Buca Eğitim Fakültesi Dergisi,
2025, sayı 63, ss. 521-548

Araştırma Makalesi

The Journal of Buca Faculty of Education,
2025, issue 63, pp. 521-548

Research Article

Dört Aşamalı Kimya Tanı Testi ve Çoktan Seçmeli Kimya Testinin Eşik Değerlerinin İncelenmesi*

Investigation of Threshold Values of Four Tier Chemistry Diagnostic Test and Multiple Choice Chemistry Test

Suat Türkoguz[1], Canan Baştürk Acar[2]

[1]Sorumlu Yazar, Prof.Dr., Dokuz Eylül Üniversitesi, Buca Eğitim Fakültesi, Matematik ve Fen Bilimleri Bölümü, Fen Bilgisi Eğitimi Anabilim Dalı, suat.turkoguz@gmail.com, (https://orcid.org/0000-0002-7850-2305)
[2] Uzman, Özel Okul Öğretmeni, cnn.basturk.18@gmail.com, (https://orcid.org/0000-0002-1184-7181)

Geliş Tarihi: *22.08.2024* ***Kabul Tarihi:****17.03.2025*

ÖZ

Öğrenciler bir test maddesiyle karşılaştıklarında test maddesine iki şekilde cevap verebilir: hızlı tahmin davranışı, çözüm davranışı. Öğrencilerin hızlı tahmin davranışı veya çözüm davranışı gösterip göstermediğini anlamak için eşik değer hesaplamak önemlidir. Bu çalışmada dört aşamalı kimya tanı testi ve çoktan seçmeli kimya testinin eşik değerlerinin incelenmesi amaçlanmıştır. Bu çalışmada, ilişkisel tarama modelinden yararlanılmıştır. Veri toplama aracı olarak, gaz basıncı konusuyla ilgili 9 maddelik Dört Aşamalı Diagnostik Kimya Testi (DADKT) ve Çoktan Seçmeli Kimya Testi (ÇSKT) kullanılmıştır. DADKT, Ünsal (2019) tarafından geliştirilmiş; ÇSKT ise aşamalı testten uyarlanıp DADKT'nin I. aşamasının test maddelerinden yararlanılmıştır. Bu çalışmada DADKT için bilimsel bilgi güvenirliği KR-20 hesaplanmış 0,460 bulunmuştur. Ayrıca DADKT için kavram yanılgısı güvenirliği KR-20 hesaplanmış ve 0,570 bulunmuştur. Bu çalışmada ÇSKT'nin KR-20 güvenirlik katsayısı 0,520 bulunmuştur. Çalışma 2020-2021 öğretim yılında Dokuz Eylül Üniversitesi, Buca Eğitim Fakültesi'nde öğrenim gören fen bilgisi öğretmen adaylarının katılımıyla gerçekleştirilmiştir. DADKT, 75 kişiye ve ÇSKT 74 kişiye uygulanmıştır. Çalışma sonucunda DADKT'nin iç geçerlik oranlarının eşik değerleri bilimsel bilgi düzeyinde 16. saniye, yanlış pozitif düzeyinde 17. saniye, kavram yanılgısı düzeyinde 28. saniye olduğu ortaya çıkmıştır. Yanlış negatif düzeyinde eşik değeri belirlenememiştir. ÇSTK'ye göre öğrencilerin test yanıtlama performansının eşik değeri 18. saniye olarak hesaplanmıştır. Bu çalışmada katılımcılara test maddelerine geri dönüş hakkı verilmediğinden dolayı her katılımcı test maddelerini bir kez cevaplamış ve cevap değiştirme hakkı olmamıştır. Test maddelerine geri dönüş hakkı verilmesinin iç geçerlik oranlarını etkileyebileceği düşünülmektedir. İleriki çalışmalarda test uygulama biçimine göre çalışmanın kapsamı genişletilebilir.

Anahtar Kelimeler: Cox-Hazard regresyon, diagnostik test, eşik değer, kavram yanılgıları.

ABSTRACT

Students encounter a test item, they respond to this test item in two ways: the first is quick guessing behavior, the other is solution behavior. It is important to calculate a threshold value to understand whether the participant shows rapid guessing behavior or solution behavior. In this study, it was aimed to determine

*Bu makale Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü Matematik ve Fen Bilimleri Eğitimi Anabilim Dalı Fen Bilgisi Öğretmenliği programının yüksek lisans tezinden üretilmiştir.

the internal validity rates of the Four-Tier Chemistry Diagnostic Test (FTCDT) and the Multiple-Choice Chemistry Test (MCCT), response time and threshold values according to the Cox-Hazard model. In this study, descriptive scanning method was used. In this context, data collection tools, the 9-item Four-Tier Chemistry Diagnostic Test (FTCDT) and the Multiple Choice Chemistry Test (MCCT) on the subject of gas pressure were used. FTCDT was developed by Ünsal (2019). MCCT, on the other hand, was adapted from the phased test and the first phase of FTCDT was used as test items. In this study, scientific information reliability KR-20 was calculated for FTCDT and was found to be 0.460. At the same time, the misconception reliability KR-20 for FTCDT was calculated and found to be 0.570. In this study, the KR-20 reliability coefficient of MCCT was found to be 0.52. The study was carried out with the participation of science teacher candidates studying at Dokuz Eylül University, Buca Faculty of Education in the 2020-2021 academic year. FTCDT was applied to 75 people and MCCT was applied to 74 people. As a result of the study, the threshold values of the internal validity rates of FTCDT are 16 seconds at the level of scientific knowledge, 17 seconds at the false positive level, and 28 seconds at the misconception level. It turned out to be. The threshold value at the false negative level could not be determined. According to MCCT, the response effort of the participants was 18. sec. It was calculated as. Since participants in this study were not given the right to return the test items, each participant had to answer the test items once and did not have the right to change the answer. It is thought that giving the right of return to test items may affect internal validity rates. In future studies, the scope of the study can be expanded depending on the test application method.

Keywords: Cox-Hazard regression, diagnostic test, misconception, threshold value.

GİRİŞ

Geçmişten günümüze ölçme değerlendirme yöntemleri gelişim ve değişim göstermiştir. Öncelikle açık uçlu sınavlarla başlayan ölçme değerlendirme yöntemleri gittikçe çoktan seçmeli testlere evrilmiştir. Çoktan seçmeli testlerin şans faktörüne sahip olmalarından dolayı dezavantajları göz önünde bulundurulup aşamalı testler uygulanmaya başlanmıştır. Aşamalı testler iki, üç ve dört aşamalı olarak zaman geçtikçe değişim göstermiştir. Dört aşamalı testlerin kavram yanılgılarını belirlemede etkili olduğu öngörülmüştür (Önsal,2016). Dört aşamalı testler; öğrencilerin bilgiye ulaşma sürecinin takibi ile kavram yanılgılarının ortaya çıkarmada diğer aşamalı testlere göre daha güvenilir sonuçlar ortaya koymuştur (Kaltakçı, 2012). Ancak zamanla aşamalı testlerde geçerlik ve güvenirlik için testin aşamaları tartışılmaya başlanmıştır. Özellikle aşamalı testlerde aşamaların kendi içinde nasıl puanlanacağı veya birbiriyle olan ilişkileri, güvenirlik ve geçerlik puanlarının nasıl hesaplanacağı ile ilgili problemler ortaya çıkmıştır.

