendirme aşamasını oluşturur.

Değerlendirme karar verme, yargıda bulunma sürecidir. Turgut (1977), değerlendirmeyi "ölçme sonuçlarının bir ölçütle kıyaslanarak bir karara varılması" olarak tanımlamaktadır. Tanımdan da anlaşıldığı gibi, değerlendirme; ölçme sonuçları, ölçüt ve karar olmak üzere üç öğeden oluşmaktadır. Ölçme, değerlendirmenin ön koşuludur. Ölçme sonuçları olmadan değerlendirme yapılması mümkün değildir. Ölçüt, ölçme işleminden elde edilen sonuçların

yorumlanmasını kolaylaştıran, değer yargısına ulaşmaya yardım eden bir kriter kıstas olarak düşünülebilir (Turgut,1983).

Değerlendirmede kullanılacak ölçütlerin özelliklerinin bilinmesi ve uygun ölçütün kullanılması önemlidir.

Eğitimde değerlendirmeye ilişkin farklı bakış açılarına göre sınıflamalar yapılabilmektedir. Bu sınıflandırmalara göre Ertürk (1993) değerlendirme türlerini; ölçüte göre değerlendirme (bağıl ve mutlak değerlendirme), amaca göre değerlendirme (Tanıma-Yerleştirmeye, Biçimlendirme ve Yerleştirmeye, Değer Biçmeye Dönük Değerlendirme) olarak sınıflandırmıştır. Aşağıdaki kısımda değerlendirme türlerine kısaca değinilmiştir.

1.1. Bir Ölçüte Göre Yapılan Değerlendirme Türleri

Bir ölçüte göre yapılan değerlendirmelerde iki türlü ölçüt kullanılmaktadır. Bunlar; “mutlak ölçüt” ve “bağıl ölçüt” tür. Mutlak ölçüt, ölçme işlemi yapılmadan önce, ölçme işlemi yapılacak olan grubun tamamını kapsayacak kesin, net standart özelliği olan ölçütlerdir. Örneğin, ölçme ve değerlendirme dersine ilişkin hazırlanmış bir başarı testinde 100 tam puan olmak üzere, “60 puan ve 60’dan yüksek alanlar başarılı olur” şeklinde ölçüt belirtilmişse mutlak ölçüt söz konusudur. Bağıl ölçüt ise, mutlak ölçütün aksine ölçme işlemi gerçekleştirildikten sonra, ölçme sonuçlarını dikkate alarak bir gruba ya da bireye göre belirlenen ölçüttür. Sınava giren bireylerin “aritmetik ortalamanın üstünde alanlar başarılı olur” şeklinde bir norma dayalı özellik belirtilmişse bağıl ölçütte örnek verilebilir.

1.2. Yapılış Amacına Göre Değerlendirme

Tanıma-Yerleştirmeye dönük değerlendirme; öğrencilerin belirli bir programa yerleşmeden önce bilgi düzeylerinin belirlenmesi amacıyla not verme amacı gütmeyen yapılan değerlendirmedir. Öğrencilerin beceri ve yeteneklerine uygun olarak seçmeli derslerden birini seçmesi veya yabancı dil kursuna kayıt yaptırmak isteyen bireylerin seviye belirleme sınavlarına girdikten sonra uygun kura yerleştirilmesi örnek olarak verilebilir.

Biçimlendirme ve Yerleştirmeye dönük değerlendirme; eğitim süreci devam ederken kazandırılması ön görülen davranışların kazandırılması aşamasında, öğrenme eksikliklerini ve yanlışlıkları belirlenerek bu eksikliklerin giderilmesi amacıyla yapılan değerlendirmedir. Öğrencilerin belirli bir konuyla ilgili bilgi, beceri ve yeterlilik düzeyini yoklamak amacıyla yapılan ölçüt referanslı ünite testleri, izleme testleri örnek verilebilir.

Değer biçmeye dönük değerlendirme; sürecin sonunda bir süreci bitirmek için yapılır. Programda kazandırılması gereken hedef davranışların kazanılma düzeyinin yoklanması ve bir

not ile yargılanması amaçlandığında yapılan değerlendirmedir. Öğretmenlerin öğrencilere uyguladıkları yılsonu başarı sınavları örnek verilebilir. İlgili kazanımlara erişme düzeyleri değerlendirilerek başarılı-başarısız, kaldı-geçti gibi kararlar alınır (Bilcan, 2014).

Değerlendirmenin doğruluğu, ölçütün uygunluğunun yanı sıra ölçme sonuçlarının güvenilirliğine ve geçerlilik bağlıdır (Baykul, 2015). Bu sebeple değerlendirmenin daha gerçekçi olabilmesi için ölçme işleminde kullanılan ölçme aracının güvenilirliğinin ve geçerliğinin yüksek olması gerekir. Bu kapsamda eğitimde çeşitli ölçme araçları kullanılmaktadır. Bu araçlardan bazıları; doğru-yanlış maddeler, eşleştirme maddeler, kısa cevaplı maddeler, açık uçlu maddeler, dereceli puanlama anahtarı, kontrol listesi ve çoktan seçmeli maddelerdir.

Bu araştırmada ölçme araçları içerisinde en yaygın olarak kullanılan çoktan seçmeli test maddeleri kullanıldığından, aşağıdaki kısımda çoktan seçmeli test maddelerinin özelliklerine kısaca değinilmiştir.

Çoktan seçmeli maddeler, cevaplayıcının madde kökünü okuduktan sonra doğru cevabı sunulan seçenekler arasından seçeceği bir ölçme aracıdır. Her bir madde, kök ve seçeneklerden oluşmaktadır. Maddenin kökü, sorunun sorulduğu bölümü, seçenekler ise cevaplayıcının olası yanıtlarını içermektedir. Seçenekler, biri doğru cevap diğerleri ise "çeldiriciler" olarak isimlendirilen yanlış cevaplardan oluşmaktadırlar. Çeldiriciler; maddede yer alan soruyla ölçülmek istenilen davranışa ya da beceriye sahip olmayan ya da doğru yanıt bilmeyenlerin aklını çelmek, onları yanıltmak amacıyla yazılmış, maddede doğru yanıt olmayan seçeneklerdir (Özçelik, 1992). Çoktan seçmeli maddelerde, soru sayısının fazla olması, ölçme sonuçlarının hem güvenilirliğini hem de kapsam geçerliğini arttıran bir unsurdur. Çoktan seçmeli maddelerin bazı sınırlılıkları da bulunmaktadır. Bunlardan biri de şans başarısıdır. Şans başarısı, cevaplayıcının maddenin doğru cevabını bilmeden ya da tam anlamıyla emin olmadan tahmin ederek bulmasına denir (Rogers 1997). Şans başarısı ve buna bağlı olarak şans puanı madde ve test istatistiklerinde hatalara sebep olur. Çoktan seçmeli maddelerin avantajları ise; birçok hedef davranışları yoklaması, maddelerin puanlamasının kolay ve çok kısa sürede yapılması, maddelere ilişkin istatistiklerin hesaplanması ve bu sayede istenilen özellikte ölçme aracı oluşturulması vb. dir. Çoktan seçmeli maddelerin bir diğer avantajı ise, madde üzerinde birçok istatistiksel işlemin yapılabilmesine olanak sağlamasıdır. Test uygulandıktan sonra madde üzerinde çeşitli istatistiksel işlemler yapılabilir ve madde üzerinde yapılan bu istatistiksel işlemlere genel olarak madde istatistiği adı verilebilir. Çoktan seçmeli maddeler için elde edilebilecek olan bazı istatistiksel, madde varyansı, madde standart sapması, madde basıklığı, madde çarpıklığı, madde güvenilirliği olmakla beraber en çok ihtiyaç duyulan iki istatistik madde güçlüğü ve madde ayırt ediciliğidir. Aşağıdaki kısımda madde ve test istatistikleri kısaca bahsedilmiştir.

1.3. Madde ve Test İstatistikleri

Madde İstatistikleri

Madde istatistikleri, belirli bir gruba ait madde özelliklerini gösterir (Baykul, 2015). Bireylerin, bir testin maddelerine verdikleri cevapları analiz edildiğinde önemli bilgiler elde edilir. Bu bilgilere dayanarak istenilen özellikte maddeler düzenlenebilir, güçlüğü istenilen derecede olmayan, ayırıcı olmayan, çeldiricileri işlemeyen kusurlu maddeler bulunabilir (Turgut, 1997).

1.3.1. Madde Güçlük İndeksi

Madde güçlük indeksi, maddeyi doğru cevaplayanların yüzdesi olarak ifade edilir. Burada maddeyi doğru cevaplayanlar arasında, madde ile yoklanan davranışa sahip olanlar ile birlikte bu davranışa sahip olmadıkları halde maddeyi şansa doğru cevaplayanlar da vardır. Madde güçlük indeksi eşitlik 1’de belirtilmiştir (Baykul, 2000);

$$P_j = \frac{n(D)}{N} \quad (1)$$

P_j = j maddesinin madde güçlük indeksi

$n(D)$ = j maddesini doğru cevaplayanların sayısı

N = Cevaplayıcı sayısı

Zamanın yeteri kadar verilmediği durumlarda öğrencilerin bir kısmı bazı soruları boş bırakmış olabilir. Boş bırakılmış maddelere erişilmemiş maddeler adı verilir. Bazı sebeplerden dolayı (sorunun uzun olması gibi) öğrencilerin bir kısmı soruları okumadan boş bırakmış olabilir erişilmemiş madde dışındaki maddelere atlanmış maddeler denir. Böyle durumlarda madde güçlük indeksi eşitlik 2’de belirtilen şekilde hesaplanır.

$$p = \frac{n(D)}{N - n(E)} \quad (2)$$

$n(E)$ = Erişilemeyen öğrenci sayısı ya da atlanmış öğrenci sayısı

Madde güçlük indeksi 1’e yaklaştıkça, doğru cevaplayan öğrenci sayısı artar, birçok öğrenci tarafından doğru yanıtlanan soru kolaydır. Madde güçlük indeksi 0’a yaklaştıkça, soruyu doğru yanıtlayan öğrenci azalır, dolayısıyla soru zordur. Madde güçlük indeksinin 0,50 yakın olması, sorunun orta güçlükte olduğunu gösterir (Atılğan, 2016).

Eğer bir maddeyi, testi alan bireylerden hiçbiri doğru yanıtlamamış ya da testi alan bireylerin tümü cevaplamışsa, bu maddenin bireyler arasındaki farkı ölçme yönünden hiçbir değer taşımadığı anlaşılır. Böyle bir maddenin, test puanlarının toplam varyansa hiçbir etkisi olmadığı için testin geçerlik ve güvenilirlik indekslerine de hiçbir katkısı yoktur (Crocker ve Algina, 1986). Madde gücü 0.00'a ve 1.00'e yaklaştıkça, test geliştiricilere bireysel farklılıklar hakkında daha az bilgi verir.

1.3.2. Madde Ayırt Edicilik Gücü İndeksi

Madde ayırt edicilik gücü indeksi; maddeyi doğru yanıtlayan ile doğru yanıtlamayan birbirinden ayırt edebilme derecesidir. Bir maddenin ayırt edicilik gücünün yüksek olması, maddeyi doğru cevaplayan üst grup ile alt grup arasındaki farka bağlıdır. Üst grupta doğru cevaplayanların sayısı ne kadar fazla ise ayırt edicilik gücü o kadar yüksektir.

Ayırt edicilik indeksi maddenin kalitesi hakkında bilgi verir ve madde puanlarıyla test puanları arasındaki ilişki olarak da tanımlanabilir. Madde ayırt edicilik gücü indeksi (r_{jx}); -1 ile +1 arasında değer alır. r_{jx} değeri yükseldikçe maddenin ayırt ediciliği artar, düştükçe azalır. Maddenin ayırt edicilik gücünün pozitif değer alabilmesi için üst grup doğru açıklama oranının, alt grup doğru cevaplama oranından yüksek olması gerekir. Bunun tersi olduğu durumda negatif değer alır. Güvenilir test, ayırt ediciliği yüksek maddelerden oluşan testtir.

Madde ayırt edicilik gücü, aynı zamanda "maddenin geçerliliği" olarak da adlandırılır. Bir madde, bilenle bilmeyeni ayırt edebiliyorsa, o madde amaca hizmet ediyor demektir. Test geliştirmede maddelerin, bilenle bilmeyeni ayırt etmesi beklenir. Bu, ölçülmek istenen niteliğe sahip olanların maddeyi doğru yanıtlamaları, ölçülmek istenilen özelliğe sahip olmayanların ise yanlış yanıtlamaları veya yanıtız bırakmaları ile eş anlamlıdır. Bireyleri ayırmak için ayırt edicilik indeksi ölçüt kabul edilir (Baykul, 2000).

