HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

ÇOKTAN SEÇMELİ TESTLERDE FARKLI PUANLAMA YÖNTEMLERİNİN TEST VE MADDE ÖZELLİKLERİNE ETKİSİNİN İNCELENMESİ

Esmanur BAŞEL KOÇAK

Yüksek Lisans Tezi

Ankara, 2024

Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiye...

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

ÇOKTAN SEÇMELİ TESTLERDE FARKLI PUANLAMA YÖNTEMLERİNİN TEST VE MADDE ÖZELLİKLERİNE ETKİSİNİN İNCELENMESİ

EXAMINING THE EFFECT OF DIFFERENT SCORING METHODS ON TEST AND ITEM PROPERTIES IN MULTIPLE CHOICE TESTS

Esmanur BAŞEL KOÇAK

Yüksek Lisans Tezi

Ankara, 2024

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Esmanur BAŞEL KOÇAK'ın hazırladığı "Çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Test Ve Madde Özelliklerine Etkisinin İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitim Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans Tezi** olarak kabul edilmiştir**.**

Jüri Başkanı	Prof. Dr. Kaan Zülfikar DENİZ	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Selahattin GELBAL	İmza
Jüri Üyesi	Doç. Dr. Sevda ÇETİN	İmza

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 13/06/2024 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ

Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu araştırmada çoktan seçmeli bir matematik başarı testinin puanlanmasında iki kategorili (1-0) puanlama, deneysel ağırlıklandırma ve güven testinin kullanılmasının madde özellikleri ile testin geçerlik ve güvenirliği üzerindeki etkileri incelenmiştir. Araştırmada kullanılan başarı testi her biri 4 seçenekli 20 maddeden oluşmaktadır. Araştırma verileri, 2023-2024 Güz döneminde İstanbul ilinin Kartal ve Pendik ilçelerinde bulunan 5 farklı ilköğretim okulunun 8. Sınıf düzeyinde okuyan 306 öğrenciye başarı testinin uygulanmasıyla toplanmıştır. Araştırma sonucunda, testi puanlarken güven testi puanlamasının birinci versiyonu kullanıldığında en düşük ortalama madde güçlük indeksine, deneysel ağırlıklandırma puanlaması kullanıldığında ise en yüksek ortalama güçlük indeksine ulaşılmıştır. Puanlama yöntemlerinden elde edilen madde ayırt edicilik indeksleri güven testi puanlamasının iki farklı versiyonunda iki kategorili (1-0) puanlamaya göre yüksek bulunurken güven testi puanlamasın birinci versiyonunda 14 maddenin ayırt edicilik indeksi ikinci versiyona göre daha yüksek bulunmuştur. Araştırmanın geçerlik analizi için ölçüt geçerliği belirlenmeye çalışılmıştır. Öğrencilerin okul akademik başarıları ölçüt olarak kabul edilmiş ve bu puanlarla araştırmada kullanılan bütün puanlama yöntemlerinden elde edilen puanlar arasında pozitif ve anlamlı ilişkili bulunmuştur. Ölçüt puanlarla ilişkisi en yüksek bulunan puanlar güven testi puanlamasının ikinci versiyonundan elde edilen puanlar olurken en düşük ilişkiye sahip puanlar ise test maddelerinin deneysel ağırlıklandırma ile puanlanmasından elde edilen puanlar olmuştur. Araştırmanın güvenirlik analizlerine göre güven testi puanlamasının birinci versiyonu ile iki kategorili (1-0) puanlamanın KR-20 katsayıları sırasıyla 0,87 ve 0,81, güven testi puanlamasının ikinci versiyonu ile deneysel ağırlıklandırma puanlanmasının Cronbach α katsayıları 0,86 bulunmuştur.

Anahtar sözcükler: çoktan seçmeli başarı testleri, madde özellikleri, geçerlik, güvenirlik, puanlama yöntemleri, deneysel ağırlıklandırma, güven testi, iki kategorili (1-0) puanlama.

Abstract

This study examined the effects of using number-right (1-0) scoring, experimental weighting, and confidence test scoring on a multiple-choice mathematics achievement test's item properties, validity, and reliability. The test comprised 20 items, each with 4 options, and was administered to 306 eighth-grade students from 5 primary schools in Istanbul's Kartal and Pendik districts during the 2023-2024 Fall semester. As a result of the research, the lowest average item difficulty index was observed with the first version of confidence test scoring, while the highest was with experimental weighting scoring. Item discrimination indexes were higher with both versions of confidence test scoring compared to number-right scoring. In the first version of confidence test scoring, 14 items showed higher discrimination indexes than the second version. For the validity analysis of the research, criterion validity was tried to be determined. Students' school academic achievement was accepted as criteria, and a positive and significant relationship was found between these scores and the scores obtained from all scoring methods used in the research. The highest correlation with criterion scores was seen in the second version of confidence test scoring, while the lowest was in experimental weighting scoring. According to the reliability analysis of the research, the KR-20 coefficients of the first version of the confidence test scoring and the number-right (1-0) scoring were found to be 0.87 and 0.81, respectively, and the Cronbach α coefficients of the second version of the confidence test scoring and the experimental weighting scoring were found to be 0.86.

Keywords: multiple-choice tests, item properties, validity, reliability, scoring methods, experimental weighting, confidence test, number right (1-0) scoring.

Teşekkür

Tez sürecimin başından sonuna kadar bana bilgi ve deneyimleriyle her zaman destek olan değerli danışman hocam Prof. Dr. Selahattin GELBAL'a çok teşekkür ederim.

Hayatımın her anında yapmak istediğim her şey için beni motive eden, maddi ve manevi destekleriyle her zaman yanımda olan, sevgilerini ve güvenlerini daima hissettiren babam Halis BAŞEL'e ve annem Emine BAŞEL'e, varlığıyla bana hep neşe veren ve her zaman destekçim olan canım kardeşim Yakuphan BAŞEL'e çok teşekkür ederim.

Hayatıma girdiği günden beri sevgisi ve emekleriyle yol arkadaşım olan, tez yazma sürecimde de anlayışı ve sabrıyla bana hep destek veren, zorlandığım her noktada kendime inancımı güçlendirmemi sağlayan sevgili eşim Fatih Onur KOÇAK'a sonsuz teşekkür ederim.

İçindekiler

Tablolar Dizini

Simgeler ve Kısaltmalar Dizini

LGS: Liseye Geçiş Sınavı

Bölüm 1

Giriş

Problem Durumu

Eğitim girdi, süreç ve çıktı öğelerinin birbiriyle etkileşim halinde oluşturdukları bir sistemdir. Bu sistem girdilerin işlenmesinden sonra geçen süreç ve oluşan çıktılarla bireylerin belirli davranışları geliştirmesini hedeflemektedir (Baykul, 2021). Girdiler sistemde belirlenen amaçlara ulaşmak için gerekli olan kaynaklardır. Sistemin bir diğer parçası olan süreç kısmı girdilerin işlendiği ve işlevsel hale getirilmeye çalışıldığı işlemler olarak tanımlanmaktadır. Çıktılar ise girdilerin amaca yönelik işlenmesinden sonra ortaya çıkan ürünlerdir. Eğitim sisteminde çıktıların her zaman istendik yönde olduğunu söylemek mümkün değildir. Çıktılar istenen ve istenmeyen çıktılar olarak ikiye ayrılabilir. Okçabol (1985) istenmeyen çıktıları "engellenmeye çalışılsa da ortaya çıkan ürünler" olarak tanımlamış, kaynakların gereksiz ve amaç dışı kullanımını bu durumlara örnek vermiştir. Eğitim sistemi için asıl önem taşıyan çıktılar ise girdilerin işlenmesinden sonra sistemin hedeflerine hizmet edecek amaca uygun ürünler olarak tanımlanmaktadır. Eğitim sürecini değerlendirmek ve geliştirmek amacıyla hedeflenen çıktılara ulaşılıp ulaşılamadığını kontrol etmek için ölçme ve değerlendirme araçları kullanılmaktadır. Ölçme ve değerlendirme süreci öğrencilerin bilgi ve beceri düzeylerinin hedeflere uygunluğunun kontrolünün sağlanmasıyla eğitim sisteminin işlevselliği hakkında bilgi sağlar. Bu bilgiler ışığında eğitim sürecinin iyileşmesi adına gerekli düzeltmeler planlanmaktadır.

"Ölçme eğitim kararlarının verilmesinde güvenilir ve geçerli bilgi sağlanmasında işe koşulur" (Baykul, 2021, s.83). Bu tanım ölçmenin eğitimdeki yerini açıkça anlatmaktadır. Eğitimde kararların verilmesi için büyük rol oynayan ölçmenin tanımı ise birçok farklı araştırmacı tarafından değişik bakış açılarıyla yapılmıştır. Weitzenhoffer (1951) ölçmeyi bir gözlemcinin fiziksel dünyada yaptığı bir işlem olarak tanımlamıştır. Buna karşın bazı bilim insanları ise ölçmenin kapsamını sadece fiziksel dünyada var olan objelerin kendileri olarak

değil onlara ait nitelikler olarak görmektedir. Buna örnek olarak Stevens (1946) ölçmeyi "Olayların ve eşyaların özelliklerine belirli kurallar dahilinde sayılar tanımlamak" olarak açıklamıştır. Turgut (1997) ise ölçmeyi "bir niteliğin gözlenip gözlem sonucunun sayı veya başka sembollerle gösterilmesidir" cümlesiyle tanımlamıştır. Ölçmenin tanımlarında bahsedilen bir niteliğin ya da davranışın gözlenmesi bu yapıların özelliklerine göre doğrudan veya dolaylı olarak yapılabilmektedir. Ölçülmek istenen nitelik doğrudan gözlenebiliyorsa ve kendisiyle doğrudan alakalı bir ölçme aracı ile ölçülebiliyorsa bu ölçme doğrudan (temel) ölçme, nitelik doğrudan gözlemlenemiyor ve ölçülemiyorsa bir başka nitelik yardımıyla ölçülmesi de dolaylı (göstergeyle) ölçme olarak tanımlanmaktadır (Güler, 2019). Eğitim alanında da hedef davranışlara göre ölçme türü değişmektedir. Özçelik (2016) okullardaki eğitimlerle ortaya çıkması beklenen özellikleri devimsel, bilişsel ve duyuşsal olmak üzere üç ayrı kategoride tanımlamıştır. Devimsel nitelikteki özellikler için doğrudan ölçmeye başvurulabileceği belirtilmiş ancak doğrudan gözlenmesi mümkün olmadığı için yönetilen sorular yoluyla kişinin hangi derecede sahip olduğu belirlenmeye çalışılan bilişsel ve duyuşsal özelliklerin ölçülebilmesi için dolaylı ölçmelere gerek duyulduğu açıklanmıştır. Ayrıca okullarda verilen eğitimlerdeki öğrenme ve öğretme süreçlerinde öğrencilere bilişsel özelliklerin kazandırılması ve öğrencilerin bu özellikleri hangi ölçüde kazandıklarının kontrol edilmesinin öneminden bahsedilmiştir (Özçelik, 2016). Bu durumda eğitim alanında genellikle dolaylı ölçme yöntemlerine başvurulduğu ifade edilebilir.

Eğitimin herhangi bir alanında kullanılacak ölçme araçları belirlenen hedeflere bağlı olarak seçilmektedir. Bu noktada önemli olan belirlenen aracın hedef davranışı ölçmeye uygun olması, geçerli ve güvenilir olmasıdır. Eğitim alanında kullanılan birçok ölçme aracı ve yöntemi bulunmaktadır. Bu araç ve yöntemlere sözlü sınavlar, yazılı sınavlar, kısa cevaplı testler, doğru yanlış testleri, çoktan seçmeli testler, performans değerlendirmeleri, portfolyo değerlendirmesi, dereceli puanlama anahtarları (rubrik) örnek verilebilir.

Çoktan seçmeli test tekniği farklı özellikleri sebebiyle en çok başvurulan ölçme araçlarından birisidir. Çoktan seçmeli testler farklı madde türlerinden oluşur. Her bir madde,

cevaplanması gereken soruyu içeren madde kökünü ve o maddeye ait cevap olabilme olasılığı bulunduran seçenekleri içermektedir. Öğrencilerden seçenekler arasından doğru olanı bulmaları beklenmektedir. Çoktan seçmeli testlerin sıkça tercih edilen birer ölçme aracı olmasının birçok farklı nedeni bulunmaktadır. Kalabalık gruplara kısa sürede sınav uygulanabilmesi kullanışlık açısından, birden fazla beceri ve üst düzey düşünme yetenekleri gerektiren maddeler hazırlanabilmesi kapsam geçerliği açısından ve test uygulamasında en önemli konulardan biri olan puanlamanın objektif olarak yapılabilmesi güvenirlik açısından olumlu olup bu nedenlerin bir arada olması testlerin sıkça tercih edilmesinde etkili olmaktadır (Çakan, 2020).

Çoktan seçmeli testler yukarıdaki sebeplerin de etkisiyle ulusal ve uluslararası yapılan birçok ölçme ve değerlendirme çalışmasında kullanılmaktadır. Bu test türü sınıf içinde de başarı testleri halinde hem öğrencilerin ders sürecinde kaydettikleri ilerlemeyi kontrol etmek, hem de ünitelerin sonuna gelindiğinde dersler başlamadan önce belirlenmiş hedef davranışları kazanıp kazanmadıklarını ölçmek adına kullanılabilmektedir. Bununla birlikte Milli Eğitim Bakanlığı 2023 yılı ölçme ve değerlendirme yönetmeliğine göre “Ülke ya da il/ilçe genelinde yapılacak sınavlar haricinde okullarda yapılan tüm sınavlar cevaplarını öğrencilerin oluşturduğu ve farklı bilişsel düzeyde kazanımları ölçen maddelerden oluşan yazılı yoklama şeklinde yapılır.” ifadesiyle okullarda çoktan seçmeli test kullanımının kısıtlandığı görülmektedir. Bu kısıtlamaların, çoktan seçmeli testlerin yüksek düzey bilişsel özellikleri ölçmede yetersiz olarak görülmesi ve öğrencilerin cevaplarını yapılandırmalarını gerektirecek becerilerinin ölçülememesi sebepleriyle yapılmış olabileceği düşünülmektedir. Diğer taraftan ara sınavların çoktan seçmeli testlerin dezavantajları nedeniyle yazılı yoklamalar yoluyla yapılması puanlamaların subjektif olması ve yeterince soru sorulamaması gibi sorunları ortaya çıkarabilir (Özçelik, 2016). Bu yüzden yönetmelik maddesinde de belirtildiği üzere “ülke ya da il/ilçe genelinde yapılacak” geniş ölçekli testlerde ve yüksek katılımlı seçme sınavlarında genellikle yazılı yoklamaların kullanımı yerine çoktan seçmeli madde türleri tercih edilmektedir. Ülkemizde şu an uygulaması

yapılan TYT (Temel Yeterlilik Testi) ve AYT (Alan Yeterlilik Testi) sınavları çoktan seçmeli standartlaştırılmış testlere örnek verilebilir.

Çoktan seçmeli testlerin puanlanması sonuçların kullanım alanlarının önemlilik düzeyi düşünüldüğünde kritik anlamlar taşımaktadır. Bu testlerin sonuçları ile kişilerin eğitim durumları ile ilgili yüksek öğretim kurumlarına yerleştirilme, alınan derslerin geçilip geçilememesi gibi kritik kararlar alınmaktadır. Bu sebeple testlerin nasıl puanlandırıldığı da kişilerin eğitim öğretim hayatlarındaki aşamalarını önemli düzeyde etkilemektedir.

Çoktan seçmeli testler genellikle iki kategorili (1-0) puanlama yöntemi ile puanlandığından bu testleri puanlamanın yalnızca bir yolu olduğu yanılgısı oluşsa da çoktan seçmeli testlerin farklı puanlama yöntemleri bulunmaktadır.

Klasik puanlama yöntemi iki kategorili (1-0) puanlamadır. Doğru cevabı puanlama yöntemi olarak da bilinen bu yöntemde doğru cevabın işaretlenmesi tam bilmeyi temsil edip 1 puana karşılık gelirken yanlış cevap ise tam bilgisizliği temsil edip 0 puana karşılık gelir (Özdemir, 2003). Bu yöntem, kişilerin puanı belirlenirken kısmi bilginin tamamen göz ardı edilmesi, puanlar arasında tamamen doğru bilmeden tamamen yanlış bilmeye direkt geçiş olması ve bireylerin herhangi bir seçeneği işaretlerken sahip olduğu bilgi düzeyinin belirlenememesi gibi birçok farklı sebeple eleştirilen bir yöntemdir (Jaradat & Tollefson, 1988). Alanda yaygın kullanımı olan iki kategorili (1-0) puanlamanın eleştirilen yönlerinin giderilebileceğinin varsayıldığı çok kategorili puanlama yöntemleri bulunmaktadır.

Çok kategorili puanlama yöntemleri, madde veya madde seçeneklerinin eşit olarak ağırlıklandırıldığı iki kategorili (1-0) puanlamanın aksine maddelerin ve seçeneklerinin farklı ağırlıklara sahip olduğu yöntemlerdir. Bu yöntemler doğrudan cevaplama yöntemleri ve cevaplayıcı kararları yöntemleri olarak ikiye ayrılmaktadır. Doğrudan cevaplama yöntemleri kendi içinde, doğruyu bulana dek cevaplama (answer untill correct), seçenek ağırlıklandırma (option weighting), çoklu doğru cevap (multiple correct response) ve IRT yöntemleri olarak sınıflandırılmaktadır. Cevaplayıcı kararlarına dayalı yöntemler ise güven

testi (confidence testing) ve alt grup seçme (subject selecting methods) yöntemlerinden oluşmaktadır (Frary, 1989).

Eğitim kararları verilirken öğrencilerin kısmi bilgilerinin ve sahip oldukları bilgi hakkındaki öz değerlendirmelerinin göz ardı edilmemesinin hedeflenen davranışlara sahip olma düzeylerini (kazanımları) gözlerken daha detaylı, tutarlı ve kararlı olunmasına katkı sağlayabileceği düşünülmektedir. Belirtilen sebeplerden dolayı kısmi bilgiyi ve öğrencilerin herhangi bir seçeneği işaretlemesinin ötesinde gerçekte sahip olduğu bilginin düzeyini ölçmeye imkan sağlayıp kolaylaştırabilecek çok kategorili puanlama yöntemlerinin test maddelerinin psikometrik özellikleri ile geçerlik ve güvenirliklerine etkisinin araştırılması önemli bulunmaktadır.

Bu araştırmada, iki kategorili (1-0) puanlama yöntemi yanı sıra çok kategorili puanlamalardan iki farklı yaklaşım incelenmiştir. Bunlar, doğrudan cevaplama yöntemlerinden biri olan seçenek ağırlıklandırmanın deneysel ağırlıklandırma yaklaşımı ile cevaplayıcı kararlarına dayalı puanlamanın alt grup seçme kategorisinde yer alan güven testi yöntemleridir. Araştırmada bu yöntemlerin çoktan seçmeli bir matematik başarı testinin maddelerinin özellikleri ile testin geçerlik ve güvenirliğine etkileri incelenmiştir.

Araştırmanın Amacı ve Önemi

Bu araştırmada çoktan seçmeli bir matematik testinin puanlanmasında iki kategorili (1-0), deneysel ağırlıklandırma ve güven testi puanlama yöntemlerinin kullanılmasının test ve madde özelliklerine etkisi üzerine bir inceleme yapılması amaçlanmıştır.

Ülkemizde çoktan seçmeli test tekniği liseye geçiş, yükseköğretime geçiş gibi seçme amaçlı ve sınıf içi başarı belirleme amaçlı sınavlarda yaygın olarak kullanılmaktadır. Kullanım alanı bu kadar geniş olan çoktan seçmeli testlerin puanlandırılması genel olarak iki kategorili (1-0) puanlama ile yapılmaktadır. İki kategorili puanlama, uygulama kolaylığı nedeniyle tercih edilse de eğitim açısından kritik olabilecek dezavantajları da bulunmaktadır. Özdemir (2002) bu yöntemin öğrencilerin bilgilerini doğrudan tahmin

etmede zayıf olmasından dolayı eleştirildiğini belirtmiştir. Abu-Sayf (1979) iki kategorili puanlama yönteminin bilgilerinden emin olamayıp maddeyi boş bırakan bireyleri tahmin yoluyla doğru seçeneği bulan bireylere göre olumsuz etkilemesi ve kısmi bilgiyi kapsamaması sebepleriyle dezavantajlı olduğunu belirtmiştir. Bu bağlamda iki kategorili puanlamanın, test uygulaması yapılan bireylerin kısmi bilgileri ile kendi bilgileri hakkındaki öz değerlendirmelerini ölçmede yetersiz kaldığı söylenebilir. Ayrıca testi alan bireyler doğru cevabı bilmese de yanlış olduğunu bildiği bazı seçenekleri eleyerek ya da tamamen tahmin yoluyla doğru cevabı şans eseri bulabilir (Turgut & Baykul, 2019). Bu sebeple çoktan seçmeli testlerde yalnızca doğru cevabın işaretlenmesi sadece yoklanan davranışın varlığını temsil etmeyebilir.

Çoktan seçmeli testler eğitim öğretim sürecinde çıktıların kontrol edilmesi için kullanıldığından bu test puanlarının kişilerin eğitim hayatlarında önemli dönüm noktalarını etkilediğini söylemek mümkündür. Bu sebeple testlerin, bireylerin bilgileri hakkında karar vermek için yalnızca tam bilmeyi temsil eden doğru cevaba 1 puan ve tam bilgisizliği temsil eden yanlış ya da boş bırakılmış cevaba ise 0 puan verildiği iki kategorili (1-0) puanlama yöntemi ile puanlandırılmasının yerine alternatif puanlama yöntemlerine başvurulması gerekli görülmektedir.

Literatürde iki kategorili puanlamanın eleştirilen yönlerinin giderilebilmesi için kullanılabilecek farklı çok kategorili puanlama yöntemleri önerilmektedir. Çok kategorili puanlanma yöntemleri; doğrudan cevaplama ve cevaplayıcı kararlarına dayalı puanlama olmak üzere iki alt gruba ayrılan puanlama yöntemleri olarak tanımlanmaktadır (Frary, 1989). Farklı puanlama yöntemlerinin kullanılmasının testlerin özelliklerine etkilerinin incelendiği çalışmalar bulunmaktadır. Özdemir (2002) çalışmasında klasik test teorisine göre iki kategorili ve önsel ağırlıklandırma ile puanlanan testin geçerlik ve güvenirliklerinde oluşan değişikleri incelemiş, geçerlik ve güvenirlikte ağırlıklı puanlama lehine artış olduğunu ifade etmiştir. Bir başka çalışmada ise Çıtak (2010), farklı puanlama yöntemlerinden, iki kategorili (1-0) puanlama, uzman görüşüne dayalı seçenek ağırlıklandırma ve deneysel

ağırlıklandırma ile puanlamanın klasik test kuramına göre güvenirlik değerlerinin üç puanlama yönteminde de çok yüksek olmadığını bulmuş, en yüksek güvenirlik değerini deneysel ağırlıklandırmanın verdiğini ifade etmiştir. Patnaik ve Traub' un (1973) çalışmalarına göre ağırlıklı puanlama ile elde edilen puanların güvenirlik değeri iki kategorili (1-0) puanlamadan elde edilenlere göre yüksek bulunmuş ayrıca ağırlıklı puanlama yöntemine ait puanların geçerliği diğer iki yönteme göre düşük çıkmıştır. Önceden yapılmış çalışmalarda çok kategorili puanlama yöntemlerinden deneysel ağırlıklandırmanın kullanıldığı görülse de öz değerlendirmeyi sağlayan güven testinin uygulamasına fazla rastlanmamıştır.

Bu araştırmada iki kategorili puanlamaya ek olarak bireylerin kısmi bilgileriyle işaretledikleri seçeneklerden puan almaları için deneysel seçenek ağırlıklandırması, bilgileri hakkındaki öz değerlendirmelerine göre puan almaları içinse güven testi yöntemi kullanılmıştır. Araştırmanın, farklı puanlama yöntemlerinin testin özellikleri üzerinde olumlu etkileri bulunması halinde, iki kategorili (1-0) puanlamadan farklı geçerli ve güvenilir ölçme sonuçları veren puanlanma yöntemlerinin de kullanımına alan açmasıyla literatüre katkı sağlayacağı düşünülmektedir.

Araştırma Problemi

Çoktan seçmeli başarı testlerini çok kategorili puanlama yöntemlerinden deneysel ağırlıklandırma, güven testi ile puanlamanın ve iki kategorili (1-0) puanlanmanın test maddelerinin özellikleri ile testin geçerlik ve güvenirliklerine etkisi nedir?

Alt Problemler

1) Matematik başarı testi iki kategorili (1-0) puanlama ve çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin özellikleri nasıl değişmektedir?

   a. Matematik başarı testi iki kategorili (1-0) puanlama ve farklı çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin güçlük indeksleri nasıl değişmektedir?

b. Matematik başarı testi iki kategorili (1-0) puanlama ve farklı çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin ayırt edicilik indeksleri nasıl değişmektedir?

2) Matematik başarı testi iki kategorili (1-0) puanlama ve çok kategorili puanlama yöntemleri ile puanlandığında testin geçerliği nasıl değişmektedir?

3) Matematik başarı testi iki kategorili (1-0) puanlama ve çok kategorili puanlama yöntemleri ile puanlandığında testin güvenirliği nasıl değişmektedir?

Sayıltılar

Araştırmada sınav uygulaması yapılan öğrencilerin güven testi puanlama yönteminde dürüst oldukları ve daha önceki senelerde uygulanmış olan LGS (Liseye Geçiş Sınavı) sınav soruları ile karşılaşmadıkları varsayılmaktadır.

Sınırlılıklar

Bu araştırma, Milli Eğitim Müdürlüğünün 2018 tarihi itibariyle belirlediği matematik dersi öğretim programında 8. Sınıf müfredatının Sayılar ve İşlemler adlı 1. ünitesinin Çarpanlar ve Katlar, Üslü İfadeler ve Kareköklü İfadeler konuları ile sınırlıdır. Araştırmada veri kaybı olmaması adına araştırma sonuçlarını etkilemeyeceği düşünülen boş cevaplar sıfır ile puanlanarak veri analizi sırasında çıkarılmamıştır. Çoktan seçmeli test maddelerinin deneysel puanlanmasında uygun olan doğru cevabı en doğru madde türü olmasına rağmen bu araştırmada tek doğru cevabı olan madde türü kullanıldığı için bir sınırlılık olarak kabul edilmiştir.

Tanımlar

Çoktan Seçmeli Test: Farklı maddelerden oluşan her bir maddeye ait madde kökü ve cevap olabilme olasılığı bulunduran seçenekleri içeren ölçme aracıdır (Çakan, 2020).

İki kategorili Puanlama: Çoktan seçmeli bir test puanlanırken bireyin anahtarlanmış doğru seçeneği işaretlemesi durumunda 1 puan, yanlış seçeneklerden birinin işaretlemesi, birden çok seçeneği işaretlemesi veya hiçbir seçeneği işaretlememesi durumlarında ise 0 puan aldığı yöntemdir.

Deneysel Ağırlıklandırma Puanlaması: Çoktan seçmeli bir testte bireylerin farklı seçenekleri işaretleme yüzdesine göre her bir seçeneğe ağırlık atanmasıyla test maddelerinin puanlandığı yöntemdir.