Aşamalı testlerin puanlanması çoktan seçmeli testleri puanlamaktan daha farklıdır. Aşamalı testlerin puanlaması ile ilgili farklı puanlama türleri önerilmiş ve bu puanlama türlerine göre iç geçerlik hesaplamaları yapılmıştır (Lee & Chen, 2011;Ranger & Kuhn; 2012; Wise & DeMars; 2006). Ayrıca aşamalı testlerin aşama sayısı değiştikçe puanlama ve iç geçerlik hesaplamaları değiştiğinden dolayı günümüzde bu testlerin puanlama, değerlendirilme ve iç geçerlik hesaplamaları ile ilgili incelemeler devam etmektedir (Gürel vd., 2015; Taber, 2017). Alan yazın incelendiğinde eşik değerin hesaplandığı çalışmalarda genellikle grafik ve %10 kesme yönteminin kullanıldığı gözlemlenmiştir. Bu çalışmada dört aşamalı kimya testinin iç geçerlik puanlarının eşik değeri Cox-Hazard yöntemiyle belirlenmiştir. Cox-Hazard yöntemiyle çoktan seçmeli testlerin eşik değerinin belirlendiği çalışmalara rastlanırken dört aşamalı testlerin iç geçerlilik puanlarına yönelik eşik değerin belirlendiği çalışmalara ulaşılamamıştır. Bu nedenle aşamalı testlerdeki geçerlik ile ilgili tartışmalara yön vermek ve eşik değer belirlemede bu çalışmanın önemli olabileceği düşünülmüştür.

1.1. Yanıtlama Sürelerinin ve Tahmin Davranışlarının Belirlenmesi

Genellikle test skorları, test yanıtlama süreleri, geçerlik ve güvenirlik değerleri gibi psikometrik özellikler testin yapısından, şeklinden, uygulanış biçiminden, katılımcının duyuşsal özelliklerinden, test ortamından etkilenir (Türkoguz, 2020). Özellikle bu testlerin zorluk düzeyleri

değişkenlik gösterir. Testlerin düşük ve yüksek riskli oluşu bu performansları etkiler. Düşük riskli testler, katılımcıların testin sonucunda bir şey kaybetmeyecekleri testlerdir. Yüksek riskli testler katılımcılar için önem arz eden sonuçlarının önemli olduğu testlerdir. Öğrenciler düşük riskli testlerde, test sonucunu dikkate almazlar. Düşük riskli testlerin katılımcılar için önemli sonuçları olmaz ve bu durum katılımcılarda düşük test motivasyonuna, düşük test puanlarına ve kısa yanıtlama sürelerine neden olmaktadır (Kong vd., 2007).

Testlerde yanıtlama süresi

Testlerde bilgisayar kullanımı, araştırmacıların test sonuçlarının istatistiksel konuları hakkında bilgi verir, madde tepki süresi gibi öğrenci davranışları hakkında yararlı veriler sağlar (Kong vd., 2007). Madde yanıtlama süresi, öğrencilerin tutumlarının kalitesini anlamak için önemlidir. Madde zorluğu ve uzunluğu gibi tepki süresini etkileyen faktörler vardır. Madde yanıtlama süresinin, katılımcıların sınavın yapıldığı derse yönelik bağlılık düzeylerinden ve tutumlarından da etkilendiği belirtilmektedir. Öğrenci; soruyu cevaplamak için yeterli bilgiye sahip değilse veya testi ciddiye almadıysa hızlı tahmin etme davranışı sergileyebilir (Kong, Wise ve Bhola, 2007). Dolayısıyla öğrencilerin hızlı tahmin etme gibi davranışları, test sonuçlarını güvenilirlik ve geçerliliklerini etkileyebilir (Wise & DeMars, 2006; Wise & Kong, 2005).

Tahmin davranışı ve tahmin davranışının belirlenmesi

Madde yanıt sürelerine dayanan efor ölçüleri, çözüm davranışı (ÇD) ve hızlı tahmin davranışı olarak adlandırılan kavramlara dayanır. Çözüm davranışı, soru maddesini etkili ve doğru bir şekilde cevaplamaya çalışmaya denir. Eğer cevap rastgele hızlı bir şekilde tahmin ediliyorsa bu duruma hızlı tahmin davranışı adı verilir.

$$ÇD = \begin{cases} 1 & YS_{ij} \geq ES_i \; ise \\ 0 \; diğer \; durumlar \; için \end{cases}$$

YS_{ij}: Herhangi bir öğrencinin bir test maddesini yanıtlama süresi

ES_i: Herhangi bir madde için belirlenen eşik süre ya da eşik değer

i: kişi temsili

j: madde temsili

ÇD: Öğrencinin test maddelerini cevaplarken hızlı tahminde bulunmayıp maddeleri çözmek için efor harcadığını gösterir.

Bilgisayar tabanlı değerlendirmelerin kullanılması araştırmacıların, rastgele tahmin etmede yanıtlama süresi bilgilerini toplamasına izin vermektedir. Rastgele tahmin etme, katılımcıların yeteneklerine bakılmaksızın, motive olmadıklarında, madde kökünü ve tüm seçenekleri dikkatli okumak için gerekli zamanı harcamamasıdır (Swerdzewski vd., 2011). Literatüre bakıldığında hızlı tahmin sonucunda oluşan madde yanıtlarını ve motivasyonu olmayan katılımcıları üç adımda belirlemek mümkündür:

1-Her bir madde için bir eşik değeri tanımlanır. Eşik değer tanımlanırken aşağıdaki özelliklere bakılır:

(a) Sınava katılan her kişi için ortak bir kriter belirlenir (Wise vd., 2004);

(b) Maddenin kelime uzunluğuna ya da kelimelerdeki karakter sayısına bakılır (Wise & Kong, 2005);

(c) Yanıtlama süresi ile ilgili örneklem içindeki frekans dağılımlarıyla ilgili grafiklere (Bilge, 2006);

(d) İki durumlu bir karışım modeli ile istatistiksel tahminler yapılır (Kong vd., 2007);

(e) Madde tepki süresinin ortalama yüzdesi (normatif eşik yöntemi) (Wise & Ma, 2012);

(f) Tepki süresi ve yanıt doğruluğu dağılımları incelenir (Ma vd., 2011).