Madde ayırt edicilik indeksini hesaplamada, sıklıkla madde toplam test puanı arasındaki ilişki katsayısı kullanılır. Madde toplam test puan ilişki katsayısıyla, bir test maddesine verilen tepkilerle, testin tümüne verilen tepkiler arasında pozitif ve doğrusal yönde bir ilişkinin varlığı varsayımı bulunur. Burada bir maddeden alınan puanlarla, tüm testten alınan puanlar pozitif yönde ve yeterince yüksek bir ilişki gösteriyorsa, o madde ayırt edici kabul edilir (Erkuş, 2003). Negatif bulunan ilişki katsayıları maddenin ters yönde ayırt edici olduğunu gösterir. Yani testin genelinde başarısız olanların, o maddeyi doğru yanıtladığı, başarılı olanların ise yanlış yanıtladığı şeklinde değerlendirilir. Bir madde için negatif madde toplam test puan ilişki katsayısı, o maddenin amacına hizmet etmediğini gösterir. Böyle bir madde ters puanlanmış, iyi yazılmamış ya da ilgili özelliği ölçmüyor olabilir.

1.4. Test İstatistikleri

Bir ölçme aracı olarak testin özellikleri, genellikle ortalaması, varyansı, çarpıklığı ve basıklığı gibi test puanları dağılımının parametreleri ile güvenilirlik ve geçerlik derecelerdir. Evren için testin bu sayısal özelliklerine test parametreleri denir. Test parametreleri, testin, deneme grubu adı verilen bir gruba uygulanması sonunda elde edilen cevaplardan kestirilir. Hesaplanan sayısal özelliklere test istatistiği denir. Test istatistikleri madde istatistiklerine bağlı olarak hesaplanabilir. Testi maddeler oluşturduğundan test istatistikleri madde istatistiğinin bir fonksiyonudur.

1.4.1. Testin Ortalaması

Bir testten bireylerin aldıkları puanların ortalaması anlamına gelir ve test puanları dağılımının yığılma eğilimi gösterdiği noktayı işaret eden sayısal bir değerdir. Madde güçlükleri indeksleri toplamı testin ortalamasını verir ve eşitlik 3 ve eşitlik 4'te belirtildiği şekilde hesaplanır.

$$\bar{X} = \frac{\sum_{i=1}^n Xi}{n} \quad \bar{X} = \sum_{j=1}^k Pj \quad (3)$$

$$\bar{X} = \frac{\sum_{i=1}^n Xi}{n} \quad \bar{X} = \sum_{j=1}^k Pj \quad (4)$$

Madde sayısına bağlı olarak testin ortalamasının küçültülmesi ya da büyütülmesi testteki maddeler değiştirilmeden istenilen düzeye gelecek biçimde testteki maddelerin çıkarılması ya da yeni madde eklenmesiyle olur. Bu da bazı sorunları beraberinde getirir. Maddenin çıkartılması testin uzunluğunu kısaltır, eklenmesi ise testi uzatır. Testin uzun olması uygulama gücü getirirken, testin güvenilirliğini, özellikle başarı testlerinde de kapsam geçerliğini düşürür. Madde güçlükleri düzeyinin küçülmesi test ortalamasını küçültürken, büyümesi test ortalamasını artırır.

1.4.2. Testin Ortalama Güçlüğü

Testte bulunan maddelerin güçlük indeksleri toplamının madde sayısına bölümü, testin ortalama güçlüğü verir. K, testte yer alan madde sayısını temsil etmek üzere, eşitlik 5'te belirtildiği gibi hesaplanır.

$$\bar{P} = \frac{\sum_{j=1}^K P_j}{K} \quad (5)$$

Testin ortalaması ve madde sayısının bilinmesi durumunda da testin ortalama güçlüğü hesaplanabilir. Bu durumda testin ortalama güçlüğü hesaplanması eşitlik 6'da gösterilmektedir.

$$\bar{P} = \frac{\bar{X}}{K} \quad (6)$$

Testin ve madde güçlük indekslerinin bilinmesi durumunda, testteki madde sayısı hesaplanabilir ve eşitlik 7'de gösterilmektedir.

$$\bar{X} = K \cdot \bar{P} \quad (7)$$

1.4.3. Test Varyansı ve Standart Sapması

Test puanları dağılımının varyansı, test varyansı olarak adlandırılır. Testin standart sapması ise, varyansın karekökü alınarak bulunur. Testin varyansı ve standart sapması, madde istatistiklerinden, madde güvenilirlik indeksi kullanılarak da bulunur.

$$S^2(X) = \frac{\sum_i^n (x_i - \bar{x})^2}{n} \quad (8)$$

Standart sapmanın karesi alındığında varyans bulunur. Güvenirliği etkilediği için testin varyansı ve standart sapmasının büyük olması istenir.

1.4.4. Testin Güvenirliği

Testin güvenirliliği, ölçme sonuçlarının seçkisiz hatalardan arınıklık derecesinin göstergesidir. Kuder ve Richardson (1937), maddelerin birbiriyle benzer olduğunu ve aynı ortalama, varyansa sahip olduğu varsayımından hareketle KR-20 geliştirmiştir. KR-20, 1-0 olarak puanlara bilen iki kategorili maddelerden oluşan testin güvenirliliğini hesaplamada

kullanılır. KR-21 ise madde güçlük indekslerinin bilinmediği durumlarda, madde güçlüğü eşit varsayımı altında geliştirilen güvenilirlik katsayısıdır (Atılgan ve diğerleri, 2016). KR-21 güvenilirlik katsayısı, madde sayısı, test ortalaması ve test varyansına bağlıdır, madde istatistiği yoktur (Baykul, 2015). Eğer test, iki kategorili puanlanabilen maddelerden oluşmuyorsa, ağırlıklı ve çoklu puanlanabilen maddelere sahipse, Cronbach Alfa yöntemi kullanılır. Cronbach Alfa madde puanlarının normal veya sürekli dağılım gösterdiğini varsayarken, KR-20 madde puanlarının Bernoulli dağılımı gösterdiğini varsayar (Atılgan ve diğerleri, 2016). Bu katsayı 1'e yaklaştığında maddelerin birbiriyle tutarlı sonuçlar ürettiği ve testin iç tutarlığın yüksek olduğu maddelerin ve dolayısıyla testin homojen olduğu ya da teste yer alan maddelerin aynı yapıyı ölçtüğünü, 0'a yaklaştıkça testin maddelerinin tutarlı olmadığı ve testin güvenilir olmadığını söyleyebiliriz.

Testin güvenirliliği KR-20 formülüyle eşitlik 9'da belirtildiği şekilde hesaplanabilir.

$$KR - 20 = \frac{K}{K - 1} \left(1 - \frac{\sum p \cdot q}{S_X^2} \right) \quad (9)$$

K : Testte bulunan madde sayısı

p : Madde güçlük indeksi

q : $1-p$

S_X^2 : Test puanları dağılımı varyansı

Test istatistikleri kullanılarak güvenilirlik eşitlik 10'da belirtilen şekilde hesaplanır.

$$KR - 21 = \frac{K}{K-1} \left(1 - \frac{K \cdot \bar{X} \cdot (\bar{X})^2}{K \cdot S_X^2} \right) \quad (10)$$

K : Testte bulunan madde sayısı

S_X^2 : Test puanları dağılımı varyansı

\bar{X} : Test puanları ortalaması

Cronbach Alfa formülüyle eşitlik 11'de belirtildiği şekilde hesaplanabilir.

$$a = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{S_X^2} \right] \quad (11)$$

1.4.5. Testin Standart Hatası

Test sonuçlarına karışan hata miktarı bir oran olarak değil de, test puanı cinsinden belirlenmek istenebilir. Test puanlarına karışan hata miktarı cinsinden;

$$S_H = S_X - (\sqrt{1 - r_{xx}}) \quad (12)$$

r_{xx} : Test puanlarına ait güvenilirlik

Şeklinde hesaplanır. Bireyin gerçek puan aralığının tahmin edilmesinde kullanılır.

1.5. Ölçme Aracı Uygulama Şekilleri

1.5.1. Kâğıt Kalem Testleri

Kâğıt kalem testleri bireylerin belli bir alana ilişkin özelliklerini ve yeteneklerini belirlemek amacıyla kullanılmaktadır. Bireylere, ölçülmek istenilen alana ilişkin özellikleri ölçen, sabit sayıda soruların bulunduğu kitapçıklar verilerek, onların belli bir süre zarfında tüm soruları yanıtlaması beklenir (Özdemir, 2015). Bu durumda bireylere sorulan sorular yazılı olarak verildiği gibi, bireylerin cevapları da yazılı olarak alınır. Yüksek ve düşük yetenek düzeylerine sahip bireylerin yetenek düzeyleri dikkate alınmadan herkese aynı sorular yönetilerek tüm maddeleri yanıtlaması beklenir (Wainer ve ark, 2000). Çoğu zaman bireyin doğru cevapladığı maddenin niteliğine (kolay-orta-zor) bakılmadan doğru sayısına göre toplam puan hesaplanır. Farklı yetenek düzeylerine sahip bireylerde testin zorluk derecesi bireyden bireye değişkenlik gösterir. Öğrenciler için sorular çok kolay ya da çok zor gelebilir bu da ölçülmek istenilen özellik hakkında kesin bilgi vermez (Özdemir,2015). Öğrencilere çok kolay bir test uygulandığında dikkatsizce davranıp hata yapabilirler ya da şaşırtmaca olduğunu düşünerek yanlış cevabı işaretleyebilirler. Çok zor bir test uygulandığında ise öğrencilerin motivasyonları düşerek soruları okumadan işaretleyebilirler veya benzer başka davranışlara yönelebilirler (Linacre, 2000). İki durumda bilgi verici olmayan sonuçlar elde edilir. Kâğıt kalem testleri ekonomiktir, ileri dönük olarak sürekli kullanılabilir olduğundan çok tercih edilir (Cömert, 2008).

Kâğıt kalem testlerinde, test araç gereçlerinin genel merkezden sınav merkezlerine taşınması, test kitapçıklarının oluşturulması, sonuçların değerlendirilmesi, güvenliğin sağlanması önemlidir.

Kâğıt-kalem testlerinde geri bildirim vermek oldukça zaman ve emek gerektirir, test gizliliğini sağlamak güçtür, bu tür sınırlılıklar bilgisayar ortamında uygulanan testleri ile aşılabılır.

Günümüzde yaygın olarak kullanılan kâğıt kalem sınavları bilgisayarın hayatımıza girmesi ve teknolojinin gelişmesiyle bilgisayar destekli sınavlara yönelim artmıştır.

1.5.2. Bilgisayar Destekli Testler (BDT)

Bilgisayar destekli testlerde (Computer Based Testing, CBT) kâğıt kalem testlerinde kullanılan soruların aynısının, yine aynı sırada bilgisayar ortamında sunulması şeklinde gerçekleştirilir (Bugbee ve Alan, 1996). Bilgisayar destekli testlerin birçok yararı bulunmaktadır. Bilgisayar destekli testler; test için harcanan zamanı azaltır, test kullanıcıları

hakkında daha fazla bilgi elde edilmesini sağlar, test sonuçlarını anında bildirir ve kâğıt kalem testlerine göre daha kolay bir şekilde kayıtları tutar (Akdemir ve Oğuz, 2008).

Bilgisayar destekli testler, özellikle öğrencilerin büyük gruplarını değerlendirirken, test dağıtım lojistiği, zaman ve iş gereksinimlerinin azaltılması gibi birçok nedenden ötürü eğitimciler için çekici bir seçenektir. Bununla birlikte, kâğıt-kalem testleri üzerinden bilgisayar destekli testleri doğrulamak için test modundaki farkın öğrenci performansı üzerinde olumsuz bir etkisi olmadığından emin olunması önemlidir (Ita M.E. , Ashley K.E. , Morin B. , 2014).

Bilgisayar destekli testlerin kullanımı son zamanlarda hızla yayılmıştır. Çünkü bu tür testler gerçek zamanlı puanlama ve anında geri bildirim sağlayabilir, bireyselleştirilmiş test yöntemlerinin kullanımını kolaylaştırabilir, test yönetimini iyileştirebilir ve test masraflarını düşürebilir (Akdemir ve Oğuz, 2008). En önemlisi, gerçek zamanlı puanlama ve anında geri bildirim sayesinde öğretmenlerin öğrencilerine rehberlik etmelerinde yardımcı olan değerli bilgiler verir (Paek, 2005). Gelişmiş BDT sistemlerinin bulunması, grafik ve multimedyanın test sorularına dâhil edilmesini mümkün kılmıştır (Segall ve ark, 2005). Bu faydalar, eğitim kurumlarını ve değerlendirme yöneticilerini, kâğıt-kalem testlerin kullanımından, BDT kullanımına geçmeye teşvik etmiştir (Mills, 2002).

BDT'nin önemli bir yararı, öğrencilerin yanlış cevaplarını etkili bir şekilde gözden geçirebilecekleri bir öğrenme ortamı yaratarak sürekli formatif geribildirim sağlamalarıdır (Martin, 2008). Etkili öğrenim, öğrencilerin gerçek zamanlı olarak test sonuçlarındaki geribildirimlerini aktif olarak incelemelerini gerektirir. Bilgisayar destekli testler, öğrencilerin geribildirimleri içselleştirmelerine ve çalışmalarıyla ilgili yargılarda bulunmak için bu bilgiyi kullanmalarına olanak tanır (Gardner, 2006).