Güven Testi Puanlaması: Çoktan seçmeli bir testin maddelerinde seçeneklerin işaretlenmesine ek olarak bireylerin cevapları ile ilgili güven düzeylerini de belirtmelerine dayalı puanlama yapılan yöntemdir.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Ölçme Araçları ve Yöntemleri

Thorndike ve Thorndike-Christ (2014) eğitim alanında eğitimcilerin, öğrencilerin ve velilerin; öğrencilerin eğitim deneyimlerini etkileyecek birçok karar almaları gerektiğini belirtmişlerdir. Eğitimciler, müfredatta belirlenen hedef kazanımların öğrencilere iyi bir şekilde nasıl aktarılabileceğiyle öğrencilerin kazanımlar kapsamında gelişimleri hakkında kararlar alırken, öğrenciler ve veliler de öğrencilerin daha yüksek eğitim kurumlarına geçişleri ve ders seçimleri gibi konular hakkında kararlar almaktadır. Bahsi geçen eğitim kararlarının verilebilmesi için bilgilere ihtiyaç vardır ve bu bilgiler ölçme araç ve yöntemleri yoluyla elde edilmektedir. Özçelik (2016) eğitim alanında hedeflerin, büyük ölçüde bilgiyi tanıma, hatırlama, bilgi hakkında akıl yürütme kavrama ve genellemeler yapma becerilerini kapsayan bilişsel özellikler hakkında olduğunu ifade etmiştir. Bu bağlamda eğitim alanında karar verilecek bilgilerin çoğunlukla bilişsel yeterliliklerin ölçülmesiyle elde edildiği söylenebilir. Bilişsel yeterliliklerin ölçülebileceği farklı ölçme araç ve yöntemleri bulunmaktadır. Bunlara sözlü sınavlar, yazılı sınavlar, kısa cevaplı testler, doğru yanlış testleri ve seçme gerektiren çoktan seçmeli testler örnek verilebilir.

Çoktan Seçmeli Testler

Seçme gerektiren testler geçmişten günümüze kullanımı çok yaygın olan ölçme araçlarıdır. Kişilerin özelliklerini ölçmek amacıyla kullanılan testlerin hem psikolojik yapılarla hem de öğrenme ile ilgili olan eğitim özellikleri ile ilgilenmekte olduğu belirtilmiştir (Turgut & Baykul, 2019). Testin diğer bir tanımı, davranışı ölçmek, davranışın anlaşılmasına katkıda bulunmak ve davranışı tahmin etmek amacıyla kullanılan bir ölçme aracı olarak yapılmıştır (Kaplan & Saccuzzo, 2005). Baykul (2021) psikoloji alanında testlerin kullanım amaçlarını zeka, yetenek, tutum, ilgi gibi yapıların özelliklerinin belirlenmesi, bu yapıların boyutlarının ortaya çıkarılması olarak belirtmiş eğitim alanında ise seçme ve yerleştirme, öğrenciler

hakkında eğitim kararları verilmesi, öğrencilerin öğrenme zorluklarının tespit edilmesi ve öğrencilerin başarı durumlarının tespit edilmesi olmak üzere dört ayrı kullanım amacı alanı tanımlamıştır.

Bir kurum ya da okulda çalışan veya öğrenci alımlarında açılan kontenjandan çok daha fazla başvuru olması durumlarında seçme yapabilmek adına testlere başvurulabilir. Bu bağlamda testler kuruma ya da okula alınacak kişilerin başvurdukları pozisyon için gerekli özellikleri taşıyıp taşımadıkları, kimlerin bu pozisyonlara daha uygun olduğuna karar verilmesi amacıyla kullanılmaktadır. Eğitim kararlarında test kullanımı ilki öğretim ikincisi ise öğrenci yönünden yapılacak incelemeler olarak iki ayrı başlıkta incelenebilir. Öğretim programlarının çıktılarının kontrollerinin yapılması ve bu sayede programların sağlamlığı hakkında bilgi edinilmesi ile öğretim hizmetinin niteliği hakkında değerlendirmelere ulaşılması test kullanımının önemli sebeplerindendir. Öğrenci yönünden test kullanımının amaçları öğrencilerin alanda eksikliklerinin saptanması, öğrenme zorluklarının belirlenmesi, öğrencinin ilgi ve yetenek alanlarına yönlendirilmesi ile öğrenme düzeylerinin belirlenmesi olarak belirtilmiştir (Baykul, 2021).

Çoktan Seçmeli Test Maddelerinin Özellikleri

Birçok farklı amaçla kullanılabilen testler maddelerden oluşmakta ve maddelerin yapısal özellikleri bulunmaktadır. Test maddeleri uygulama yapılan kişilerin yanıtlaması gereken uyarıcılar olarak tanımlanmış, bu maddelerin yanıtlarının puanlanabileceği ya da derecelendirilebileceği belirtilmiştir (Kaplan & Saccuzzo, 2005). Test maddelerinin farklı bölümleri bulunmaktadır. Bu bölümler öğrencilerin cevaplamasının beklenen sorunun yer aldığı madde kökü, sorunun doğru yanıtı olma olasılığı olan seçenekler, sorunun doğru yanıtını içeren anahtarlanmış cevap ve soruya dair yanlış bilgisi olan ya da bilgisi olmayan öğrencileri çekmesi için verilen çeldirici seçenekler olarak açıklanmaktadır (Güler, 2019). Uygulanan testin maddelerindeki çeldiricilerin anlamlı olabilmesi için doğru cevabı bilmeyen öğrenciler tarafından bariz yanlış seçenekler olarak algılanmaması gereklidir. Çoktan seçmeli test maddelerinde çeldiricilerin iyi çalışabilmesi için çeldirici hazırlama sürecinde

öğrencilerin sıklıkla yaptıkları hataların ve yanlış öğrenmelerin dikkate alınması gerektiği ifade edilmiştir (Çakan, 2020). Büyük gruplara uygulama kolaylığı aynı zamanda puanlama objektifliği ve kolaylığı gibi olumlu özelliklerinden dolayı sıkça tercih edilen çoktan seçmeli test yöntemi, bilişsel düzeyde sadece düşük seviyeli özellikleri ölçebildiği üst düzey bilişsel özellikleri yoklayamadığı ve maddenin doğru yanıtını bilmeyen öğrencilerin seçeneklerden birini işaretleyerek doğru cevabı bulma olasılığının olması yönleriyle eleştirilmektedir (Turgut & Baykul, 2019).

Çoktan Seçmeli Test Maddelerinin Avantajları

Çoktan seçmeli testlerin sıkça tercih edilmesinin birçok farklı sebebi bulunmaktadır. Bu sebeplerden ilki bu test türünün büyük gruplara uygulama yapılırken pratiklik sağlamasıdır. Üniversiteye giriş sınavları gibi büyük çaplı, ülke genelinde uygulanacak sınavlarda çoktan seçmeli test tercih edilmesi birkaç saat içinde binlerce kişiyi sınav yapabilme olanağı sağlamaktadır. Pratik bir şekilde büyük gruplara sınav yapıldıktan sonra sınav kağıtlarının puanlanması cevap anahtarları yardımı ile hızlıca tamamlanabilir hatta optik form okuyucu makineler yardımı ile de puanlama yapılabilmektedir (Çakan, 2020; Turgut & Baykul, 2019). Çoktan seçmeli testlerin önemli avantajlarından bir diğeri ise öğrenciler soruları yanıtlarken belirli seçeneklerin arasından seçim yaptığından puanlama objektifliğinin sağlanması olacaktır (Turgut & Baykul, 2015). Başka bir deyişle çoktan seçmeli testlerle puanlama güvenirliği hatasız olduğundan özellikle eğitim ve kariyer kararları alınan liseye geçiş, yükseköğretime giriş, tıpta uzmanlık, kamu personeli seçme sınavı gibi sınavlarda kullanılmaktadır. Uygulama kolaylığı ve puanlama süresinin kısalığı kullanışlık açısından önemli özellikler olarak belirtilmiştir (Güler, 2019). Öğrencilerin uygulanan sınavdaki zamanlarını cevap yazmaya ayırmamalarının da etkisiyle sınav yapılan alanla ilgili daha fazla soru sorulabilmesi de bu ölçme aracının kapsam geçerliğini ve puanların güvenirliğini artırmaya fayda sağlamaktadır (Çakan, 2020; Güler, 2019). Çoktan seçmeli test maddelerinin genellikle hatırlama düzeyinde bilgileri ölçtüğü eleştirilerinin yanı sıra soruların madde yazımı konusunda uzman olan kişiler tarafından

hazırlanması halinde üst düzey bilişsel becerilerin yoklanabildiği görüşleri de bulunmaktadır (Güler, 2019; Turgut & Baykul, 2019). Çoktan seçmeli test türünün bilindik avantajlarından bir diğeri ise farklı öğrenim düzeylerine uygulanabilir olmasıdır. Bu noktada ilköğretimin başlangıç sınıf düzeyindeki öğrencilerin kendilerini ifade etme özelliklerini yeni yeni kazanmalarından dolayı onlara çoktan seçmeli test tekniğinin uygulanmaması gerektiği yönünde görüşler bulunmaktadır (Güler, 2019). Çoktan seçmeli maddelerde seçenek sayısı da öğrenim düzeyi arttıkça artacak şekilde planlanmaktadır.

Çoktan Seçmeli Test Maddelerinin Dezavantajları

Çoktan seçmeli testlerin avantajlarının yanı sıra dezavantajları da bulunmaktadır. Bu ölçme türündeki önemli problemlerden birisi şans başarısıdır. Öğrencilerin belirli seçenekler arasından doğru cevabı hiç bilmeseler de şans yoluyla doğru seçeneği işaretleyebilmeleri şans başarısı olarak tanımlanmaktadır (Çakan, 2020). Şans başarısı seçenek sayısı arttıkça azalmaktadır. Test puanlarını şans başarısı hatasından arındırmak için istatistiksel bir düzeltme formülü kullanılmaktadır. Düzeltme formülü uygulanınca alınan puan düzeltilmiş puan olarak anılmaktadır. Turgut ve Baykul (2015), düzeltme formülünün belirli varsayımlar altında kullanıldığını belirtmişlerdir. Bu varsayımlardan ilki, öğrencilerin cevaplarını yönergeye uygun şekilde kaydetmiş olmalarıdır. Başka bir ifadeyle, öğrencilerin maddelerin yanıtlarını belirlerken kaydırma yapmamış ve bir maddenin yanıtını başka bir maddeye işaretlemedikleri varsayılmaktadır. İkinci varsayım ise öğrencilerin tüm yanlışlarına yanlış cevaplamış olma şanssızlığının sebep olmuş olmasıdır. Burada tartışılan konu öğrencilerin yanlış yanıtlarının hepsinin şanssızlıkla olduğunun kabul edilmesidir. Oysaki öğrencilerin yanlış yanıtı tercih etmesi bilgi eksikliği, yanlış hesaplama, yanlış öğrenme gibi sebeplerden kaynaklı olabilir. Son varsayım öğrencilerin şansla yanıtladığı tüm maddelerdeki seçeneklerin öğrencilere eşit çekicilikte görünmüş olmasıdır.

Bu varsayımlarla düzeltilmiş puan formülü aşağıdaki gibidir (Turgut & Baykul, 2019):

$$P = n(D) - \frac{n(Y)}{a-1}$$

$n(D)$: $Doğru\ sayısı$

$n(Y)$: $Yanlış\ sayısı$

a : $Seçenek\ Sayısı$

Düzeltme formülünün uygulanacağı sınavlarda öğrencilere bu durumun mutlaka açıklanmasının ve öğrencilerin emin olmadıkları soruları cevapsız bırakmaları konusunda uyarılmalarının gerektiği belirtilmiştir (Çakan, 2020).

Çoktan seçmeli testlerin bir diğer olumsuz özelliği ise madde hazırlama sürecinin uzun ve nispeten zor olmasıdır. Kaliteli bir çoktan seçmeli test maddesi yazabilmek için madde kökü ve çeldirici yazımında dikkat edilmesi gerekenlerin bilinmesi ve konuda uzman olunması gereklidir (Güler, 2019; Turgut & Baykul, 2019). Çoktan seçmeli testlerin literatürde üzerinde durulan bir diğer genel dezavantajı ise bu yöntemle ileri düzey bilişsel becerileri yoklamanın zorluğu olarak tanımlanmaktadır. Öğrenciler çoktan seçmeli test maddelerini yanıtlarken yapılandırılmış olan seçeneklerden birini seçtiği için yaratıcılık düzeyinde becerilerini sergilemelerinin mümkün olmayacağı açıklanmıştır (Çakan, 2020; Turgut & Baykul, 2019).

Çoktan Seçmeli Testleri Puanlama Yöntemleri

Çoktan seçmeli testleri puanlarken iki kategorili (1-0) puanlama veya çok kategorili puanlama yöntemleri kullanılabilmektedir.

İki Kategorili Puanlama Yöntemi (Number Right Scoring)

Doğru cevabı puanlama bir diğer adıyla iki kategorili (1-0) puanlama çoktan seçmeli testlerin puanlanmasına en sık kullanılan yöntemdir. Bu puanlama yönteminde maddenin doğru cevabını işaretleyen öğrenciler 1 tam puan alırken yanlış seçeneklerden birini (çeldiriciler) işaretleyen, birden çok seçeneği işaretleyen veya maddeyi boş bırakan her öğrenci de o madde özelinde 0 puan almaktadır. 1 puanın tam bilmeyi (complete information) temsil ettiği ve 0 puanın tam yanlış bilmeyi (misinformation) temsil ettiği

belirtilmektedir (Özdemir, 2003). İki kategorili puanlamanın eleştirilen yönlerinden biri, doğru cevabı bilerek işaretleyen bir öğrenci ile doğru cevabı tahmin yoluyla işaretleyen bir öğrenciyi aynı kategoriye yerleştirmesidir. Ek olarak, yanlış seçeneklere yönelen öğrenciler arasında hiçbir bilgisi olmayanları, kısmi bilgisi olanları ve tahmin yoluyla yanlış seçeneği seçenleri aynı kategoriye koyması da eleştirilmektedir (Jaradat & Tollefson, 1988). Öğrencilerin bilgi düzeylerini net bir şekilde ölçemediği için eleştirilen iki kategorili (1-0) puanlama yönteminin yükseköğretim programlarına yerleşme gibi önemli sınavlarda kullanılmasının problem oluşturduğu yönünde görüş bildirilmiştir (Özdemir, 2003). Bu bağlamda, iki kategorili (1-0) puanlamanın öğrencilerin bilgi düzeyleri arasındaki farkları yeterince ayırt edememesi görüşü, alternatif puanlama yöntemlerinin geliştirilmesine ve önerilmesine yol açmıştır.

Çok Kategorili Puanlama Yöntemleri

Kısmi bilgi puanlama yöntemleri olarak da bilinen bu yöntemler doğrudan cevaplama yöntemleri ve cevaplayıcı kararlarına dayalı yöntemler olmak üzere iki gruba ayrılmaktadır.

Doğrudan Cevaplama Yöntemleri

Doğrudan cevaplama yöntemleri, doğru cevabı bulana kadar işaretleme, seçenek ağırlıklandırma, çoklu doğru cevap yöntemi ve IRT yöntemleri olmak üzere dört başlık altında kategorize edilmektedir (Frary, 1989).

Doğru Cevabı Bulana Kadar İşaretleme Yöntemi. (Answer Until Correct Method). Bu yöntemde sınava giren kişilerden doğru olduğunu düşündükleri seçeneği işaretlemeleri istenir ancak işaretleme işlemi kişi doğru cevabı seçene kadar devam eder (Frary, 1989). Bu puanlama tekniğinde madde puanının seçilmeyen seçenek sayısına eşit olması gerekmektedir. Örneğin n çeldirici sayılı bir maddede doğru cevabı ilk seferinde bulan kişi (n) puan alırken ikinci seferinde bulan kişi (n-1), r. seferinde bulan kişi maddeden (n+1-r) puan alacaktır (Brown, 1965). Bilgisayar ortamında yapılmadığında uygulama

açısından zahmetli olan bu yöntem pratik değildir. Bilgisayar ortamında kullanılması ise maliyet açısından sorun olabilmektedir.

Çoklu Doğru Cevap Yöntemi. (Multiple Correct Option Method). Çoktan seçmeli testlerde birden çok doğru cevabı olan maddelerin yazılması mümkündür. Bu testlerde maddelere ait doğru cevap sayısının değişebileceği yönergesinin öğrencilerle paylaşılması gerekmektedir. Bu bağlamda doğru cevap sayısının maddeden maddeye değişebilme ihtimaline karşı cevaplayıcıların tüm seçeneklerin doğruluğunu incelemesi gerekmektedir (Özdemir, 2003). Bu durumda her maddenin seçeneklerinin yapısı nedeniyle doğru yanlış formatına dönüştüğü aktarılmaktadır (Frary, 1989). Bahsi geçen türde maddeleri olan bu testlerin maddelerini puanlamada kullanılan yöntem doğru cevaplanan seçenek sayısından yanlış işaretlenen seçenek sayısını çıkartmak olarak açıklanmıştır (Kurz, 1999).

IRT Yöntemleri (Item Response Theory Methods). Bu yöntemin ilk olarak Samejima tarafından 1969 yılında ortaya konulduğu belirtilmektedir (Frary, 1989)**.** Samejima tarafından ortaya atılan modelin yalnızca seçenekleri doğruluk açısından sıralanabilen maddelerde kullanılabileceği aktarılmış olup, 1972 yılında Bock tarafından seçenek sıralama sınırı olmadan kullanılabilecek nominal cevap modelinin geliştirildiği de belirtilmiştir (Özdemir, 2003).

Seçenek Ağırlıklandırma Yöntemleri. Seçenek ağırlıklandırma yöntemleri önsel ve deneysel ağırlıklandırma olarak ikiye ayrılmaktadır. Temelde test maddelerine ilişkin seçeneklere ait puanların önceden bir uzman tarafından ağırlıklandırılması veya seçenek puanlarının deneysel olarak ağırlıklandırılması mantığına dayanır (Frary, 1989). Ağırlıklandırma puanlamasının tek doğru cevabı olan ve diğer cevapların yanlış olduğu durumda kullanılması ölçme sonuçlarında hataya yol açabilmektedir. Ağırlıklı puanlama yönteminin cevabı en doğru olan madde formlarında kullanılması uygun olacaktır.

Önsel Ağırlıklandırma (Uzman Kanısına Dayalı Ağırlıklandırma). Bir çalışmada önsel ağırlıklandırma yönteminde öncelikle uzmanların inceledikleri test maddelerinin seçeneklerini tamamen doğru bilgiden tamamen yanlış bilgiye doğru sıraladığı ardından

farklı uzmanların seçeneklere verdiği puanların ortalamasının hesaplanması ile seçeneklerin ağırlıklandırıldığı ifade edilmiştir (Kurz, 1999). Önsel ağırlıklandırmanın uzmanların kararlarından dolayı hata içerebileceği ve maliyet açısından negatif yönleri olabileceği açıklanmıştır (Echternacht, 1973). Devrim (2002) önsel ağırlıklandırmanın kullanıldığı çalışmasında uzmanların her zaman tutarlı olmaması sebebiyle puanlamada zorluklar yaşandığı belirtmiş, zaman ve maliyet faktörlerinin de yöntemin dezavantajları arasında olduğunu ifade etmiştir.

Deneysel Ağırlıklandırma (Görgül Ağırlıklandırma). Seçenek ağırlıklandırmanın bir diğer alt grubu deneysel ağırlıklandırmadır. Bu ağırlıklandırma için farklı formül ve yöntemler geliştirilmiştir. Önsel ağırlıklandırmadan farklı olarak burada seçeneklere uygulama öncesi puan atanmaz. Sınavın uygulandığı öğrencilerin seçenekleri işaretleme yüzdelerine dayalı olarak maddelerin seçenekleri ağırlıklandırılır (Frary, 1989). Deneysel ağırlıklandırmanın oran ağırlıklandırması yapılan bir versiyonunda sınava katılan öğrencilerin içinden belirlenen üst gruplara göre madde seçeneklerine puan atandığı ifade edilmektedir (Claudy, 1978). Guttman tarafından geliştirilen ağırlıklandırma sisteminde ise, her bir seçeneğin ağırlığı, o seçeneği işaretleyen sınav katılımcılarının toplam test puanlarının ortalaması ile orantılıdır (Kurz, 1999). Literatürde deneysel ağırlıklandırma çalışmalarında kullanılan alternatif yöntemlerden biri iki serili ağırlıklandırmadır. Çift serili ağırlıklandırmada boş bırakılan bir seçenek dahil olmak üzere her bir seçeneğe puan atanır. Bu ağırlıklandırma da Brogden Biserial veya Clemans Lambda korelasyon katsayıları kullanılır Bu aslında bir ayırt edicilik indeksidir. Teorik olarak, ağırlıklar (-1) ile (+1) arasında herhangi bir değer alabilir (Claudy, 1978).

Cevaplayıcı Kararları Yöntemleri

Cevaplayıcı kararları yöntemleri alt grup seçme ve güven testi yöntemleri olmak üzere iki gruba ayrılmaktadır.

Alt Grup Seçme Yöntemleri. Bu yöntemler kapsayarak ve eleyerek puanlama olarak tanımlanmaktadır. Kapsayarak puanlama öğrencilere doğru olduklarına inandıkları

tüm seçenekleri işaretlemeleri, eleyerek puanlama ise yanlış olduğuna inandıkları bütün seçenekleri işaretlemeleri yönergelerinin verildiği puanlama yöntemleridir (Frary, 1989). Bu yöntemlerde öğrencilerin tüm çeldiricileri eleyip sadece anahtarlanmış doğru seçeneği seçmeleri durumunda tam bilgiye sahip oldukları, çeldiricilerin bazılarını elediklerinde kısmi bilgiye, çeldiricilerle birlikte doğru cevabı da elediklerinde kısmi yanlış bilgiye ve sadece anahtarlanmış doğru cevabı elediklerinde tam yanlış bilgiye sahip oldukları son olarak tüm seçenekleri eledikleri ya da ilgili maddeyi boş bıraktıklarında ise bilgi yokluğu durumunda oldukları ifade edilmektedir (Özdemir, 2003). Eleme puanlamasında Coombs tarafından kullanılan teknikte öğrenciler yanlış olduğunu düşünerek işaretlediği her bir çeldirici için 1 puan almaktadır. Yanlış olduğuna inanılıp işaretlenen seçeneklerden biri doğru seçenek olduğunda işaretlenen seçeneklerden bir puan çıkarılıp puanlandırma yapıldığı aktarılmıştır (Frary, 1989). Kapsayarak ve eleyerek puanlama yöntemleri ile yapılan çalışmalarda hem geçerlik hem güvenirlik değerlerinde artış bulunduğu ancak uygulama yönergelerinin öğrenci gruplarının kafasını karıştırdığı belirtilmiş ve bu durumun psikometrik artıştan daha kritik bir önem taşıyor olabileceği aktarılmıştır (Kurz, 1999).

Güven Testi Yöntemi. Cevaplayıcı kararlarına dayalı puanlama yöntemlerinden olan güven testi öğrencilerin maddeleri cevaplamaya ek olarak işaretledikleri seçenekle ilgili güven düzeylerini veya derecelerini ifade etmelerini gerektiren puanlama yöntemidir (Frary, 1989). Bu yöntemde herhangi bir maddede aynı seçeneği işaretleyen öğrencilerin kendi cevaplarına güvenme düzeylerine göre o madde özelinde farklı puanlar alabilecekleri ifade edilmiştir. Ayrıca güven testi uygulaması yapıldığında, bir maddeyi tahmin yoluyla ya da tamamen şans eseri doğru işaretleyen öğrencilerin, güven düzeylerini dürüst bir şekilde ifade etmeleri durumunda, tam bilgiye sahip öğrencilerle aynı puanı almayacakları göz önünde bulundurulursa, gerçek bilgiye sahip olanları belirlemek için de kullanışlı bir yöntem olabileceği düşünülmektedir. Ebel ’in 1965 yılında yaptığı çalışmada ise bu yöntem kullanıldığında güven ile ilgili olan ancak başarı ile ilgili olmayan bir değişkenin ölçüm hatası varyasyonunda artışa yol açabileceğini belirttiği açıklanmıştır (Kurz, 1999). Sadece

hatırlama düzeyinde olan bir madde de dahi anahtarlanmış doğru seçeneği işaretleyen öğrencilerin ve yanlış seçeneklerden birini işaretleyen öğrencilerin cevaplarından emin olma düzeylerinin farklılık gösterebileceği aktarılmıştır (Dressel & Schmid, 1953). Sonuç olarak bilgisi eksik olduğundan ya da tahmin yoluyla işaretlediğinden değil kendinden emin olamadığından güven düzeyi düşük olan öğrenciler açısından da negatif etkileri olabileceği ifade edilmiştir.

Güven testi yönteminde, öğrencilerin eminlik düzeylerini belirlemeleri soru çözümünden ayrı bir vakit aldığından, uygulama süresini uzatması bir dezavantaj olarak kabul edilebilir. Güven testinin başka bir dezavantajı puanlama aşamasında öğrencilerin her bir test maddesinden alacağı puanın tek tek kontrol edilip belirlenmesi ardından toplam puanın bulunmasının zaman alması ve pratik olmaması olarak aktarılmaktadır (Dressel & Schmid, 1953). Kurz (1999) testlerde güven testi yöntemi uygulandığında geçerlik ve güvenirlikte önemli düzeyde artış bulunmaması, kişilik özellikleri değişkenin sürece dahil olması, test puanlama ve testi uygulama süresinin standart bir teste göre fazla olması durumlarının ortaya çıktığını ve bu faktörler sebebiyle güven testi yönteminin kullanım alanının kısıtlı olduğunu belirtmiştir.

Geçerlik

Geçerlik ölçme araçlarında ölçülmek istenen özelliğin başka özelliklerle karıştırılmadan ölçülme derecesi olarak tanımlanmaktadır (Turgut ve Baykul, 2013). Geçerliğe ait bir diğer tanımlama test puanlarının kullanımlarının ve yorumlarının değerlendirilmesi şeklinde yapılmıştır (Cronbach,1988). Assossiation vd. (2014) ise geçerliği kanıtların ve teorinin test puanlarının yorumlarını ne ölçüde desteklediğini ifade etmesi şeklinde tanımlamış ve bundan dolayı geçerliğin testlerin geliştirilmesinde ve testlerin değerlendirilmesinde en temel nokta olduğunu belirtmiştir.

Kapsam, yapı ve ölçüt geçerliği olmak üzere farklı geçerlik kanıtı sağlama yöntemleri bulunmaktadır.

Kapsam Geçerliği (Content Validity)

Kapsam geçerliği hazırlanan bir ölçme aracının maddelerinin ölçmesi hedeflenen davranışları ne derece kapsadığı ile ilgilidir. Kapsam geçerliğinin sağlanabilmesi ve kontrol edilebilmesi için ölçme aracının ölçülmek istenen özelliklerinin belirlenmiş ve tanımlanmış olması gerekir (Turgut & Baykul, 2019). Bir ölçme aracında kapsam bir diğer deyişle içerik geçerliğinin varlığını kontrol ederken dikkate alınacak nokta bu ölçme aracının maddelerinin ilgili konuda sorulabilecek tüm soruları örnekleyebilmiş olmasıdır (Salkind, 2013). Test içeriğinin içerik alanını ne kadar temsil ettiği alanında uzman kişiler tarafından kontrol edilebilir. Bir alan testi geliştirilirken o alanda ölçülmek istenen özellikler belirlenip uzmanlardan test maddelerinin belirtilen özelliklerden hangisine ait olduğunu seçmeleri istenebilir. Bu aşamadan sonra uzmanlar seçilen maddelerin belirlenen özellikleri temsil etme durumunu değerlendirebilir (Assossiation vd., 2014).

Ölçüt Dayanaklı Geçerlik (Criterion-Related Validity)

Bir ölçme aracından elde edilen sonuçların başka değişkenlerle ilişkini kontrol etmek geçerlik kanıt sağlama yöntemlerinden biridir (Assossiation vd., 2014). Bir ölçme aracının sonuçlarının yeterli geçerlik ve güvenirlikte olduğu bilinen başka bir ölçme aracının sonuçları ile korelasyonuna bakılarak hesaplanan kanıt toplama yöntemi ölçüt dayanaklı geçerlik olarak tanımlanmaktadır (Güler, 2019) . Geçmişte, şimdilerde ölçüt geçerliğinin anlamı olan “psikolojik testlerin uygulanmasından sonra puanların belirli bir ölçütle karşılaştırılmasıyla bulunan ilişki” geçerliğin genel tanımı olarak kabul edilmiştir (Kelecioğlu & Şahin, 2014).