2-Eşik değerleri belirlendikten sonra, yanıtlama sürelerine bakılır. Yanıtlama süreleri eşik değerlerden daha düşük olan madde yanıtları belirlenir, doğru yanıtlansa dahi "0" olarak puanlanır. Her öğrenci için yanıtlama süresine dayalı bireysel yanıtlama performansı (BYP) hesaplanabilir (Wise & Kong, 2005).

$$BYP_i = \frac{\sum ÇD_{ij}}{k}$$

BYP: "BYP" yanıtlama süresine bağlı yanıtlama eforu anlamına gelmektedir.

k: Testteki madde sayısıdır.

3-Katılımcılar BYP'ye göre isteksiz ya da başarısız olarak tanımlandıktan sonra, araştırmacılar doğru analiz için bu katılımcıların verileri çıkarır veya saklar (Wise & DeMars, 2006; Wise & Kong, 2005).

Wise ve Kong (2005), her bir madde için çözüm davranışını hesaplarken, Wise (2006) tüm maddeler için çözüm davranışını hesaplamıştır. Bir madde için ortalama çözüm davranışını bulduktan sonra, bu ölçüme madde bazlı yanıtlama süresi performansı (MYP) olarak tanımlamıştır.

$$MYP_j = \frac{\sum_{j=1}^{n} ÇD_{ij}}{N}$$

MYP: "MYP" yanıtlama süresi doğruluğu anlamına gelmektedir.

N: Teste katılan örneklemi gösterir.

$$TYP \quad = \frac{\sum ÇD_{ij}}{N}$$

TYP: Test yanıtlama performansıdır. Öğrencilerin tüm testte gösterdikleri çözüm davranışının ortalama değeridir.

Eşik Değer

Sınavlarda teste katılan kişilerin akademik başarısını ve yanıtlama süresini test maddesinin özellikleri etkileyebilir. Bir maddeyi yanıtlamak için gereken süre maddenin zorluğunu ölçmek için kullanılabilir (Hanley,1962). Madde zorluğu, madde ayırt edicilik indeksi ile test maddesinin yanıtlama süresi tahmin edilebilir (Halkitis vd., 1996; Lunz vd., 1994 ; Schnipke, 1997). Örneğin, zorluk indeksi yüksek olan, soru maddesindeki kelime sayısının fazla olduğu, grafik, tablo, resim gibi görselliğe sahip olan maddeler yanıtlanırken daha fazla zamana ihtiyaç duyulur. Ancak zorluk indeksi düşük olan maddelere daha kısa sürede daha doğru cevaplar verildiği kanıtlanmıştır. Test maddesinde görselliğin ön planda olması, kelime değişikliği yapılması, mantıksal düzenlemelere gidilmesi yanıtlama süresinde önemli değişikliklere yol açtığı gözlemlenmiştir (Swanson vd., 2001). Soru maddelerinin tablo, grafik, resimlerle sunulması yanıtlama süresine ve katılımcıların başarısına etki edebilmektedir (Rayner, 1998).

Test maddesindeki yanıtlama süresi ile test maddesine verilen iki modlu (0-1) tepkiden yola çıkarak dağılım eğrisi yapılır. Bu dağılım eğrisinden belirlenen kritik nokta veya başka bir deyişle eşik değere göre zorluk indeksi yüksek olan maddeler ve zorluk indeksi düşük olan maddeler ayırt edilir ve çözüm davranışında bulunan öğrenciler ile hızlı tahmin davranışında bulunan öğrenciler tespit edilir (Bolsinova, de Boeck veTijmstra, 2016). Madde yanıtlama süresi

ve maddeye verilen cevabın doğruluğu arasındaki ilişki teste katılan kişiler hakkında çıkarımlarda bulunmayı sağlamaktadır.

Teste katılan kişiler bazen test maddelerini cevaplamaktan kaçınabilirler. Bu durum, test maddesinin zor, uzun, karmaşık olması veya test katılımcısının konuyla ilgili yeterli bilgisinin olmaması durumunda gerçekleşir. Boş bırakılan veya yanlış cevap verilen maddeleri yanlış olarak kodlamak geçerlik ve güvenirlik hesapları için hataya sebebiyet verirken aynı zamanda test katılımcısının yetenek tahmini için de negatif bir yanlılık ortaya koyar (Guo, Rios, Haberman, Liu vd., 2016). Bu tür hataları önlemek için boş bırakılan veya hızlı tahmin davranışı gösterilmiş olan maddeler için yanıtlama süresine bakılarak puan ve güvenirlik hesabı yapılabilir (Bulut, 2015). Sadece yanıtlama süresine göre testlerin geçerlik ve güvenirliği hesaplanabilmektedir. Bu yüzden testlerin yanıtlama sürelerine göre veri eksiltme işlemleri yapılabilir (Guo vd., 2016). Veri eksiltme yapılabilmesi için öncelikle test yanıtlama süresine yönelik bir eşik değer belirlenir. Bu eşik değer genellikle araştırmalar arasında değişkenlik gösterir. Eşik değer, testlerin türüne, madde niteliklerine göre değişebilir.

Bazı araştırmalarda eşik değer dağılım grafiklerine göre yaklaşık olarak belirlenmiştir. Bazılarında ise toplam süresi ve toplam puana bakılarak tahminde bulunulmuştur. Eşik değer belirleme hususunda araştırmalar devam etmektedir. Bazı araştırmalarda tüm test maddeleri için ayrı ayrı eşik değer belirlenmesi önerilmiştir. Bundan dolayı eşik değerlerin değişken bir yapıya sahip olduğu ile ilgili tartışmalar ortaya çıkmıştır. Eşik değere göre hızlı tahmin davranışı gösterilen, boş bırakılan veya atılan cevaplar için farklı kodlamalar yaparak analiz edilmesi tavsiye edilmiştir. Bu şekilde güvenirlik hakkında daha doğru ve daha olumlu bilgiler elde edilebilir (Weeks vd., 2016).

Streiner'a (2003) göre bazı durumlarda güvenirlik düşmektedir. Bu durumlar: süre sınırı olan testler, soru maddelerinin kolaydan zora doğru hazırlandığı testler, bir soru maddesine verilen cevap başka bir soru maddesine verilen cevaba bağılıysa, testin kapsamının farklı boyutlardan oluşması şeklinde sıralanabilir.

Güvenirliği arttırmak için de bazı değişiklikler yapılabilir. Örneğin, sınırlı zamanda yapılan bir testteki madde sayısını arttırmak veya madde sayısını sabit bırakıp zaman sınırını değiştirmek güvenirliği arttırabilir (Semmes vd., 2011).