Bilgisayar destekli testler özellikle biçimlendirici değerlendirme ve öz değerlendirmeye uygundur ve öğrencileri kendi kendine yönlendiren öğrenciler olarak yetkilendirir (Nicol ve Macfarlane, 2006). Buna ek olarak, BDT; doğru derecelendirme ve gerçek zamanlı puanlama gibi işlevler sunmaktadır (Paek 2005, Segall ve ark. 2005, Martin, 2008). Ayrıca videolar, animasyonlar ve simülasyonlar gibi dinamik özelliklerin kullanılması için önemli bir potansiyel sunarlar. Bu nedenle, BDT'ler önemli etkilere sahip olabilecek ses veya diğer duyu uyaranlarını (Thewall, 2000, Huang ve diğerleri, 2009) kullanmayı mümkün kılmaktadır. Örneğin, sesli ve görüntülü geribildirim, öğrencilerin eksik veya çözülmemiş test sorularını daha iyi anlamalarına yardımcı olur (Martin 2008). BDT'nin başka avantajları ise test süresinin kısa olması, öğrenci yeteneklerinin belirlerken kolaylık sağlaması, maddelerin tekrar tekrar kullanılması, hızlı kodlanabilmesi ve eğitim sistemi hakkında daha fazla bilgi vermesi (Bjainsson 2008).

Günümüzde kullanılan bilgisayar destekli ölçme ve değerlendirme araçlarından Socrative, kullanıcıların istedikleri zamanda hem sınavı gerçekleştirip hem de sonuçlarını anlık olarak görebilecekleri online sınav araçlarındandır. Kahoot sınav aracında öğretmenler soru ve

cevapları hazırlayarak öğrencilerin sınav sonuçlarını başarı sırasına göre görmelerini ve kendi aralarında rekabet oluşturmalarını sağlayan bir ölçme aracıdır. Quizizz, yine öğretmenlerin soruları ve cevapları kendileri bu sisteme girerek öğrencilere uyguladıkları bir sınav uygulama platformudur. Bu sınav platformları tarayıcı üzerinden ücretsiz olarak online kullanılmaktadır. TAO, açık kaynak kodlu çevrimiçi sınav yapmaya olanak sağlayan, birçok soru çeşidi (çoktan seçmeli, eşleştirme, çoklu ortam, sıralama, açık uçlu, doya yükleme, kısa cevaplı sorular, bölge seçmeli sorular, boşluk doldurma, açılır menüden seçerek boşluk doldurma, uzun cevaplı sorular) hazırlanabilen, soru bankası oluşturulabilen, sonuçların rapor olarak alınabildiği bir sınav aracıdır. Katılımcılara “testtaker” adı verilmektedir. Tao diğer sınav platformlarından farklı olarak web tarayıcısı üzerinden değil bilgisayara kurulum yapılarak kullanılmaktadır.

1.5.3. Bilgisayar Ortamında Bireye Uyarlanmış Test (BOBUT)

“Bilgisayar Ortamında Bireye Uyarlanmış Test” (BOBUT) (Computerized Adaptive Testing CAT), bireyin yeteneğine bağlı olarak madde güçlük düzeyinin, bilgisayar ortamında eşleştirilmesidir. Ölçme sürecinde Madde Tepki Kuramı'nın bilgisayar ortamında kullanıldığı ölçme metoduna Bilgisayar Uyarlamalı Test denmektedir.

Rudner (1998), bilgisayar uyarlamalı test sürecinin akışı aşağıdaki gibi belirtmiştir.

1. Soru havuzu içinde bulunan sorulmamış soruların belirlenen yetenek kestirimine en uygun olanı seçilir.
2. Yetenek kestirimine en uygun olarak belirlenen soru sorulur ve testi alanın yanıtı alınır.
3. Tüm cevaplanan sorulara dayanılarak yeni yetenek kestirimi yapılır.
4. Sonlandırma kriterine ulaşılan kadar 1'den 3'e kadar olan adımlar tekrar edilir (Akt. Eroğlu, 2013).

Uyarlamalı testte eşit güvenilirliği sağlamak için genellikle Klasik Test Teorisi kullanıldığı soru havuzunun yarısı kadar maddeye ihtiyaç vardır (Mehrens, 1991). Daha az madde ile benzer güvenilirliğin sağlanması test süresini aynı oranda azaltacaktır.

Günümüzde kullanılan bilgisayar uyarlamalı test örneklerini göz önüne alındığında uygun soru havuzunun hazırlanması ve testin ayarlanması gibi süreçlerin güçlüğü nedeniyle bilgisayar uyarlamalı test daha çok yaygın olarak kullanılan ve merkezi olarak yönetilen geniş ölçekli test uygulamalarında tercih edildiği görülmektedir.

Kullanılan bilgisayar uyarlamalı test örnekleri; A.B.D. savunma bakanlığı tarafından geliştirilen ASVAB (Armed Services Vocational Aptitude Test Battery), Eğitimsel Test Ofisi tarafından geliştirilen işletme doktorasına hazırlık testi GMAT (The Graduate Management Admission Test) ve GRE (the Graduate Record Examination), Vienna Test Sisteminde 15 yaş

üstü kişilere uygulanabilen The Adaptive Matrices Test (AMT), Japoncanın kullanımında yeterliliğini ölçmek için kullanılan J-CAT (Japanese Computerized Adaptive Test) olarak sıralanabilir.

Bilgisayarın eğitimde kullanılmasıyla bireyler hakkında birçok bilgi elde etmek de mümkündür. Bunlardan biri de bireylerin maddeyi cevaplarken harcadığı sürenin ölçülmesidir.

1.6. Tepki Süresi

Tepki süresi, uyarıcının bireye sunulduktan sonra bireyin verilen uyarıcıya tepkisini sonlandırınca kadar geçen süredir.

Doğru yanıtlama süreleri, test geçerliliği ve etkili sınav yapmayı sağlamak için eğitimde geniş kapsamlı değerlendirmelerde test soruları oluşturmak için gereklidir. Kitapçığı hazırlarken yerine getirilmesi gereken temel hedeflerden biri, kitapçığın önceden belirlenmiş test süresi içerisinde hazırlanmasıdır. Bu yüzden, kişilerin bir testi tamamlaması için ihtiyaç duyduğu ortalama zaman olarak tanımlanan test yanıtlama süresinin bilinmesi çok önemlidir. Test yanıtlama süresi çeşitli yollardan elde edilebilir. Test yanıtlama süreleri pilot çalışmalarda doğrudan ölçme ile elde edilebilir. Ancak bu yaklaşım, genellikle zahmetli, zaman alıcı ve masraflıdır. Bunun yerine, test yanıtlama süreleri test hazırlanması ve geliştirilmesi aşamasında çok defa eğitim uzmanları tarafından ölçülür. Ancak, uzmanlar arasındaki doğruluk ve tutarlılık dereceleri düşük olabilir. Alternatif bir diğer seçenek ise, yanıt sürelerini test edilmeden erişilebilen verilerden hesaplamaktır (Örneğin belirli bir testteki sözcük sayısı).

Halkitis, Jones ve Pradhan (1996) madde yanıt süresinin madde güçlüğü, madde ayırt edici özelliği ve bir lisans sınavında sözcük sayımı ile ilişki derecesi üzerinde çalışma yapmıştır. Bütün göstergeler birlikte madde yanıtlama süresinin değişkenin yarısını oluşturduğunu göstermiştir. En güçlü gösterge sözcük sayımı ($R^2 = 27.2\%$), onu takip eden madde güçlüğü ($R^2 = 16.2\%$) ve madde ayırt etme özelliği ($R^2 = 6.8\%$) şeklinde oluşmuştur.

Test soruları hazırlamak kendi başına bir sonuç değildir, daima belirli bir öğrenci kitlesini test etmek amacıyla yapılır. Test soruları farklı alt gruplarda farklı işlev göstereceğinden; yanıtlama süreleri, test soruları tasarlama da değerli bilgiler sağlayabilir. Dolayısıyla, bu bilgi farklı zaman gereksinimleri olan alt grupların ihtiyacına uygun test soruları oluşturmak için faydalıdır. Madde özellikleri hakkında yapılan araştırmalardan daha çok kişiler arası ilişki özellikleri ve yanıtlama süreleri hakkında yapılan araştırmalar üzerinde durmaktadır (Lee ve Chen, 2011; Schnipke ve Scrams, 2002). Buna rağmen, öğrenci karakter özelliklerinin yanıtlama sürelerini nasıl etkilediği, yeteneklerin hesaplanabilmesi için yanıtlama süreleri hesaplamalarını bir yardımcı bilgi kaynağı olarak gören araştırmalar farklı bir açıdan karşımıza çıkıyor (Wang ve Hanson, 2005).

Yapılan çalışmalarda (Klein Entink, Fox, ve van der Linden, 2009) yetenek ile hız arasında olumsuz bir ilişki bulunmuştur; ayrıca yetenekli öğrencilerin daha fazla zaman harcadıklarını (Goldhammer ve Klein Entink, 2011) gösterirken, diğerleri bunun tersini bildirmiştir (Davison ve diğerleri, 2011).

Yanıt süresi ayrıca, uygun zaman sınırlarını belirlemek, hızlanma sorunlarını incelemek, olağandışı test performansını belirlemek ve madde tepki kuramı madde parametrelerini daha doğru bir şekilde düzeltmek için verileri süzmek için kullanılmıştır.

Yanıt süresi, sınav yapanların test edici davranışlarının ana göstergesi olarak sıkça kullanılmıştır (Goldhammer ve ark., 2014). Tepki süresi tipik olarak, bir teste ya da sınava harcanan toplam zaman olarak ifade edilebilir. Bu, sınav yapılan bireyin ilgili görev üzerinde çalışmayı bitirdikten sonra incelediği zamandan çıkarılarak hesaplanır. Görevin araştırma amacına bağlı olarak ne anlam ifade ettiği konusunda küçük farklılıklar olabilir. Bazı araştırmacılar tek bir soruyu bir görev olarak görürken bazıları bir öğrenme değerlendirmesinin bir parçasını bir görev olarak görmüş ve buna göre yanıt sürelerini hesaplamışlardır.

Araştırmacılar, test tasarımını, test yönetimini ve kalite kontrolünü iyileştirmek için tepki süresi analizi yapmaktadırlar. Lee ve Chen'in (2011) ifadesine göre, bilgisayar uyarlamalı testlerin tasarımını iyileştirmek için sorulara yanıt verme süresi incelenmiştir (Van der Linden, 2008). Operasyonel ortamlarda, tepki süresi analizi, tahsis edilen test süresinin sınav yapılan kişilerin acele etmeden sorulara cevap vermesi için yeterli olmasını sağlamak gibi test idarelerine ilişkin kararlar için kullanılmıştır. Bireylerin gerçek yeteneklerini göstermeleri için yeterli test süresi önemlidir. Zamanın kısıtlı olması nedeniyle sınava giren kişiler maddeleri hızlı yanıtlarlar bu da kişilerin gerçek yeteneklerini temsil etmeyebilir ve bu nedenle, alınan puanlar geçerliliği tehdit edebilir (Lu Sireci, 2007).

1.7. Araştırmanın Amacı

Bu araştırmanın amacı, çoktan seçmeli maddelerden oluşan bir ölçme aracında, bireylerin maddeyi yanıtlamak için harcadıkları süre ile Klasik Test Teorisi (KTK) madde istatistikleri (p ve r_{bis}) arasındaki ilişkiyi bireylerin yer aldığı sıralama konumuna göre araştırmaktır.

1.8. Araştırmanın Önemi

Bu çalışmada, madde istatistikleri ve yanıtlama süresi arasındaki ilişki incelenmek istenmektedir. Test oluşturma (assembly) sürecinin önemli bir parçasını da bireylere, testte yer alan maddeleri yanıtlamak için tanınacak olan *zaman* oluşturur. Yanıtlama süresinin belirlenmesinde, madde kökünün uzunluğu ve karmaşıklığının yanında, madde istatistiklerinin de önemsenmesi gerektiği düşünülmektedir. Bu araştırma gerek orijinalliği gerekse çalışmadan üretilecek olan bilginin ölçme değerlendirme alanının test oluşturma süreçlerine getireceği somut katkılarından dolayı önem arz etmektedir.

1.9. Problem

KTK' a göre elde edilen madde istatistikleri ile madde yanıtlama süresi arasında nasıl bir ilişki bulunmaktadır?