Ölçüt kavramı belirli bir kararın doğruluğunu ölçmek için kullanılan yeterli geçerlik ve güvenirliğe sahip olduğu bilinen ölçme aracı olarak tanımlanırken, geçerliği belirlenecek olan ölçme aracı ise yordayıcı olarak tanımlanmıştır (Turgut & Baykul, 2015). Geçerli kabul edilip ölçüt olarak belirlenen test sonuçları ile yordayıcı olarak belirlenen test sonuçlarının korelasyon katsayısı incelendiğinde, katsayının +1’e yakınlığı yordayıcı aracın da geçerliği

olduğuna, 0'a yakınlığı yordayıcı aracın düşük düzeyli bir geçerliğe sahip olduğuna, -1'e yakınlığı ise yordayıcı ile ölçütün ölçtükleri değişkenlerin birbirinden farklı olduğuna ve bu değişkenlerin ters yönde çalıştığına dair bilgi vermektedir (Baykul, 2021).

Ölçüt dayanaklı geçerlik türü uyum ve yordama geçerliği olarak iki başlık altında incelenmektedir.

Uyum Geçerliği

Uyum (uygunluk) geçerliği ölçüt olarak belirlenen testin yordayıcı test ile aynı zamanda ya da daha erken yapılmış olduğu durumlarda incelenir. Ölçüt test daha önceden ya da eş zamanlı uygulanmış olduğundan puanları mevcuttur. Bu durumda yordayıcı testin sonuçlarının bu test ile aynı değişkeni ölçtüğü bilinen geçerli ve güvenilir ölçüt test sonuçları ile korelasyon katsayısı incelenir. Korelasyon katsayısı yordayıcı testin ölçüt teste göre istenen değişkeni ölçmeye ne kadar uygun olduğunu göstermiş olacaktır (Baykul, 2021; Güler, 2019).

Yordama Geçerliği

Yordama ya da tahmin geçerliği ölçülen bir davranışın, özelliğin gelecekteki durumu hakkında tahminde bulunmak için kullanılır. Ölçüt olarak belirlenen puan henüz mevcut değildir. Yordayıcı puan kullanılarak daha sonraki performans hakkında tahmin yapılmasıdır (Baykul, 2021). Yordayıcı puanları belirleyen ölçme aracı uygulandıktan bir süre sonra ölçüt olarak puanların elde edilmesiyle bu puanların arasındaki korelasyon katsayısı incelenir.

Ölçütü Belirleyen Faktörler. Thorndike ve Thorndike-Christ (2014) ölçütü belirlemek için istenen özellikleri 4 başlıkta açıklamışlardır. Bunlardan ilki ilgili olma maddesidir. Burada tahmin edilmeye çalışan özellik ile ölçütün neyi ölçtüğünün uyumlu olması gerektiğinden bahsedilmiştir. Eğitim üzerinden düşünülecek olursa yordayıcı aracın ölçme istediği özellik ile ölçütün öğrenme hedeflerinin benzer olması gerekmektedir. Ayrıca ilgili olma faktörünün bir bakıma ölçütün içerik geçerliği anlamına gelebileceğini de belirtmişlerdir. İkinci faktör ise önyargıdan bağımsızlıktır. Önyargıdan bağımsızlık ile

kastedilen ölçme yapılan grupların eşit fırsatlara sahip bir şekilde ölçülebilmesidir. Başka bir deyişle aynı düzeydeki kişilerin ölçme hataları dışında eşit puanı alabilmesinden bahsedilmektedir. Benzer bir grubun eşit olmayan koşullarda elde ettiği puanların arasındaki korelasyon katsayılarının anlamlı bir sonuç vermeyebileceğine dikkat çekmişlerdir. Üçüncü faktör olarak güvenirlikten bahsedilmiştir. Bir ölçme aracı ile belirli bir özellik hakkında tahminde bulunabilmek için belirlenen özelliğin tutarlı olması gerekmektedir. Ölçüt olarak belirlenen puan sürekli farklılaşıyor ve tutarsız değişimler gösteriyorsa bir yordayıcı tarafından tahmin edilemeyeceği belirtilmiştir. Ölçüt belirlemek için bahsedilen son özellik kullanılabilirlik ve uygunluktur. Burada her bir bireyin ölçüt puanının elde edilmesi için ne kadar beklenebileceğinden bahsedilmektedir (Thorndike & Thorndike-Christ, 2014). Ölçüt puanlarının pratik bir şekilde elde edilmesi zamandan ve masraf olabilecek durumlardan kaçınmak için önemlidir.

Yapı Geçerliği (Construct Validity)

Yapı terimi psikolojik yapıyı yani insan davranışlarının doğrudan gözlenemeyen ve ölçülemeyen bir yönünün teorik olarak kavramsallaştırılmasını ifade eder (Ebel ve Frisbie, 1991). Zeka, tutum, güdü gibi doğrudan gözlenemeyen özellikler psikolojik yapı olarak tanımlanmaktadır. Cronbach ve Meehl (1955)'e göre ölçülecek özelliğin tanımlanabilmesi adına herhangi bir ölçüt veya içerik tamamen yeterli olmadığında yapı geçerliğinin incelenmesi gerekmektedir.

Yapı geçerliğini değerlendirmek için tek bir yöntem bulunmamakla birlikte, bu süreç için birkaç farklı adım önerilmektedir. İlk olarak, yapıyı operasyonel ve kuramsal tanımlarını ortaya koymak için farklı yapılarla ilişkilendirmek gerekir. Bu tanımlar yapıldıktan sonra, belirli hipotezler geliştirilmeli ve bu hipotezleri kontrol edecek ölçme araçları bulunmalı veya mevcut araçlar kullanılmalıdır. Belirlenen ölçme araçları bir gruba uygulandıktan sonra veri toplama süreci başlar ve elde edilen verilerin hipotezleri destekleyip desteklemediğine karar

verilir. Eğer elde edilen verilerin analizi, hipotezleri destekliyorsa, yapısal özellikler belirlenebilir; desteklemiyorsa, farklı hipotezler üzerinde çalışılabilir (Baykul, 2021).

Bir ölçme aracında yapı geçerliği ile ilgili kanıt toplamak için kullanılabilecek farklı yöntemler bulunmaktadır.

Gruplar Arasındaki Farklar

Psikolojik yapılar farklı tutum ve davranışları ortaya çıkardığı için, bir ölçme aracı belirli bir yapıyı ölçmeye yönelik farklı davranışlara sahip gruplara uygulandığında, grupların puanlarının farklılaşması beklenir. Bu durum, grup puanları arasında düşük bir korelasyonun oluşmasına neden olabilir. İki farklı grubun düşük korelasyon katsayısının ortaya çıkması, yapı geçerliğine dair bir kanıt olarak değerlendirilebilir. Ancak, dikkat edilmesi gereken önemli bir nokta, bu grupların sahip oldukları özelliklerin güçlü olmasıdır. İncelenen yapıya ait davranışların, yapıyla ilgili olmayan dış etkenlerden kaynaklanması, korelasyon katsayısının yapı geçerliğinin varlığını doğrulama açısından kullanılmasını engelleyebilir (Baykul, 2021).

Korelasyon

Bir yapıyı değerlendirmek amacıyla kullanılan ölçme aracının geçerliği önceden kanıtlanmış bir ölçme aracı ile olan korelasyon katsayısının incelenmesi, yapı geçerliği hakkında kanıt toplamak için kullanılacak bir başka yöntem olarak öne çıkmaktadır. İki ölçme aracı arasındaki yüksek korelasyon katsayısı, bu araçların aynı yapıyı ölçtüğü anlamına gelir ve bu da çalışma kapsamındaki ölçme aracının yapı geçerliğine dair bir kanıt olarak kabul edilebilir. Öte yandan, düşük korelasyon, üzerinde çalışılan ölçme aracının yapı geçerliğinin olmadığını işaret edebilir (Baykul, 2021).

Çoklu Yapı Çoklu Yöntem Modeli

Çoklu Yapı Çoklu Yöntem Modeli, bir yapıyı ölçmek amacıyla geliştirilen ölçme araçlarının uygunluğunu belirlemek için Campbell ve Fiske (1959) tarafından geliştirilmiştir.

Bu modelde, farklı yapılar belirlenen yöntemlerle ortaya konur. İkinci aşama, her yöntemle her yapı için ölçüm yapılması olarak tanımlanmaktadır (Ünsal Özberk & Doğan, 2014).

Faktör Analizi

Ölçme araçlarında ölçülmek istenen bir yapı veya yapılar belirlenir. Bu yapılar, ölçme araçlarındaki değişkenlerin sayıları ve değişkenlerin toplam test puanları üzerindeki etkileri aracın ölçmeye çalıştığı temel özellikleri açıklar (Baykul, 2021). Faktör analizi, bu özellikleri ortaya çıkarmak için kullanılan bir yöntemdir. Bu analiz, ölçme aracındaki her bir ölçümün altında yatan yapıları ve ilişkileri anlamak için kullanılır. Faktör analizi, özellikle karmaşık yapıları açıklamak ve ölçme aracının geçerliliği hakkında daha derinlemesine bir anlayış geliştirmek için önemli bir araçtır (Baykul, 2021).

Güvenirlik

Thorndike ve Thorndike-Christ (2014), bir ölçme aracının güvenirliğinin, ne ölçtüğünden ziyade ölçülmek istenen herhangi bir özelliği ne kadar doğru bir şekilde ölçtüğü ve aynı kişilerin birkaç kez ölçülmesinde alınan puanların ne kadar tutarlı olacağıyla ilgili olduğunu vurgulamışlardır. Başka bir deyişle ölçmede güvenirliğin hazırlanan ölçme aracının ölçülmesi hedeflenen özelliği ne derece duyarlı ölçtüğünün ve ölçmeden elde edilen puanların şans başarısından ne kadar arınık olduğunun göstergesi olduğu söylenebilir (Özbek, 2020). Ölçme için oldukça önem taşıyan güvenirlik kavramına ait bir başka tanımlama ise “Bir ölçme sonucu, içindeki tesadüfi hataların azlığı oranında güvenilirdir.” (Turgut & Baykul, 2015, s.123)

Güvenirlik, geçerliğin ön koşuludur. Bu nedenle güvenilir olmayan bir teste ait geçerlik kanıtları kontrol edilmez, bu kanıtlar anlamsız olacaktır.

Literatürde tesadüfi hataların türlerine göre ölçme aracında güvenirliğin değişik anlamlara gelebileceği belirtilmiştir. Bunlardan ilki tutarlılıktır. Tutarlılık ölçme aracının tekrarlı kullanımında ulaşılan sonuçların benzerlik derecesi olarak tanımlanmaktadır (Güler, 2019). İkinci kavram ise kararlılıktır. Bu kavram aynı ölçme aracının bir gruba farklı

zamanlarda uygulanmasında grubun sonuçlarının benzer olması ile ilgilidir. Bu sonuçlar bariz şekilde değişiyorsa ölçme sonuçlarının kararsız olduğu yani güvenilir olmadığı kanısına varılır (Baykul, 2021). Üçüncü kavram duyarlılıktır. Duyarlılık ölçme aracının ölçmeyi hedeflediği özelliği hangi seviyede hassas ölçtüğü ile ilgilidir. Birimi daha küçük olan ölçme aracının birimi daha büyük olan bir ölçme aracına göre daha duyarlı olduğu ve bu sebeple daha güvenilir olduğu söylenebilir (Baykul, 2021)

Güvenirlikle ilgili ilk çalışmalar, Spearman'ın (1904) kitabında ortaya çıkmıştır ve bu kavram daha sonra Kuder ve Richardson (1937) tarafından farklı güvenirlik katsayıları ile tanımlanmıştır. Bu bağlamda, güvenirliğin, ölçme aracının güvenilir bir şekilde ve istikrarlı bir biçimde ölçüm yapma kapasitesini değerlendirme sürecinde temel bir öneme sahip olduğu ifade edilmektedir (Thorndike, 2014; Kuder & Richardson, 1937).

Test Puanı Teorisinin Temelleri

Klasik Test Teorisine göre ölçmeye hiç hatanın karışmadığı durumlarda kişilerin gözlenen puanların kişilerin gerçek puanı olduğunu varsayılmaktadır. Ancak hiçbir ölçme aracı mükemmel değildir, bu yüzden de gözlenen puan her zaman kişinin gerçek puanından biraz farklı olabilmektedir. İki puan arasındaki bu fark ölçme hatasıdır. Gerçek puan (X), Gözlenen puan (T) ve hatanın (e) matematiksel ilişkileri ve sembolik gösterimleri aşağıda verilen şekildedir:

$$X = T + e$$

Klasik test teorisinin büyük bir kabulü ölçme hatalarının tesadüfi olmasıdır. Tesadüfi hata kaynağı ve yönü belli olmayan hata olarak tanımlanmaktadır (Thorndike & Thorndike-Christ, 2014).

Ölçmenin Standart Hatası

Yapılan ölçümlerin güvenirliğini değerlendirmenin bir yöntemi, bir kişiye ait tekrarlı ölçümlerin beklenen varyans miktarının ele alınmasıdır. Bu kapsamda, Thorndike ve Thorndike-Christ (2014) tarafından bir kişinin ağırlığının tartıda 200 kez ölçülmesi örneği

verilmiştir. Kişinin ölçülen ağırlıklarının frekans dağılımı alınarak, bu dağılımın ortalaması belirlenir. Frekans dağılımının ortalaması, kişinin gerçek ağırlığına yaklaşık bir tahmin olarak düşünülebilir. Aynı zamanda, bu dağılım, ölçümlerin ortalamaları etrafındaki yayılma veya dağılımı tanımlayan bir standart sapma içerir. Zira bu dağılım, ölçüm hatalarından kaynaklanan ve ölçümlerdeki değişkenliği yansıtan bir standart sapmayı içermektedir. Bu standart sapma aynı zamanda ölçmenin standart hatası olarak adlandırılmaktadır (Thorndike & Thorndike-Christ, 2014).

Güvenirlik Katsayısı

Güvenirlik, tekrarlı ölçümlerde bireylerin puanlarının genellikle benzer seviyede kalmasını ifade eder (Thorndike & Thorndike-Christ, 2014). Bir ölçme aracının güvenirliğinin göstergesi, aracın ilk uygulamasında yüksek puan alan bir kişinin ikinci uygulamada da benzer bir düzeyde yüksek puan almasıdır. İki paralel ölçüm arasındaki korelasyon katsayısı genellikle güvenirlik katsayısı olarak adlandırılır (Baykul, 2021). Güvenirlik katsayısı, matematiksel olarak ifade edilen gerçek puanların varyansının, gözlenen puanların varyansına oranı olarak da tanımlanabilir (Kaplan & Saccuzzo, 2005). Bu tanımlama ile güvenirlik katsayısının formülü aşağıda verildiği şekildedir:

$$\mathrm{r} = \frac{\sigma_T^2}{\sigma_X^2}$$

Güvenilirlik katsayısı ile ölçmenin standart hatası arasında bir yakınlık bulunmaktadır. Daha önce sembolik gösterimde belirtildiği gibi, gözlenen puan, gerçek puan ile ölçme hatalarının bir varyansı olarak tanımlanmaktadır. Ölçme araçları bir grup bireye uygulandığında, gözlenen puanların varyansı ortaya çıkar ve bu, grup içindeki dağılımın bir göstergesidir. Ölçme aracının uygulandığı gruptaki bireylerin tamamının aynı gerçek puana sahip olmadığı ve hataların bağımsız ve rastgele olduğu varsayımıyla, gözlenen puanlardaki varyansın, gerçek puanlardaki varyans ile ölçme hatalarındaki varyansın toplamı olduğu gözlemlenmektedir (Thorndike & Thorndike-Christ, 2014).

$$SD_X^2 = SD_T^2 + SD_e^2$$

Bu bağlamda, güvenirlik katsayısı, gözlenen puanlardaki varyansın, bireyler arasındaki gerçek farklardan kaynaklanan oranı olarak tanımlanır. Aynı zamanda, ölçmenin standart hatasının, gözlenen puanların standart sapması ve güvenirlik katsayısının bir fonksiyonu olduğu gösterilmiştir.

$$SD_e = SD_X(\sqrt{1} - r_{tt})$$

Belirli bir değişkenlik düzeyine sahip bir grupta, güvenirlik katsayısı yükseldikçe, ölçmenin standart hatası azalır. Yüksek güvenirliği olan bir testte bireyler arasındaki puan farklılıklarının genellikle ölçülen kişiler arasındaki gerçek farklardan kaynaklandığı belirtilmiştir (Thorndike & Thorndike-Christ, 2014).

Güvenirlik Kestirme Yöntemleri

Ölçme araçlarının ne ölçüde güvenilir olduklarını değerlendirmek için birden fazla veri toplama yöntemi bulunmaktadır. Güvenirlik kestirim yöntemlerinde korelasyon hesabından faydalanılmaktadır. Güvenirlik [0,+1] aralığında değer almaktadır (Özbek, 2020).

Bu yöntemlerin benzerlik ve farklılıkları üzerinden güvenirlik değerlerinin ne amaçla kullanılacağına göre uygun olana karar verilmektedir. Bu veri toplama yöntemleri:

1. Aynı testin veya ölçümün tekrar edilmesi (test-tekrar test)

2. Testin ikinci bir “eşdeğer” formunun uygulanması (paralel veya alternatif test formları)

3. Tek bir uygulamadan elde edilen eşdeğer yarılar (testi iki yarıya bölme) ile iç-tutarlılık güvenirliği şeklindedir (Thorndike & Thorndike-Christ, 2014).

Test Tekrar Test Metodu

Bireylerin ölçme sonuçlarının zaman içindeki kararlılığını incelemek amacıyla kullanılan bir güvenirlik kestirim yöntemidir. Bir gruba uygulanan ölçme aracı, belirli bir zaman diliminden sonra aynı gruba tekrar uygulandığında elde edilen puanlar arasındaki

Pearson Momentler Çarpımı Korelasyon katsayısı hesaplanır. Test-tekrar test prosedüründen elde edilen korelasyon katsayısı, kararlılık katsayısı olarak adlandırılmaktadır (Crocker & Algina, 2006). Bu yöntemin davranış örneklemi ya da uygulanan testin maddeleri ile ilgili bir varyans olmadığında diğer bir deyişle zeka gibi zaman içinde değişime yatkın olmadığı düşünülen istikrarlı özelliklerin ölçüldüğü durumlarda kullanıma uygun olacağı belirtilmiştir (Kaplan & Saccuzzo, 2005; Özbek, 2020). Bu bağlamda güvenirlik kestirimi için test tekrar test metodunun kullanılmasında dikkat edilmesi gereken bazı durumlar vardır.

Bu durumların ilki, aktarma etkisi (carryover effect) olarak adlandırılır. Bu etki kişilerin ilk testteki maddeleri hatırlamasını ve ilk test uygulamasının ikinci testin sonuçlarını etkilemesini açıklar (Kaplan & Saccuzzo, 2005). Testin uygulandığı gruptaki öğrenciler ilk uygulamadan sonra hatırladıkları bir sorunun doğru yanıtını öğrenebilir ve bu sayede ikinci uygulamada daha yüksek puan alabilir. Bu durumda güvenirlik gerçekte olduğundan daha yüksek çıkacaktır. İkinci durum ise alıştırma etkisi olarak bilinir ve kişilerin zamanla gerçek bilgilerinin artması anlamına gelir örneğin yalnızca teste girmekle uygulama yapılan gruptaki öğrencilerin becerileri gelişebilir, testin konusu hakkında düşünüp doğru bilgiyi ikinci uygulamaya kadar öğrenebilirler (Kaplan & Saccuzzo, 2005).

Bahsi geçen bu durumlardan yola çıkarak iki testin uygulanması arasında geçecek zamanın seçimi konusunda dikkatli olunması gerektiği bilinmektedir. İki uygulama arasındaki zaman dilimi, aktarma veya alıştırma etkilerinin kaybolmasına izin verecek kadar uzun olmalı, ancak uygulama yapılan grubun gerçek puanlarında farklı faktörlerin etki edebileceği değişikliklerin meydana gelmesine izin verecek kadar uzun olmamalıdır (Crocker & Algina, 2006). Uygulama yapılan grubun gerçek puanlarında oluşabilecek değişiklik yine grubun özelliklerine bağlıdır. Küçük yaş gruplarının daha önceki test uygulamasındaki maddeleri ve yanıtlarını hatırlama ihtimalleri daha düşük, uygulama yapılan yetişkin grubun ise ilk testi hatırlama ve aktarma etkisi ile önceki puanların daha farklı puanlara ulaşma ihtimalleri daha yüksektir. Test tekrar test yöntemi kullanımında

zaman faktörünün yanı sıra potansiyel sorunlar arasında, testlerin uygulandığı grupların iki uygulama arasında değişen motivasyon düzeyleri ve sağlık durumlarındaki değişimler bulunmaktadır. Ayrıca, uygulamaların yapıldığı ortamlardaki ışık, ses gibi ortamın sınava uygunluğunu etkileyecek özelliklerin değişimlerinin de dikkate alınması gerektiği de açıklanmıştır (Özbek, 2020). Bu faktörler, test sonuçlarını dolayısıyla testin güvenirliğini etkileyebilecek potansiyel dışsal değişkenleri içermektedir.

Eşdeğer (Paralel) Formlar Metodu

Birbirine paralel ve madde güçlük düzeylerinin benzer olduğu aynı özellikleri ölçen iki farklı testin aynı gruba uygulaması metodudur (Kaplan & Saccuzzo, 2005).

Paralel formlar güvenirlik kestiriminde de test tekrar test metodundan olduğu gibi iki testten alınan puanların arasındaki Pearson Momentler Çarpımı Korelasyon katsayısı hesaplanır ve elde edilen bu katsayıya denklik katsayısı denir (Crocker & Algina, 2006). Denklik katsayısı iki test formunun da güvenirliğini temsil eder (Turgut & Baykul, 2019). Bu güvenirlik kestirimi metodunda iki test genellikle aynı gün içinde uygulanır ancak bazen farklı günlerde uygulama yapılması da mümkündür. Kaplan ve Saccuzzo (2005) paralel bir diğer adıyla eşdeğer formlar yönteminde iki testin uygulamasının aynı anda yapıldığında zamanın ve uygulama koşullarının oluşturabileceği varyans etkilerinin ortadan kaldırıldığını belirtmişlerdir. İki test aynı gün içinde uygulandığında varyasyon oluşturabilecek durumların yalnızca rastgele hatalar ve eşdeğer formlar oluşturulurken seçilen içerik örnekleminden kaynaklı farklılıklar olabileceği de belirtilmiştir (Crocker & Algina, 2006; Kaplan & Saccuzzo, 2005).

Paralel formlar yönteminin de bazı dezavantajları bulunmaktadır. Aynı özellikleri ölçen birbirine denk iki paralel testi geliştirmenin zorlayıcı olması ve harcanan zamanın farklı yöntemlere göre iki kat daha fazla iş yükü anlamına gelmesi bu yöntemin kullanılma oranını düşürmektedir (Özbek, 2020).

Testi İki Yarıya Bölme ve İç-Tutarlılık Katsayıları

Uzmanların iki farklı uygulamaya dayalı test tekrar test ve paralel formlar yöntemlerine göre kullanmaya daha yatkın oldukları güvenirlik kestirim yöntemleri tek uygulamaya dayalı testi iki yarıya bölme ve diğeri iç tutarlılık katsayıları hesaplama yöntemleridir.

Eşdeğer Yarılar (Testi İki Yarıya Bölme) Metodu

Bu güvenirlik kestiriminde tek bir test ile bir uygulama yapılır, test ikiye bölünür ve bu iki bölüm ayrı puanlanır. Sonrasında bu iki bölümün puanları arasındaki korelasyon katsayısı hesaplanır (Özbek, 2020). Testin nasıl ikiye bölüneceğine farklı durumlara göre karar verilmektedir. Eğer yeterli sayıda test maddesi bulunuyorsa en iyi yöntemin soruları rastgele ikiye bölmek olduğu belirtilmiştir. Bazı uzmanlar ise hesaplama kolaylığı açısından testi ortadan ikiye ayırmaya yatkındır. Ancak bu durumun test maddelerinin giderek zorlaştığı formatta hazırlandığında sorunlara yol açabileceği de belirtilmiştir. Giderek zorlaşan test maddesi düzeni olan testler için öneri soruları tek ve çift sayılı olanlar olarak ayırıp puanlamaların buna göre yapılması olmuştur (Kaplan & Saccuzzo, 2005).

Eşdeğer yarılar metodunda korelasyon hesabı yapılırken uygulanan testteki madde sayısı yarıya düştüğünden güvenirlik katsayısı diğer yöntemlere nazaran düşük çıkacaktır, testin iki yarısının birbirine eşit olduğu kabul edilerek uygulanan testten hesaplanacak korelasyon katsayısına Spearman Brown güvenirlik katsayısı denir (Baykul, 2021). Spearman Brown katsayısı aşağıda verilen eşitlikle hesaplanır:

$$r = \frac{r_{1,2}}{1 + r_{1,2}}$$

r değeri testin bütününe ait güvenirlik katsayısını temsil ederken $r_{1,2}$ değeri testin ikiye bölünen eşit yarılarından birinin güvenirlik katsayısını temsil eder. Tüm testin güvenirlik katsayısı eş değer yarıların güvenirlik katsayısından yüksek olacak ancak bu yarılara ait güvenirlik katsayısının iki katından da düşük olacaktır (Baykul, 2021). Bu formülle elde edilen güvenirlik katsayısının yüksek olması hem testin güvenilir olduğuna hem de testin belirlenen iki yarısının eşdeğer olduğuna kanıt kabul edilmektedir (Baykul, 2021). Bu

yöntemle elde edilen güvenirlik katsayısının düşük olması ise testin güvenirliğinin düşük olduğunu gösterebilir ancak yarıya bölünen kısımların eşdeğer olmadığının göstergesi de olabilir. Bu durumda yarıların eşdeğer olup olmadığının hipotez testi ile kontrol edilmesi ve sonuca göre testin güvenirliğine dair bir kanıya varılması gerektiği belirtilmiştir (Baykul, 2021; Özbek, 2020).

İç Tutarlılık Katsayıları

İç tutarlılık katsayılarının hesaplanması testin güvenirliğinin tek bir uygulama ile kestirilmesi için kullanılan bir diğer yöntemdir. Bu yöntemde testi iki yarıya bölme metodundan farklı olarak testin bütün maddelerinin birbirleri ile tutarlılıkları incelenmektedir. İç tutarlılık katsayıları KR-20 ve Kr-21 formülleri ile hesaplanmaktadır.

KR-20 Formülü.KR-20 Formülü, Kuder ve Richardson (1937) tarafından testin tek uygulamasından güvenirlik tahminlerinin geliştirilmesi amacıyla farklı yöntemlerle işlenmiştir. Bu yöntemin, testin iki yarısının varyanslarının farklı olabileceği ve iki yarının ayrı ayrı puanlanması gerekliliği gibi dezavantajlarını ortadan kaldırarak, maddelerin bölünme sürecindeki tüm olası yolları hesaba kattığı açıklanmıştır (Kaplan & Saccuzzo, 2005). Kr-20 formülü, sadece doğru cevaba 1, yanlış ya da boş cevaba ise 0 puan verilen diğer bir deyişle iki kategorili (1-0) puanlama yapılan testlerde uygulanabilir. Ayrıca, bu formülün kullanılabilmesi için maddelerin güçlük düzeylerinin bilinmesi ya da hesaplanabilmesi gerektiği vurgulanmaktadır (Özbek, 2020). KR-20 katsayısı aşağıda verilen eşitlikle hesaplanmaktadır:

$$KR-20 = \frac{K}{K-1}\left[1-\frac{\Sigma s_x^2}{S_x^2}\right]$$

K : $Testte\ yer\ alan\ madde\ sayısı$

s_x^2 : $Her\ maddenin\ puan\ varyansı$

S_x^2: $Testin\ varyansı$

KR-21 Formülü .KR-21 formülü de test maddelerinin birbiriyle tutarlılığını kontrol eder. KR-20 formülünden farkı ise maddelerin güçlük katsayılarının eşit varsayılması ile hesaplanmasıdır. KR-21 hesaplanırken test maddelerinin güçlük katsayılarının bilinmek zorunda olmadığı belirtilmiştir (Özbek, 2020). KR-21 katsayısı aşağıda verilen eşitlik yardımıyla hesaplanmaktadır:

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

K : $Testte\ yer\ alan\ madde\ sayısı$

$\bar{X}$: $Testin\ ortalaması$

S_x^2: $Testin\ varyansı$

KR-21 formülü ile elde edilen güvenirlik değeri yüksek olduğunda hem testin maddelerinin güçlük düzeylerinin birbirine yakın olduğuna hem ölçülmek istenen özelliğin tek boyutlu olduğuna hem de testin güvenilir olduğu yorumlarına ulaşılabileceği belirtilmiştir (Baykul, 2021). Testin güvenirlik katsayısı düşük çıkması ise madde güçlük düzeylerinin birbirinden farklı olması, testin çok boyutlu olması ya da testin güvenirliğinin düşük olması ihtimallerini içermektedir.