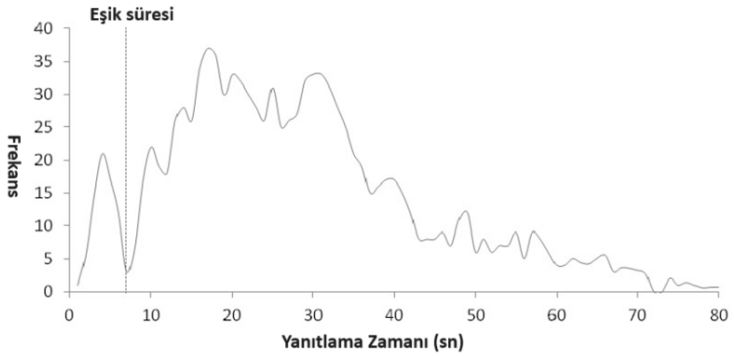
Eşik Değer Belirleme Yöntemleri

a) Grafik

Schnipke ve Scrams (1997) ile Wise ve Kong (2005) yaptıkları çalışmalarda eşik değer belirlemek için ilk olarak zamana bağlı dağılım frekanslarını kullanmışlardır. Bu frekans grafiklerine göre mod değerlerinden sonra gelen minimum frekans değerini eşik değer noktası olarak kabul etmişlerdir. Onlara göre bu eşik değerler 3 ile 7 saniye gibi değerler olmalıdır. Madde yanıtlama sürelerine (RT) göre frekans dağılımı grafiği Şekil 1'de verilmiştir ve bu grafiğin eşik değer noktası belirleme konusunda yardımcı olabileceği düşünülmüştür.

Şekil 1

Bir Test Maddesi İçin Varsayımsal Bir Yanıtlama Zaman Eğrisi Dağılımı (Lee & Chen, 2011)



b) Cox-Hazard modeline göre eşik belirleme

Yaşam sürdürme analizlerde en çok kullanılan model Cox'un da önerdiği "Cox-HM" dir. Bu regresyon modeli Cox tarafından 1972 yılında geliştirilmiş ve yaşam çözümlenmesi çalışmaları için de önemli adımlar atılmasına sebep olmuştur (Yetkin, 2006). Birçok araştırmacı bu regresyon modelini incelemenin faydalı olacağını düşünmüştür (Ata vd., 2007).

Cox Orantılı Hazard Regresyon Modeli'nde ortak bağımlı değişken (yaşam süresi), hasta olan bireylerin ölene kadar geçen takip zamanları (ölüm), kullanılan cihazın bir süre sonra bozulma süreleri, öğrencilerin soru maddesiyle karşılaştığı andan itibaren verdiği tepki süresi olabilirken; açıklayıcı değişkenler ve değişken üzerinde etkili olan faktör değişkenler (cinsiyet, yaş, tedavi çeşidi, öğretim tekniği vb.) olabilir. Sürekli veriye bağlı bağımlı değişkenler ile kategorik veriye bağlı bağımsız değişkenler arasındaki sebep-sonuç bağlantısının ortak bağımlı değişkenlerle ortaya çıkarmak amacıyla kullanılan regresyon yöntemine Cox regresyon yöntemi denilmektedir (Yetkin, 2006). Çünkü lineer doğrusal regresyon analizleri, sürekli bağımsız değişkenler ile sürekli bağımlı değişkenler arasındaki ilişkileri açıklayabilmektedir. Cox-HM'de ise ikili kategorik değişkenler aracılığı ile yaşam süresine ortak bağımlı değişkenler kümesi oluşturulmakta ve açıklayıcı değişkenin etkisi açıklanmaktadır (İnceoğlu, 2013; Zacks, 1992).

Tablo 1

Değişkenlerin Örneklerle Tanımlanması

Denek	**Tedavi Yöntemi (Açıklayıcı değişken)**	**Sürekli Bağımlı Değişken (Yaşam Süresi, t/ay)**	**Kesikli Bağımsız değişken (Netice)**	**Ortak bağımlı değişken (X_i)**	**Ortak bağımlı değişken (Matris kümesi x)**
1	İlaç	2	Sağ	-	$x = \|X_1, X_2, X_3 \ldots X_p\|$
2	Işın	3	Ölü	$X_1=3$	
3	İlaç	6	Sağ	-	
4	İlaç	12	Ölü	$X_2=12$	
-	Işın	8	Ölü	$X_3=8$	
-	İlaç	6	Sağ	-	
N	Işın	12	Sağ	-	

Tablo 1'i bu çalışmaya göre uyarladığımızda; Açıklayıcı değişken tedavi yönteminde "İlaç tedavisi" ve "Işın tedavisi"nin yerine "Üzerinde veri eksiltmesi yapılmamış ham veriler" ve "Grafiklerden faydalanarak belirlenmiş olan t eşik değer noktasından Δt süre sonrasına veri eksiltilerek oluşturulan yeni veriler kümesi elemanları" şeklinde iki grup kullanılabilir. Sürekli Bağımlı Değişken Yaşam Süresi (t) yerine öğrencinin test maddelerine verdiği cevaplar

yazılabilir. Kesikli bağımsız değişken için tedavi sonunda ölen hasta için soru maddesine doğru cevap veren öğrenciler alınabilir. Ortak bağımlı değişken için soru maddelerine doğru cevap veren öğrencilerin yanıtlama süreleri ve bunların oluşturdukları matris kümeleri yazılabilir. Bu nedenle veri kümesi ikili kategorik veri kümesinden oluştuğu için doğrulamada parametrik istatistiklerden ziyade yarı-parametrik Cox-HM kullanılmıştır. Bu çalışmada Cox-HM'de zamandan bağımsız (sabit) ve zamana bağlı açıklayıcı değişkenleri kullanarak eşik değer noktasının en küçük hatayla doğrulanmasına çalışılmıştır.

Bireyin izlenmeye başlandığı andan, olayın gerçekleştiği ana kadar geçen süreye "yaşam sürdürme süresi" veya "başarısızlık süresi", olayın yaşandığı ana ise "başarısızlık zamanı" denir (Cox & Oakes, 1984). Bu çalışmada orijinal veri ile eşik değer noktasında Δt anından sonra düzenlenmiş veriler arasındaki risk analizinin yapılabilmesi için başarısızlık anı öğrencinin soru maddelerine doğru cevap verdiği an başarısızlık anı kabul edilmiştir. Yanlış cevap verdiği an da başarısızlık anı olarak tanımlanabilir. Yorumlama sırasında bu duruma dikkat edilmelidir.

c) Cox-HM'nin Tanımlanması

Cox-HM'de, 1'de görüldüğü gibi $\boldsymbol{X_i}$ neticesi belli olan ortak bağımlı değişkenler vektörü $\boldsymbol{x}$ ve yaşam süresi $\boldsymbol{t}$ olsun. $\boldsymbol{X_i}$ belirli bir tedaviden sonra gerçekleşen ölüm, bir cihazın alındığından itibaren bozulması, bir soruya verilen doğru yanıt vs. durumlar olabilir. $\boldsymbol{X_i}$ neticesi belli olan ortak bağımlı değişkenlerine göre hazard (risk) fonksiyonu $\boldsymbol{h(t;x)}$ şeklinde yazılabilir. Burada risk fonksiyonu $\boldsymbol{h(t;x)}$ "yaşama süresi ise ölme riski", "doğru yanıt verme süresi ise yanlışa düşme riski", "yanlış verme süresi ise doğru yapma riski" olarak tanımlanabilir. Buna göre orantısal hazard (risk) modeli; $\boldsymbol{h(t;x)= h_0(t).exp(\beta' x)}$ olarak yazılır.