1.9.1. Problemler

1. Bireylerin madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

a) Alt yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

b) Orta yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

c) Üst yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

d) Tüm gruptaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

e) Alt ve Üst yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

f) Alt ve Orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

g) Üst ve Orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

2. Bireylerin madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

a) Alt yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

b) Orta yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

c) Üst yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

d) Tüm gruptaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

e) Alt ve Üst yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

f) Alt ve Orta yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

g) Üst ve Orta yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

1.10. Araştırmanın Sayılıtları

Bu araştırma aşağıdaki sayılıtlar üzerine temellendirilmiştir:

- Testi alan bireyler yanıtlama sürecinde maksimum performanslarını ortaya koymuşlardır.
- Testi alan bireyler yanıtlama sürecini içtenlik ve ciddiyetle yapmışlardır.
- Öğrencilerin bilgisayar kullanma becerileri yanıtlama sürecini etkilememiştir.

1.11. Araştırmanın Kapsam ve Sınırlılıkları

- 1-Araştırma İngilizce kelime bilgisi alt testi ile sınırlıdır.
- 2- Araştırma Klasik Test Teorisi ile sınırlıdır.
- 3- Araştırma Klasik Test Teorisi'nin madde istatistiklerinden olan madde güçlük indeksi ve madde ayırt edicilik gücü indeksi ile sınırlıdır.
- 4- Araştırma Bilgisayar Destekli Testler (BDT) ile sınırlıdır.

2. İLGİLİ ARAŞTIRMALAR

İmamoğlu (2007), 15 kişiden oluşan 3 gruba İngilizce testini Kâğıt kalem testi ve Bilgisayar Destekli testi olarak uygulamış ve bireylerin başarıları ve sınav bitirme sürelerini karşılaştırmıştır. Öğrencilerin İngilizce ve bilgisayar bilgilerinin eşit olduğu varsayılmış. İngilizce testi bir gruba kâğıt kalem testi olarak uygulanırken iki gruba bilgisayar ortamında uygulamış. Araştırmada sınavlar arasında fark olmadığı, sınav bitirme süresi olarak incelendiğinde öğrencilerin kâğıt kalem testini daha kısa sürede tamamladıkları elde edilmiştir. Bilgisayar destekli test sonuçlarını etkileyecek 3 değişken belirleyerek bunlar arasındaki korelasyonu incelemiştir.

Akdemir ve Oğuz (2008), iki farklı bölümlerde okuyan 43 öğrenci üzerinde yaptığı araştırmada öğrenci puanlarının kâğıt-kalem testi ve bilgisayar ortamında uygulanan test arasında farklılık olup olmadığını incelemiştir. Araştırmada ölçme ve değerlendirme dersi için 30 maddelik çoktan seçmeli başarı testi kullanılmış. Analiz sonucunda öğrencilerin her iki formdan elde ettikleri puanlar arasında anlamlı bir fark olmadığı elde edilmiştir.

Kezer ve Koç (2014), 1166 öğrenci ile yaptıkları çalışmada bilgisayar ortamında bireye uyarlanmış test (BOBUT) yöntemi ile kâğıt kalem testini karşılaştırmıştır. Ölçme aracı olarak İngilizce kelime testi kullanılmış. Bireylerin yetenek kestirimlerinde madde sayısı ve süre olarak BOBUT yönteminin daha avantajlı olduğu elde edilmiştir.

Ayberk (2012), 136 öğrenci üzerinde yaptığı çalışmada 60 maddelik genel yetenek testini hem kâğıt kalem testi hem de bilgisayar ortamında uygulamış ve testin psikolojik özelliklerini, cinsiyete göre ve öğrenci performansına göre karşılaştırmıştır. Öğrencilere aynı zaman da bilgisayara yönelik tutum ölçeği ve bilgisayar aşinalığı ölçeği de uygulamış cinsiyete göre karşılaştırmıştır. Araştırma sonucunda kâğıt kalem testi ve bilgisayar ortamında uyguladığı test puanlarının benzer, bilgisayar aşinalığı ve cinsiyete göre bilgisayar ortamından elde etti puanla örtüştü bulguları elde ederken, tutum puanları ile bilgisayar ortamından elde ettiği puanların farklı çıkmıştır.

Vein, Bergstrom, Gershon, ve Lunz (1994), madde metni uzunluğu, (göreceli) madde zorluğu, madde sırası ve doğru cevap düzeninin (çoktan seçmeli sorularda) birbiriyle ilişkili göstergeler olarak tanımlamaktadır. Dahası, ayrı resimli kitapçık uygulamalarından kaynaklanmış olabilmesine rağmen, bir figürün varlığı tepki süreleri üzerinde güçlü bir etkiye sahip olmuştur. Tıp lisansı sınavından elde edilen verilere göre madde yanıtlama süresi değişken oranının yaklaşık %45'i madde zorluğu, resim varlığı/yokluğu ve sözcük sayısı ile açıklanmıştır. Araştırmacılar, madde zorluğundaki logit değişkenin +14 saniye, resim kullanımının +12 saniye ve her sözcüğün yaklaşık olarak 0.5 saniye eklediğini bildirmiştir (Akt. Swanson ve diğerleri, 2001).

Hornke (2000), bilgisayar ortamında bireye uyarlanmış testte (CAT), yanıtlama süresi bilgisinden ne gibi ilave anlamların çıkarılabileceğini araştırmış. Tepki süreleri ile test puanlarının ilişkisini detaylı olarak incelemek için, genel zihinsel yeteneklerin bilgisayar ortamında bireye uyarlanmış test ile yapılan daha geniş bir çalışmadan elde edilen veriler kullanmıştır. Bilgisayar ortamında bireye uyarlanmış testte $n = 5912$ Alman askeri gençten elde edilen verileri kullanmıştır. Maddeler bireysel performansa göre 456 maddelik büyük bir havuzdan alınmıştır. Herhangi bir maddeye cevap verme süresi 180 saniye ile sınırlamıştır ve sabit bir test süresi yoktur. Bununla birlikte, teoride maksimum 30 maddeyle test süresi 90 dakika ile sınırlandırılmıştır. Her bir adımda test sırasında en fazla bilgiyi sağlayacak öğeyi bulmak için madde parametreleri, madde gücü ve ayırt ediciliği kullanılmıştır. 0,38 veya daha düşük bir standart ölçüm hatasına (SEM) ulaşıldığında test durdurulmuştur. Bu, her durumda ve herhangi bir puan düzeyinde 0.85 güvenilirliğine eşdeğerdir. 1930'ların başından beri, cevap süreleri puanlardan farklılaştırılması gereken kişilik özelliklerinin göstergesi olarak kabul edilmiştir. Bu fikir tartışılmış, bu fikre destekleyen ve karşı çıkan argümanlar sunmuştur. Ayrıca güncel modelleme yaklaşımlarını önermiştir.

Wise ve Kong (2005), düşük riskli değerlendirmeler uygulandığında, sınava girenlerin en iyi çabayı gösterme derecesi genellikle belirsizdir ve sonuçta ortaya çıkan test puanlarının geçerliliği ve yorumlanması karmaşılaşır. Bu çalışma, bilgisayar destekli test maddelerinde sınava girme çabalarını ölçmek için madde yanıt süresine dayalı yeni bir yöntem sunmaktadır. Yanıt süresi çabası (RTE) olarak adlandırılan bu ölçüm, bir madde okunduğunda, motive edilmeyen sınav katılımcılarının çok hızlı bir şekilde cevap vereceği (yani, maddeyi okumak ve tam olarak değerlendirmek için zamanları olmadan) çok fazla cevap vereceği hipotezine dayanmaktadır. RTE puanlarının psikometrik özellikleri ampirik olarak incelenmiş ve puan güvenilirliği ve geçerliliği için destekleyici kanıtlar bulunmuştur. RTE puanlarının potansiyel uygulamaları ve sonuçları tartışılmaktadır.

Van der Linden (2008), eğitimsel ve psikolojik testlerde anormal tepki-zaman ikilisini belirlemek için, sınava girenlerin çalışma hızını öğelerin istediği zamandan ayırmanın önemli olduğunu belirtmiş. Anormal tepki sürelerini tespit etmeye yönelik bir Bayesian prosedürünü türetmek için bu özelliğe sahip tepki süreleri için lognormal bir model kullanılmıştır. Ayrıca, hiyerarşik bir çerçevede normal cevap modeliyle cevap-zaman modelinin bir kombinasyonu, teste girenlerin hızı hakkındaki teminat bilgisinin cevap vektörlerinden elde edildiği, anormal cevap zamanlarının tespiti için alternatif bir prosedürde kullanılmıştır. Prosedürler Graduate Management Admission Test® (GMAT®) için bir veri seti kullanılarak gösterilmektedir. Ek olarak, uyarlamalı bir testte benzetilmiş hile yapma davranışı kullanılarak bir güç çalışması yapılmıştır.

3. YÖNTEM

3.1. Yöntem

Bu bölümde araştırmanın problemine uygun olarak seçilen araştırma modeli, araştırma grubu, veri toplama araçları ve verilerin analiz edilmesi ile ilgili bilgiler alt başlıklarla verilmiştir.

3.2. Araştırmanın Modeli

Bu çalışmada, İngilizce kelime testinden elde edilen madde istatistikleriyle öğrencilerin maddeleri yanıtlama süresi ilişkilendirilmeye çalışıldığından söz konusu araştırma ilişkisel (korelatif) bir araştırma olarak değerlendirilebilir.

3.3. Araştırma Grubu

Araştırma için kullanılan kelime bilgisi testinin deneme uygulaması, 2017-2018 eğitim-öğretim yılında Mersin Üniversitesi Eğitim Fakültesinde öğrenim gören 166 öğrenci üzerinde yapılmıştır. Asıl uygulamaya ilişkin veriler ise aynı dönemde Mersin Üniversitesi Eğitim Fakültesinde yabancı dil dersi alan amaçlı örnekleme yöntemi ile belirlenen toplam 212 öğrencilerden elde edilmiştir.

3.4. Veri Toplama Araçları

Araştırma kapsamında ölçme aracı olarak kelime bilgisi testi kullanılmıştır. Kullanılan kelime bilgisi testini oluşturmak için şu aşamalar izlenmiştir. Teste yer alan yabancı dil alanındaki maddeler Yabancı dil dersinde okutulan Network Starter Student Book kitabındaki kelimelerden oluşturulmuştur. Belirlenen maddeler farklı uzmanlar tarafından tekrar incelenmiş ve maddeler üzerinde gerekli düzeltmeler yapılmıştır. Deneme amaçlı 30 maddeden oluşan kelime bilgisi testi, Ölçme ve Değerlendirme Merkezi Bilgisayar Laboratuvarı'nda TAO sınav platformuna aktarılmıştır. Kelime bilgisi testi Eğitim Fakültesinde öğrenim gören 166 öğrenciye uygulanmıştır. Elde edilen veriler TAO veri tabanından alınarak düzenlenmiştir. Maddelere ilişkin madde ve test istatistiklerinin belirlenmesi aşamasında ITEMAN 3.5 programı kullanılmıştır. Deneme uygulamasında yer alan madde güçlük indeksi ve madde ayırt edicilik indeksine ilişkin bilgiler Tablo 3.1'de sunulmuştur. Hesaplamalar sonucunda madde ayırt edicilik indeksi 0.30'un altında olan 6 madde tespit edilmiştir. Geriye kalan maddelerin yeterli

ayrıt edicilik indekslerine sahip olduğu gözlenmiştir. Madde güçlük indeksi 0.15-0.91 arasında değişmektedir.

Tablo 3.1.
Deneme Uygulamasındaki Maddelere İlişkin Madde Ayırıt Edicilik İndeksleri ve Madde Güçlük İndeksleri

Kelime Bilgisi Testi					
Madde No	p	r_{bis}	Madde No	p	r_{bis}
1	0.17	0.02	16	0.30	0.08
2	0.69	0.01	17	0.86	0.68
3	0.58	0.63	18	0.62	0.63
4	0.27	0.45	19	0.55	0.50
5	0.58	0.58	20	0.27	0.12
6	0.61	0.46	21	0.31	0.25
7	0.38	0.73	22	0.54	0.47
8	0.80	0.74	23	0.15	0.17
9	0.64	0.59	24	0.34	0.35
10	0.73	0.41	25	0.21	-0.02
11	0.69	0.68	26	0.87	0.77
12	0.52	0.72	27	0.91	0.74
13	0.81	0.54	28	0.16	0.44
14	0.77	0.48	29	0.74	0.70
15	0.28	0.50	30	0.62	0.72

Deneme uygulamasına ait test istatistikleri ise Tablo 3.2'de yer almaktadır.

Tablo 3.2.
Deneme Uygulamasına Ait Test İstatistikleri

Kelime Bilgisi Testi	
Madde Sayısı	30
Katılımcı sayısı	166
Ortalama	15.988
Varyans	21.325
Standart Sapma	4.618
Çarpıklık	0.043
Basıklık	-0.175
Minimum	5
Maksimum	30
Medyan	17
KR 20	0.751
Ölçmenin Standart Hatası	2.304
Ortalama Güçlük	0.533

Tablo 3.2 incelendiğinde ortalama ve medyanın birbirine yakın olduğu, çarpıklık ve basıklık katsayılarının 0.00 ile 1.00 arasında değer aldığı görülmektedir. Güvenirlik katsayısı ise

0.751 olarak bulunmuştur. Güvenirlik katsayısı çok yüksek olmamakla birlikte testin yeterli güvenilirliğe sahip olduğu söylenebilir. Ortalama güçlüğü 0.533 olarak bulunan testin, uygulanan gruba göre orta güçlükte olduğu söylenebilir.