Cronbach Alfa. Kuder ve Richardson'ın geliştirdiği KR-20 ve KR-21 formüllerinin iki kategorili (1-0) puanlanan testlerinin güvenirliğini kestirmede kullanıldığı bilinmektedir. Ancak bu katsayılar tutum ölçekleri gibi likert tipi puanlanan, maddelerin doğru ya da yanlış olmadığı testlerde kullanıma uygun değildir (Güler, 2019). Likert tipi puanlama seçeneği olan ölçeklerin güvenirliklerinin tespitinde Cronbach Alfa katsayısından faydalanılır (Özbek, 2020).

Cronbach'ın geliştirdiği bu formül hem likert tipi puanlanan hem de iki kategorili (1-0) puanlanan testlerin güvenirlik kestirimleri için kullanılabilir bu özelliğiyle de KR-20 ve KR-21 formüllerinden daha genel bir kullanım alanı olduğu belirtilmiştir (Kaplan & Saccuzzo, 2005).

Cronbach Alfa formülü aşağıdaki verildiği gibidir:

$$r = \alpha = (\frac{k}{k-1})(\frac{\sigma_X^2 - \Sigma\sigma_i^2}{\sigma_X^2})$$

α : $Alfa\ g\ddot{u}venirlik\ katsayısı$

k: $Madde\ sayısı$

σ_i^2: $Madde\ varyansı$

$\Sigma\sigma_i^2$: $Madde\ varyanslarının\ toplamı$

σ_X^2: $Testin\ varyansı$

Testlerde bulunan maddelerin aynı özelliği ölçtüğü durumlarda güvenirlik yüksek olur ve alfa katsayısı 1'e yaklaşır. "Güvenirlik katsayısı $\alpha \geq 0{,}9$ ise mükemmel, $0{,}7 \leq \alpha < 0{,}9$ ise iyi $0{,}6 \leq \alpha < 0{,}7$ kabul edilebilir $0{,}5 \leq \alpha < 0{,}6$ ise zayıf ve $\alpha < 0{,}5$ ise kabul edilemez olarak yorumlanır." (Özbek, 2020, s.56)

Puanlayıcı Güvenirliği

İki kategorili puanlama yöntemi dışında, birden fazla puanlayıcının değerlendirdiği ve subjektif olarak puanlanabilen testlerin güvenirliği ile ilgili, farklı puanlayıcıların puanlarından hareketle kestirimlerde bulunmak mümkündür.

Güler (2019), puanlayıcı güvenirliği kestirimi için iki farklı senaryodan bahsetmektedir. İlk senaryo, ölçülmek istenen özelliğin süreksiz ve matematiksel olarak puanlanamayan bir yapıda olması durumudur. Bu durumda, puanlayıcıların puanlama benzerliklerinin yüzdeliklerinin hesaplanması gerektiğini belirtmektedir. İkinci senaryo ise ölçülmek istenen özelliğin sürekli bir yapıda olması durumudur. Bu durumda ise iki puanlayıcının verdiği puanlar arasındaki korelasyon katsayısının incelenmesi gerektiğini belirtmektedir (Güler, 2019).

Güvenirliği Etkileyen Faktörler

Testteki Madde Sayısı

Kaplan ve Saccuzzo (2005) örnekleme modeline göre bir ölçme aracındaki her bir maddenin ölçülmek istenen özelliğin bir örneği olduğunu belirtmişlerdir. Eğer madde sayısı artarsa ölçülmek istenen özelliğin gerçek temsili de artmış olacaktır. Bu durum çok kısa başarı testlerinin evren puanlarının kaba bir tahminini yapması olarak açıklanmakta ve testlerdeki madde sayısı arttıkça bireylerin ölçülmek istenen özellikteki gerçek farklılıklarının ortaya çıkacağı belirtilmektedir (Özbek, 2020). Bu bağlamda sorunsuz maddelerin sayısı arttıkça ölçme aracının da güvenirliği artacağı açıklanmıştır. Fazla maddeli bir testinin az maddeli bir teste oranla daha yüksek güvenirlik kestirimi yapacağı belirtilse de belli bir madde sayısından sonra yeni eklenecek maddelerin güvenirliğe katkısı sınırlı olacaktır (Kaplan & Saccuzzo, 2005) . Testlerdeki madde sayısında değişiklik yapılmasının testin güvenirliğine yapacağı etki Spearman-Brown formülü ile hesaplanmaktadır.

$$r_{XX'} = \frac{kr_{XX}}{1 + (k-1)r_{XX}}$$

Formülde bulunan

r_{xx}: $Testin\ ilk\ halinin\ g\ddot{u}venirlik\ katsayısı$

$r_{XX'}$: $Madde\ sayısı\ artırılmış\ yeni\ testin\ g\ddot{u}venirlik\ katsayısı$

k : $Testin\ uzatıldığı\ kat\ sayısı$

Maddelerin Özellikleri

Test maddeleri sadece sayıları itibariyle güvenirliği etkilemez maddelerin güçlük ve ayırt edicilik düzeyleri de güvenirlik üzerinde etki sahibidir. Test maddelerinin aşırı kolay veya aşırı zor olması puan dağılımlarının iki uca yayılmasına sebebiyet verir bu durumda testin uygulandığı bireylerin ölçülmek istenen özellik açısından farklılaşmama sorunu ortaya çıkar (Özbek, 2020). Testin uygulandığı bireylerin aşırı zor maddeleri rastgele cevaplama ihtimallerinin daha yüksek olduğu da ortaya konmuştur (Turgut & Baykul, 2015). Literatüre göre orta güçlükte ve ayırıcılığı yüksek maddelerin bulunduğu testlerin güvenirlik kestirimleri daha yüksektir.

Testin Uygulanma Koşulları

Testlerin uygulandığı ortam, o ortamın sıcaklık düzeyi, ortamın ışık düzeyi, uygulama yapıldığı anda ortamda gürültü olması, bireylere verilen zamanın eşit olmaması gibi faktörler bireylerin puanlarına hata karışmasına neden olmaktadır (Özbek, 2020). Testin güvenirliğinin iyi olduğundan bahsedebilmek için tüm bu faktörlerin standart bir biçimde testin uygulandığı bireylere sağlanması gerekmektedir.

Puanlama Objektifliği

Testlerin objektif olarak puanlanması güvenirlik üzerinde ciddi bir öneme sahiptir. Birbirinden farklı puanlayıcıların maddelere aynı puanı vermesi, puanlama yapılırken puanlayıcının yanlı olmaması gibi durumlar güvenirliği artıracaktır. Puanlayıcıların puanlama yaptıkları test maddesinin yanıtına göre inisiyatif alabildikleri durumlarda güvenirlik objektif puanlamanın olduğu bir test türüne göre düşük olacaktır (Özbek, 2020; Turgut & Baykul, 2015).

Birey Özellikleri

Testin uygulandığı bireylerin sınav günündeki ruhsal ve fiziksel durumları da testin güvenirliğini etkileyen faktörlerdendir. Bireylerin sınav konusu hakkındaki tutumları, sınav günü hasta ya da sağlıklı olmaları, uygulama yapılacak testin konusuna yakın zamanda çalışmaları gibi durumların testin güvenirliğini etkilediği belirtilmiştir (Özbek, 2020). Uygulama yapılan grubun testle ilgili yeterince güdülmeleri gerekir ki maddelere dikkatle yanıt versinler, aksi halde sorulara üzerine düşünülmeyen rastgele cevaplar verilebilir ya da soruları boş bırakabilirler. Bu durumların da testin güvenirliğini düşürme sebebi olduklarından bahsedilmiştir (Turgut & Baykul, 2019).

İlgili Araştırmalar

Bu bölümde geçmişten günümüze farklı puanlama yöntemleri ile puanlanan çoktan seçmeli testlerin psikometrik özelliklerinin değişiminin incelendiği araştırmalarla ilgili sonuçların ortaya konulması amaçlanmıştır. Çoktan seçmeli testlerin farklı puanlama

yöntemleri ile puanlanmasının geçerlik ve güvenirlik değerlerine etkisinin incelendiği çalışmalarda değişik sonuçlar bulunmaktadır kimi araştırmalar geçerlik ve güvenirlik değerlerinde kayda değer yükselişlerden bahsederken kimi araştırmalar da kayda değer farklılıklar olmadığını ortaya koymaktadır.

Dressel ve Schmid'in (1953) çalışmalarında öğrencilerin cevaplarından emin olma düzeylerine göre puan aldığı bir güven testi uygulaması bulunmaktadır. Emin olma seviyeleri için belirledikleri 4 düzeyli ölçeğe göre puanlama yapmışlardır. Ölçek "kesin, oldukça kesin, rasyonel tahmin ve seçim için savunulabilir bir temel yok" ifadelerinden oluşmaktadır. Anahtarlanmış doğru yanıtın işaretlenmesi durumunda emin olma düzeyinin artışına göre pozitif yükselen, yanlış yanıtlardan birinin işaretlenmesi durumunda emin olma düzeyinin artışına göre negatif yükselen puanlar alınacak şekilde puanlama sistemini oluşturmuşlardır. Bu çalışmanın sonucunda güven testi yönteminin üstün, ortalama ve düşük düzeydeki öğrencileri yaklaşık olarak eşit düzeyde ayırt ettiğini belirtmişlerdir.

Patnaik ve Traub (1973) çalışmalarında 69 maddelik çoktan seçmeli bir testin geleneksel iki kategorili (1-0) puanlama, doğru yanıtlardan yanlış yanıtların bir kısmının çıkarılmasından elde edilen formül puanlaması ve maddelerin seçeneklerine uzman kanısına dayalı ağırlıklı puanlar atanmasıyla elde edilen ağırlıklı puanlama ile puanlanmasından elde edilen geçerlik ve güvenirlik değerlerini incelemişlerdir. Güvenirlik için testi yarıya bölme iç tutarlık katsayısı, geçerlik içinse ortalama okul başarıları ile farklı yöntemlerden elde edilen test puanlarının arasındaki korelasyon katsayıları incelenmiştir. Çalışmanın sonuçlarına göre ağırlıklı puanlama ile elde edilen puanların güvenirlik değeri hem iki kategorili (1-0) puanlama hem formül puanlamasından anlamlı ölçüde yüksek çıkmıştır diğer yandan ağırlıklı puanlama ile elde edilen puanlar diğer iki yönteme göre daha az geçerli bulunmuştur. İki kategorili puanlama ve formül puanlaması arasında güvenirlik ve geçerlik açısından önemli ölçüde fark bulunmadığı da belirtilmiştir

Claudy (1978) çalışmasında deneysel ağırlıklandırma yönteminin farklı türleri olan çift serili ağırlıklandırma, Guttman ağırlıklandırması, oran ağırlıklandırmasını ayrıca iki

kategorili (1-0) puanlama ile şans başarısı için düzeltme formüllerini kullanarak çoktan seçmeli testleri puanlamıştır. Çalışmada deneysel ağırlıklandırma yöntemleri ile elde edilen puanların iki kategorili (1-0) puanlama ile elde edilen puanlardan daha güvenilir sonuçlar verdiği sonucuna ulaşılmıştır. Claudy (1978) en yüksek güvenirlik katsayısının 0,86 ile çift serili ağırlıklandırmaya ait olduğunu raporlamış ayrıca 0,73 güvenirlik katsayısı ile en düşük değeri düzeltme formülü ile elde edilen puanların verdiğini de belirtmiştir. Bu çalışmada öğrencilerin işaretledikleri seçeneklere göre oran ağırlıklandırması da yapılmıştır. İki kategorili puanlamaya göre puanlanan testlerde öğrencilerin sıralaması alınmış %25 ve %50'lik iki üst gruba göre seçenek ağırlıklandırması yapılmıştır. Bu ağırlıklandırmaya göre de sırasıyla güvenirlik değerleri 0,80 ve 0,81 bulunmuştur. Ağırlıklandırmalar üst gruplara göre yapıldığı için iki güvenirlik değeri arasında yüksek farklar bulunmadığı söylenebilir.

Jaradat ve Tollefson (1988) çalışmalarında 4 tane paralel formlu testin şans başarısını düzeltme formülü ile puanlama, eleyerek puanlama ve kapsayarak puanlama yöntemleri ile puanlanmaları ile elde edilen geçerlik ve güvenirlik değerlerini karşılaştırmayı amaçlamışlardır. Çalışma 54 lisansüstü öğrencisine 104 maddelik soru havuzundan rastgele oluşturulan 4 farklı formun uygulanması ile gerçekleştirilmiştir. Güvenirlik kestirimleri için iç tutarlık katsayılarından KR-20 ve Cronbach Alfa (α) değerleri hesaplanmış, geçerlik kontrolü içinse öğrencilerin açık uçlu test formu puanları, vize notu puanları ve ders kapsamında yaptıkları projelerden aldıkları puanlar ölçüt kabul edilerek bu değerlerin arasındaki korelasyon katsayıları incelenmiştir. Çalışmanın sonuçlarına göre hem güvenirlik hem geçerlik yönünden farklı puanlama türleri arasında anlamlı farklılıklar bulunmamıştır. Bu çalışmada testlere ek olarak öğrencilere uygulanan puanlama yöntemlerinden hangilerinin bilgilerini ölçmede daha iyi olacağını düşündükleri sorulmuştur. Popülasyonun %69'u eleyerek ve kapsayarak puanlamanın bilgilerini tek cevabı işaretledikleri yöntemlere göre daha iyi ölçtüğünü düşündüklerini %52'si ise iki kategorili (1-0) puanlamayı tercih ettiklerini ifade etmiştir.

Akkuş ve Baykul (2001) bir matematik testinin madde ve test istatistiklerinin, iki kategorili (1-0) puanlama, önsel ağırlıklandırma ve Zinger puanlamaları (Z1, Z2) olmak üzere dört farklı puanlama türüne göre puanlandığında nasıl değiştiğini incelemişlerdir. Tüm puanlama yöntemlerinde güvenirlik kestirimi için Cronbach alfa katsayıları elde edilmiştir. Bu çalışmada farklı puanlama türlerinin en yüksek katsayının 0,90 ile 1-0 puanlamaya, en düşük katsayısının 0,86 ile önsel ağırlıklandırmaya ait olduğu belirtilmiş ancak katsayıların Fischer'ın z istatistiğine dönüştürüldüğünde aralarında anlamlı bir fark bulunamadığı ortaya konmuştur.

Özdemir (2002) çalışmasında iki kategorili (1-0) puanlama ile önsel ağırlıklandırma yöntemleri ile puanlanan çoktan seçmeli bir testin klasik test teorisi ve örtük özellikler teorisine göre psikometrik özelliklerini incelemiştir. Örtük özellikler teorisinde ağırlıklı puanlama ile marjinal güvenirlik katsayısından elde edilen güvenirlik katsayısını 0,68 olarak klasik test teorisindeki 1-0 puanlama ile KR-20 güvenirlik katsayısından elde edilen güvenirlik katsayısını ise 0,49 olarak raporlamıştır. Bu çalışmada sunulan bir diğer sonuç, örtük özellikler teorisinde 1-0 puanlama ile bulunan Lord'un güvenirlik kestiriminin 0,78 katsayısının, klasik test teorisinde ağırlıklı puanlamadan elde edilen 0,54 Cronbach Alfa katsayısından daha yüksek olmasıdır.

Özdemir (2002) ve Akkuş ve Baykul (2001) çalışmalarında ortak olarak ortaya atılan sonuçlardan biri de önsel diğer adıyla uzman kanısına dayalı puanlandırma yapabilmek için test maddesi hazırlamanın çok uzun süren bir süreç olduğudur. Ayrıca Özdemir (2002) uzmanların test maddelerinin seçeneklerini ağırlıklandırırken tutarlı olmadıkları durumlar sebebiyle zorluklar yaşandığını bu nedenle farklı çalışmalarda önsel ağırlıklandırma yerine deneysel ağırlıklandırmanın kullanılabileceğini ifade etmiştir.

Gözen Çıtak (2010) çalışmasında çoktan seçmeli bir sözel yetenek testini üniversite öğrencilerine uygulamış ve testi iki kategorili (1-0) puanlama, uzman kanısına dayalı seçenek ağırlıklandırma ve deneysel seçenek ağırlıklandırma yöntemlerine göre puanlamış ve bu yöntemlerin testin geçerlik ve güvenirliğine etkisini incelemiştir. Çalışmada testin

güvenirlik değerleri iki kategorili 1-0 puanlama da 0,64 uzman kanısına dayalı seçenek ağırlıklandırma da 0,68 ve deneysel seçenek ağırlıklandırma yönteminde 0,69 bulunmuştur. Değerlerde net bir niceliksel fark bulunmadığı belirtilmiştir. Bu çalışmada öğrencilerin ortak aldıkları sözel bir derse ait puanları ölçüt kabul edilerek bu puanlarla testin farklı şekillerde puanlanmasından elde edilen puanların arasındaki korelasyon katsayıları hesaplanarak ölçüt geçerliği kontrol edilmiştir. Klasik test kuramına göre iki kategorili (1-0) puanlama yönteminde geçerlik katsayısının 0,55 ile geçerlik katsayısı 0,52 bulunan ağırlıklı puanlama yöntemlerinden daha yüksek değer verdiği raporlanmıştır. Çıtak'ın çalışmasında yapı geçerliğini kontrol etmek adına üniversitede bölümüne sözel puanla yerleşmiş öğrencilerle bölümüne sayısal puanla yerleşmiş öğrencilerin sözel yetenek testinden aldıkları puanlar karşılaştırılmıştır. İki grubun test puanı ortalamalarının karşılaştırılmalarında, bütün puanlama yöntemleri için sözel puanla yerleşen öğrenciler lehine 0,01 düzeyinde anlamlı farklılık bulunduğu ifade edilmiştir. Bu sonuca göre üç puanlama yöntemiyle de iki grup arasındaki ayrım yapılabildiği için puanlama yöntemlerinin geçerli olduğu sonucuna varılmıştır.

Diedenhofen ve Musch (2019) çalışmalarında deneysel seçenek ağırlıklandırması, uzman kanısına dayalı ağırlıklandırma ve iki kategorili (1-0) puanlama yöntemlerinin çoktan seçmeli bir testin geçerlik ve güvenirliğine etkilerini incelemişlerdir. 675 kişiye 27 soruluk bir çoktan seçmeli testin uygulandığı çalışmada deneysel ağırlıklandırmanın iki kategorili (1-0) puanlamaya göre testin güvenirliğini arttığını, uzman kanısına dayalı ağırlıklandırmanın ise güvenirliği artırmadığı belirtilmiştir. Geçerlikte ise puanlama türleri arasındaki farkların istatistiksel olarak anlamlı olmadığı sonucuna varılmıştır.

Yaşar vd. (2021) çalışmalarında bir testin maddelerinin, iki kategorili (1-0) puanlama ve madde güçlüğüne göre ağırlıklandırılmış olarak puanlanmasının öğrencilerin dersi geçme ve kalma durumlarını nasıl değiştirdiğini incelemeyi amaçlamıştır. Çalışmanın sonuçları 34 maddelik çoktan seçmeli bir testin 431 kişilik bir gruba uygulanmasından elde edilen verilere göre ortaya çıkmıştır. İki yöntemden elde edilen puanlara göre Mcdonald's

Omega iç tutarlık katsayıları sırasıyla 0,73 ve 0,72 olarak bulunmuş, birbirlerine yakın oldukları belirtilmiştir. 50 ile 60 puan ölçüt alınıp öğrencilerin başarılı-başarısız olma durumları incelendiğinde ise soruların madde güçlüğüne göre ağırlıklandırıldığı yöntemde daha fazla öğrencinin başarısız kategorisinde olduğu bulunmuştur. Bu çalışmanın bir diğer sonucu olarak madde güçlüğüne göre ağırlıklandırma yapılan puanlama yönteminde bireyler arasındaki farkın daha iyi ortaya koyulduğu ifade edilmiştir

İlgili araştırmalar incelendiğinde güven testi puanlama yöntemine sık rastlanmadığı ve Dressel ve Schmid (1953)'in çalışmalarında güven testi yönteminin farklı seviyelerdeki öğrencileri eşit düzeyde ayırdığı sonucuna varılmıştır. Seçenek ağırlıklandırma üzerine yapılan çalışmalarda ise Patnaik ve Traub (1973) uzman kanısına dayalı ağırlıklandırma ile güvenirlikte diğer puanlama türlerine göre yüksek sonuçlara ulaşıldığını belirtirken Gözen Çıtak (2010)'ın uzman kanısına dayalı ağırlıklandırma ile deneysel ağırlıklandırmayı karşılaştırdığı çalışmasında ise puanlama türleri arasında net bir fark bulunamadığı görülmektedir. Claudy (1978) ve Diedenhofen ve Musch (2019)'un çalışmalarında ağırlıklı puanlamadan elde edilen güvenirlik değerleri iki kategorili puanlamadan elde edilen güvenirlik değerlerine göre yüksek bulunmuştur. Jaradat ve Tollefson (1958) ile Diedenhofen ve Musch (2019)'un çalışmalarında geçerlik için farklı puanlama türlerinde anlamlı fark bulunamadığı belirtilirken Patnaik ve Traub (1973)'un çalışmalarında iki kategorili (1-0) puanlamadan elde edilen puanların ağırlıklı puanlamadan elde edilenlere göre daha geçerli olduğu bulunmuştur.

Bölüm 3

Yöntem

Araştırmanın Türü

Bu araştırma, 8. sınıf düzeyine uygulanmak üzere hazırlanan 20 soruluk bir matematik başarı testinin farklı puanlama yöntemleri ile puanlanmasının test ve madde özelliklerinin belirlenmesi ve karşılaştırılması üzerine olması nedeniyle betimsel araştırma türündedir. Çalışmanın bağımlı değişkenleri madde özellikleri ile geçerlik ve güvenirlik olup bağımsız değişkenleri farklı puanlama yöntemlerinden, iki kategorili (1-0) puanlama, deneysel ağırlıklandırma ve güven testi olarak belirlenmiştir.

Araştırmanın Çalışma Grubu

Araştırmada farklı puanlama türlerinin kullanılmasının çoktan seçmeli testlerin madde özellikleri, geçerlikleri ve güvenirliklerine etkilerinin incelenmesine yönelik çalışıldığından evrene genelleme yapılmayacaktır. Bu bağlamda çalışma grubu oluşturmak için amaca uygun olacak kadar yeterli kişi sayısı belirlenmiş, araştırma İstanbul ilinin Kartal ve Pendik ilçelerinden rastgele seçilen 5 farklı ilköğretim okulunda 8. Sınıfta öğrenim gören toplam 486 öğrencinin uygulamaya katılması ile gerçekleştirilmiştir.

Verilerin bilgisayar ortamına aktarılması sürecinde 180 öğrencinin sorulara tamamen rastgele işaretlemeler yaptığı fark edilmiş, bu öğrencilere ait cevap kağıtları analize dahil edilmemiştir. 5 farklı okuldan olmak üzere toplam 306 öğrencinin cevap kağıdı üzerinden analizler yapılmıştır. Farklı okullarda test uygulamalarına katılıp analizlere dahil edilen öğrencilerin dağılımı Tablo 1' de sunulmuştur.

Tablo 1

Araştırmaya Katılan Öğrencilerin Okullara ve Cinsiyete Göre Dağılımı

Okul No	Kız	Erkek	Toplam
Okul 1	0	50	50
Okul 2	16	14	29
Okul 3	53	74	128
Okul 4	91	0	91
Okul 5	4	4	8
Toplam	164	142	306

Araştırmaya katılan öğrencilerin okullara göre cinsiyet dağılımı Tablo 1'de sunulmuştur. Tablo 1'e göre araştırmaya Okul 1'de 50 erkek öğrenci, Okul 2'de 16 kız ve 14 erkek öğrenci, Okul 3'te 53 kız 74 erkek öğrenci, Okul 4'te 91 kız öğrenci ve Okul 5'te 4 kız 4 erkek öğrenci katılmıştır.

Veri Toplama Süreci

Araştırmanın uygulaması için öncelikle İstanbul Valiliği İl Milli Eğitim Müdürlüğünden gerekli izinler alınmış ardından Pendik ve Kartal ilçelerinden rastgele seçilen 5 farklı ilköğretim okulunun okul yöneticileri ve matematik öğretmenleri ile görüşülmüştür. Her bir okulda araştırmanın yapılacağı günler belirlenmiştir. Veriler 20 Aralık 2023- 12 Ocak 2024 tarihleri arasında toplanmıştır. Uygulama yapılmadan önce Veli Onam Formu ve Çocuk Ergen Bilgilendirme formları okul yöneticileri aracılığıyla velilere ve öğrencilere iletilmiş, araştırma hakkında gerekli bilgilendirmeler yapılmıştır.

Veriler tüm okullarda başarı testinin bir ders saati içinde 50 dakika süreyle öğretmenlerin ve araştırmacının gözetiminde uygulanması yoluyla toplanmıştır. Her bir okulda testler dağıtıldıktan sonra sınav yönergeleri öğrencilere araştırmacı tarafından sözlü olarak aktarılmıştır. Testin puanlanmasında kullanılacak yöntemlerden biri olan güven testi puanlaması için her soru maddesinin alt kısmında bulunan "eminim" ve "emin değilim" ifade

kutucuklarından birinin o madde özelinde kendi kararlarına göre işaretlenmesi gerektiği tüm öğrencilere aktarılmıştır.

Veri Toplama Araçları

Araştırmada kullanılan başarı testi Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Müdürlüğü tarafından hazırlanıp daha önceki senelerde uygulanmış, ve kamuoyuna açık erişimle paylaşılmış olan Liselere Geçiş Sistemi (LGS) sınav soruları içinden seçilmiştir. Başarı testi oluşturulması sürecinde 2018, 2019, 2020, 2021 ve 2022 olmak üzere 5 seneye ait LGS sayısal bölümü matematik sınav sorularının tamamı taranmıştır. Başarı testinin hazırlanması sürecinde bir matematik eğitimi alanı bilim uzmanı ve bir matematik öğretmeninin uzman görüşlerine başvurulmuştur. Araştırma da kullanılacak başarı testi 8. sınıf müfredatı kapsamında Çarpanlar ve Katlar, Üslü İfadeler ve Kareköklü İfadeler konularına ait soruların arasından bu konuların alt kazanımlarını yeterince temsil ettiği düşünülen 20 farklı sorunun seçilmesi ile oluşturulmuştur. Seçilen sorularda kullanılan dilin açık ve anlaşılır olduğu kontrol edilmiştir. Farklı anlamlar çıkarılabilecek hikayeli sorulara yer verilmemeye çalışılmıştır. Öğrenme dışı faktörlerin etkili olmaması adına soru hikayeleri incelenmiş, test konusu dışında kalan ancak kullanımı gerekli olabilecek bilgilerin verilmiş olmasına dikkat edilmiştir. Uzmanların dönütlerine göre soruların güçlük düzeyleri, konu dağılımı, öğrencilerin maddeleri çözmesi için öngörülen süreler gibi faktörler göz önünde bulundurularak sorular sıralanmış ve test haline getirilmiştir.

Veri Toplama Araçlarını Puanlama Yöntemleri

Testlerin uygulanmasından elde edilen veriler puanlandırılırken farklı puanlama yöntemlerinden yararlanılmıştır. Testi puanlama da iki kategorili (1-0) puanlama, deneysel ağırlıklandırma ve güven testi yöntemleri kullanılmıştır.

İki kategorili Puanlama (1-0). İki kategorili puanlamada soruların anahtarlanmış doğru cevabını işaretleyen öğrenciler ilgili maddeden “1”, yanlış seçeneklerden birini

işaretleyen, maddeyi yanıtsız bırakan ya da birden çok seçeneği işaretleyen öğrenciler ise ilgili maddeden “0” puan almıştır.