Bu modelde; $\boldsymbol{\beta'}$, regresyon katsayıları vektörü; $\boldsymbol{h_0(t),\ x=0}$ olduğunda temel (baseline) risk fonksiyonudur. Örneğin iki farklı tedavi sonucunda hastaların ölüm durumları temel durum olarak alınabilir. Bu temel durumdan tedavi yöntemlerinin yaşam süresi üzerindeki riskleri ortaya konabilir. Bu fonksiyonların anlamlı olabilmesi için Cox-HM'nin temel varsayımlarının bilinmesi gerekir. Bağımsız değişkenlerin risk (hazard) fonksiyonu üzerindeki etkilerinin loglineer olması ve bağımsız değişkenlerin loglineer fonksiyonu ile risk fonksiyonu arasındaki ilişkinin çarpımsal olması gerekmektedir (Özdamar, 2001). Bu iki varsayıma ek olarak gözlemlerin birbirinden bağımsız olmaları ve risk oranının zamana göre değişmemesi, yani sabit olması ya da bir bireyin hazard fonksiyonun diğer bireyin hazard fonksiyonuna orantılı olması gerekmektedir (Arı & Önder, 2013; Ata vd., 2007; Yay vd., 2007).

Bu fonksiyondaki **x** matrisinin tek ya da çok değişkenli olmasına göre $\boldsymbol{\beta}$ regresyon katsayıları aşağıdaki modellere göre tahmin edilir.

Tek değişkenli Cox-HM'de; $\boldsymbol{h(t,x)=h(t) = [h_0(t)].e^{\beta x_1}}$ *biçiminde yazılır.*

Çok değişkenli Cox-HM'de ise, $\boldsymbol{h(t,x)=h(t) = [h_0(t)].e^{(\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4+\cdots\ldots\beta_p x_p)}}$ biçiminde yazılır.

Bu eşitlikte $\boldsymbol{X_1, X_2, ..., X_p}$ ortak bağımlı değişkenlerdir. Ortak değişkenler, yaşam süresine etkide bulunan yaş, kan basıncı ya da orijinal verilerle belirli eşik değere göre ikili karşılaştırması yapılması gereken değişkenler olabilir.

Yaşam sürdürme analizinde iki farklı grubu karşılaştırmak için hazard oranları kullanılabilir.

$\boldsymbol{X_1^*, X_2^*, X_3^*, X_4^*}$ ... ... $\boldsymbol{X_P^*}$ ve $\boldsymbol{X_1, X_2 X_3 X_4 \ldots\ldots X_p}$ iki ayrı gruba ait ortak bağımlı değişkenler olmak üzere hazard oranı;

$$\boldsymbol{HR = \frac{h(t, X^*)}{h(t, X)} = \frac{h_0(t).exp\ (\sum_{i=1}^{p} \beta_i X_i^*)}{h_0(t).exp\ (\sum_{i=1}^{p} \beta_i X_i)} = \emptyset}$$

şeklinde tanımlanır. $\boldsymbol{\theta}$, sabit bir değerdir. Hazard oranı formülü, temel hazard fonksiyonu $\boldsymbol{h_0(t)}$ lerin sadeleştirilmesiyle

$$\boldsymbol{HR = exp\,(\sum_{i=1}^{p} \beta_i (X_i^* - X_i)}$$

üstel gösterim şekliyle yazılabilir.

Cox-HM'nin temel varsayımı olan orantılı hazard varsayımı, hazard oranının zamana karşı sabit olması ya da bir bireyin hazard fonksiyonunun diğer bireyin hazard fonksiyonuna orantılı olması anlamına gelmektedir (Therneau & Grambsch, 2000).

d) β katsayılarının tahminlenmesi

Orantılı hazard modelinin bilinmeyen parametreleri olan $\boldsymbol{\beta}$ katsayıları en çok olabilirlik yöntemi kullanılarak tahmin edilebilir. Bu yöntemin kullanılabilmesi için örneklem verisinin olabilirliğinin bilinmesi gerekir. Modeldeki bilinmeyen parametreler olan $\boldsymbol{\beta}$ ların bir fonksiyonu olarak düşünülen örneklem verisinin olabilirliği, çalışmadaki birim ya da bireylerin gözlemlenmesiyle elde edilen verinin ortak olasılığıdır.

Gözlenen n tane yaşam süresi arasından k tanesi sıralanmış olarak $\boldsymbol{(t_1<t_2<....<t_k)}$, risk sonucu olan verileri göstersin. Bir R_i setinde, $\boldsymbol{t_i}$ zamanında değerleri saptanan $\boldsymbol{x_i}$ ortak bağımlı değişken vektörü belirlenmiş olsun. Bir neticesi belli olan değişkenin, yaşam süresi üzerine etkide bulunan tüm değişkenler dikkate alınarak belirlenecek genel risk içindeki oranı riskler oranı biçiminde belirlenir.

$\boldsymbol{HR = exp(\beta_i' x_i) \,/\, \sum_{j \in R_i} exp\,(\beta_j' x_j)}$ e eşitliği ile verilir.

Farklı yaşam sürelerinin bu oranlarla çarpımı, kısmi olabilirlik fonksiyonunu verir. Regresyon katsayıları bu kısmi olabilirlik fonksiyonu yardımı ile tahmin edilirler. Kısmi olabilirlik fonksiyonu $\boldsymbol{L(\beta)}$ ile gösterilir ve;

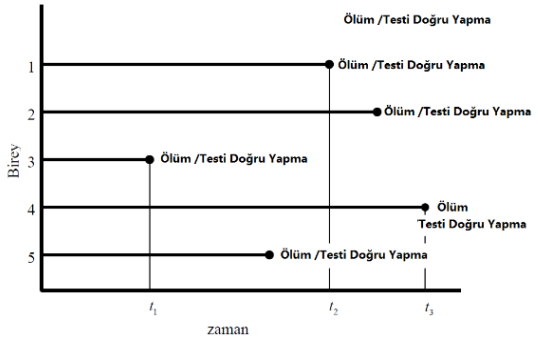
$$\boldsymbol{L(\beta) = \prod_{i=1}^{k} exp\,(\beta_i' x_i) / \sum_{j \in R_i} exp\,(\beta_j' x_j)}$$

Şeklinde ifade edilir.

Olabilirlik fonksiyonun daha iyi açıklayabilmek için 1 den 5 e kadar numaralanmış bireylerden oluşan bir örneği ele alalım.