Asıl kelime bilgisi testi, Ölçme ve Değerlendirme Merkezi Bilgisayar Laboratuvarı'nda TAO sınav platformunda, Eğitim Fakültesi'nde öğrenim gören ve Yabancı Dil dersi alan 212 öğrenciye uygulanmıştır. Elde edilen veriler TAO veri tabanından alınarak düzenlenmiştir. Teste yer alan maddelere ilişkin madde ve test istatistiklerinin belirlenmesinde ITEMAN 3.5 programı kullanılmıştır. Asıl test uygulamasında yer alan madde güçlük indeksi ve madde ayırt edicilik indeksine ilişkin bilgiler Tablo 3.3 'te sunulmuştur.

Tablo 3.3.
Asıl Kelime Bilgisi Testi Uygulamasındaki Maddelere İlişkin Madde Ayırt Edicilik İndeksleri ve Madde Güçlük İndeksleri

Kelime Bilgisi Testi					
Madde No	p	r_{bis}	Madde No	p	r_{bis}
1	0.26	0.32	16	0.95	0.72
2	0.78	0.04	17	0.92	0.47
3	0.53	0.63	18	0.67	0.72
4	0.27	0.30	19	0.49	0.36
5	0.56	0.43	20	0.26	0.10
6	0.49	0.40	21	0.34	0.18
7	0.32	0.50	22	0.48	0.28
8	0.71	0.61	23	0.26	0.63
9	0.67	0.55	24	0.51	0.53
10	0.74	0.46	25	0.27	0.15
11	0.63	0.75	26	0.82	0.58
12	0.48	0.73	27	0.89	0.74
13	0.75	0.62	28	0.20	0.29
14	0.79	0.68	29	0.79	0.61
15	0.24	0.34	30	0.74	0.51

Asıl uygulamaya ait test istatistikleri Tablo 3.4'te yer almaktadır.

Tablo 3.4.
Asıl Uygulamaya Ait Test İstatistikleri

Kelime Bilgisi Testi	
Madde Sayısı	30
Katılımcı sayısı	212
Ortalama	16.807
Varyans	20.637
Standart Sapma	4.543
Çarpıklık	-0.048
Basıklık	-0.483
Minimum	4
Maksimum	28
Medyan	17
KR 20	0.742
Ölçmenin Standart Hatası	2.306
Ortalama Güçlük	0.560

Tablo 3.4 incelendiğinde ortalama ve medyanın birbirine yakın olduğu, çarpıklık ve basıklık katsayılarının 0.00 ile 1.00 arasında değer aldığı görülmektedir. Güvenirlik katsayısı ise 0.742 olarak bulunmuştur. Ortalama güçlüğü 0.560 olarak bulunan asıl testin, uygulanan gruba göre orta güçlükte olduğu söylenebilir.

212 Öğrencinin puanları, puanları en düşükten en yükseğe sıralandıktan sonra ilk %27'sini oluşturan 64 kişi alt yetenek grubu, son %27' sini oluşturan 66 kişi üst yetenek grubu, geriye kalan 82 kişi ise orta yetenek grubu olarak belirlenmiştir.

Madde kökünde yer alan kelime sayısının karıştırıcı etki yaratmaması için mümkün olduğu kadar eşit olmasına dikkat edilmiştir.

3.5. Uygulama

Kelime Bilgisi Testi olarak oluşturulan ölçme aracı Mersin Üniversitesi Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi Bilgisayar Laboratuvarı'nda yer alan bilgisayarlarda Bilgisayar Destekli Test sistemi olan TAO sınav platformuna yüklenmiştir. 45 öğrenci kapasitesi olan bu laboratuvarda kelime bilgisi testi farklı oturumlarda uygulanmıştır. Uygulama yapılmadan önce bilgisayar laboratuvarında gerekli düzenlemeler yapılmıştır. Öğrenciler sınav salonuna alındıktan sonra sınavla ilgili bilgiler verilip gerekli uyarılar yapılmıştır. TAO programında önceden sınava girecek öğrencilerin okul numaraları tanımlanmış olup adaylar okul numaralarını kullanıcı adı olarak girmiştir ve şifreleri de tarafımızca verilmiştir. Öğrenciler aynı anda kullanıcı adı ve şifreleri ile giriş yaptıktan sonra ekranda çıkan "Teste Başla" butonuna basmalarıyla cevaplamaları gereken maddeler ekrana

gelmiştir. Adaylara maddeleri cevaplamaları için 30 dakika süre verilmiştir. Sınav sonunda ekranda çıkan "Sınavı Bitir" butonu ile sınavı sonlandırmışlardır.

3.6. Verilerin Analizi

Araştırmadan elde edilen veriler, ITEMAN 3.5, Microsoft Office Excel 2007 ve R programlama dili yazılımları ile analiz edilmiştir.

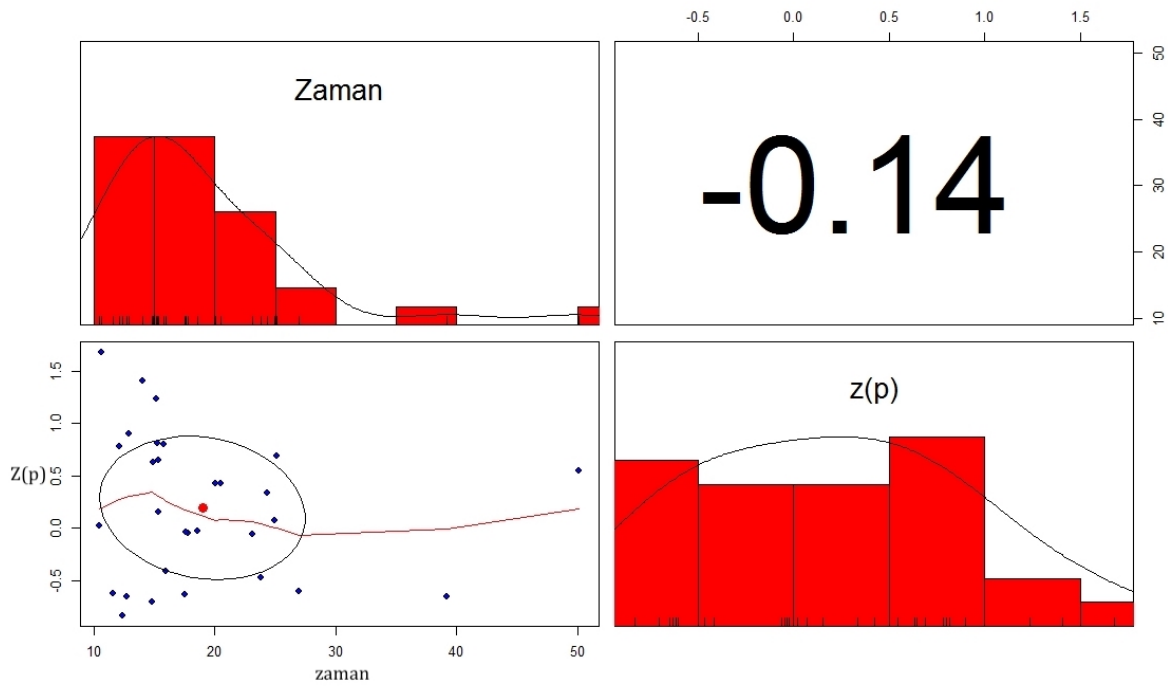
Bilgisayar ortamından elde edilen madde yanıtlama süreleri R programlama dili kullanılarak düzenlenmiştir. Bireylerin madde tepkilerinden hareketle her madde için madde güçlük indeksi ve madde ayırt edicilik indeksi hesaplanmıştır. Korelasyon hesaplanmadan önce p değerleri Z değerine dönüştürülmüştür. Buna ek olarak madde toplam test korelasyonu olarak kullanılan r_{bis} değerleri Fisher Z değerlerine dönüştürülmüştür. Bireyler testten aldıkları puanlara göre sıralandıktan sonra %27 gruplara göre alt-orta-üst gruplara ayrılmıştır. Gruplara ayrılan bireylerin her bir madde için harcadıkları ortalama süre hesaplanmıştır. Elde edilen ortalama süre ile madde güçlük indeksi ve madde ayırt edicilik indeksi arasında korelasyon katsayısı hesaplanan korelasyon katsayıları Fisher'in Z' sine dönüştürülmüştür. İkili gruplar için Fisher' in Z'sine dönüştürülen korelasyon katsayıları arasındaki farkın manidar olup olmadığı hesaplanmıştır. Korelasyonlar arası fark ise R' da yazılan Fisher Z testine göre gerçekleştirilmiştir.

4. BULGULAR ve YORUM

1. Bireylerin madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

a) Alt yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

Alt yetenek grubunun; maddelere verdikleri madde yanıtlama süreleri ile madde güçlük indeksleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.1'de belirtilmiştir.

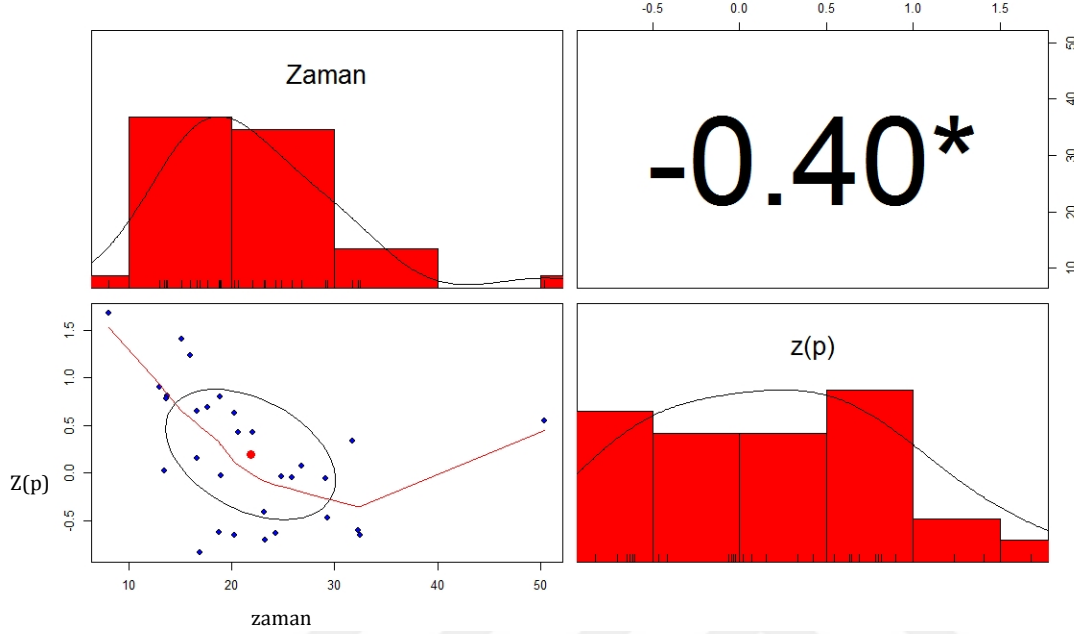


Şekil 4.1. Alt Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.1'de yer alan alt yetenek grubundaki bireylerin, maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında korelasyon katsayısı -0.14 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.05 manidarlık düzeyinde anlamlı olmadığı görülmüştür. Alt yetenek grubundaki bireylerin, maddeleri yanıtlama süreleri ile madde güçlük düzeyleri arasında anlamlı bir ilişkinin bulunmadığı söylenebilir. Buna göre kolay, orta ve zor maddelere daha fazla zaman ayırmıştır ya da daha az zaman ayırmıştır çıkarımı yapılamamaktadır.

b) Orta yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

Orta yetenek grubunun; maddelere verdikleri madde yanıtlama süreleri ile madde güçlük indeksleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.2' de belirtilmiştir.