Güven Testi ile Puanlama Yöntemleri. Çok kategorili puanlama yöntemlerinden olan güven testi ile puanlamada iki farklı yöntem kullanılmıştır. Güven testi puanlamasının birinci versiyonu için test maddelerinin her birinin altında konumlandırılmış “eminim” ve “emin değilim” ifadeleri baz alınmıştır. Bir maddede hem anahtarlanmış doğru cevabı hem de eminim seçeneğini işaretleyen öğrenciler “1” tam puan alırken, anahtarlanmış doğru cevabı işaretleyip emin değilim seçeneğini işaretleyen öğrenciler “0” puan almışlardır. Diğer tüm durumlar için de öğrencilere “0” puan verilmiştir. Güven testi puanlamasının ikinci versiyonunda ise maddede hem anahtarlanmış doğru cevabı hem de eminim seçeneğini işaretleyen öğrenciler “2” tam puan alırken, anahtarlanmış doğru cevabı işaretleyip emin değilim seçeneğini işaretleyen öğrenciler “1” puan almışlardır. Diğer tüm durumlar için öğrencilere “0” puan verilmiştir.

Deneysel Ağırlıklandırma Yöntemi. Çok kategorili puanlama yöntemlerinden biri olan deneysel ağırlıklandırma yönteminde ise oran ağırlıklandırması uygulanmıştır. Özdemir (2003) deneysel ağırlıklandırmada öğrencilerin her bir maddeye verdiği yanıtların dağılımından yararlanıldığını ve seçeneklerin işaretlenme yüzdelerine göre ağırlıklandırıldığını belirtmiştir. Bu araştırmada da uygulama yapılan çalışma grubunun her bir maddenin seçeneklerinden en çok işaretlenen seçeneğinden en az işaretlenen seçeneğine doğru seçenek ağırlıklandırması yapılmıştır. En çok seçilen seçeneği işaretleyen öğrencilere 4 puan verilmiş, ardından en az işaretlenenlere doğru işaretledikleri seçenek özelinde 3, 2 ve 1 puan verilmiştir. Maddeyi boş bırakan öğrenciler o madde özelinde “0” puan almıştır.

Verilerin Analizi

Verilerin analizlerinin yapılması için Microsoft Office Excel, IBM SPSS 26 ve Test Analysis Program (TAP) adlı programlar kullanılmıştır. Öncelikle öğrencilerin test cevap

kağıtlarından her bir maddede işaretledikleri seçenekler Excel programına aktarılmış daha sonra kullanılan puanlama yöntemlerine özel puanlamalar yapılmıştır. Madde ve test istatistiklerinin hesaplanmasında Excel, SPSS ve TAP programlarından faydalanılmıştır.

Tablo 2'de uygulanan testin iki kategorili (1-0) puanlamaya göre elde edilen betimsel istatistikleri sunulmuştur.

Tablo 2

İki Kategorili Puanlama Yöntemine Göre Testin Betimsel İstatistikleri

Puanlama Yöntemi	Aritmetik Ortalama	Standart Sapma	Ortanca	En yüksek puan	En düşük puan	Çarpıklık	Basıklık
İki kategorili Puanlama (1-0)	11,11	4,30	11	20	2	0,162	-0,747

Tablo 2 incelendiğinde 20 maddelik uygulanan test 1-0 puanlama göre puanlandığında aritmetik ortalamanın 11,11, ortancanın 11 olduğu görülmektedir. Testin çarpıklık katsayısı 0,162, basıklık katsayısı ise-0,747 olarak elde edilmiştir. Çarpıklık ve basıklık katsayıları +1,-1 aralığında olduğu için normal dağıldığı söylenebilir

Araştırmanın birinci alt probleminin birinci kısmı doğrultusunda her bir puanlama yönteminde madde özelliklerinin nasıl değiştiğini incelemek adına bütün maddelerin madde güçlük indeksleri (p_j) hesaplanmıştır. Madde güçlük indeksi, maddeyi doğru cevaplayanların sayısının toplam cevaplayıcı sayısına oranı olarak tanımlanmaktadır. [0,1] aralığında değer alabilen güçlük indeksinin madde puanlarının dağılımını da betimleyen bir değer olduğu da ifade edilmektedir (Baykul, 2021). "Maddeyi doğru cevaplayan öğrenci sayısı arttıkça değer 1'e yaklaşacak madde kolaylaşacak, yanlış cevaplayan öğrenci sayısı arttıkça değer 0'a yaklaşacak yani madde zorlaşacaktır." (Baykul, 2021, s.220). Madde güçlük indeksinin aralığı 0,00-0,29 aralığındaki sorular zor, 0,30-0,69 aralığındaki sorular normal, 0,70-1,00 aralığındaki sorular ise kolay sorular olarak tanımlanmaktadır (Özçelik,

2013). İki kategorili puanlama ile güven testi puanlamasının birinci versiyonunda madde güçlük indeksi aşağıda verilen eşitlik yardımıyla hesaplanmıştır:

$$Madde\ güçlük\ indeksi\ (p_j) = \frac{D_J}{N}$$

D_j: $j\ maddesini\ doğru\ cevaplayan\ kişi\ sayısı$

N: $Cevaplayıcı\ sayısı$

Güven testi puanlamasının ikinci versiyonunda madde güçlük indeksi aşağıda verilen eşitlik yardımıyla hesaplanmıştır:

$$Madde\ güçlük\ indeksi\ (p_j) = \frac{\bar{X}_J}{2}$$

$\bar{X}_j$: $j\ maddesiniden\ alınan\ puanların\ aritmetik\ ortalaması$

Deneysel ağırlıklandırma puanlamasında madde güçlük indeksi aşağıda verilen eşitlik yardımıyla hesaplanmıştır:

$$Madde\ güçlük\ indeksi\ (p_j) = \frac{\bar{X}_J}{4}$$

$\bar{X}_j$: $j\ maddesiniden\ alınan\ puanların\ aritmetik\ ortalaması$

Farklı puanlama yöntemlerinden elde edilen madde güçlük indekslerinin arasında anlamlı farklılık olup olmadığını belirlemek için Friedman testi yapılmıştır. Ardından puanlama yöntemlerine ait madde güçlük indekslerinin ikili karşılaştırmalarını yapmak için Wilcoxon işaretli sıralar testi yapılmıştır. Farklı puanlama yöntemlerinden elde edilen güçlük indekslerinin aralarındaki ilişki bir de Pearson korelasyon katsayısı hesaplanarak incelenmiştir.

Araştırmanın birinci alt probleminin ikinci kısmı doğrultusunda dört farklı puanlama türüyle puanlanan maddelerin ayırt edicilik indeksleri hesaplanmıştır. “Madde ayırt edicilik indeksi, testin ölçmek istediği özelliğe sahip olanlar ile olmayanları birbirinden ne ölçüde ayırdığını gösteren bir korelasyon katsayısıdır” (Turgut & Baykul, 2015, s.227). Ayırt edicilik

indeksi 0,19'dan küçük olan maddelerin testten çıkarılması gerektiği, 0,20-0,29 aralığında olan maddelerin teste düzeltilerek koyulması gerektiği ve 0,30 dan yüksek değeri olan maddelerin olduğu gibi teste konulabileceği belirtilmiştir (Turgut & Baykul, 2019). İki kategorili puanlama yöntemine ait madde ayırt edicilik indeksi aşağıda verilen eşitlik yardımıyla hesaplanmıştır:

$$Madde\ ayırt\ edicilik\ indeksi\ (r_b) = \left(\frac{\mathrm{Y_p} - \mathrm{Y_q}}{S_Y}\right)\frac{pq}{y}$$

p : Sürekseg değişkenin birinci kategorisindeki ölçüm sayısının toplam içindeki oranı

q : Sürekseg değişkenin ikinci kategorisindeki ölçüm sayısının toplam içindeki oranı

$\mathrm{Y_p}$ ve $\mathrm{Y_q}$: Yapay sürekseg değişkene ait iki kategorinin sürekli değişken için ortalamalar

S_Y : Sürekli değişkene ait tüm ölçümlerin standart sapması

y : Normal dağılım eğrisi altında kalan alanda p ve q'yu ayıran noktanın ordinat yüksekliği

Güven testi puanlamaları ile deneysel ağırlıklandırma puanlamalarında ise madde ayırt edicilik indeksleri toplam test puanları ile her madde arasındaki Pearson korelasyon katsayısı hesaplanarak elde edilmiştir.

Farklı puanlama yöntemlerinden elde edilen madde ayırt edicilik indekslerinin arasında anlamlı farklılık olup olmadığını belirlemek için Friedman testi yapılmıştır. Ardından puanlama türlerine ait madde ayırt edicilik indekslerinin ikili karşılaştırmalarını yapmak için Wilcoxon işaretli sıralar testi yapılmıştır.

Araştırmanın ikinci alt problemi doğrultusunda geçerlik analizleri için öncelikle öğrencilerin okullarından bu araştırmada kullanılan testin konularına paralel konularla ilgili sınavlarının sonuçları alınmıştır. Öğrencilerin okul akademik başarılarının geçerli olduğu varsayılmış ve ölçüt kabul edilmiştir. Farklı puanlama yöntemlerinden elde edilen toplam puanlar ile okul puanlarının ilişkileri Pearson korelasyon katsayısı hesaplanarak incelenmiştir. "Pearson korelasyon katsayısı iki değişkenin arasındaki ilişkiyi açıklamak ve ilişkinin miktarını bulup yorumlamak amacıyla kullanılır" (Büyüköztürk, 2021, s.31). Ölçüt

puanları iki kategorili (1-0) puanlama ile elde edilmiştir. Bu sebeple araştırmanın iki kategorili (1-0) puanlama ile elde edilen puanları ile aralarında yüksek korelasyon çıkabileceği düşünülmektedir. İkinci alt problemin geçerlik analizleri için okul akademik başarıları ile korelasyonların hesaplanmasına ek olarak farklı puanlama yöntemlerinden elde edilen puanların birbirleri için ölçüt kabul edildiklerinde aralarındaki ilişki Pearson korelasyon katsayıları hesaplanarak kontrol edilmiştir.

Araştırmanın üçüncü alt problemi doğrultusunda testin güvenirlik kestirimleri kullanılan puanlama türlerinin yapılarına göre yapılmıştır. Test iki kategorili (1-0) puanlama ve güven testi puanlamasının birinci versiyonu ile puanlandığında Kuder Richardson formüllerinden KR-20 güvenirlik katsayısı aşağıdaki eşitlik yardımıyla hesaplanmıştır (Turgut & Baykul, 2019, s.123):

$$KR-20 \;=\; \frac{K}{K-1}\left[\frac{\sum_{j=1}^{K} P_j(1-P_j)}{S_x^2}\right]$$

$K:\ Testte\ yer\ alan\ madde\ sayısı$

$P_J: j\ maddesini\ doğru\ cevaplandıranların$ öğ$renci\ sayısına\ oranı$

$S_x^2:\ Testin\ varyansı$

Test deneysel ağırlıklandırma ve güven testi puanlamasının ikinci versiyonu ile puanlandığında ise Cronbach Alfa değerleri hesaplanmıştır. Güvenirlik kestirimleri aşağıda verilen eşitlik yardımıyla hesaplanmıştır (Baykul, 2021, s.145):

$$r \;=\; \alpha \;=\; (\frac{K}{K-1})(1-\frac{\sum_{j=1}^{K}\sigma^2\,(X_j)}{\sigma^2(X)})$$

$\alpha:\ Alfa\ güvenirlik\ katsayısı$

$K: Testte\ yer\ alan\ madde\ sayısı$

$\sigma^2(X_j): Maddelerin\ varyansı$

$\sigma^2(X): Testin\ varyansı$

Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde araştırma verilerinin analizine dayalı olarak alt problemlere cevap olabilecek bulgulara yer verilmiştir. Bulguların sunumunda alt problemlerdeki sıra takip edilmiştir. Önce alt problem tekrar yazılmış, alt probleme çözüm olabilecek bulgulara yer verilmiş ve bulgulara yönelik yorumlar yapılmıştır.

Tablo 3'te alt problemlerin daha iyi incelenebilmesi için bütün farklı puanlama türlerine ilişkin betimsel istatistiklere yer verilmiştir. İki kategorili (1-0) puanlama, güven testi ile puanlama ve deneysel ağırlıklandırma ile puanlama yöntemlerinden elde edilen test puanlarının aritmetik ortalamaları, test puanlarından elde edilen ortalama güçlükler, standart sapmaları ve her bir puanlama yöntemine göre alınabilecek en yüksek puanlar ile her birine ait en yüksek ve en düşük puanlar verilmiştir.

Tablo 3

Uygulama Yapılan Grubun Test Puanlarına İlişkin Betimsel İstatistikler

Puanlama Yöntemi	Aritmetik Ortalama	Ortalama Güçlük	Standart Sapma	Alınabilecek En Yüksek Puan	En Yüksek Puan	En Düşük Puan
İki kategorili (1-0)	11,11	0,55	4,30	20	20	2
Güven testi (birinci versiyon)	8,84	0,44	4,90	20	20	0
Güven testi (ikinci versiyon)	19,91	0,49	9,00	40	40	3
Deneysel Ağırlıklandırma	60,99	0,76	13,00	80	80	21
Toplam	25,21	0,56	7,80	80	80	0

Tablo 3 incelendiğinde 20 puan üzerinden değerlendirilen iki kategorili (1-0) ve güven testi yönteminin birinci versiyonu ile puanlama yöntemlerinde aritmetik ortalamanın

sırasıyla 11,11 ve 8,84 puan, 40 puan üzerinden değerlendirilen güven testi yönteminin ikinci versiyonu ile puanlamada aritmetik ortalamanın 19,91 ve son olarak 80 puan üzerinden değerlendirilen deneysel ağırlıklandırma ile puanlama yönteminde ise aritmetik ortalamanın 60,99 olduğu görülmektedir.

Tablo 3'e göre araştırmada kullanılan testin farklı puanlama yöntemleri ile puanlanmasından elde edilen ortalama güçlük değerleri iki kategorili (1-0) de 0,55 güven testi yönteminin birinci versiyonunda 0,44 güven testi yönteminin ikinci versiyonunda 0,49 ve son olarak deneysel ağırlıklandırmada ise 0,76 bulunmuştur. Bu durumda uygulanan test iki kategorili (1-0) puanlama yöntemi ve güven testi yönteminin iki farklı versiyonu ile puanlandığında ortalama güçlükleri 0,30-0,69 değer aralığında olduğundan testin normal güçlük kategorisinde olduğu, deneysel ağırlıklandırma puanlamasında ise 0,70-1,00 değer aralığında olduğundan testin kolay güçlük kategorisinde olduğu söylenebilir.

Alt Problemlere Göre Bulgular ve Yorumlar

1.Alt Problem "Matematik başarı testi iki kategorili (1-0) puanlama ve farklı çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin özellikleri nasıl değişmektedir?"

1a) Matematik başarı testi iki kategorili (1-0) puanlama ve farklı çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin güçlük indeksleri nasıl değişmektedir? Bu alt probleme cevap bulabilmek için iki ve çok kategorili puanlamalara göre maddelerin güçlük indeksleri hesaplanmış ve tablo 4'te sunulmuştur.

Tablo 4

Farklı Puanlama Yöntemlerine Göre Hesaplanan Maddelerin Güçlük İndeksleri

	Madde Güçlük İndeksi			
Madde No	İki Kategorili Puanlama	Deneysel Ağırlıklandırma	Güven Testi (birinci versiyon)	Güven Testi (ikinci versiyon)
Madde 1	0,93	0,97	0,86	0,89
Madde 2	0,88	0,96	0,80	0,84
Madde 3	0,75	0,90	0,66	0,70
Madde 4	0,52	0,75	0,44	0,48

Madde 5	0,40	0,75	0,29	0,33
Madde 6	0,59	0,75	0,38	0,49
Madde 7	0,57	0,77	0,40	0,48
Madde 8	0,51	0,74	0,39	0,45
Madde 9	0,58	0,75	0,43	0,51
Madde 10	0,33	0,67	0,21	0,27
Madde 11	0,86	0,94	0,76	0,81
Madde 12	0,34	0,68	0,22	0,28
Madde 13	0,17	0,60	0,09	0,13
Madde 14	0,81	0,91	0,78	0,80
Madde 15	0,44	0,63	0,23	0,34
Madde 16	0,50	0,72	0,41	0,46
Madde 17	0,52	0,75	0,43	0,48
Madde 18	0,50	0,69	0,42	0,46
Madde 19	0,63	0,77	0,48	0,56
Madde 20	0,28	0,56	0,17	0,23
Ortalamalar	0,56	0,76	0,44	0,50

Tablo 4'e göre farklı puanlama türlerinin madde güçlük düzeylerinin ortalamaları hesaplandığında en yüksek ortalama 0,76 ile deneysel ağırlıklandırmaya ait bulunurken en düşük ortalamanın 0,44 ile güven testi yönteminin birinci versiyonuna ait olduğu görülmektedir.

Tablo 4 incelendiğinde iki kategorili (1-0) puanlanan maddelerin güçlük indekslerinin 0,17 ile 0,93 arasında değiştiği görülmektedir. Güven testi yöntemi ile puanlanan maddelerin güçlük indekslerinin birinci versiyonda 0,09 ile 0,86 arasında değiştiği, ikinci versiyonda ise 0,13 ile 0,89 arasında değiştiği görülmektedir. Son olarak deneysel ağırlıklandırma ile puanlanan maddelerin güçlük indekslerinin 0,56 ile 0,97 arasında değiştiği görülmektedir.

Bütün puanlama yöntemlerinde en yüksek madde güçlük indeksi madde 1'e aittir. Farklı puanlama yöntemlerine göre hesaplandığında madde 1'in en düşük güçlük indeksi 0,86 ile güven testinin birinci versiyonu ile puanlandığında en yüksek güçlük indeksi ise 0,97 ile deneysel ağırlıklandırma ile puanlandığında elde edilmiştir. Bu sorunun bütün puanlama yöntemlerinde 0,70'ten büyük güçlük indeksi değeri olması sebebiyle kolay soru kategorisinde olduğu söylenebilir.

Testin güçlük indeksi en düşük olan maddesi, iki kategorili (1-0) puanlama, güven testinin birinci versiyonu ve güven testinin ikinci versiyonu olmak üzere üç puanlama

yönteminde de madde 13 olarak belirlenmiştir. Maddenin iki kategorili (1-0) puanlamada güçlük indeksi 0,17 güven testinin birinci versiyonunda güçlük indeksi 0,09 güven testinin ikinci versiyonunda güçlük indeksi 0,13'tür. Bu yöntemlerde 13. maddenin güçlük indeksinin 0,00-0,29 aralığında olması sebebiyle zor soru kategorisinde olduğu söylenebilir.

Maddelerin zorluk ve kolaylık düzeylerine göre farklı puanlama türlerinde değişimleri incelendiğinde, kolay maddelerden biri olan 1. maddenin en yüksek güçlük düzeyi ile en düşük güçlük düzeyi arasındaki fark 0,11 bulunmuştur. Tüm puanlama yöntemlerinde kolay bulunan 14. maddenin maksimum ve minimum güçlük düzeyleri arasındaki fark ise 0,13'tür. Testte güçlük düzeyi en düşük olan maddelerden 13. madde de ise en yüksek güçlük indeksi ile en düşük güçlük indeksi arasındaki fark 0,51 olarak bulunmuştur. Farklı puanlama türlerinde güçlük düzeyi zor olarak bulunan madde 20 de ise bu farkın 0,39 olduğu görülmektedir. Bu maddelerin güçlük düzeyleri arasındaki farklar incelendiğinde, farklı puanlama türlerine göre kolay maddelerin güçlük indeksleri arasındaki değişimlerin zor maddelerin güçlük indeksleri arasındaki değişimlere göre daha az olduğu görülmektedir.

Puanlama yöntemlerine ait ortanca güçlük indeksi değerleri ise iki kategorili (1-0) puanlama da 0,52 güven testi ile puanlamanın birinci versiyonunda 0,42 güven testi ile puanlamanın ikinci versiyonunda 0,48 ve deneysel ağırlıklandırma ile puanlamada ise 0,75 olarak bulunmuştur.

Tablo 4 daha detaylı incelendiğinde, farklı puanlama yöntemlerine göre güçlük kategorisi değişen maddelerin bulunduğu görülmektedir. Madde 3 güven testinin birinci versiyonunda 0,66 güçlük indeksi ile normal kategorisinde diğer puanlama yöntemlerinde ise kolay kategorisindedir. Madde 4, 5, 6, 7, 8 ve 9 deneysel ağırlıklandırma puanlamasında kolay ancak diğer puanlama yöntemlerinde normal kategorisindedir. Madde 10 güven testi puanlamasının iki versiyonunda da zor kategorisinde diğer yöntemlerde ise normal kategorisindedir. Madde 12 iki kategorili (1-0) puanlama ve deneysel ağırlıklandırma ile puanlandığında normal kategorisinde diğer puanlama yöntemlerinde ise zor kategorisinde bulunmaktadır. Madde 13 ise yalnızca deneysel ağırlıklandırma yönteminde normal, diğer

puanlama yöntemlerinde zor kategorisinde bulunmuştur. Madde 14 tüm puanlama yöntemlerinde kolay kategorisinde bulunmuştur. Madde 15 güven testinin birinci versiyonuyla puanlandığında zor kategorisinde diğer puanlama yöntemleriyle puanlandığında ise normal kategorisindedir. Madde 16, 17 ve 19 deneysel ağırlıklandırma puanlaması ile puanlandığında kolay kategorisinde diğer puanlama yöntemleri ile puanlandığında ise normal kategorisinde bulunmuştur. Madde 18 tüm puanlama yöntemlerinde normal güçlük düzeyindedir. Madde 20 ise deneysel ağırlıklandırma ile puanlandığında normal diğer puanlama yöntemleri ile puanlandığında ise zor kategorisinde bulunmuştur. Bu farklılıklar her puanlama yönteminde anahtarlanmış doğru cevaba verilen puanların değişmesi ile açıklanabilir.

Farklı puanlama yöntemleri ile hesaplanan madde güçlük indekslerinin karşılaştırılması için yapılan Friedman testine ilişkin bulgular Tablo 5'te sunulmuştur.

Tablo 5

Farklı Puanlama Yöntemlerine Göre Hesaplanan Madde Güçlük İndekslerinin Karşılaştırılmasına ilişkin Friedman Testi Sonuçları

	Ortalama Sıralar	Ki-kare	P
İki kategorili	3,00		
Deneysel Puanlama	4,00		
Güven Testi (birinci versiyon)	1,00	57,00	0,000
Güven Testi (ikinci versiyon)	2,00		

*** P<0,01*

Tablo 5 incelendiğinde farklı puanlama türlerine göre hesaplanan maddelerin güçlük indeksleri küçükten büyüğe doğru sıralandığında elde edilen sıra puanları ortalamasına göre 4 aritmetik ortalama ile deneysel puanlama yönteminden elde edilenlerin daha kolay olduğu anlaşılmaktadır. Bu ortalamalara göre madde güçlükleri en zor olanın güven testi yönteminin birinci versiyonuna göre puanlanan maddeler olduğu görülmektedir. Bunu güven testinin ikinci versiyonu takip etmekte üçüncü sırada ise iki kategorili (1-0) puanlanan

maddeler gelmektedir. Analiz sonuçlarına göre en az iki grup arasında istatistiksel olarak anlamlı fark olduğu görülmektedir (X^2= 57,00, p<0,01).

Puanlama yöntemlerine göre elde edilen güçlük indekslerinin ikili olarak karşılaştırmaları için yapılan Wilcoxon İşaretli Sıralar Testi Tablo 6'da sunulmuştur.

Tablo 6

Farklı Puanlama Yöntemlerine Göre Hesaplanan Madde Güçlük İndekslerinin Wilcoxon İşaretli Sıralar Testi Sonuçları

		N	Ortalama Sıralar	Sıralar Toplamı	Wilcoxon Testi (Z)	P
Deneysel ile İki Kategorili (1-0)	Negatif Sıralar	0[a]	,00	,00		
	Pozitif Sıralar	20[b]	10,50	210,00	-3,922	0,000
	Eşitlik	0[c]				
	Toplam	20				
Güven Testi (birinci versiyon) ile İki Kategorili (1-0)	Negatif Sıralar	20[d]	10,50	210,00		
	Pozitif Sıralar	0[e]	,00	,00	-3,927	0,000
	Eşitlik	0[f]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile İki Kategorili (1-0)	Negatif Sıralar	20[g]	10,50	210,00		
	Pozitif Sıralar	0[h]	,00	,00	-3,944	0,000
	Eşitlik	0[i]				
	Toplam	20				
Güven Testi (birinci versiyon) ile Deneysel	Negatif Sıralar	20[j]	10,50	210,00		
	Pozitif Sıralar	0[k]	,00	,00	-3,922	0,000
	Eşitlik	0[l]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile Deneysel	Negatif Sıralar	20[m]	10,50	210,00		
	Pozitif Sıralar	0[n]	,00	,00	-3,922	0,000
	Eşitlik	0[o]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile Güven Testi (birinci versiyon)	Negatif Sıralar	0[p]	,00	,00		
	Pozitif Sıralar	20[q]	10,50	210,00	-3,939	0,000
	Eşitlik	0[r]				
	Toplam	20				

a. Deneysel <İki kategorili
b. Deneysel> İki kategorili
c. Deneysel = İki kategorili
d. Güven Testi (birinci versiyon) < İki kategorili
e. Güven Testi (birinci versiyon) > İki kategorili
f. Güven Testi (birinci versiyon) = İki kategorili
g. Güven Testi (ikinci versiyon) < İki kategorili
h. Güven Testi (ikinci versiyon) > İki kategorili
i. Güven Testi (ikinci versiyon) = İki kategorili
j. Güven Testi (birinci versiyon) < Deneysel
k. Güven Testi (birinci versiyon) > Deneysel
l. Güven Testi (birinci versiyon) = Deneysel
m. Güven Testi (ikinci versiyon) < Deneysel
n. Güven Testi (ikinci versiyon) > Deneysel
o. Güven Testi (ikinci versiyon) = Deneysel
p. Güven Testi (ikinci versiyon) < Güven Testi (birinci versiyon)
q. Güven Testi (ikinci versiyon) > Güven Testi (birinci versiyon)
r. Güven Testi (ikinci versiyon) = Güven Testi (birinci versiyon)

Tablo 6'da sunulan sonuçlar, deneysel ağırlıklandırma ve iki kategorili (1-0) puanlama yöntemleri kullanıldığında madde güçlük indekslerinin arasında anlamlı bir fark olduğunu göstermektedir ($z=-3,922$; $p<0,01$). Fark puanlarının sıra ortalaması ve toplam incelendiğinde, gözlenen farkın deneysel ağırlıklandırma puanlaması lehine olduğu görülmektedir. Sonuçlar, test deneysel ağırlıklandırma yöntemi ile puanlandığında yirmi maddenin tamamının (pozitif sıralar, N=20) güçlük indekslerinin iki kategorili (1-0) puanlamaya göre yüksek olduğunu göstermektedir. Deneysel ağırlıklandırma yönteminden elde edilen güçlük indekslerinin yüksek çıkma nedeni, iki kategorili (1-0) puanlama da yalnızca doğru yanıtlara puan verilirken bu yöntemde öğrencilerin seçeneklere yönelme oranına göre bütün seçeneklere puan atanması olarak görülmektedir.

Tablo 6 incelendiğinde, güven testi yönteminin birinci versiyonu ve iki kategorili (1-0) puanlamanın kullanıldığı durumda madde güçlük indekslerinin arasında anlamlı bir fark olduğunu göstermektedir ($z=-3,927$; $p<0,01$). Fark puanlarının sıra ortalaması ve toplamlara göre gözlenen farkın iki kategorili (1-0) puanlama lehine olduğu görülmektedir. Test iki kategorili (1-0) puanlama ile puanlandığında yirmi maddenin tamamının (negatif sıralar N=20) güçlük indeksleri güven testinin birinci versiyonuna göre daha yüksek bulunmuştur. Bu durum güven testi yönteminin puanlama yapısıyla açıklanabilir. Bu yöntemde iki kategorili puanlamadan farklı olarak öz değerlendirme faktörü de puanlama kriterlerine dahil

olduğundan doğru seçeneği işaretlese bile “emin değilim” ifadesini seçen öğrenciler tam puan alamamış dolayısıyla maddelerin güçlük indeksleri düşmüştür.