Şekil 2

Bireyler için Ölüm veya Testi Doğru Yapma Durumları



Şekil 2'de yaşam süresini ölüm üzerinden değerlendirilecek olursa t_1 zamanında 3. Birey, t_2 zamanında 1., 3. ve 5. Birey, t_3 zamanında tüm bireyler ölmüştür. Her bir zaman kesiti için R_i kümesi oluşmuştur. t_1 süresi için R_1(1 kişi), t_2 süresi için R_2(3 kişi) ve t_3 süresi için R_3(5kişi) 'dir. Şekil 2, yaşam süresini testi doğru yapma üzerinden değerlendirilecek olursa t_1 zamanında 3. Birey, t_2 zamanında 1., 3. ve 5. Birey, t_3 zamanında tüm bireyler testi doğru yapmıştır. Her bir zaman kesiti için R_i kümesi oluşmuştur. t_1 süresi için R_1(1 kişi), t_2 süresi için R_2(3 kişi) ve t_3 süresi için R_3(5kişi) 'dir. Örnek te görüldüğü gibi yaşam süreleri bu şekilde uyarlanabilir.

t1 zamanı için $\boldsymbol{HR_1 = exp(\beta_1' x_1) \,/\, exp(\beta_1' x_1)}$

t2 zamanı için $\boldsymbol{HR_2 = exp(\beta_2' x_2) \,/ (exp(\beta_1' x_1) + exp(\beta_3' x_3) + exp(\beta_5' x_5))}$

t3 zamanı için $\boldsymbol{HR_3 = exp(\beta_3' x_3) \,/ (exp(\beta_1' x_1) + exp(\beta_2' x_2) + exp(\beta_3' x_3) + exp(\beta_4' x_4) + exp(\beta_5' x_5))}$

$\boldsymbol{L(\beta) = HR_1 * HR_2 * HR_3}$

Olabilirlik fonksiyonunun belirlenmesiyle, $\boldsymbol{\beta}$ parametrelerinin tahmin edilmesi için işlem kolaylığı açsından, olabilirlik fonksiyonunun logaritması alınır. Logaritması alınmış ifadenin her bir $\boldsymbol{\beta}$ parametresine göre kısmi türevi alınarak denklem sistemi elde edilir. Elde edilen denklem sisteminin çözümü $\boldsymbol{\beta}$ parametrelerini verir.

e) β katsayılarının önemliliğinin test edilmesi

$\boldsymbol{\beta}$ katsayılarının önemliliği için $\boldsymbol{H_0: \beta=0}$ hipotezi test edilir. Bu amaçla üç test yöntemi ileri sürülmüştür. Bunlar; Wald testi, Benzerlik Oranı (Likelihood, ***LR***) testi, Score testidir.

Her üç önemlilik testi için ortak hipotez kurulur. Karar kriterlerinde ise farklı dağılımlardan yararlanılır. Her üçü için geçerli olan hipotezler;

$\boldsymbol{H_0 : \beta i = 0}$ hipotezi test edilir. Kurulan modelin geçerli olmadığını ifade eder.

$\boldsymbol{H_1 : \beta i \neq 0}$ kurulan modelin geçerli olduğunu ifade eder.

Eğer alınan karar $\boldsymbol{H_0}$ ret ise bu durumda $\boldsymbol{\beta i \neq 0}$ için kurulan model geçerlidir.

(Lee & Wang, 2003; Terzi, 2003).

f) Wald testi

Wald testi en büyük benzerlik tahminlerinin (MLE) normal dağıldığı varsayımına dayanır (Klein & Moeschberger, 2003). Regresyon katsayısının, standart hatasına oranı;

$$\boldsymbol{z = (\beta i / SH_{\beta i})}$$

Wald istatistiği olarak adlandırılır. Bu durumda Wald istatistiği Standart Normal Dağılım (SND) gösterir ve SND'nin kritik değerleri ile karşılaştırılarak önemliliği belirlenir. Aynı zamanda Wald istatistiği,

$$\boldsymbol{w = z^2 = (\beta i / SH_{\beta i})^2}$$

olarak da kullanılabilir. Bu durumda ise Wald test istatistiği 1 serbestlik dereceli ki-kare dağılımı gösterir ve 1 serbestlik dereceli ki-kare dağılımının kritik değerleri $\boldsymbol{(x_{kr}^2)}$ ile karşılaştırılarak önemliliği belirlenir (Bal, 1997).

Elde edilen Wald istatistik değeri ile anlamlı sonuç elde edilmesi için;

$W > x_{kr}^2$ olduğunda hipotezi ret edilerek oluşturulan modelin geçerli olduğu kanısına varılır.

$W < x_{kr}^2$ olduğunda ise H_0 hipotezi kabul edilerek alınan değişkenlerin hazard fonksiyonunu açıklamada yeterli olmadıkları düşünülmektedir (Klein vd, 2003; Lee & Wang, 2003; Özdamar, 2003; Scheike vd,, 2014; Terzi, 2003, Yetkin, 2006).

ARAŞTIRMA PROBLEMİNİN İFADESİ

Bu kapsamda çalışmanın problem cümlesi "Dört Aşamalı Kimya Tanı Testi ve Çoktan Seçmeli Kimya Testinin eşik değerleri nasıldır?" olarak belirlenmiştir. Ana problemin çözümüne ilişkin dört tane alt problem belirlenmiştir. Bunlar:

1-DADKT ile ÇSKT'nin eşik değerleri TYP grafiğine göre nasıldır?

2- DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için eşik değerleri zamana dayalı grafiğe göre nasıldır?

3- DADKT ile ÇSKT'nin yanıtlama eşik değerleri TYP'leri Cox- Hazard modeline göre göre nasıldır?

4- DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için eşik değerleri Cox-Hazard analizine göre nasıldır?

YÖNTEM

3.1.Araştırma Modeli

Bu çalışmada, nicel araştırma yöntemlerinden genel tarama modellerinden olan ilişkisel tarama modeli kullanılmıştır. İki veya daha fazla değişken arasındaki değişimi veya değişim seviyesini belirlemeyi amaçlayan araştırma modeline ilişkisel tarama modeli denir (Karasar,2005).

3.2.Çalışma Grubu / Katılımcılar

Çalışma; 2020-2021 öğretim yılında Dokuz Eylül Üniversitesi, Buca Eğitim Fakültesi fen bilgisi öğretmenliği bölümünde öğrenim görmekte olan 149 öğretmen adayının katılımıyla gerçekleştirilmiştir. Katılımcıların; 24'ü erkek, 125'i kadındır. Çalışmaya katılan öğretmen adaylarının 14'ü 2.sınıf, 49'u 3.sınıf, 86'sı 4.sınıf seviyesinde öğrenim görmektedir. Çalışmaya katılan öğretmen adayları; üniversiteye gelmeden önce 83 (%55,71)'ü Ege, 25(%16,78)'i Marmara, 22(%14,77)'si Akdeniz, 8(%5,37)'i İç Anadolu, 6(%4,03)'sı Güneydoğu Anadolu, 4(%2,69)'ü Doğu Anadolu, 1(%0,67)'i Karadeniz Bölgesinde ikamet ediyorlardı.