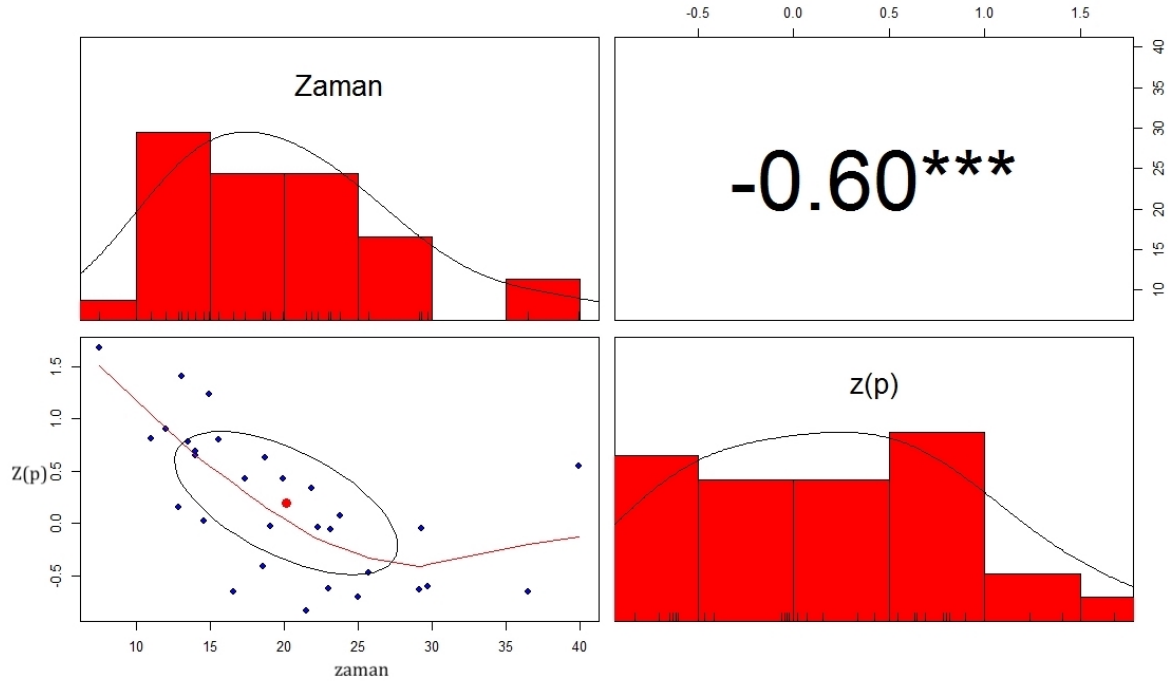


Şekil 4.2. Orta Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.2' de yer alan orta yetenek grubundaki bireylerin, maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında korelasyon katsayısı -0.40 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.05 manidarlık düzeyinde anlamlı olduğu görülmüştür. Orta yetenek grubundaki bireylerin, maddeleri yanıtlama süreleri, madde güçlük düzeylerine göre anlamlı olarak değişim göstermektedir. Orta yetenek grubundaki bireylerin kolay maddelere daha az zaman ayırırken zor maddelere daha fazla zaman ayırdığı gözlenmiştir.

c) Üst yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

Üst yetenek grubunun; maddelere verdikleri madde yanıtlama süreleri ile madde güçlük indeksleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.3' te belirtilmiştir.

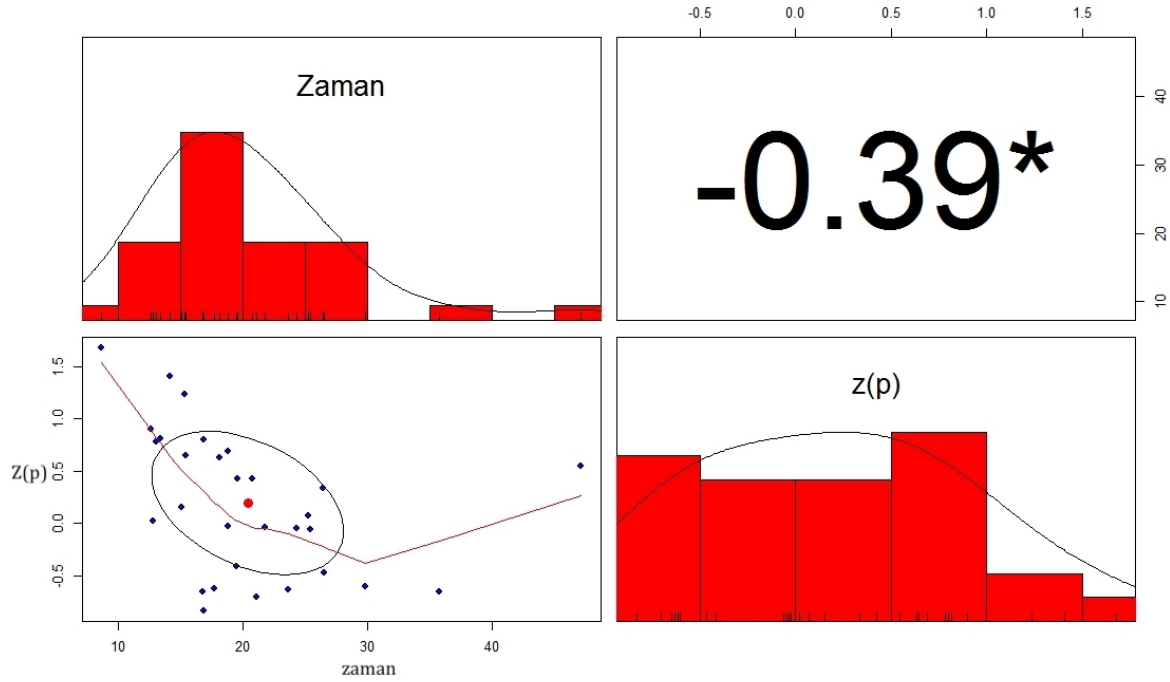


Şekil 4.3. Üst Yetenek Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.3' te yer alan üst yetenek grubundaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında korelasyon katsayısı -0.60 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.001 manidarlık düzeyinde anlamlı olduğu görülmüştür. Üst yetenek grubundaki bireylerin, maddeleri yanıtlama süreleri, madde güçlük düzeylerine göre anlamlı olarak değişim göstermektedir. Üst yetenek grubundaki bireyler kolay maddeleri daha kısa zamanda çözerken, zor maddelerde daha fazla zaman harcadığı gözlemlenmiştir.

d) Tüm gruptaki bireylerin; madde yanıtlama süreleri ile madde güçlük indeksleri arasında nasıl bir ilişki vardır?

Tüm grubun; maddelere verdikleri madde yanıtlama süreleri ile madde güçlük indeksleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.4'te belirtilmiştir.



Şekil 4.4. Tüm Grubunun; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Güçlük İndeksleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.4' te yer alan tüm gruptaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında korelasyon katsayısı -0.39 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.05 manidarlık düzeyinde anlamlı olduğu görülmüştür. Tüm grubun, maddeleri yanıtlama süreleri, madde güçlük düzeylerine göre anlamlı olarak değişim gösterdiği söylenebilir.

e) Alt ve Üst yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Alt ve üst yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında korelasyon katsayıları sonuçları Tablo 4.1'de verilmiştir.

Tablo 4.1.
Alt ve Üst Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları

	r	N	Z	P
Alt	-0.14	30	-2.028	0.042
Üst	-0.60	30		

Alt ve üst yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde güçlük indeksleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir farkın olduğu saptanmıştır ($Z_{30} = -2.028$; $p < .05$).

f) Alt ve Orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Alt ve orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında korelasyon katsayıları sonuçları Tablo 4.2’de verilmiştir.

Tablo 4.2.
Alt ve Orta Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları

	r	N	Z	p
Alt	-0.14	30	-1.038	0.298
Orta	-0.40	30		

Alt ve orta yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde güçlük indeksleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir fark bulunamamıştır ($Z_{30} = -1.038$; $p > .05$).

g) Üst ve Orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında hesaplanan korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Üst ve orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında korelasyon katsayıları sonuçları Tablo 4.3’te verilmiştir.

Tablo 4.3.
Üst ve Orta Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Güçlük İndeksleri Arasında İlişki Farkı Sonuçları

	r	N	Z	p
Üst	-0.60	30	0.990	0.322
Orta	-0.40	30		

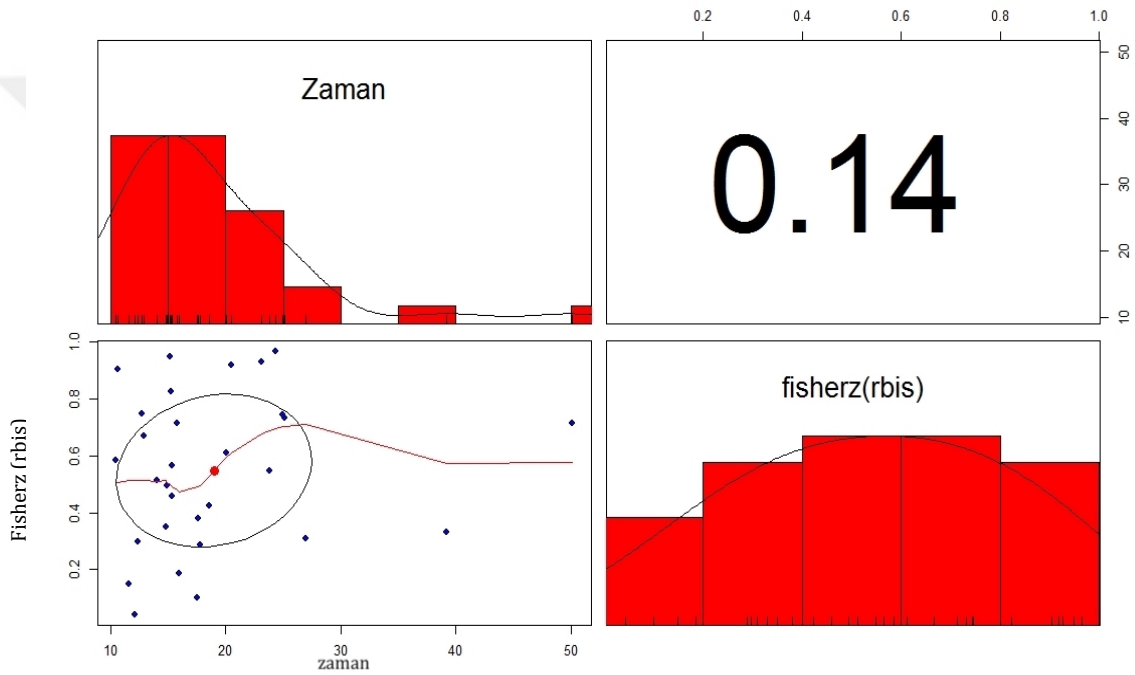
Üst ve orta yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde güçlük indeksleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir fark bulunamamıştır ($Z_{30} = -0.990$; $p > .05$).

2. Bireylerin madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

a) Alt yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

Alt yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.5'te belirtilmiştir.

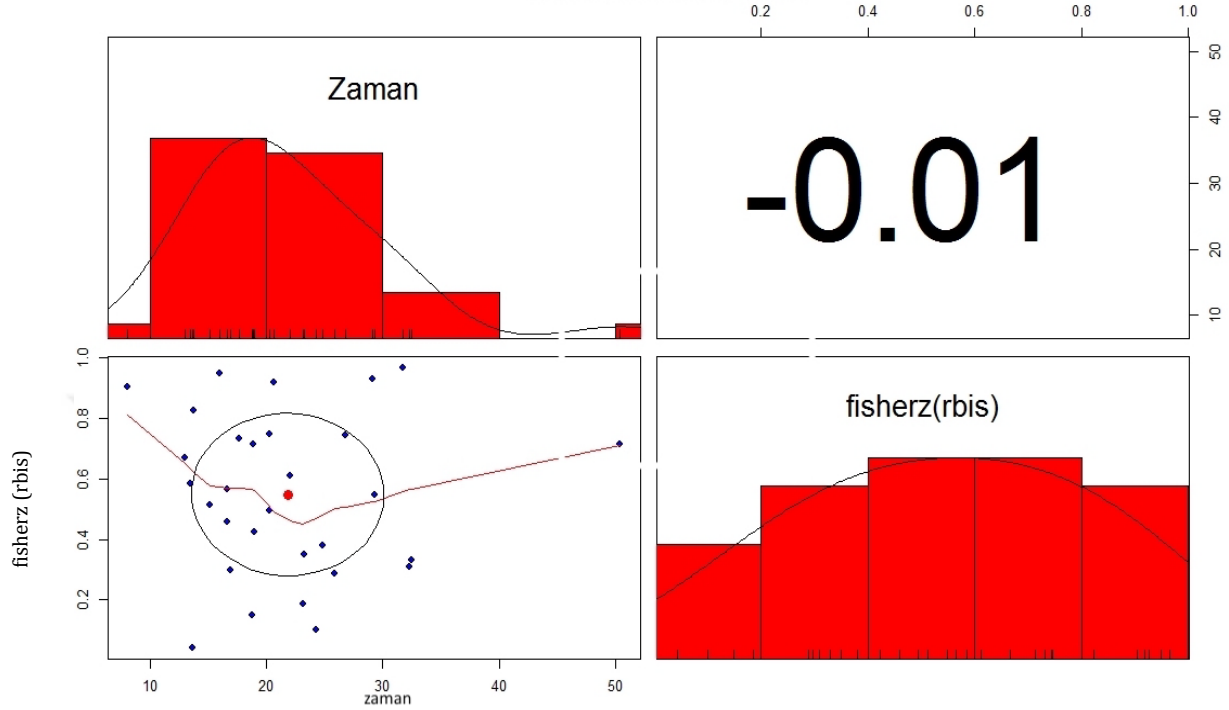


Şekil 4.5. Alt Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.5' te yer alan alt yetenek grubundaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde ayırt edicilik düzeyleri arasında korelasyon katsayısı 0.14 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.05 manidarlık düzeyinde anlamlı olmadığı görülmüştür. Alt yetenek grubundaki bireyler, madde ayırt edici ise daha fazla zaman harcıyor ya da madde ayırt edici değil ise daha az zaman harcıyor diye bir çıkarım yapılamamıştır.

b) Orta yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

Orta yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.6' da belirtilmiştir.