Tablo 6’da sunulan sonuçlar, güven testi yönteminin ikinci versiyonu ve iki kategorili (1-0) puanlamanın kullanıldığı durumda madde güçlük indekslerinin arasında anlamlı bir fark olduğunu göstermektedir (z=-3,944; p<0,01). Fark puanlarının sıra ortalaması ve toplamlara göre gözlenen farkın iki kategorili (1-0) puanlama lehine olduğu görülmektedir. Test iki kategorili puanlama ile puanlandığında yirmi maddenin tamamının (negatif sıralar N=20) güçlük indekslerinin güven testi yönteminin ikinci versiyonuna göre daha yüksek olduğu görülmektedir. Bu iki puanlama yöntemi arasındaki fark, güven testinin ikinci versiyonunda da birinci versiyonu gibi maddelerde öz değerlendirme kriterinin etkisiyle daha az öğrencinin tam puan alabilmesiyle açıklanabilir.

Tablo 6 incelendiğinde güven testinin birinci versiyonu ve deneysel ağırlıklandırma karşılaştırmasında madde güçlük indekslerinin arasında anlamlı bir fark olduğunu görülmektedir (z=-3,922; p<0,01). Fark puanlarının sıra ortalaması ve toplam incelendiğinde gözlenen farkın deneysel ağırlıklandırma lehine olduğu görülmektedir. Bu sonuçlara göre, test deneysel ağırlıklandırma ile puanlandığında yirmi maddenin tamamının (negatif sıralar N=20) maddelerin güçlük indekslerinin güven testinin birinci versiyonuna göre daha yüksek olduğu bulunmuştur. Deneysel ağırlıklandırma puanlamasının yapısı nedeniyle işaretlenen her seçeneğe bir miktar puan verildiğinden maddelerin güçlük indekslerinin yüksek çıktığı görülmektedir.

Tablo 6’da sunulan sonuçlar, test güven testinin ikinci versiyonu ve deneysel ağırlıklandırma ile puanlandığında madde güçlük indekslerinin arasında anlamlı bir fark olduğunu göstermektedir (z=-3,922; p<0,01). Fark puanlarının sıra ortalaması ve toplam incelendiğinde, gözlenen farkın deneysel ağırlıklandırma lehine olduğu görülmektedir. Bu sonuçlara göre, test deneysel ağırlıklandırma ile puanlandığında yirmi maddenin tamamının (negatif sıralar, N=20) güçlük indeksi güven testinin ikinci versiyonuyla puanlanmasına göre

daha yüksektir. Bu karşılaştırmadaki fark yine deneysel ağırlıklandırma yönteminde tüm seçeneklere puan atanmasıyla açıklanabilir.

Tablo 6'da sunulan sonuçlar, test güven testinin ikinci versiyonu ve güven testinin birinci versiyonu ile puanlandığında madde güçlük indekslerinin arasında anlamlı bir fark olduğunu göstermektedir ($z=-3,939$; $p<0,01$). Fark puanlarının sıra ortalaması ve toplam incelendiğinde gözlenen farkın güven testinin ikinci versiyonu lehine olduğu görülmektedir. Bu sonuçlara göre, test güven testinin ikinci versiyonu ile puanlandığında yirmi maddenin tamamının (pozitif sıralar N=20) maddelerin güçlük indeksleri güven testinin birinci versiyonuyla puanlanmasına göre daha yüksektir. Güven testi yönteminin iki versiyonunda da öz değerlendirme kriteri olsa da birinci versiyonda, bir maddeyi doğru işaretleyen öğrenciler "emin değilim" ifadesini kullandığında hiç puan alamazken, ikinci versiyonda aynı durumda olan öğrenciler kısmi puan almıştır. Bu durum güven testinin ikinci versiyonunda madde güçlük indekslerinin daha yüksek olmasını sağlamıştır.

Tablo 7'de başarı testinin farklı puanlama yöntemleriyle puanlanmasından elde edilen güçlük indekslerinin korelasyon değerleri sunulmuştur.

Tablo 7

Farklı Puanlama Yöntemlerine Göre Hesaplanan Güçlük İndekslerinin Pearson Korelasyon Analizi Sonuçları

	1	2	3	4
1.İki Kategorili	1			
2. Deneysel Ağırlıklandırma	,954**	1		
3. Güven Testi (birinci versiyon)	,980**	,964**	1	
4.Güven Testi (ikinci versiyon)	,995**	,959**	,995**	1
N	20	20	20	20

*** P<0,01*

Tablo 7 incelendiğinde farklı puanlama türlerinin kullanılmasıyla hesaplanan madde güçlük indekslerinin birbirleriyle korelasyonları incelendiğinde tüm korelasyon değerlerinin 0,950'in üzerinde ve 0,01 düzeyinde anlamlı olduğu görülmektedir. Bu durumda tüm

puanlama yöntemlerinde güçlük indekslerinin pozitif ve yüksek düzeyde ilişkisi olduğu söylenebilir. Başka bir deyişle güçlük indeksleri tüm puanlama yöntemlerinde aynı yönde artma ve azalma eğilimine sahiptir. İki kategorili (1-0) puanlama yöntemi ölçüt olarak kabul edildiğinde, en yüksek ilişkinin 0,99 katsayı ile güven testinin ikinci versiyonuna ait olduğu, bu yöntemi 0,98 katsayı ile güven testinin birinci versiyonunun takip ettiği ve en düşük ilişkinin ise 0,95 katsayı ile deneysel ağırlıklandırmaya ait olduğu bulunmuştur.

1b) Matematik başarı testi iki kategorili (1-0) puanlama ve farklı çok kategorili puanlama yöntemleri ile puanlandığında test maddelerinin ayırt edicilik indeksleri nasıl değişmektedir? Bu alt probleme cevap bulabilmek için iki ve çok kategorili puanlamalara göre maddelerin ayırt edicilik indeksleri hesaplanmış ve tablo 8'de sunulmuştur.

Tablo 8

Farklı Puanlama Yöntemlerine Göre Hesaplanan Maddelerinin Ayırt Edicilik İndeksleri

Madde No	Madde Ayırt Edicilik İndeksleri			
	İki Kategorili Puanlama	Deneysel Ağırlıklandırma	Güven Testi (birinci versiyon)	Güven Testi (ikinci versiyon)
Madde 1	0,32	0,25	0,47	0,43
Madde 2	0,32	0,20	0,47	0,42
Madde 3	0,46	0,35	0,52	0,50
Madde 4	0,55	0,47	0,61	0,60
Madde 5	0,42	0,44	0,49	0,47
Madde 6	0,50	0,56	0,59	0,57
Madde 7	0,49	0,53	0,62	0,58
Madde 8	0,41	0,47	0,57	0,52
Madde 9	0,57	0,62	0,62	0,62
Madde 10	0,38	0,41	0,47	0,44
Madde 11	0,45	0,43	0,47	0,49
Madde 12	0,40	0,58	0,50	0,47
Madde 13	0,33	0,52	0,40	0,38
Madde 14	0,40	0,34	0,42	0,42
Madde 15	0,49	0,63	0,51	0,53
Madde 16	0,54	0,67	0,62	0,60
Madde 17	0,59	0,61	0,66	0,65
Madde 18	0,66	0,72	0,67	0,68
Madde 19	0,56	0,62	0,63	0,62
Madde 20	0,55	0,71	0,57	0,59
Ortalamalar	0,47	0,51	0,54	0,53

Tablo 8'e göre farklı puanlama türlerinin madde ayırt edicilik indekslerinin ortalamaları hesaplandığında en yüksek ortalama 0,54 ile güven testinin birinci versiyonuna ait bulunurken en düşük ortalamanın 0,47 ile iki kategorili puanlamaya aittir.

Tablo 8' de bulunan veriler incelendiğinde iki kategorili (1-0) puanlamada maddelerin ayırt edicilik indekslerinin 0,32 ile 0,66 arasında değiştiği görülmektedir. Ayırt edicilik gücü en düşük olan maddelerin 0,32 katsayı ile madde 1 ve madde 2 olduğu ayırt edicilik indeksi en fazla olan maddenin ise 0,66 katsayı ile madde 18 olduğu görülmektedir.

Diğer puanlama yöntemlerinden elde edilen madde ayırt edicilik değerleri incelendiğinde ise deneysel ağırlıklandırma yönteminde madde ayırt edicilik indekslerinin 0,20 ile 0,72 arasında değiştiği, güven testinin birinci versiyonunda ise 0,40 ile 0,67 arasında son olarak güven testinin ikinci versiyonunda 0,38 ile 0,68 arasında değiştiği görülmektedir. Deneysel ağırlıklandırma yönteminde hiçbir maddenin çıkarılacak düzeyde olmadığı ancak 0,20 katsayı ile madde 2'nin ve 0,25 katsayı ile madde 1'in düzeltilerek teste eklenmesi gerektiği bulunmuştur. 0,72 katsayı ile ayırt edicilik indeksi en yüksek çıkan maddenin 18. madde olduğu görülmektedir. Güven testinin birinci versiyonunda da hiçbir maddenin testten çıkarılacak ya da düzeltilecek düzeyde olmadığı görülmektedir. Bu yöntemde de 0,67 katsayı ile ayırt edicilik indeksi en yüksek maddenin 18. Madde olduğu görülmektedir. Güven testinin ikinci versiyonuna ait değerler incelendiğinde ise hiçbir maddenin testten çıkarılması veya düzeltilmesi gerekmediği sonucuna varılmıştır. Bu yönteme göre ayırt edicilik indeksi en düşük olan madde 0,38 katsayı ile madde 13 ayırt edicilik indeksi en fazla olan madde ise 0,68 ile 18.maddedir.

Tüm puanlama yöntemlerinde ayırt edicilik indeksi en yüksek olan madde 18. Madde olarak bulunmuştur. Farklı puanlama yöntemlerine göre genel olarak madde 1, madde 2 ve madde 13'ün teste düzeltilerek koyulması ya da çıkarılması gerektiği sonuçlarına varılmıştır. Tüm puanlama yöntemlerinde ayırt edicilik indeksi düşük çıkan 13.madde deneysel ağırlıklandırma yönteminde 0,52 ile nispeten yüksek bir sonuç vermiştir. Bu durum deneysel ağırlıklandırma yönteminde seçenekler ağırlıklandırırken öğrencilerin madde 13'te anahtarlanmış doğru cevabı yanlış bir seçenekten daha az seçmeleri sonunda yanlış cevabın doğru cevaptan daha fazla puan almış olmasıyla açıklanabilir.

Maddelerin ayırt edicilik düzeylerine göre farklı puanlama türlerinde değişimleri incelendiğinde, ayırıcılığı yüksek olan 18. maddenin en yüksek ayırt edicilik indeksi ile en düşük ayırt edicilik indeksi arasındaki fark 0,06 bulunmuştur. Tüm puanlama yöntemlerinde ayırt edicilik indeksi nispeten yüksek olan 9. maddenin maksimum ve minimum ayırt edicilik indeksi arasındaki fark ise 0,03'tür. Testte ayırt edicilik indeksi en düşük olan maddelerden 1. madde de ise maksimum ve minimum ayırt edicilik indeksi arasındaki fark 0,22 olarak bulunmuştur. Farklı puanlama türlerinde yine ayırt edicilik indeksi düşük bulunan madde 2 de ise bu farkın 0,27 olduğu görülmektedir. Bu maddelerin ayırt edicilik indeksleri arasındaki farkların ayırt edicilik indeksleri yüksek olan maddelerde ayırt edicilik indeksi düşük olan maddelere göre daha az olduğu görülmektedir.

Farklı puanlama yöntemleri ile hesaplanan madde ayırt edicilik indekslerinin karşılaştırılması için yapılan Friedman testine ilişkin bulgular Tablo 9'da sunulmuştur.

Tablo 9

Farklı Puanlama Yöntemlerine Göre Hesaplanan Madde Ayırt Edicilik İndekslerinin Karşılaştırılmasına ilişkin Friedman Testi Sonuçları

	Ortalama Sıralar	Ki-kare	P
İki kategorili	1,3	30,57	0,000
Deneysel	2,38		
Güven Testi (birinci versiyon)	3,43		
Güven Testi (ikinci versiyon)	2,9		

Tablo 9 incelendiğinde farklı puanlama türlerine göre hesaplanan maddelerin ayırt edicilik indeksleri küçükten büyüğe doğru sıralandığında elde edilen sıra puanları ortalamasına göre 3,43 aritmetik ortalama ile güven testinin birinci versiyonundan elde edilenlerin en yüksek ayırt ediciliğe sahip olduğu görülmektedir. Bunu güven testinin ikinci versiyonu takip etmekte üçüncü sırada ise deneysel ağırlıklandırma ile puanlanan maddeler gelmektedir. Ortalamalara göre ayırt ediciliği en az olanın iki kategorili puanlama ile puanlanan maddeler olduğu görülmektedir. Analiz sonuçlarına göre en az iki grup arasında istatistiksel olarak anlamlı fark olduğu görülmektedir (X^2= 30,57; p<0,01).

İkili grupların karşılaştırmaları için yapılan Wilcoxon İşaretli Sıralar Testi Tablo 10'da sunulmuştur.

Tablo 10

Farklı Puanlama Yöntemlerine Göre Hesaplanan Madde Ayırt Edicilik İndeksleri için Wilcoxon İşaretli Sıralar Testi Sonuçları

		N	Ortalama Sıralar	Sıralar Toplamı	Wilcoxon Testi (Z)	P
Deneysel ile İki Kategorili (1-0)	Negatif Sıralar	6[a]	10,83	65	-1,496	0,135
	Pozitif Sıralar	14[b]	10,36	145		
	Eşitlik	0[c]				
	Toplam	20				
Güven Testi (birinci versiyon) ile İki Kategorili	Negatif Sıralar	0[d]	0	0	-3,928	0,000
	Pozitif Sıralar	20[e]	10,5	210		
	Eşitlik	0[f]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile İki Kategorili (1-0)	Negatif Sıralar	0[g]	0	0	-3,931	0,000
	Pozitif Sıralar	20[h]	10,5	210		
	Eşitlik	0[i]				
	Toplam	20				
Güven Testi (birinci versiyon) ile Deneysel	Negatif Sıralar	6[j]	10,5	63	-1,289	0,197
	Pozitif Sıralar	13[k]	9,77	127		
	Eşitlik	1[l]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile Deneysel	Negatif Sıralar	6[m]	10,75	64,5	-0,915	0,360
	Pozitif Sıralar	12[n]	8,88	106,5		
	Eşitlik	2[o]				
	Toplam	20				
Güven Testi (ikinci versiyon) ile Güven Testi (birinci versiyon)	Negatif Sıralar	14[p]	10,21	143	-2,533	0,011
	Pozitif Sıralar	4[q]	7	28		
	Eşitlik	2[r]				
	Toplam	20				

a. Deneysel <İki kategorili
b. Deneysel> İki kategorili

c. Deneysel = İki kategorili
d. Güven Testi (birinci versiyon) < İki kategorili
e. Güven Testi (birinci versiyon) > İki kategorili
f. Güven Testi (birinci versiyon) = İki kategorili
g. Güven Testi (ikinci versiyon) < İki kategorili
h. Güven Testi (ikinci versiyon) > İki kategorili
i. Güven Testi (ikinci versiyon) = İki kategorili
j. Güven Testi (birinci versiyon) < Deneysel
k. Güven Testi (birinci versiyon) > Deneysel
l. Güven Testi (birinci versiyon) = Deneysel
m. Güven Testi (ikinci versiyon) < Deneysel
n. Güven Testi (ikinci versiyon) > Deneysel
o. Güven Testi (ikinci versiyon) = Deneysel
p. Güven Testi (ikinci versiyon) < Güven Testi (birinci versiyon)
q. Güven Testi (ikinci versiyon) > Güven Testi (birinci versiyon)
r. Güven Testi (ikinci versiyon) = Güven Testi (birinci versiyon

Tablo 10'da sunulan sonuçlar, test deneysel ağırlıklandırma ve iki kategorili (1-0) puanlama ile puanlandığında madde ayırt edicilik indekslerinin arasında istatiksel olarak anlamlı bir fark olmadığını göstermektedir (z=-1,496; p>0,01). Deneysel ağırlıklandırma puanlaması kullanıldığında 14 maddenin ayırt edicilik indeksinin iki kategorili puanlamaya göre yüksek olduğu, 6 maddenin ayırt edicilik indeksinin de iki kategorili puanlamada daha yüksek olduğu bulunmuştur.

Tablo 10 incelendiğinde, test güven testinin birinci versiyonu ve iki kategorili (1-0) puanlama ile puanlandığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark olduğunu göstermektedir (z=-3,928; p<0,01). Fark puanlarının sıra ortalaması ve toplam incelendiğinde, gözlenen farkın güven testinin birinci versiyonu lehine olduğu görülmektedir. Bu sonuçlara göre, test güven testinin birinci versiyonu ile puanlandığında yirmi maddenin tamamının (pozitif sıralar N=20) ayırt edicilik indeksleri iki kategorili (1-0) puanlamaya göre daha yüksek bulunmuştur.

Tablo 10'da sunulan sonuçlar, güven testinin ikinci versiyonu ve iki kategorili (1-0) puanlama yöntemleri kullanıldığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark olduğunu göstermektedir (z=-3,931; p<0,01). Fark puanlarının sıra ortalaması ve toplam incelendiğinde, gözlenen farkın güven testinin ikinci versiyonu lehine olduğu görülmektedir. Bu sonuçlara göre, test güven testinin ikinci versiyonu ile puanlandığında

yirmi maddenin tamamının (pozitif sıralar N=20) ayırt edicilik indekslerinin iki kategorili (1-0) puanlamaya göre daha yüksek olduğu görülmektedir.

Güven testinin iki farklı versiyonunun iki kategorili (1-0) puanlama ile karşılaştırılmasından elde edilen sonuçlar, bu yöntemlerin iki kategorili puanlamaya göre test maddelerinin ayırt ediciliğini anlamlı bir farkla artırdığını göstermektedir. Bu durumda iki kategorili puanlanan testlerin puanlamasına öz değerlendirme değişkeninin eklenmesinin test puanlarının varyansını artırması nedeniyle ayırt edicilik gücünü artırdığı yorumu yapılabilir.

Tablo 10 incelendiğinde, güven testinin birinci versiyonu ve deneysel ağırlıklandırma puanlamaları kullanıldığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark olmadığı görülmektedir ($z=-1,289$; $p>0,01$). Test güven testinin birinci versiyonuna göre puanlandığında 13 maddenin ayırt edicilik gücünün deneysel ağırlıklandırmaya göre yüksek çıktığı, 6 maddenin ayırt ediciliğinin ise deneysel ağırlıklandırma ile puanlandığında daha yüksek çıktığı bulunmuştur. 1 maddenin ise iki farklı yöntemde de aynı ayırt edicilik indeksine sahip olduğu görülmektedir.

Tablo 10'da sunulan sonuçlar, test güven testinin ikinci versiyonu ve deneysel ağırlıklandırma ile puanlandığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark olmadığını göstermektedir ($z=-0,915$; $p>0,01$). Test güven testinin ikinci versiyonuna göre puanlandığında 12 maddenin ayırt edicilik gücünün deneysel ağırlıklandırmaya göre yüksek çıktığı, 6 maddenin ayırt ediciliğinin ise deneysel ağırlıklandırma ile puanlandığında daha yüksek çıktığı bulunmuştur.

Tablo 10'a göre, test güven testinin ikinci versiyonu ve güven testinin birinci versiyonu ile puanlandığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark olduğu görülmektedir ($z=-2,533$; $p<0,05$). 14 maddenin ayırt edicilik indeksinin birinci versiyonda daha yüksek 4 maddenin ayırt edicilik indeksinin ise ikinci versiyonda daha yüksek olduğu bulunmuştur. Kalan 2 maddenin iki yöntemde de eşit ayırt edicilik indeksine sahip olduğu bulunmuştur.

Tablo 11'de başarı testinin farklı puanlama yöntemleriyle puanlanmasından elde edilen güçlük indekslerinin korelasyon değerleri sunulmuştur.

Tablo 11

Farklı Puanlama Yöntemlerine Göre Hesaplanan Ayırt Edicilik İndekslerinin Pearson Korelasyon Analizi Sonuçları

	1	2	3	4
1.İki Kategorili	1			
2. Deneysel Ağırlıklandırma	,782**	1		
3. Güven Testi (birinci versiyon)	,880**	,655**	1	
4.Güven Testi (ikinci versiyon)	,965**	,744**	,969**	1
N	20	20	20	20

*** P<0,01*

Tablo 11'de farklı puanlama türlerinin kullanılmasıyla hesaplanan madde ayırt edicilik indekslerinin birbirleriyle korelasyonları incelendiğinde tüm korelasyon değerlerinin 0,01 düzeyinde anlamlı olduğu görülmektedir. Tüm puanlama yöntemlerinde güçlük indekslerinin pozitif düzeyde ilişkisi olduğu söylenebilir. Başka bir deyişle güçlük indeksleri tüm puanlama yöntemlerinde aynı yönde artma ve azalma eğilimine sahiptir. İki kategorili (1-0) puanlama yöntemi ölçüt olarak kabul edildiğinde, en yüksek ilişkinin 0,96 katsayı ile güven testinin ikinci versiyonuna ait olduğu, bu yöntemi 0,88 katsayı ile güven testinin birinci versiyonunun takip ettiği ve en düşük ilişkinin ise 0,78 katsayı ile deneysel ağırlıklandırmaya ait olduğu bulunmuştur.

2.Alt Problem "Matematik başarı testi iki kategorili (1-0) puanlama ve çok kategorili puanlama yöntemleri ile puanlandığında testin geçerliği nasıl değişmektedir?"

Araştırmanın ikinci alt probleminin geçerlik analizleri öğrencilerin okullarında uygulanan sınavların sonuçlarının ölçüt kabul edilmesi ve bu sonuçlar ile araştırmada uygulanan testin iki kategorili (1-0) ve çok kategorili puanlama yöntemleri ile

puanlanmasından elde edilen toplam puanların ilişkisinin Pearson korelasyon katsayısı hesaplanarak incelenmesi yoluyla hesaplanmıştır.

Tablo 12'de öğrencilerin okul akademik başarısı ile farklı puanlama yöntemleri ile puanlanan test sonuçları arasındaki korelasyon katsayıları verilmiştir.

Tablo 12

Okul Akademik Başarıları ile Farklı Puanlama Yöntemleri ile Puanlanan Test Sonuçları Arasındaki Pearson Korelasyon Analizi Sonuçları

	İki Kategorili (1-0)	Deneysel Ağırlıklandırma	Güven Testi (Birinci versiyon)	Güven Testi (İkinci versiyon)
Okul akademik başarısı	0,725**	0,612**	0,723**	0,740**
P	0,000	0,000	0,000	0,000
N	306	306	306	306

***P<0,01*

Tablo 12 incelendiğinde ölçüt kabul edilen okul akademik başarısı ve iki kategorili (1-0) puanlama yönteminin arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,725; p<0,01). Okul akademik başarısı ve deneysel ağırlıklandırma ile puanlanan testin arasında orta düzeyde, pozitif ve anlamlı bir ilişki vardır (r=0,612; p<0,01). Okul akademik başarısı ve güven testi yönteminin birinci versiyonu ile puanlanan testin arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki vardır (r=0,723; p<0,01). Okul akademik başarısı ve güven testi yönteminin ikinci versiyonu ile puanlanan testin arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki vardır (r=0,740; p<0,01).

Ölçüt kabul edilen okul akademik başarısına göre iki kategorili (1-0) puanlama yöntemi ve çok kategorili puanlama yöntemleri ile puanlanan yordayıcı testin geçerli olduğu söylenebilir. Başka bir deyişle Okul akademik başarısına göre iki ve çok kategorili puanlama yöntemleri ile puanlanan yordayıcı testin çarpanlar ve katlar, üslü ifadeler ve kareköklü ifadeler konularını ölçebildiğini söylemek mümkündür. Araştırmada kullanılan testin iki kategorili (1-0) puanlandığında ve çok kategorili puanlama yöntemleri ile puanlandığında ölçüt kabul edilen okul akademik başarıları ile korelasyon katsayılarının değiştiği görülmektedir. Okul akademik başarısı ile en yüksek ilişkinin 0,740 katsayısı ile test güven

testinin yönteminin ikinci versiyonuna, en düşük ilişkinin ise 0,612 katsayı ile test deneysel ağırlıklandırma yöntemine ait olduğu görülmektedir. Her puanlama yönteminde pozitif ve anlamlı ilişki çıktığı ve katsayılar nicel olarak hem iki kategorili (1-0) puanlamanın katsayısından hem de birbirinden çok farklılaşmadığı için bu yöntemlerin testin geçerliğine olumsuz bir etkisi olmadığı sonucuna varılabilir.

Patnaik ve Traub (1973) yürüttükleri çalışmada farklı puanlama yöntemlerinin geçerlik analizleri için öğrencilerin okul başarıları ile test puanlarının arasındaki korelasyon katsayılarını hesaplamış, bu çalışmanın bulgularına benzer olarak ölçüt puanla en düşük korelasyon katsayısını seçenek ağırlıklandırma yönteminin verdiğini raporlamışlardır. Gözen Çıtak (2010) da çalışmasında Okul akademik başarısı ile farklı puanlama yöntemlerinden elde edilen puanların korelasyonunu incelemiş bu araştırmanın sonucuna destek olacak şekilde deneysel ağırlıklandırma puanlamasının iki kategorili puanlamaya göre daha düşük korelasyon ilişkisi olduğunu belirtmiştir.

2. Alt problemin geçerlik analizleri için ölçüt puan olarak kabul edilen Okul akademik başarısı puanları ile farklı puanlama yöntemlerinin korelasyon katsayılarının hesaplanmasına ek olarak farklı puanlama yöntemlerinden elde edilen puanların birbirleri için ölçüt kabul edilerek aralarındaki Pearson korelasyon katsayıları hesaplanmıştır.

Tablo 13'te öğrencilerin iki kategorili (1-0) ve çok kategorili puanlama yöntemleri ile puanlanan test sonuçlarının aralarındaki korelasyon katsayıları verilmiştir.

Tablo 13

İki Kategorili ve Çok Kategorili Puanlama Yöntemleri ile Puanlanan Test Sonuçları Arasındaki Pearson Korelasyon Analizi Sonuçları

	1	2	3	4
1.İki Kategorili	1			
2. Deneysel Ağırlıklandırma	0,864**	1		
3. Güven Testi (birinci versiyon)	0,911**	0,731**	1	
4.Güven Testi (ikinci versiyon)	0,975**	0,813**	0,978**	1

N	20	20	20	20

** *P<0,01*

Tablo 13 incelendiğinde iki kategorili (1-0) puanlama ve deneysel ağırlıklandırma puanlamalarından elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,864; p<0,01). İki kategorili (1-0) puanlama ve güven testi puanlamasının birinci versiyonundan elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,911; p<0,01). İki kategorili puanlama ve güven testi puanlamasının ikinci versiyonundan elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,975; p<0,01). Deneysel ağırlıklandırma ve güven testi puanlamasının birinci versiyonu ile elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,731; p<0,01). Deneysel ağırlıklandırma ve güven testi puanlamasının ikinci versiyonu ile elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,813; p<0,01). Güven testi puanlamasının birinci ve ikinci versiyonuyla elde edilen test sonuçları arasında yüksek düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir (r=0,978; p<0,01).

Bu sonuçlara göre, farklı puanlama türleri arasında yüksek düzeyde, pozitif ve anlamlı ilişkiler bulunduğu için birbirleri için ölçüt kabul edildiklerinde de hepsinin geçerli sonuçlar verdiği söylenebilir. İkili grupların arasında en yüksek ilişki 0,978 korelasyon katsayısı ile güven testi puanlamasının birinci ve ikinci versiyonu arasında, en düşük ilişki ise 0,731 korelasyon katsayısı ile deneysel ağırlıklandırma puanlaması ve güven testi puanlamasının birinci versiyonu arasında bulunmuştur. En yüksek ilişkinin güven testinin farklı versiyonları arasında bulunması iki yöntemde de “eminim” ve “eminim değilim” ifadelerine göre puanlama yapılmasından kaynaklı olduğu düşünülmektedir. Deneysel ağırlıklandırmada maddeler puanlanırken bütün seçeneklere işaretlenme yüzdesine göre puan verilmiş ancak güven testinin birinci versiyonunda herhangi bir madde için puanlama yapılırken doğru seçenek işaretlense bile “eminim değilim” ifadesi seçildiğinde o madde özelinde öğrenciye sıfır puan verilmiştir. Bu durum testin toplam puanları arasında büyük

fark oluşturduğundan en düşük korelasyon katsayısının deneysel ağırlıklandırma ile güven testinin birinci versiyonu arasında çıktığı düşünülmektedir.