3.3.Veri Toplama Süreci ve Araçları

Dört Aşamalı Diagnostik Kimya Testinin (DADKT) Geçerlik ve Güvenirlik Bilgileri

Ünsal (2019), gaz basıncı konusuyla ilgili 9 maddeden oluşan Dört Aşamalı Diagnostik Kimya Testini (DADKT) fen bilgisi öğretmen adaylarıyla geliştirmiştir. Ünsal (2019), öğretmen adaylarına testi yanıtlamaları için 27 dk vermiştir. Ünsal (2019) testi geliştirirken açık uçlu soru oluşturma, çoktan seçmeli formata dönüştürme ve aşamalı formata dönüştürme aşamalarını izlemiştir. Öncelikle testi geliştirirken öğretmen adaylarına açık uçlu 52 soru sorulmuş, daha sonra yanıtlara göre açık uçlu sorular çoktan seçmeli teste dönüştürülmüştür. Sonrasında çoktan seçmeli testin maddelerine verilen cevaplardaki seçtikleri seçeneklerin seçilme gerekçesini açıklamalarını isteyen bir boşluk bırakılarak öğretmen adaylarından (n:88) yanıtlarına gerekçeler istenmiştir. Öğretmen adaylarının test maddelerine verdikleri yanıtların gerekçeleri incelenerek

çoktan seçmeli teste gerekçelerin olduğu II. aşama eklenmiştir. Öğretmen adaylarının verdikleri cevaplara güven düzeyini belirlemek için test maddesinin soru kökünün olduğu I. aşamadan sonra "Verdiğiniz yanıttan emin misiniz?" ve I. aşamanın soru köküne verilen yanıtın gerekçesinin seçildiği II. aşamadan sonra da yine "Verdiğiniz yanıttan emin misiniz?" soruları yöneltilmiştir. Böylelikle gaz basıncıyla ilgili çoktan seçmeli test dört aşamalı diagnostik kimya testine (DADKT) dönüştürülmüştür.

Tablo 2

DADKT'nin Puanlama Tablosu

	I. Aşama	**II. aşama**	**III. aşama**	**IV. aşama**
	Cevap	**Güven düzeyi**	**Cevap**	**Güven güzeyi**
Bilimsel bilgi	Doğru	Eminim	Doğru	Eminim
I.tip bilgi eksikliği	Doğru	Eminim	Doğru	Emin Değilim
	Doğru	Emin Değilim	Doğru	Eminim
	Doğru	Emin Değilim	Doğru	Emin Değilim
Yanlış pozitif	Doğru	Eminim	Yanlış	Eminim
II.tip bilgi eksikliği	Doğru	Eminim	Yanlış	Emin Değilim
	Doğru	Emin Değilim	Yanlış	Eminim
	Doğru	Emin Değilim	Yanlış	Emin Değilim
Yanlış negatif	Yanlış	Eminim	Doğru	Eminim
III.tip bilgi eksikliği	Yanlış	Eminim	Doğru	Emin Değilim
	Yanlış	Emin Değilim	Doğru	Eminim
	Yanlış	Emin Değilim	Doğru	Emin Değilim
Kavram yanılgısı	Yanlış	Eminim	Yanlış	Eminim
IV.tip bilgi eksikliği	Yanlış	Eminim	Yanlış	Emin Değilim
	Yanlış	Emin Değilim	Yanlış	Eminim
	Yanlış	Emin Değilim	Yanlış	Emin Değilim

Bu çalışmada kullanılan DADKT'nin bilimsel bilgi güvenirliği KR-20 (tüm aşamaların doğru olması durumunda 1 puan alma şartına göre) 0,460; kavram yanılgısı güvenirliği KR-20 (I. ve III. Aşamaya yanlış cevap verilmesi, II. ve IV. Aşamada emin olunması şartına göre) 0,570 olarak hesaplanmıştır. 9 maddelik DADKT'nin yapı geçerliği için açıklayıcı ve doğrulayıcı faktör analizleri gerçekleştirilmiştir. DADKT'nin 4 faktörlü yapı sergilediği bu analizler doğrultusunda doğrulanmıştır. DADKT'nin s1, s6 ve s9 maddeleri İdeal Gaz kavramını, s3 ve s7, maddeleri Gaz-sıvı basınçları ilişkisini, s4 ve s5 maddeleri kapalı kaplardaki gaz sistemlerini s2 ve s8 maddeleri barometre kavramlarını içermektedir. DADKT'nin test maddeleri için fen eğitimi alanında 4 uzman görüşüne başvurulmuştur. Uzmanlar, 9 maddelik DADKT'de yer alan test maddelerini öğretmen adaylarına uygunluğuna ve kimya ders içeriğine 'uygun', 'uygun değil, düzeltilmesi gerekiyor' ve 'uygun değil' şeklinde incelemiştir. DADKT'nin tüm maddeleri için I. ve III. aşamanın kapsam geçerlilik indeksi (KGI) sırasıyla 0,98 ve 0,95 olarak hesaplanmıştır (Lawshe, 1975). Ayrıca bu çalışmanın içerik geçerliliği için yanlış pozitif (YP) ve yanlış negatif (YN) değerleri de hesaplanmıştır. Hestenes ve Halloun (1995), aşamalı diagnostik testlerde dış geçerliliğin kanıtı olarak YP ve YN tanımını önermiştir. Hestenes ve Halloun (1995) YP'yi yanlış bir nedene dayalı olarak kendinden emin bir tutumla test maddesine doğru yanıt olarak tanımlarken, YN'yi doğru nedene dayalı olarak kendinden emin bir tutumla test maddesine verilen yanlış yanıt olarak tanımlamışlardır. Aşamalı diagnostik testlerde dış geçerlilik için YN, yüzde 10'dan az olmalıdır (Gürçay & Gülbaş, 2015). Ancak, aşamalı diagnostik testlerde YP'yi azaltmak zordur. Bilgi eksikliği olan öğrenciler çoktan seçmeli testlerde doğru cevabı tahmin etme şansına sahip olurlar ve test maddesinin çeldiricilerinden doğru seçeneği seçmeleri olasıdır (Peşman & Eryılmaz, 2010). Tablo 2'nin puanlamasında, 9 maddelik DADKT'nin YP ve YN oranları sırasıyla %17,9 ve %11,3 olarak hesaplanmıştır. Yanlış negatif oranlar, Hestenes ve Halloun'un

(1995) görüşlerine göre 9 maddelik DADKT'nin geçerli bir araç olduğunu kabul edilebilir. DADKT'nin yapı geçerliğinin çalışmanın örnekleminde doğrulanıp doğrulanmadığının incelenmesi amacıyla doğrulayıcı faktör analizi (DFA) gerçekleştirilmiştir. DFA, AMOS 16 (Arbuckle, 2007) programında en yüksek olabilirlik yöntemiyle (Maximum Likelihood) gerçekleştirilmiştir. Faktöriyel yapının gözlemlenen değerlerle uyum derecesinin belirlenebilmesi amacıyla CMIN/df <5, RMSEA<0,08 ve RMR<0,08 uyum indekslerinin değerleri hesaplanmıştır (Çokluk vd., 2012; Kline, 2011). DFA, DADKT'nin dört faktörlü DFA modeli çalışma kapsamında elde edilen verilerle iyi bir uyum gösterdiği saptanmıştır (CMIN/DF = 1,490; RMSEA = 0,079; RMR = 0,021).