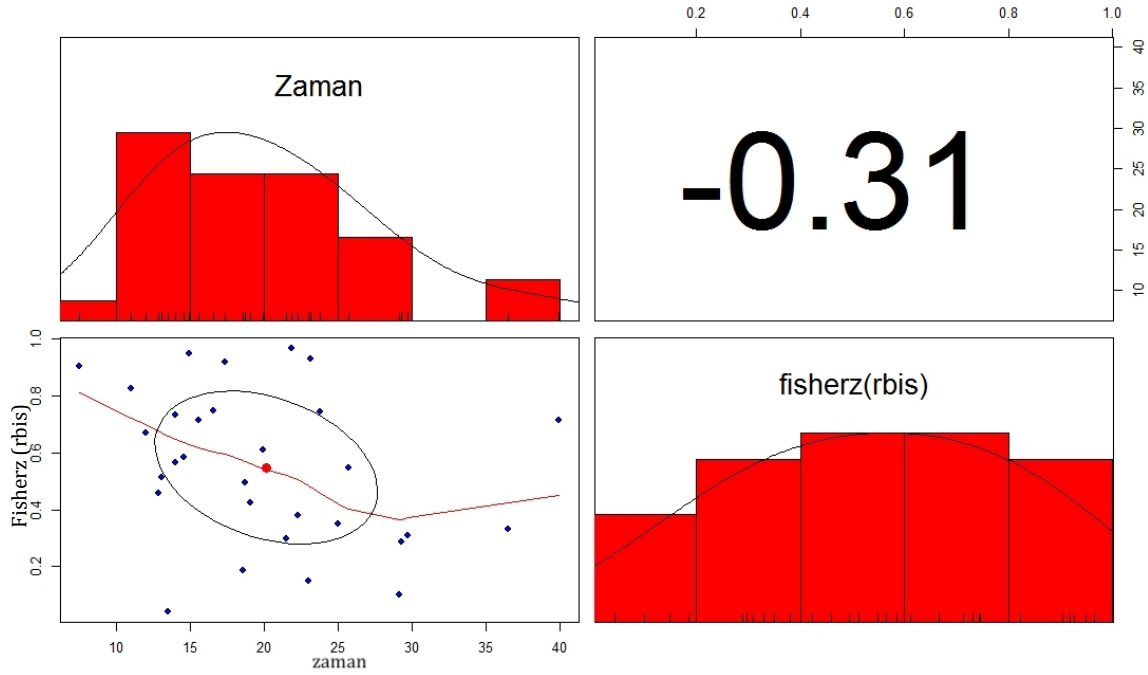


Şekil 4.6. Orta Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.6' da yer alan orta yetenek grubundaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde ayırt edicilik düzeyleri arasında korelasyon katsayısı -0.01 olarak elde edilmiştir. Elde edilen analiz sonucunun 0.05 manidarlık olmadığı görülmüştür. Orta yetenek grubundaki bireyler, madde ayırt edici ise daha fazla zaman harcıyor ya da madde ayırt edici değil ise daha az zaman harcıyor diye bir çıkarım yapılamamıştır.

c) Üst yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

Üst yetenek grubundaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.7' de belirtilmiştir.

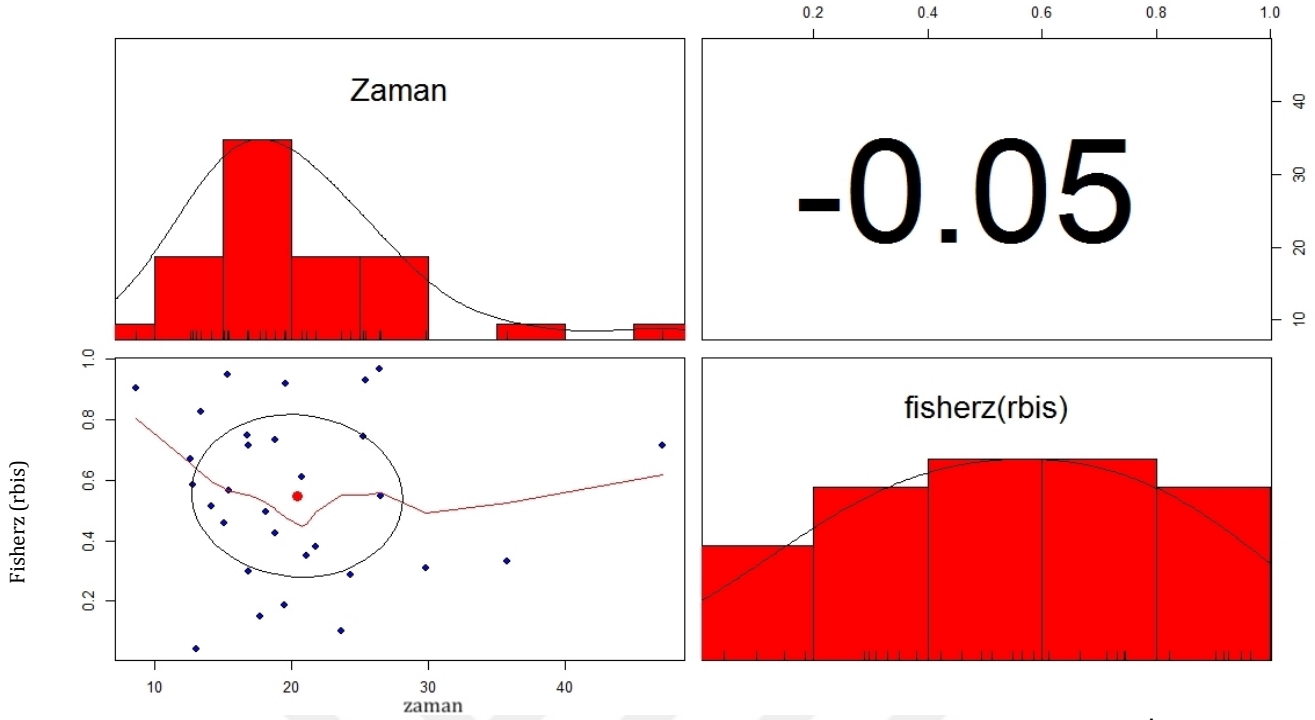


Şekil 4.7. Üst Yetenek Grubundaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.7' de yer alan üst yetenek grubundaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde ayırt edicilik düzeyleri arasında korelasyon katsayısı -0.31 olarak elde edilmiştir. Elde edilen analiz sonucunda ilişkinin 0.05 manidarlık düzeyinde anlamlı olmadığı görülmüştür. Üst yetenek grubundaki bireyler, madde ayırt edici ise daha fazla zaman harcıyor ya da madde ayırt edici değil ise daha az zaman harcıyor diye bir çıkarım yapılamamıştır.

d) Tüm gruptaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında nasıl bir ilişki vardır?

Tüm gruptaki bireylerin; madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında anlamlı bir ilişki gösterip göstermediğine ilişkin sonuçlar Şekil 4.8' de belirtilmiştir.



Şekil 4.8. Tüm Gruptaki Bireylerin; Maddelere Verdikleri Madde Yanıtlama Süreleri İle Madde Ayırt Edicilik Düzeyleri Arasında İlişkinin Analiz Sonuçları

Şekil 4.8' de yer alan tüm gruptaki bireylerin maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde ayırt edicilik düzeyleri arasında korelasyon katsayısı -0.05 olarak elde edilmiştir. Elde edilen analiz sonucunda ilişkinin 0.05 manidarlık düzeyinde anlamlı olmadığı görülmüştür. Tüm gruptaki bireyler için, madde ayırt edici ise daha fazla zaman harcıyor ya da madde ayırt edici değil ise daha az zaman harcıyor diye bir çıkarım yapılamamıştır.

Üst, Orta, Alt ve Tüm gruplara bakıldığında madde yanıtlama süreleri ile madde ayırt edicilik indeksi arasında anlamlı bir ilişki bulunamamıştır.

e) Alt ve Üst yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Alt ve Üst yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri arasında korelasyon katsayıları sonuçları Tablo 4.4'te verilmiştir.

Tablo 4.4.
Alt ve Üst Yetenek Grupları İçin Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları

	r	N	Z	p
Alt	0.14	30		
Üst	-0.41	30	-1.695	0.0899

Alt ve üst yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir fark bulunamamıştır ($Z_{30} = -1.695$; $p > .05$).

f) Alt ve Orta yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Alt ve orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında korelasyon katsayıları sonuçları Tablo 4.5'te verilmiştir.

Tablo 4.5.

Alt ve Orta Yetenek Grupları için Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları

	r	N	Z	p
Alt	0.14	30	-0.554	0.579
Orta	-0.01	30		

Alt ve orta yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir fark bulunamamıştır ($Z_{30} = -0.554$; $p > .05$).

g) Üst ve Orta yetenek grupları için madde yanıtlama süreleri ve madde ayırt edicilik düzeyleri korelasyon katsayıları arasında anlamlı bir fark var mıdır?

Üst ve orta yetenek grupları için madde yanıtlama süreleri ve madde güçlük indeksleri arasında korelasyon katsayıları sonuçları Tablo 4.6'da verilmiştir.

Tablo 4.6.

Üst ve Orta Yetenek Grupları için Madde Yanıtlama Süreleri ve Madde Ayırt Edicilik Düzeyleri Arasında İlişki Farkı Sonuçları

	r	N	Z	p
Üst	-0.31	30	-1.141	0.253
Orta	-0.01	30		

Üst ve orta yetenek grubundaki bireylerin, madde yanıtlama süreleri ile madde ayırt edicilik düzeyleri arasında hesaplanan korelasyon katsayı değerleri anlamlılığı araştırılmış ve karşılaştırma için Fisher Z testi kullanılmıştır.

Bulunan korelasyon katsayıları arasında anlamlı bir fark bulunamamıştır ($Z_{30} = -1.141$; $p > .05$).

5. TARTIŞMA SONUÇ ve ÖNERİLER

5.1. Tartışma ve Sonuç

Araştırmada, kelime bilgisi testi bilgisayar ortamına aktarılarak, farklı yetenek düzeyine ait öğrencilerin madde yanıtlama süresi ile madde istatistikleri arasındaki ilişki incelenmiştir.

Alt yetenek grubundaki bireylerin, maddelere ilişkin madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında anlamlı bir ilişki bulunmaz iken yetenek grubu farklılaştığında yani orta ve üst grup bireylerde madde yanıtlama süreleri ile maddelerin madde güçlük indeksleri arasında anlamlı ilişkiler gözlenmiştir.

Bireylerin madde yanıtlama süreleri ile madde ayırt edicilik indeksi arasında ilişki incelendiğinde anlamlı bir ilişki bulunamamıştır. Farklı gruplar (üst, orta, alt) incelendiğinde aynı sonuçlar elde edilmiştir.

Test hazırlama süreçlerinde birçok araştırmada yanıtlama süresinin belirlenmesinde; en güçlü gösterge olarak madde kökünün uzunluğu gösterilirken, onu takiben madde güçlük indeksi sonra ise madde ayırt edicilik düzeyi yer almaktadır (Halkitis, Jones ve Pradhan 1996). Bu çalışmada ise madde kökünün eşit olmasına özen gösterilmiş, madde güçlüğü olarak bakıldığında çalışmalarla benzerlik gösterirken, ayırt edicilik düzeyi ile süre arasında ilişki bulunamamıştır.

Üst yetenek grubundaki bireyler kolay maddeleri daha kısa zamanda çözerken, zor maddelerde daha fazla zaman harcadığı gözlemlenmiştir. Üst yetenek grubundaki bireyler testi daha uzun sürede tamamlamıştır. Elde edilen bu bulgu, (Goldhammer ve Klein Entink, 2011; Klein Entink, Fox, ve van der Linden, 2009); yaptığı çalışma ile benzerlik göstermektedir. Goldhammer ve Klein Entink (2011) ve Klein Entink, Fox, ve van der Linden (2009) yaptıkları araştırmalarda yetenekli öğrencilerin daha fazla zaman harcadıklarını gözlemlemişlerdir.

Elde edilmiş olan korelasyon katsayılarının gruplar arasında değişip değişmediğini incelemek üzere yapılan Fisher Z karşılaştırmalarında ise sadece alt ve üst grupların cevaplama süreleri ile madde güçlükleri arasında hesaplanan korelasyonların anlamlı olarak farklılaştığı görülmüştür. Diğer fark karşılaştırmalarında anlamlı bir farklılık elde edilememiştir.