3.Alt Problem Matematik başarı testi iki kategorili (1-0) puanlama ve çok kategorili puanlama yöntemleri ile puanlandığında testin güvenirliği nasıl değişmektedir?

Araştırmanın üçüncü alt problemine cevap bulabilmek için test iki kategorili (1-0) puanlama ve güven testi yönteminin birinci versiyonu ile puanlandığında güvenirlik kestirimleri için KR-20, deneysel ağırlıklandırma ve güven testinin ikinci versiyonu ile puanlandığında Cronbach alfa katsayısı hesaplanmıştır.

Tablo 14'te testin iki kategorili (1-0) puanlama yöntemi ile puanlanmasından ve çok kategorili puanlama yöntemleri ile puanlanmasından elde edilen güvenirlik katsayısı değerleri verilmiştir.

Tablo 14

İki Kategorili ve Çok Kategorili Puanlama Yöntemleri ile Puanlamaya Göre Güvenirlik Değerleri

Puanlama Türü	N	Güvenirlik Kestirim Yöntemi	Güvenirlik Katsayısı
İki kategorili (1-0)	306	KR-20	0,81
Deneysel	306	Cronbach Alfa (α)	0,86
Güven Testi (Birinci versiyon)	306	KR-20	0,87
Güven Testi (İkinci versiyon)	306	Cronbach Alfa (α)	0,86

Tablo 14 incelendiğinde araştırmada kullanılan testin iki kategorili (1-0) puanlamayla ve farklı çok kategorili puanlama yöntemleriyle puanlandığında elde edilen güvenirlik katsayılarının değiştiğini görülmektedir. İki kategorili puanlama ve güven testi puanlamasının birinci versiyonu ile elde edilen puanlara ait KR-20 değerleri incelendiğinde iki kategorili (1-0) puanlamanın güvenirliğinin 0,81 güven testinin birinci versiyonunun güvenirliğinin 0,87 olduğu görülmektedir. Deneysel ağırlıklandırma ile puanlamadan elde

edilen α katsayısının 0,86 güven testinin ikinci versiyonu ile puanlamadan elde edilen α katsayısının ise 0,86 olduğu bulunmuştur.

Özçelik (2013) grup karşılaştırmaları yapılacak sınavlarda 0,60 ile 0,80 aralığında güvenirlik değeri, bireyler hakkında kararlar alınacak sınavlarda 0,80 üzerinde, ciddi kararların alınacağı sınavlarda ise 0,90'ın üzerinde güvenirlik değeri beklendiğini belirtmiştir. Bu durumda bu araştırma da kullanılan tüm puanlama yöntemlerine göre uygulanan testin güvenilir olduğu söylenebilir.

Testin en yüksek güvenirlik kestiriminin 0,87 KR-20 katsayısı ile güven testinin birinci versiyonu ile puanlandığında elde edildiği, en düşük güvenirliğinin ise 0,81 KR-20 katsayısı ile iki kategorili (1-0) puanlandığında elde edildiği görülmektedir.

Güvenirlik katsayıları incelendiğinde tüm çok kategorili puanlama yöntemlerinden elde edilen değerlerin iki kategorili (1-0) puanlamadan elde edilen değere göre yüksek olduğu bulunmuştur. Bu durumda 1-0 yerine dereceli anahtarlama ile puanlama yapıldığında puanların heterojenliği arttığından güvenirliğin de arttığı yorumu yapılabilir. Claudy (1978) çift serili ağırlıklandırma, Guttman ağırlıklandırması, oran ağırlıklandırmasını ayrıca iki kategorili (1-0) puanlama ile şans başarısı için düzeltme formüllerini kullanarak çoktan seçmeli testleri puanladığı çalışmasında araştırma bulgularını destekleyecek şekilde en yüksek güvenirlik kestiriminin çift serili ağırlıklandırma puanlamasından, en düşük güvenirlik kestiriminin 1-0 puanlamaya yapılmış düzeltme formülü puanlamasından elde edildiğini ifade etmiştir. Bu çalışmanın bulgularını destekleyecek şekilde Diedenhofen ve Musch (2019) çalışmalarında deneysel ağırlıklandırmanın iki kategorili (1-0) puanlamaya göre şekilde çoktan seçmeli bir testin güvenirliğini artırdığını belirtmişlerdir.

Bölüm 5

Sonuç ve Öneriler

Bu bölümde 306 kişinin yanıtladığı çoktan seçmeli bir matematik başarı testinin iki kategorili (1-0) puanlanma yöntemi, güven testi yönteminin iki farklı versiyonu ve deneysel seçenek ağırlıklandırma yöntemi kullanılarak puanlanmasının testin madde özellikleri ile geçerlik ve güvenirliğine etkilerinin incelenmesi için yapılan analizlerden elde edilen sonuçlar paylaşılmıştır. Ayrıca birtakım öneriler de sunulmuştur.

Sonuçlar

1) Farklı puanlama türlerine göre hesaplanan güçlük indekslerine göre en düşük güçlük indeksleri güven testi puanlamasının birinci versiyonunda ve en yüksek güçlük indeksleri ağırlıklı puanlamada görülmektedir.

2) Farklı puanlama yöntemlerine göre hesaplanan güçlük indeksleri arasındaki farklar maddeler kolaylaştıkça azalmakta ve zorlaştıkça artmaktadır.

3) Farklı puanlama yöntemlerine göre maddelerin güçlük indekslerinin istatistiksel olarak anlamlı farklılaşıp farklılaşmadığını kontrol etmek için yapılan Friedman testi sonuçlarına göre, farklı puanlama türlerinden elde edilen madde güçlük indekslerinin sıra puanları ortalamasına göre maddelerin deneysel puanlama yöntemiyle puanlandığında daha kolay düzeyde olduğu, güven testi yönteminin birinci versiyonuna göre puanlanan maddelerin güçlük düzeylerinin daha zor olduğu bunu güven testinin ikinci versiyonu takip ederken üçüncü sırada ise iki kategorili (1-0) puanlanan maddelerin geldiği bulunmuştur

4) Farklı puanlama yöntemlerine göre hesaplanan madde güçlük indekslerinin ikili karşılaştırmalarında puanlama yöntemlerindeki tüm ikili grupların arasında anlamlı fark olduğu bulunmuştur. Deneysel ağırlıklandırma ve iki kategorili (1-0) puanlama kullanıldığında madde güçlük indeksleri deneysel ağırlıklandırma da daha yüksektir. Güven testinin birinci versiyonu ve iki kategorili (1-0) puanlamanın kullanıldığı durumda madde güçlük indeksleri iki kategorili (1-0) puanlamada daha yüksektir. Güven testinin ikinci

versiyonu ve iki kategorili (1-0) puanlamanın kullanıldığı durumda madde güçlük indeksleri iki kategorili (1-0) puanlamada daha yüksektir. Güven testinin birinci versiyonu ve deneysel ağırlıklandırma karşılaştırmasında madde güçlük indeksleri deneysel ağırlıklandırmada daha yüksektir. Güven testinin ikinci versiyonu ve deneysel ağırlıklandırma puanlamaları kullanıldığında madde güçlük indeksleri deneysel ağırlıklandırmada daha yüksektir. Güven testinin ikinci versiyonu ve güven testinin birinci versiyonu ile puanlandığında madde güçlük indeksleri güven testinin ikinci versiyonunda daha yüksektir.

5) Farklı puanlama yöntemlerine göre madde ayırt edicilik indekslerinin güven testinin birinci versiyonunda en yüksek olduğu görülmektedir. Bunu güven testinin ikinci versiyonu takip ederken üçüncü sırada ise deneysel ağırlıklandırma en düşük madde ayırt ediciliklerinin ise iki kategorili puanlamadan elde edildiği görülmektedir.

6) Farklı puanlama yöntemlerinden elde edilen madde ayırt edicilik indekslerinin ikili karşılaştırmalarında deneysel ağırlıklandırma ve iki kategorili (1-0) puanlama kullanıldığında madde ayırt edicilik indekslerinin arasında istatiksel farklılaşma bulunmamıştır. Güven testinin birinci versiyonu ve iki kategorili (1-0) puanlamadan elde edilen madde ayırt edicilik indeksleri güven testinin birinci versiyonunda anlamlı bir farkla daha yüksektir. Güven testinin ikinci versiyonu ve iki kategorili (1-0) puanlama yöntemleri kullanıldığında madde ayırt edicilik indeksleri güven testinin ikinci versiyonunda anlamlı bir farkla daha yüksektir. Güven testinin birinci versiyonu ve deneysel ağırlıklandırma puanlamaları kullanıldığında madde ayırt edicilik indekslerinin arasında anlamlı bir fark yoktur. Güven testinin ikinci versiyonu ve deneysel ağırlıklandırma puanlanmalarına göre madde ayırt edicilik indekslerinin arasında anlamlı bir fark yoktur. Güven testinin ikinci versiyonu ve güven testinin birinci versiyonu ile puanlama yapıldığında madde ayırt edicilik indeksleri güven testinin birinci versiyonunda anlamlı bir farkla yüksek bulunmuştur.

7) Farklı puanlama yöntemlerine göre hesaplanan ayırt edicilik indeksleri arasındaki farklar maddelerin ayırt edicilik indeksleri azaldıkça azalmakta ve ayırt edicilik indeksleri arttıkça artmaktadır.

8) Ölçüt geçerliği kapsamında iki kategorili (1-0) puanlama ile çok kategorili puanlama yöntemlerinin Okul akademik başarısı ile korelasyon katsayıları hesaplanmıştır. Sonuçlara göre Okul akademik başarısı puanları ile araştırmada kullanılan bütün puanlama türlerinden elde edilen puanlar arasında pozitif ve istatiksel olarak anlamlı ilişkiler olduğu bulunmuştur. Okul akademik başarısı ile en yüksek ilişki güven testinin ikinci versiyon puanlamasına, en düşük ilişki ise deneysel ağırlıklandırma puanlamasına aittir.

9) Farklı puanlama yöntemleri birbiri için ölçüt kabul edildiğinde aralarındaki ilişkiler incelenmiş bütün yöntemlerin arasında pozitif, yüksek ve anlamlı ilişki bulunmuştur. En yüksek ilişki güven testi puanlamasının birinci ve ikinci versiyonu arasında, en düşük ilişki ise deneysel ağırlıklandırma puanlaması ve güven testi puanlamasının birinci versiyonu arasındadır.

10) Farklı puanlama yöntemlerinin güvenirlik kestirimlerine göre testin en yüksek güvenirlik değeri güven testinin birinci versiyonuna ait en düşük güvenirlik değeri ise iki kategorili puanlamaya aittir.

Öneriler

Uygulamaya Yönelik Öneriler

Araştırma sonuçlarına göre uygulamada kullanılan test maddelerinin güçlük indekslerinin farklı puanlama yöntemlerinde anlamlı değiştiği ve ayırt edicilik düzeylerinin de birkaç puanlama yöntemi arasında anlamlı farklılaştığı görülmüştür. Güven testi yönteminin birinci versiyonunda soruların güçlük düzeylerinin yükseldiği ve ayırt edicilik düzeylerinin de diğer yöntemlere oranla daha yüksek olduğu bulunmuştur. Bu sonuçlara dayanarak öğrencilerin maddeleri yanıtlama süreçlerinde öz değerlendirme yapmalarıyla, çoktan seçmeli bir testte tahmin ile işaretleme motivasyonunun azalmasından kaynaklı şans başarısının düşmesi ve testlerin uygulanma amacına hizmet edecek şekilde madde ayırt ediciliklerinin artması sağlanabilir.

Araştırmada kullanılan testin farklı puanlama türlerine ait test güvenirliklerinden en yüksek olanının güven testi yönteminin birinci versiyonuna ardından da güven testi yönteminin ikinci versiyonuna ait olduğu bulunmuştur. Ayrıca ölçüt geçerliğinde en yüksek katsayının güven testi yönteminin ikinci versiyonuna ait olduğu görülmektedir. Bu sonuçlara göre çoktan seçmeli bir testin puanlanmasına öz değerlendirme faktörünün dahil olmasıyla daha geçerli ve güvenilir sonuçların elde edilmesi sağlanabilir.

Bu araştırmada uygulanan testte deneysel ağırlıklandırma puanlama türüne uygun olabilecek madde türü yerine tek doğru cevabı olan maddeler kullanıldığından bu puanlama yönteminin geçerlik katsayısının diğer yöntemlere oranla daha düşük bulunduğu söylenebilir, bu sonuca dayalı olarak deneysel ağırlıklandırma puanlama türünde cevabı en doğru olan madde türlerinin kullanılması önerilmektedir.

Farklı Araştırmalara Yönelik Öneriler

Bu araştırmada kullanılan güven testi yöntemlerinde öğrencilerin öz değerlendirme yapmaları için yalnızca iki seçenek arasından seçim yapmaları istenmiştir. Sonraki çalışmalarda öğrencilerin bilgi düzeyleri hakkında daha rahat ve doğru karar verebilecekleri dereceli güven aralığının bulunduğu ölçekler kullanılabilir.

Benzer alanda yapılacak bir çalışmaya bu araştırmada kullanılan puanlama yöntemlerine ek olarak literatürde bulunan eleyerek puanlama ve kapsayarak puanlama yöntemleri de eklenerek tüm yöntemlerin karşılıklı etkilerinin incelenmesi önerilebilir.

Bu araştırmada uygulama yalnızca devlet okullarındaki öğrencilere yapılmıştır. İlerideki çalışmalarda hem devlet okullarında hem özel okullarda benzer test uygulamalarının yapılmasıyla bu iki grubun arasındaki farklar güven testi yöntemleri özelinde incelebilir.

Araştırmada ölçüt geçerliği kontrollerinin yapılması için farklı okullardan öğrencilerin sınav puanları alınmıştır. Her ne kadar sınav puanları benzer konuların işlendiği aynı sayıda sorudan oluşan çoktan seçmeli testlerden elde edilse de her öğrenci için ölçüt puanlar

tamamen aynı sınavdan elde edilmemiştir. Farklı çalışmalarda ölçüt puanları okullardan alınacak olursa ülke genelinde ortak yapılan dönem içi sınavlara ait puanların ölçüt olarak kabul edilmesi önerilebilir.

Kaynaklar

Association, A. E. R., Association, A. P., Education, N. C. O. M. I., & Testing, J. C. O. S. F. E. a. P. (2014). Standards for educational and psychological testing.

Abu-Sayf, F. K. (1979). The Scoring of Multiple-Choice Tests: A Closer Look. *Educational Technology*, *19*(6), 5–15. http://www.jstor.org/stable/44421466

Baykul, Y. (2021). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (4. bs). Pegem Akademi.

Brown, J. (1965). Multiple response evaluation of discrimination. *British Journal of Mathematical and Statistical Psychology*, *18*(1), 125-137. https://doi.org/10.1111/j.2044-8317.1965.tb00696.x

Büyüköztürk, Ş. (2021). *Sosyal bilimler için veri analizi el kitabı* (29. bs). Pegem Akademi.

Claudy, J. G. (1978). Biserial Weights: A New Approach to Test Item Option Weighting. *Applied Psychological Measurement*, *2*(1), 25-30. https://doi.org/10.1177/014662167800200102

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Cengage Learning.

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 3–17). Lawrence Erlbaum Associates, Inc.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. https://doi.org/10.1037/h0040957

Çakan, M. (2020). Eğitim sistemimizde yaygın olarak kullanılan sınav türleri. İçinde S. Tekindal (Ed.), *Eğitimde ölçme ve değerlendirme* (6. bs, ss. 87-122). Pegem Akademi.

Diedenhofen, B., & Musch, J. (2019). Option weights should be determined empirically and not by experts when assessing knowledge with multiple-choice items. *International*

Journal of Selection and Assessment, *27*(3), 256–266. https://doi.org/10.1111/ijsa.12252

Dressel, P. L., & Schmid, J. (1953). Some Modifications of the Multiple-Choice Item. *Educational and Psychological Measurement*, *13*(4), 574-595. https://doi.org/10.1177/001316445301300404

Echternacht, G. (1973). A comparison of various item option weighting schemes. *ETS Research Bulletin Series*, *1973*(1). https://doi.org/10.1002/j.2333-8504.1973.tb00202.x

Frary, R. B. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education*, *2*(1), 79-96. https://doi.org/10.1207/s15324818ame0201_5

Gözen Çıtak, G. (2010). *Klasik test ve madde tepki kuramlarına a göre çoktan seçmeli testlerde farklı puanlama yöntemlerinin karşılaştırılması*. *9*(1), 170-187.

Güler, N. (2019). *Eğitimde ölçme ve değerlendirme* (14. bs). Pegem Akademi.

Jaradat, D., & Tollefson, N. (1988). The Impact of Alternative Scoring Procedures for Multiple-Choice Items on Test Reliability, Validity, and Grading. *Educational and Psychological Measurement*, *48*(3), 627-635. https://doi.org/10.1177/0013164488483006

Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (6th ed). Thomson/Wadsworth.

Kelecioğlu, H., & Şahin, S. G. (2014). Geçmişten Günümüze Geçerlik. *Journal of Measurement and Evaluation in Education and Psychology*, *5*(2), Article 2. https://doi.org/10.21031/epod.41706

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160. https://doi.org/10.1007/BF02288391

Kurz, T. B. (1999). *A Review of Scoring Algorithms for Multiple-Choice Tests.*

MEB. (2018). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınava ait soru ve cevap anahtarları*. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2018_06/03153730_SAYISAL_BYLYM_A_kitapYY.pdf

MEB. (2019). Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınava ait soru ve cevap anahtarları. https://www.meb.gov.tr/meb_iys_dosyalar/2019_06/02130019_2019_SAYISAL_BOLUM.pdf

MEB. (2020). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınava ait soru ve cevap anahtarları*. https://www.meb.gov.tr/meb_iys_dosyalar/2020_06/21195513_2020_sayisal_bolum_a.pdf

MEB. (2021). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınava ait soru ve cevap anahtarları*. https://cdn.eba.gov.tr/icerik/lgs/2021_SAYISAL_BOLUM_A_.pdf

MEB. (2022). *Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezi sınava ait soru ve cevap anahtarları*. https://cdn.eba.gov.tr/icerik/lgs/2022_sayisal_bolum_a_kitapcigi_ve_cevap_anahtari.pdf

Milli Eğitim Bakanlığı. (2023, Eylül). *Millî eğitim bakanlığı ölçme ve değerlendirme yönetmeliği*. https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=40317&MevzuatTur=7&MevzuatTertip=5

Okçabol, R. (1985). Sistem kavramı ve eğitim sistemimiz. *Eğitim ve Bilim, 9*(53).

Özbek, Y. (2020). Ölçme araçlarında bulunması istenen nitelikler. İçinde S. Tekindal (Ed.), *Eğitimde ölçme ve değerlendirme* (6. bs, ss. 41-85). Pegem Akademi.

Özçelik, D. A. (2013). *Test hazırlama kılavuzu* (5. bs). Pegem Akademi.

Özçelik, D. A. (2016). *Ölçme ve değerlendirme* (5. bs). Pegem Akademi.

Özdemir, D. (2002). *Çoktan seçmeli testlerin klasik test teorisi ve örtük özellikler teorisine göre hesaplanan psikometrik özelliklerinin iki kategorili ve ağırlıklandırılmış puanlanması yönünden karşılaştırılması* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Özdemir, D. (2003). Çoktan seçmeli testleri puanlama yöntemlerine bir bakış. *Eğitim Araştırmaları Dergisi*, *4*(12), 1-8.

Patnaik, D., & Traub, R. E. (1973). Differentıal weighting by judged degree of correctness. *Journal of Educational Measurement*, *10*(4), 281-286. https://doi.org/10.1111/j.1745-3984.1973.tb00805.x

Salkind, N. J. (2013). Test & measurement for people who (think they) hate tests & measurement (2. bs). SAGE publications.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72. https://doi.org/10.2307/1412159

Thorndike, R. M., & Thorndike-Christ, T. (2014). Measurement and evaluation in psychology and education (8th. ed). Pearson.

Turgut, M. F. (1997). Eğitimde *Ölçme ve değerlendirme metotları*. (10. bs). Nüve Matbaası

Turgut, M. F., & Baykul, Y. (2015). *Eğitimde ölçme ve değerlendirme* (7. bs). Pegem Akademi.

Turgut, M. F., & Baykul, Y. (2019). *Eğitimde ölçme ve değerlendirme* (8. bs). Pegem Akademi.

Ünsal Özberk, E. B., & Doğan, N. (2014). Çoklu Özellik-Çoklu Yöntem Analizlerinde Kullanılan Farklı Modellere İlişkin Sonuçların İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *5*(2). https://doi.org/10.21031/epod.01666

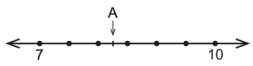
Weitzenhoffer, A. M. (1951). Mathematical structures and psychological measurements. *Psychometrika*, *16*(4), 387–406. https://doi.org/10.1007/bf02288802

Yaşar, M., Kartal, S., & Aybek, E. C. (2021). Scoring Methods for Multiple Choice Tests: How does the Item Difficulty Weighted Scoring Change Student's Test Results? Bartın Üniversitesi Eğitim Fakültesi Dergisi, 10(2), 309-324. https://doi.org/10.14686/buefad.878504

EK-A: Matematik Başarı Testi

Adı Soyadı: Sınıfı/No:

1)



Yukarıdaki sayı doğrultusunda 7 ile 10'a karşılık gelen noktaların arası 6 eş parçaya ayrılmıştır.

Buna göre A noktasına karşılık gelen sayı aşağıdakilerden hangisi olabilir?

A)$\sqrt{94}$ B)$\sqrt{88}$ C)$\sqrt{79}$ D)$\sqrt{68}$

☐ **Eminim**
☐ **Emin Değilim**

2) *a ≠0 ve m, n tam sayılar olmak üzere*

$a^n.a^m = a^{n+m}$ *ve* $\frac{a^m}{a^n} = a^{m-n}$

$(a^n)^m = a^{n.m}$ *dir.*

Aşağıda sadece ön yüzlerinde birer üslü ifadenin yazılı

Mavi Kartlar

2^{-2}	2^3	2^{-1}	2^4

Kırmızı Kartlar

4^{-1}	4^{-3}	4^2	4^0

olduğu 4 mavi ve 4 kırmızı kart verilmiştir.

Mavi kartlardaki her bir üslü ifade kırmızı kartlardaki kendisine denk olmayan her bir üslü ifade ile birer kez çarpılarak yeni üslü ifadeler elde ediliyor.

Elde edilen bu üslü ifadelerden ikisinin birbirine oranı <u>en çok</u> kaçtır?

A) 2^{12} B) 2^{15} C) 2^{16} D)2^{17}

☐ **Eminim**
☐ **Emin Değilim**

3) *a ≠0 ve m, n tam sayılar olmak üzere*

$a^n.a^m = a^{n+m}$ *ve* $\frac{a^m}{a^n} = a^{m-n}$ *dir.*

Aşağıda, her bir hücresinde 2'nin birbirinden farklı tam sayı kuvvetlerinin yazılı olduğu iki sütunlu bir tablo verilmiştir. Tabloda bu üslü ifadelerden ikisi E ve F harfleriyle gösterilmiştir.

I. Sütun	II. Sütun
2^{-1}	2^{-2}
E	**F**
2^3	2^1

I. sütundaki üç üslü ifadenin çarpımı tam kare pozitif bir tam sayıya ve II. sütundaki üç üslü ifadenin çarpımı da tam kare pozitif bir tam sayıya eşittir.

Buna göre E + F <u>en az</u> kaçtır?

A)33 B)17 C)9 D)3

☐ **Eminim**
☐ **Emin Değilim**

4) *Bir ondalık gösterimin, basamak değerleri toplamı şeklinde yazılmasına ondalık gösterimin çözümlenmesi denir.*

Bir basketbol takımındaki beş oyuncunun boy uzunluklarının çözümlenmiş şekli aşağıdaki tabloda verilmiştir.

Tablo: Oyuncuların Boylarının Uzunlukları

İsim	Boy Uzunluğu (cm)
Ayça	$2{\cdot}10^2 + 1{\cdot}10^0 + 1{\cdot}10^{-1}$
Beyza	$1{\cdot}10^2 + 7{\cdot}10^1 + 5{\cdot}10^0 + 5{\cdot}10^{-1}$
Ceyda	$1{\cdot}10^2 + 8{\cdot}10^1 + 4{\cdot}10^0$
Derya	$1{\cdot}10^2 + 8{\cdot}10^1 + 7{\cdot}10^0 + 2{\cdot}10^{-1}$
Esra	$1{\cdot}10^2 + 8{\cdot}10^1 + 5{\cdot}10^0 + 6{\cdot}10^{-1}$

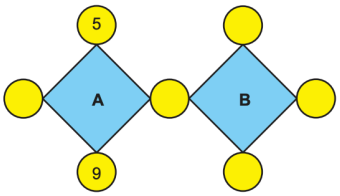
Takımın antrenörü, boyu 185 santimetreden kısa olan oyunculardan birini oyun kurucu olarak oynatacaktır.

Buna göre verilen oyuncular arasında oyun kurucu oynayabilecek kaç oyuncu vardır?

A)4 B)3 C)2 D)1

☐ **Eminim**
☐ **Emin Değilim**

5)



Yukarıdaki şekilde verilen her bir dairenin içine birbirinden farklı birer doğal sayı yazılacaktır. Bu sayılardan ikisi şekilde verilmiştir. Bulundukları dörtgenin köşelerindeki dairelerde yazan dört sayının çarpımına eşit olan A ve B sayıları aralarında asaldır.
Buna göre A + B <u>en az</u> kaçtır?

A)162 B)191 C)258 D)289

☐ **Eminim**
☐ **Emin Değilim**

6) a, b, c, d, birer doğal sayı olmak üzere

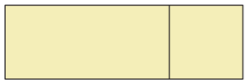
$a\sqrt{b} = \sqrt{a^2 b}$

$a\sqrt{b} + c\sqrt{b} = (a+c)\sqrt{b}$

$a\sqrt{b} - c\sqrt{b} = (a-c)\sqrt{b}$

$a\sqrt{b} - c\sqrt{b} = (a-c)\sqrt{b}$

$a\sqrt{b} \,.\, c\sqrt{d} = (a.c)\sqrt{b.d}$ dir.



Çevresinin uzunluğu $\sqrt{800}$ cm olan dikdörtgen şeklindeki kâğıt, yukarıdaki gibi dikdörtgen şeklinde iki parçaya ayrılıyor.

Kare şeklindeki parçanın bir kenarının uzunluğu $\sqrt{8}$ cm olduğuna göre dikdörtgen şeklindeki parçanın bir yüzünün alanı kaç santimetrekaredir?

A)16 B)24 C)32 D)40

☐ **Eminim**
☐ **Emin Değilim**

7) *a, b birer doğal sayı olmak üzere*

$a\sqrt{b} = \sqrt{a^2 b}$

Yukarıda çapı KL doğru parçası olan daire şeklinde bir karton eş bölmelere ayrılmış 10 santimetrelik bir cetvel verilmiştir. KL doğru parçası, K noktası 2 'ye karşılık gelecek şekilde cetvelin kenarı ile çakıştırıldığında L noktası 6 ile 7 arasında, 7'ye daha yakın bir noktaya karşılık gelmektedir.

Buna göre KL doğru parçasının uzunluğu, santimetre cinsinden aşağıdakilerden hangisi olabilir?