5.2. Öneriler

- Araştırma İngilizce kelime bilgisi testi üzerine kurgulanmıştır. Farklı alanlardaki ders içeriklerine göre benzer araştırmalar yürütülebilir.

- Araştırmada bireylerin bilgisayar kullanma becerileri yanıtlama sürecini etkileyebileceğinden, bilgisayar becerisi gelişmiş ve az gelişmiş bireyler üzerinde bir araştırma yapılabilir.
- Maddelere görseller eklenerek tepki süresi incelenebilir.
- Daha az örnekleme kağıt-kalem testi doğrudan ölçüm yapılabilir ve bilgisayar ölçümü arasında fark olup olmadığı incelenebilir.
- Şans başarısı dikkate alınmamıştır. Şans başarısı dikkate alınarak bir çalışma yürütülebilir.
- Yetenek gruplarının ortalama cevaplama süreleri karşılaştırılabilir.



KAYNAKLAR

- [1]. Baykul, Y. (25-27 Kasım 1999). Ülkemizde eğitimde ölçme ve değerlendirme uygulamaları hakkında bazı düşünceler, *V.21.Yüzyılın Eşiğinde Türk Eğitim Sistemi Ulusal Sempozyumu Bildiri Kitapçığı*. Ankara: Hacettepe Üniversitesi Eğitim Fakültesi.
- [2]. Magnosson, D. (1967). *Test Theory*. Addison-Wesley Publishing Company.
- [3]. Lord, M. F ve Novic, R. M. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company, New York,
- [4]. Arıcı, H. (1972). *İstatistik Yöntemler ve Uygulama*. Ankara: Hacettepe Üniversitesi Basımevi.
- [5]. Turgut, M. F. (1983). *Eğitimde ölçme ve değerlendirme ve metodları*. Ankara: Nüve Matbaası.
- [6]. Tekin, H. (1993). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Yargı Yayınları.
- [7]. Turgut, M. F. (1997). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Yargıcı Matbaası.
- [8]. Ertürk, S. (1993). *Eğitimde program geliştirme*. Ankara: Meteksan.
- [9]. Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test teorisi ve Uygulaması*. Ankara: Pegem Akademi
- [10]. Özçelik, D.A. (1992). *Eğitimde ölçme ve değerlendirme*. (2. baskı). Ankara: ÖSYM
- [11]. Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM
- [12]. Atılgan, H. (Ed.). Kan, A. Doğan, N. (2016). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı.
- [13]. Crocker, L. ve Algine, J. (1986). *İntroduction to classical and modern test theory*. New York: Holt Rinehart and Winston.
- [14]. Erkuş, A. (2003). *Psikometri üzerine yazılar*(1. Baskı). Ankara: Türk Psikologlar Derneği, No: 24.
- [15]. Özdemir, B. (2015). *Madde Düzeyinde Boyutluluk Modellerinin Bilgisayar Ortamında Bireyselleştirilmiş Test Yöntemleri Üzerindeki Etkisinin İncelenmesi*. Yayınlanmış doktora tezi. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara
- [16]. Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., et al. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Erlbaum.
- [17]. Linecra, J. M. (2000). *Computer-Adaptive Testing: a methodology whose time has come*, 21 Ağustos 2017 tarihinde [Çevrim-ici: <http://www.rasch.org/memo69.htm>], adresinden erişildi.
- [18]. Cömert, M. (2008). *Bireye Uyarlanmış Bilgisayar Destekli Ölçme ve Değerlendirme Yazılımı Geliştirilmesi*. Yayınlanmış Yüksek Lisans Tezi, Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- [19]. Alan C., Bugbee Jr.,(1996). The Equivalence of Paper-and-Pencil and Computer-Based Testing. *Journal of Research on Technology in educational*, 3, 282-299.
- [20]. Akdemir, O. ve Oğuz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education*. 51,1198-1204.doi:10.1016/j.compedu.2007.11.007
- [21]. Mills, C. N. (Ed.). (2002). *Computer-based testing: Building the foundation for future assessment*. NJ: Lawrence Erlbaum
- [22]. Mazzeo, J., & Harvey, A. I. (1988). *The Equivalence of Scores from Automated and Conventional Educational and Psychological Tests*. College Board Report No. 88-8, NY: College Entrance Examination Board.
- [23]. Kyllonen, P. C. (1991). Principles for creating a computerized test battery. *Intelligence*, 15(1), 1-15.
- [24]. Ita M.E, Ashley K.E. , Morin B., (2014). *Comparing Student Performance on Computer-Based vs. PaperBased Tests in a First-Year Engineering Course*. American Society for Engineering Education.
- [25]. Gardner, J. (2006). *Assessment and learning*, London: Sage Publications.
- [26]. Martin, R. (2008). New possibilities and challenges for assessment through the use of technology. In: F.S. Scheuermann and A.G. Pereira, eds. *Toward a research agenda on computer*

- based assessment: challenges and needs for European educational measurement*. Ispra (VA), Italy: European Commission.
- [27]. Paek, P. (2005). *Recent trends in comparability studies: pearson educational measurement*, 25 Ağustos 2018 tarihinde (http://www.pearsonolutions.com/downloads/research/TrendsComparabilityStudies_rr0505.pdf) adresinden erişildi.
- [28]. Segall, N., Doolen, T.L., Porter, J.D. (2005). A usability comparison of PDA-based quizzes and paper-and-pencil quizzes. *Computers & Education*, 45 (4), 417–432.
- [29]. Thelwall, M. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education*, 34 (1), 37–49.
- [30]. Nicol, D.J., Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31 (2), 199–218.
- [31]. Huang, Y.M., Lin, Y.T., Cheng, S.C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers & Education*, 52 (1), 53–67.
- [32]. Björnsson, J.K., (2008). Changing Icelandic national testing from traditional paper and pencil based tests to computer based assessment: some back ground, challenges and problems to overcome. In: F.S. Scheuermann and A.G. Pereira, eds. *Toward a research agenda on computer based assessment: challenges and needs for European educational measurement*. Ispra (VA), Italy: European Commission, 6–9.
- [33]. Swanson, D. B. Case, S. M. Ripkey, D. R. Clauser, B. E. Holtman, M. C. (2001). Relationships among item characteristics, examinee characteristics, and response times on USMLE Step 1. *Academic Medicine*, 76
- [34]. Lee, Y. H. & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*.
- [35]. Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J. Steinberg, L., Thissen, D. (2000). *Computerized adaptive testing: A primer*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- [36]. Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* New York, NY: Academic Press, Inc. pp. 179–203.
- [37]. Rudner, L. (1998) An On-line, Interactive, Computer Adaptive Testing Mini-Tutorial. *ERIC Clearinghouse on Assessment and Evaluation*.
- [38]. Davison, M. L. Semmes, R. Huang, L. Close, C. N. (2011). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement*
- [39]. Schnipke, D. L., Scrams, D. J. (2002). *Exploring issues of examinee behavior: Insights gained from response-time analyses*.
- [40]. Çıkrıkçı-Demirtaşlı, N. (1999). Psikometride Yeni Ufuklar: Bilgisayar Ortamında Bireye Uyarlanmış Test. *Türk Psikoloji Bülteni*.
- [41]. Lord, F.M., Stocking, M.L. (1988). Item response theory. In J.P. Keeves (ed.), *Educational Research, methodology, and measurement: an international handbook*. New York: Pergamon press.
- [42]. Van Lent, G. (2008). Important considerations in e-assessment. In F. Scheuermann, & A. Guimarães Pereira (Eds.), *Towards a research agenda on computer-based assessment. Challenges and needs for European educational measurement*
- [43]. Erkuş, A. (2006). *Ölçme ve Değerlendirme*. Ankara: Ekinoks.
- [44]. Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-I*. Ankara: Pegem.
- [45]. Eroğlu, M. G. (2013). *Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının Ölçme Kesinliği ve Test Uzunluğu Açısından Karşılaştırılması*. Hacettepe Üniversitesi, Ankara.
- [46]. Goldhammer, F. , Klein Entink, R. H. (2011). Speed of reasoning and its relation to reasoning ability. *Intelligence*,
- [47]. Başol, G. (2015). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- [48]. Erkan, S., Gömleksiz, M. (Ed.) (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Nobel.

- [49]. Keray-Dinçel, B. (2016). *Türkçe Öğretmenlerinin Anlama Becerilerinde Ölçme Ve Değerlendirmeye Yönelik Görüşleri ilişkin bir araştırma*. Doktora tezi. Gazi Üniversitesi, Ankara.
- [50]. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments*. Mahwah, NJ: Lawrence Erlbaum Associates. 237–266.
- [51]. Kezer, F. , Koç, N. (2014). Bilgisayar Ortamında Bireye Uyarlanmış Test Stratejilerinin Karşılaştırılması. *Eğitim Bilimleri Araştırmaları Dergisi*,4,145-174.
. *Eğitim Bilimleri Araştırmaları Dergisi*,4,145-174.
- [52]. Van der Linden, W.J. (2008). Bayesian Procedures For Identifying Aberrant Response-Time Patterns In Adaptive Testing. *Psychometrika*, 73(3), 365–384. doi: 10.1007/s11336-007-9045-8
- [53]. Hornke, L.F. (2000). Item Response Times in Computerized Adaptive Testing. *Psicologica*, 21, 175-189
- [53]. Wise, S.L. ve Kong X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests, *Applied Measurement in Education*, 18(2), 163-183. doi: 10.1207/s15324818ame1802_2



EKLER

EK: Kelime Bilgisi Testi

1. I don't have any clean clothes. I want to _____ the laundry today.
a) clean b) go c) wash d) do
2. Let's _____ the movies after the school.
a) go to b) see c) watch d) visit
3. I am always busy at work. I _____ a lot of meetings.
a) answer b) attend c) watch d) listen
4. We _____ married in 2015 before we _____ to İstanbul.
a) had / carried b) got/ moved c) had / moved d) got / carried
5. Mertcan grew _____ in Ankara.
a) out b) up c) in d) on
6. Can I use your computer? I want to _____ an e-mail to my brother.
a) attend b) send c) check d) read
7. My sister is _____. She is at the beach with her family for two weeks.
a) at home b) on vacation c) outside d) on business
8. My sister is fourteen years older than I am. I am twenty-five. My sister is _____.
a) thirty-eight b) thirty-five c) thirty-nine d) thirty-seven
9. This house has a big _____ with a new stove and a refrigerator.
a) bathroom b) kitchen c) sitting room d) bedroom
10. The car is in the _____. I don't have to worry about the storm.
a) yard b) garage c) park d) street
11. There are a lot of nice stores near my school. Sometimes I _____ after class and buy new clothes.
a) go walking b) go shopping c) go dancing d) go running
12. Every morning, I get dressed at 8:00 a.m. At 8:30, I _____ because my bus comes at 8:40.
a) take a shower b) have breakfast c) watch TV d) leave home
13. In my free time, I _____. I really like stories.
a) read books b) play games c) watch TV d) listen to music
14. I like fruit. I eat a lot of _____.
a) fish b) peas c) mangos d) pasta
15. I cannot eat dairy products like _____.
a) fries b) meat c) butter d) corn
16. My sister is a _____. She works in a dress shop.
a) cashier b) nurse c) cook d) tailor
17. My son wants to be a(n) _____. He wants to be on TV.
a) pilot b) architect c) scientist d) actor
18. She does not want to eat. She has a(n) _____.
a) headache b) stomachache c) backache d) earache
19. Lina is pretty, and she has _____ hair.
a) medium-length b) average height c) average width d) medium-depth
20. I don't like my job very much. I am going to _____ and look for another one.
a) fire b) finish c) resign d) retire
21. She doesn't have brothers or sisters. She is the _____ child.
a) only b) alone c) lonely d) single
22. Do you live in a house or a(n) _____.
a) building b) village c) apartment d) home
23. Breakfast, lunch and dinner are _____.
a) meals b) food c) eatings d) foodtimes
24. Happy is the _____ of sad.
a) oppose b) opposite c) opposite d) opposed
25. Do you want a(n) _____ or a return ticket?
a) one-journey b) lonely c) single d) double
26. Jack is at the airport waiting for Katie. Her _____ is late.
a) train b) subway c) plane d) coach
27. Jane goes out to eat a lot. She does not like _____.

- a) painting b) driving c) cooking d) swimming
28. Our teacher does not _____ us use our mobile phones in class.
a) make b) let c) forbid d) allow
29. The weather was great. It was really _____. We had a good time in the country.
a) windy b) stormy c) sunny d) rainy
30. What time is it? It is 8:15 – a _____ past eight.
a) half b) fifteen c) quarter d) fourth



ÖZGEÇMİŞ

Adı ve Soyadı :Fatma ALTUNER

Doğum Tarihi :15 Aralık 1991

E-mail :altunerfatma91@gmail.

Öğrenim Durumu :

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Ekonometri	Marmara Üniversitesi	2010-2014
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2015-2019

Görevler :

Görev Ünvanı	Görev Yeri	Yıl
Sürekli İşçi	Mersin Üniversitesi	2015-