A) $2\sqrt{5}$ B) $2\sqrt{6}$ C) $3\sqrt{3}$ D)$4\sqrt{3}$

☐ **Eminim**
☐ **Emin Değilim**

8) 400 metrelik düz bir yarış pistine başlangıç noktasına uzaklıkları metre cinsinden 2'nin pozitif tam sayı kuvvetleri olacak şekilde yerleştirilebilecek en fazla sayıda engel yerleştiriliyor. Bu pistte 8 atletin yarıştığı bir engelli koşusunda yarışmacılardan biri 20.metrede, bir diğeri 50.metrede yarışı bırakıyor.

Diğer yarışmacılar yarışı tamamladığına göre yarış bittiğinde atletleriin her birinin üzerinden atladığı engel sayılarının toplamı kaçtır?

A)57 B)63 C)64 D)72

☐ **Eminim**
☐ **Emin Değilim**

9) Altan ve Can, defterlerine uzunlukları santimetre cinsinden doğal sayı olan birer kare çiziyorlar. Altan'ın çizdiği karenin alanı kenar uzunlukları 7 cm ve 9 cm olan bir dikdörtgenin alanından büyük, Can'ın çizdiği karenin alanı ise bu dikdörtgenin alanından küçüktür.

Buna göre Altan ile Can'ın çizdiği karelerin alanları arasındaki fark <u>en az</u> kaç santimetrekaredir?

A)8 B)15 C)32 D)39

☐ **Eminim**
☐ **Emin Değilim**

10) *$a \neq 0$ ve m, n tam sayılar olmak üzere*

$a^n.a^m = a^{n+m}$ ve $\frac{a^m}{a^n} = a^{m-n}$ dir.

Bir kenarının uzunluğu 5^4 cm olan kare şeklindeki kâğıdın bir yüzüne aşağıdaki gibi 12 eş dikdörtgen ve 1 kare çizilmiştir. Bu şekillerden kare ve 2 eş dikdörtgen kırmızıya boyanmıştır.

Buna göre kırmızı bölgelerin alanları toplamı kaç santimetredir?

A)2.5^7 B)5^7 C)2.5^6 D)5^6

☐ **Eminim**
☐ **Emin Değilim**

BAŞARILAR…

***Her sorunun yanıtlanması gerekmektedir.**

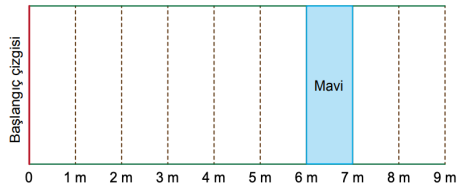
***Her sorunun cevap şıkkı işaretlendikten sonra cevapların altında bulunan eminim/emin değilim kutucuklarından birinin işaretlenmesi gerekmektedir.**

Adı Soyadı:

Sınıfı/No:

1) a, b, c, d, birer doğal sayı olmak üzere $a\sqrt{b} = \sqrt{a^2 b}$ *dir.*

Bir bilye atma oyununa ait, kısa kenar uzunluğu 1 m olan dokuz eş dikdörtgensel bölgeden oluşan oyun parkuru aşağıda verilmiştir.



Başlangıç çizgisinden atış yapan bir oyuncunun attığı bilye, parkurda gösterilen mavi bölgede kalmıştır.

Buna göre bu bilyenin başlangıç çizgisine uzaklığı metre cinsinden aşağıdakilerden hangisi olamaz?

A)$2\sqrt{10}$ B)$3\sqrt{5}$ C)$4\sqrt{3}$ D$2\sqrt{13}$

☐ **Eminim**
☐ **Emin Değilim**

2) *$a \neq 0$, $b \neq 0$ ve k, m, n tam sayılar olmak üzere*

$(a^n)^m = a^{n.m}$ ve $(a \,.\, b)^k = a^k . b^k$ dir.

25^0	81^2	25^2
5^4	36^{10}	1^{10}
10^1	3^8	6^{20}

Yukarıda verilen dokuz adet kutudan her birine bir üslü ifade yazılmıştır. Bu üslü ifadelerden birbirine denk olanların bulunduğu kutular aynı renge boyanacaktır.

Buna göre, <u>boyanmayan</u> kutudaki üslü ifade aşağıdakilerden hangisidir?

A)81^2 B)6^{20} C)25^0 D10^1

☐ **Eminim**
☐ **Emin Değilim**

3) *|a|, 1 veya 1'den büyük, 10'dan küçük bir gerçek sayı ve n bir tam sayı olmak üzere $a.10^n$ gösterimi "bilimsel gösterim" dir.*

Aşağıdaki tabloda bir bitkinin aylık uzama miktarları verilmiştir.

Tablo: Bitkinin Aylara Göre Uzama Miktarı

Ay	Uzama Miktarı (mm)
Nisan	$0{,}081 \cdot 10^4$
Mayıs	$0{,}19 \cdot 10^3$
Haziran	$0{,}0025 \cdot 10^5$

Buna göre, bu bitkinin tablodaki üç aylık toplam uzama miktarının milimetre cinsinden bilimsel gösterimi aşağıdakilerden hangisidir?

A)$1{,}25.\ 10^3$ B)$1{,}25.10^4$ C)$2{,}735.10^{12}$ D) $2{,}735.10^{11}$

☐ **Eminim**
☐ **Emin Değilim**

4)Aşağıdaki tabloda Ordu, Giresun ve Trabzon şehirlerini ziyaret eden turistlerin sayıları verilmiştir.

Tablo: Şehirleri Ziyaret Eden Turistlerin Sayıları

Şehirler	Turist Sayısı
Ordu	$0{,}125 \cdot 10^6$
Giresun	$9{,}5 \cdot 10^4$
Trabzon	$x \cdot 10^7$

Trabzon'u ziyaret eden turistlerin sayısı, Ordu'yu ziyaret eden turistlerin sayısından az ve Giresun'u ziyaret eden turistlerin sayısından fazladır.

Buna göre x'in alabileceği değerlerden biri aşağıdakilerden hangisidir?

A) 10^{-3} B) $3 \,.\, 10^{-3}$ C) 10^{-2} D)$3 \,.\, 10^{-2}$

☐ **Eminim**
☐ **Emin Değilim**

7) Alanı 118 ^{m2} olan bir evin dikdörtgen biçimindeki odaları ve salonu dışındaki bölümlerinin toplam alanı 34 m^2 dir. Salonun alanı, metrekare cinsinden bir tamkare sayıdır ve odaların alanları toplamından küçüktür.

Bu salonun kısa kenarının uzunluğu $\sqrt{18}$ m olduğuna göre uzun kenarın uzunluğu <u>en fazla</u> kaç metredir?

A) $7\sqrt{2}$ B) $6\sqrt{2}$ C) $4\sqrt{2}$ D)$3\sqrt{2}$

☐ **Eminim**
☐ **Emin Değilim**

8)

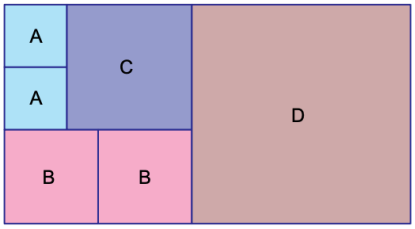
Zeynep parasının yarısı ile paketi 30 lira olan A marka ve diğer yarısı ile paketi 50 lira olan B marka kedi mamalarından alıyor. Bu paketlerden markası aynı olan 6 tanesini evinde beslediği kedileri için ayırdıktan sonra kalan paketleri bir hayvan barınağına veriyor.

Zeynep'in hayvan barınağına verdiği A marka ve B marka mamaların paketlerinin sayıları eşit olduğuna göre Zeynep mamalar için toplam kaç lira harcamıştır?

A)300 B)600 C)700 D)900

☐ **Eminim**
☐ **Emin Değilim**

9) Dikdörtgen şeklindeki bir kâğıt, alanları santimetrekare cinsinden 10'dan büyük birer tam kare pozitif tam sayıya eşit olan karesel bölgelere aşağıdaki gibi ayrılmıştır.

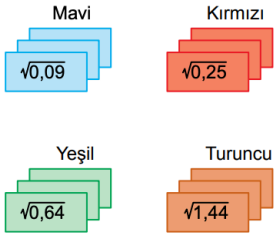


Eşit alanlı bölgeler aynı harf ile gösterildiğine göre dikdörtgen şeklindeki bu kâğıdın bir yüzünün alanı <u>en az</u> kaç santimetrekaredir?

A)168 B)255 C)364 D)392

☐ **Eminim**
☐ **Emin Değilim**

10) Aşağıda dört farklı renkteki kartların her birinden üçer adet verilmiştir. Aynı renkteki kartların üzerinde aynı kareköklü ifade yazmaktadır.



Buna göre Eymen <u>en fazla</u> kaç kart seçmiştir?

A)8 B)9 C)10 D)11

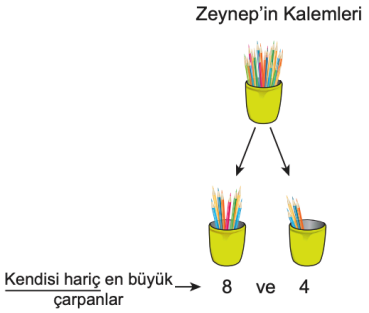
☐ **Eminim**
☐ **Emin Değilim**

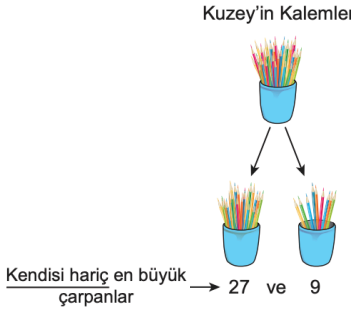
***Her sorunun yanıtlanması gerekmektedir.**

***Her sorunun cevap şıkkı işaretlendikten sonra cevapların altında bulunan eminim/emin değilim kutucuklarından birinin işaretlenmesi gerekmektedir.**

BAŞARILAR…

5) Zeynep'in kalem sayısının çarpanlarından kendisi hariç en büyük iki çarpanı ile Kuzey'in kalem sayısının çarpanlarından kendisi hariç en büyük iki çarpanı aşağıda gösterilmiştir.





Zeynep ve Kuzey, yukarıda verilen çarpanların toplamı kadar kalemi arkadaşlarına vermiştir.

Buna göre, Zeynep ve Kuzey'in toplam kaç kalemi kalmıştır?

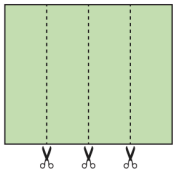
A)22 B)48 C)49 D)64

☐ **Eminim**
☐ **Emin Değilim**

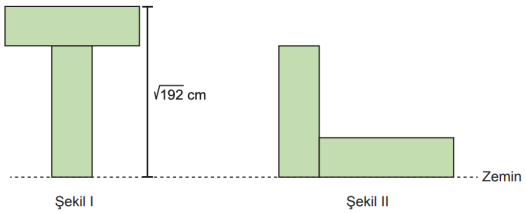
6) *a, b, c, birer doğal sayı olmak üzere*

$a\sqrt{b} = \sqrt{a^2 b}$

$a\sqrt{b} + c\sqrt{b} = (a + c)\sqrt{b}$



Dikdörtgen şeklindeki bir kâğıt, yukarıdaki gibi kesilerek dikdörtgen şeklinde dört eş parça elde edilmiştir. Bu parçaların kısa kenarları ile uzun kenarları çakıştırılarak aşağıdaki gibi iki farklı şekil oluşturulmuştur.



Şekil I'in yüksekliği $\sqrt{192}$ cm ve Şekil II'in yüksekliği $28\sqrt{3}$ cm'dir.

Buna göre başlangıçta verilen dikdörtgen şeklindeki kâğıdın bir yüzünün alanı kaç santimetrekaredir?

A)288 B)144 C)96 D)72

☐ **Eminim**
☐ **Emin Değilim**

7) Alanı 118 ^{m2} olan bir evin dikdörtgen biçimindeki odaları ve salonu dışındaki bölümlerinin toplam alanı 34 m^2 dir. Salonun alanı, metrekare cinsinden bir tamkare sayıdır ve odaların alanları toplamından küçüktür.

Bu salonun kısa kenarının uzunluğu $\sqrt{18}$ m olduğuna göre uzun kenarın uzunluğu <u>en fazla</u> kaç metredir?

A) $7\sqrt{2}$ B) $6\sqrt{2}$ C) $4\sqrt{2}$ D)$3\sqrt{2}$

☐ **Eminim**
☐ **Emin Değilim**

8)

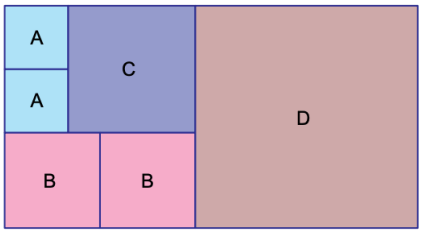
Zeynep parasının yarısı ile paketi 30 lira olan A marka ve diğer yarısı ile paketi 50 lira olan B marka kedi mamalarından alıyor. Bu paketlerden markası aynı olan 6 tanesini evinde beslediği kedileri için ayırdıktan sonra kalan paketleri bir hayvan barınağına veriyor.

Zeynep'in hayvan barınağına verdiği A marka ve B marka mamaların paketlerinin sayıları eşit olduğuna göre Zeynep mamalar için toplam kaç lira harcamıştır?

A)300 B)600 C)700 D)900

☐ **Eminim**
☐ **Emin Değilim**

9) Dikdörtgen şeklindeki bir kâğıt, alanları santimetrekare cinsinden 10'dan büyük birer tam kare pozitif tam sayıya eşit olan karesel bölgelere aşağıdaki gibi ayrılmıştır.

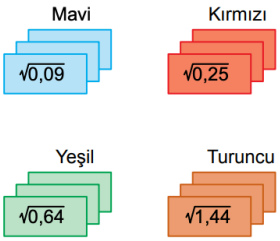


Eşit alanlı bölgeler aynı harf ile gösterildiğine göre dikdörtgen şeklindeki bu kâğıdın bir yüzünün alanı <u>en az</u> kaç santimetrekaredir?

A)168 B)255 C)364 D)392

☐ **Eminim**
☐ **Emin Değilim**

10) Aşağıda dört farklı renkteki kartların her birinden üçer adet verilmiştir. Aynı renkteki kartların üzerinde aynı kareköklü ifade yazmaktadır.



Buna göre Eymen <u>en fazla</u> kaç kart seçmiştir?

A)8 B)9 C)10 D)11

☐ **Eminim**
☐ **Emin Değilim**

***Her sorunun yanıtlanması gerekmektedir.**

***Her sorunun cevap şıkkı işaretlendikten sonra cevapların altında bulunan eminim/emin değilim kutucuklarından birinin işaretlenmesi gerekmektedir.**

BAŞARILAR…

EK-B: Veli Onam Formu

T.C
HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
SOSYAL VE BEŞERİ BİLİMLER ETİK KURULU
VELİ BİLGİ VE ONAM FORMU

Araştırmayı destekleyen kurum: Hacettepe Üniversitesi

Araştırmanın adı: Çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Testlerin Geçerlik ve Güvenirliğine Etkisinin İncelenmesi

Araştırmacı: Esmanur BAŞEL KOÇAK

Adresi:

E-posta adresi:

Telefonu:

Danışman: Prof. Dr. Selahattin GELBAL

Adresi:

E-posta adresi:

Telefonu:

Araştırma Konusu: Eğitim ve öğretim sürecinin en önemli kısımlarından olan çıktıların kontrolünün sağlanması ve daha sonraki süreçlerde neler yapılacağına karar verilmesi için ölçme ve değerlendirme çok önemli bir rol oynamaktadır. Çoktan seçmeli test tekniği okullarda öğretim çıktılarının kontrolü için en çok başvurulan ölçme yöntemlerinden birisidir. Bu çalışmada 8. Sınıf öğrencilerine uygulanan iki çoktan seçmeli testin farklı puanlama yöntemlerinden, iki kategorili puanlama (1,0), çok kategorili puanlamanın doğrudan cevaplama yöntemlerinden deneysel ağırlıklandırma (görgül), cevaplayıcı kararları yöntemlerinden ise güven testi (confidence testing) ile puanlandırılmasının testlerin güvenirliğine ve geçerliğine etkisinin neler olacağının araştırılması amaçlanmaktadır. Araştırmada İstanbul ilinin belirlenen birkaç farklı ilçesinden ilköğretim okullarının 8. Sınıf düzeyindeki öğrencilerine daha önce yapılmış olan LGS matematik sınav soruları içinden Çarpanlar ve Katlar, Üslü İfadeler ve Kareköklü İfadeler konularını kapsayanların seçilmesi ile oluşturulan iki sınıf içi başarı testi uygulanacak olup veriler bu yol ile toplanacaktır.

Araştırma Uygulaması: Başarı testi şeklindedir.

Onam: Sayın veliler,

Ben, Esmanur Başel Koçak, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Bölümü'nde yüksek lisans öğrencisiyim. Yürüttüğüm tez çalışmam kapsamında, 8. sınıf öğrencilerine uygulanan testlerin farklı puanlama yöntemleri ile puanlandırılmasının testlerin geçerlik ve güvenirliklerine etkisini inceleyeceğim. Bu amaçla İstanbul ilinin farklı ilçelerinde bulunan farklı ilköğretim okullarındaki 8. Sınıf öğrencilerinden veri toplayacağım.
Verileri farklı yıllarda çıkmış olan LGS sorularından oluşan 10 soruluk iki farklı test aracılığı ile elde edeceğim. İki test aynı gün içerisinde Aralık ayında ara tatil sonrası uygulanacaktır. Araştırma uygulamasına katılım tamamıyla gönüllülük esasına dayalıdır. Çocuğunuz çalışmaya katılıp katılmamakta özgürdür, katılımı tamamen onun ve sizin isteğinize bağlıdır, reddedebilir ya da herhangi bir aşamasında ayrılabilirsiniz. Çocuğunuz çalışmaya katıldıktan sonra istediği an vazgeçebilir. Böyle bir durumda veri toplama aracını uygulayan kişiye, çalışmayı tamamlamayacağını söylemesi yeterli olacaktır. Araştırmaya katılma, katılmamama veya katıldıktan sonra vazgeçme durumunda öğrencilerin

akademik başarıları, okul ve öğretmenleriyle olan ilişkileri etkilemeyecektir. Çalışmada öğrencilerden kişisel veri olarak yalnızca isim, soyisim, şube ve numara bilgileri alınacaktır fakat bu veriler çalışmanın hiçbir aşamasında paylaşılmayacaktır. Cevaplar tamamıyla gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir. Araştırma çocuğunuz için herhangi bir istenmeyen durum ya da risk içermemektedir. Araştırma için Hacettepe Üniversitesi Sosyal ve Beşeri Bilimler Araştırma Etik Kurulundan onay alınmıştır. T.C. Millî Eğitim Bakanlığı'nın ve okul yönetiminin de izni ile gerçekleşmektedir. Uygulanacak olan testler matematik sorularından oluşmakta olup kişisel rahatsızlık verecek durumlar ve sorular içermemektedir. Fakat katılım sırasında herhangi bir sebepten çocuğunuz kendisini rahatsız hissederse, testi yarıda bırakıp çıkabilir. Böyle bir durumda, rahatsızlığın giderilmesi için yardımcı olunacak ve testleri tamamlamamak çocuğunuza herhangi bir sorumluluk getirmeyecektir. Onay vermeden önce araştırma hakkında daha detaylı bilgi sahibi olmak isterseniz çekinmeden sizlere sağlanılan iletişim adreslerinden bize ulaşabilirsiniz. Çalışma sonrasında da herhangi bir sorunuz için bizimle iletişim kurabilirsiniz. Saygılarımızla,

Araştırma hakkında soru sormak ya da ek bilgi almak için araştırmacı Esmanur BAŞEL KOÇAK ile iletişime geçebilirsiniz (Mail: Tel: Adres:)

Öğrencinin velisi.....olarak yukarıdaki metni okudum, anladım ve çalışma hakkında soru sorma imkanı buldum. Velisi bulunduğum öğrencinin bu çalışmayı istediği zaman ve herhangi bir neden belirtmek zorunda kalmadan bırakabileceğini ve bıraktığı takdirde herhangi olumsuzluk ile karşılaşmayacağımı anladım.

Bu koşullarda söz konusu araştırmaya velisi olduğum öğrencinin katılmasına, hiçbir baskı ve zorlama olmaksızın, izin veriyorum.

Formun bir örneğini aldım / almak istemiyorum.

- **Tarih:**

Öğrenci Velisi:	**Araştırmacı:**
Adı, soyadı:	Adı, soyadı: Esmanur Başel Koçak
Adres:	Adres:
Tel:	
İmza:	Tel:
	e-posta:
	İmza:

EK-C: Çocuk/Ergen Formu

T.C.
HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
SOSYAL VE BEŞERİ BİLİMLER ETİK KURULU
ÇOCUK/ERGEN BİLGİ VE ONAM FORMU

Araştırmacı: Esmanur BAŞEL KOÇAK

Adresi:

E-posta adresi:

Telefonu:

Sevgili Öğrenci,

Ben, Esmanur Başel Koçak, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Bölümü'nde yüksek lisans öğrencisiyim. Yürüttüğüm tez çalışmam kapsamında, 8. sınıf öğrencilerine uygulanan testlerin farklı puanlama yöntemleri ile puanlandırılmasının testlerin geçerlik ve güvenirliklerine etkisini inceleyeceğim. Bu amaçla İstanbul ilinin farklı ilçelerinde bulunan farklı okullardaki 8. Sınıf öğrencilerinden veri toplayacağım. Çalışmaya katılman durumunda senin farklı yıllarda çıkmış olan LGS sorularından oluşan 10 soruluk iki farklı testi çözmeni isteyeceğim.

Araştırma uygulamasına katılımın tamamıyla gönüllülük esasına dayalıdır. Çalışmaya katılıp katılmamakta özgürsün, katılım tamamen sizin isteğinize bağlıdır, reddedebilir ya da herhangi bir aşamasında ayrılabilirsin. Çalışmaya katıldıktan sonra istediğin an vazgeçebilir, böyle bir durumda veri toplama aracını uygulayan kişiye, çalışmayı tamamlamayacağını söylemen yeterli olacaktır. Araştırmaya katılma, katılmamama veya katıldıktan sonra vazgeçme durumunda akademik başarın, okul ve öğretmenlerinle olan ilişkilerin etkilemeyecektir. Çalışmada senden kişisel bilgi olarak yalnıza isim ve soy isim alınacaktır fakat bu veriler çalışmanın hiçbir aşamasında paylaşılmayacaktır. Cevaplar tamamıyla gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir. Araştırma senin için herhangi bir istenmeyen durum ya da risk içermemektedir. Araştırma için Hacettepe Üniversitesi Sosyal ve Beşerî Bilimler Araştırma Etik Kurulundan onay alınmıştır. T.C. Millî Eğitim Bakanlığı'nın ve okul yönetiminin de izni ile gerçekleşmektedir. Uygulanacak olan testler matematik sorularından oluşmakta olup kişisel rahatsızlık verecek durumlar ve sorular içermemektedir. Fakat katılım sırasında herhangi bir sebepten kendisini rahatsız hissedersen, testi yarıda bırakıp çıkabilirsin. Böyle bir durumda, rahatsızlığın giderilmesi için yardımcı olunacak ve testleri tamamlamamak sana herhangi bir sorumluluk getirmeyecektir. Onay vermeden önce araştırma hakkında daha detaylı bilgi sahibi olmak istersen çekinmeden araştırmacıya iletişim adreslerinden ulaşabilirsin. Aklına şimdi gelen veya daha sonra gelecek olan soruları çekinmeden istediğin zaman yukarıda belirttiğim telefon numaramdan veya e-posta adresimden bana ulaşarak sorabilirsin.

Yukarıdaki metni okudum, anladım ve çalışma hakkında soru sorma imkanı buldum. Bu koşullarda söz konusu araştırmaya katılmayı, hiçbir baskı ve zorlama olmaksızın, kabul ediyorum.

- **Tarih:**

Öğrenci: Adı, soyadı: İmza:	**Araştırmacı:** Adı, soyadı: İmza:

EK-Ç: Araştırma Etik Komisyonu Onay Bildirimi

T.C.
HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler Araştırma Etik Kurulu

Tarih: 27/09/2023 16:47
Sayı: E-66777842-300-00003101642
00003101642

Sayı : E-66777842-300-00003101642
Konu : Etik Kurul (Esmanur BAŞEL KOÇAK)

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 12.09.2023 tarihli ve E-51944218-300-00003064884 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı Yüksek Lisans öğrencisi **Esmanur BAŞEL KOÇAK**'ın **Prof. Dr. Selahattin GELBAL** sorumluluğunda yürüttüğü **"Çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Testlerin Geçerlik ve Güvenirliğine Etkisinin İncelenmesi"** isimli tez çalışması Üniversitemiz Sosyal ve Beşeri Bilimler Araştırma Etik Kurulunun **26 Eylül 2023** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. İsmet KOÇ
Kurul Başkanı

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: A2C870DD-C68A-4957-A37E-D4D27A1FE5B2 Belge Doğrulama Adresi: https://www.turkiye.gov.tr/hu-ebys

Adres:
E-posta: Elektronik Ağ: www.hacettepe.edu.tr
Telefon: Faks:
Kep:

Bilgi için: Meryem KÖSE
Bilgisayar İşletmeni
Telefon: 03122977367

EK-D: Araştırma İzni

T.C.
İSTANBUL VALİLİĞİ
İl Millî Eğitim Müdürlüğü

Sayı : E-59090411-44-91181432 04.12.2023
Konu : Anket ve Araştırma İzni (Esmanur BAŞEL KOÇAK)

HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜNE
(Eğitim Bilimleri Enstitüsü Müdürlüğü)

İlgi : a) Yenilik ve Eğitim Teknolojileri Genel Müdürlüğünün 21.01.2020 tarihli ve 2020/2 sayılı genelgesi.
b) Valilik Makamının 04.12.2023 tarihli ve E-59090411-20-91109821 sayılı oluru.

Valilik Makamının Anket ve Araştırma İzni konulu ilgi (b) oluru ve kullanılması uygun görülen ölçme araçlarının Müdürlüğümüzce mühürlenen örnekleri ekte gönderilmiştir.

İlgi (a) genelgenin 28. maddesinde; "Araştırma uygulama izni alan kamu kurum ve kuruluşları, uluslararası kuruluşlar, üniversiteler, sivil toplum kuruluşları ve araştırmacılar tamamladıkları bilimsel araştırma ile ilgili sonuç raporlarını, izni aldıkları ilgili birime çalışma bitiminden itibaren 30 gün içerisinde göndereceklerdir." ifadesi yer almaktadır.

Olur gereğince işlem yapılması ve araştırma sonuç raporunun ekte sunulan örneğe göre Müdürlüğümüz Strateji Geliştirme Şubesine gönderilmesi hususlarında gereğini arz ederim.

Ayşe Nur ÇALIKÇI
İl Millî Eğitim Müdürü a.
Şube Müdürü

Ek:
1- Valilik Oluru (1 Sayfa)
2- Rapor Örneği
3- Ölçekler

EK-E: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

* tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
* görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
* başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
* atıfta bulunduğum eserlerin bütününü kaynak olarak gösterdiğimi,
* kullanılan verilerde herhangi bir tahrifat yapmadığımı,
* bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

03/07/2024

(İmza)

Esmanur BAŞEL KOÇAK

EK-F: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

03/07/2024

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: Çoktan Seçmeli Testlerde Farklı Puanlama Yöntemlerinin Test Ve Madde Özelliklerine

Etkisinin İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
03/07/2024	80	124235	13/ 06/2024	%7	2411975894

Uygulanan filtreler:
1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Esmanur Başel Koçak

Öğrenci No.: N21135109

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

(Prof. Dr. Selahattin GELBAL)

EK-G: Thesis/Dissertation Originality Report

03/07/2024

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: Examining the Effect of Different Scoring Methods on Test and Item Properties in Multiple

Choice Tests

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
03/07/2024	80	124235	13/06/2024	%7	2411975894

Filtering options applied:
1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Esmanur Başel Koçak

Student No.: N21135109 Signature

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED
(Prof. Dr. Selahattin GELBAL)

EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım haklan bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

..... /..... /.....

(imza)

Esmanur BAŞEL KOÇAK

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

(1) Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma *sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü Üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) Madde 6. 2. *Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3.şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) Madde 7. *1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü Üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*

Madde 7.2. *Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir*

**Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.*