ÇSKT'nin Güvenirlik ve Geçerlik Bilgileri:

Bu çalışmada gaz basıncıyla ilgili Çoktan Seçmeli Kimya Testi de kullanılmıştır. Aslında ÇSKT, Ünsal (2019) tarafından geliştirilen gaz basıncıyla ilgili 9 maddelik DADKT'nin I. aşamasıdır. ÇSKT, DADKT'den diğer aşamalar çıkartılarak kullanılmıştır. ÇSKT'nin KR-20 güvenirlik analizi için doğru yanıtlara 1 diğer yanıtlara 0 verilmiştir. Bu çalışma için ÇSKT'nin KR-20 güvenirlik katsayısı 0,520 bulunmuştur. ÇSKT'nin kapsam geçerliği için uzman görüşlerine başvurulmuştur. Bu nedenle, Yüksek Öğretim Kurumu'nun Eğitim Fakültelerinin Fen Bilgisi Öğretmenliği Programının Kimya ders içeriğine göre 9 maddelik ÇSKT'nin kapsam ve görünüş geçerliliği dokuz uzman (Fen eğitiminden dokuz yüksek lisans öğrencisi) tarafından yeniden değerlendirilmiştir. Uzmanlar, 9 maddelik ÇSKT'de yer alan test maddelerini öğretmen adaylarına uygunluğuna ve kimya ders içeriğine 'uygun', 'uygun değil, düzeltilmesi gerekiyor' ve 'uygun değil' şeklinde incelemiştir. Varsa ek görüşlerini test maddesinin yanında bırakılan boş alana yazmışlardır. ÇSKT 'nin her bir maddesi için madde kapsam geçerlilik oranları (KGO_i) ve tanı testinin tüm maddeleri için test kapsam geçerlilik indeksi (KGI), Lawshe (1975) formülleri kullanılarak hesaplanmıştır: ÇSKT'IN tüm maddeleri için kapsam geçerlilik indeksi (KGI) 0,98 hesaplanmıştır (Lawshe, 1975). ÇSKT'nin yapı geçerliğinin çalışmanın örnekleminde doğrulanıp doğrulanmadığının incelenmesi amacıyla doğrulayıcı faktör analizi (DFA) gerçekleştirilmiştir. DFA, AMOS 16 (Arbuckle, 2007) programında en yüksek olabilirlik yöntemiyle (Maximum Likelihood) gerçekleştirilmiştir. Faktöriyel yapının gözlemlenen değerlerle uyum derecesinin belirlenebilmesi amacıyla CMIN/df <5, RMSEA<0,08 ve RMR<0,08 uyum indekslerinin değerleri hesaplanmıştır (Çokluk, Şekercioğlu ve Büyüköztürk, 2012; Kline, 2011). DFA ÇSKT'nin dört faktörlü DFA modeli çalışma kapsamında elde edilen verilerle iyi bir uyum gösterdiği saptanmıştır (CMIN/DF = 1,344; RMSEA = 0,069; RMR = 0,020).

3.4.Verilerin Toplanması ve Analizi

Verilerin Toplanma Süreci

Veri toplama, Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Fen Bilgisi Öğretmenliği bölümünde öğrenim görmekte olan öğrencilere uygulanan DADKT ve ÇSKT ile gerçekleştirilmiştir. Çalışmanın veri toplama işleminin yapılacağı sırada COVID-19 pandemisi meydana gelmiş ve dersler online olarak bilgisayar üzerinden yürütülmüştür. Bu yüzden bu çalışmanın verileri bilgisayar ile online test ortamlarında toplanmıştır. Online test ortamlarında yanıtlanma süresi toplam testin yanıtlanmasına göre belirlenmekte her test maddesi için kayıt tutulmamaktadır. Bu nedenle online test ortamlarında her test maddesinin yanıtlanma süresinin belirlenmesi önemli bir problem teşkil etmektedir. Online test ortamları olarak Microsoft Office uygulaması olan Teams-Form ortamı ve Dokuz Eylül Üniversitesi'nin SAKAI uzaktan eğitim portalı kullanılmıştır. Her test maddesi için ayrı ayrı online test formu oluşturulmuş ve bu test formları SAKAI uzaktan eğitim portalının test uygulama ortamında tek tek sırayla öğrencilere sunulmuştur. Böylelikle her test maddesi için yanıtlama performansı ve süresi kaydedilmiştir. Katılımcıların geri dönüt olarak tekrar test maddelerini görme ve yanıtlama şansı olmamıştır. Her iki test türü için testin ilerleme yönü doğrusal olup geriye dönük değildir. Son olarak SAKAİ

programında her test maddesi için yanıtlama süresine ulaşılabildiğinden dolayı bu programla toplanan veriler araştırma için kodlama ve analiz edilme aşamasına geçilmiştir.

Verilerin Kodlama ve Puanlama Süreci Test puanlarının İç Geçerlik Oranlarına Göre Kodlanması

Bu çalışmada bilimsel bilgi, yanlış pozitif, yanlış negatif, bilgi eksikliği ve kavram yanılgısı üzerinden puanlar hesaplanmıştır. Tüm testteki soruların her bir aşamasında doğru cevaplar "1" ve yanlış cevaplar "0" olarak kodlanmıştır. Bir soruya bilimsel bilgi sonucunu çıkarabilmek için öğrencilerin sorunun I. ve III. aşamasına doğru, II. ve IV. aşamasına da "eminim" cevabını vermiş olması gerekmektedir. Bilimsel bilgi puanlaması "1-1-1-1" şeklinde olmalıdır. Bir soru için pozitif yanlış yani yanlış sebepli doğru sonucuna ulaşmak için öğrencinin sorunun I. aşamasına doğru, II. ve IV. aşamasına "eminim" cevabını verip cevabın nedeninin sorulduğu III. aşamaya yanlış cevap vermesi gerekmektedir. Pozitif yanlış için puanlama "1-1-0-1" şeklinde olmalıdır (Hestenes & Halloun; 1995). Bir sorunun negatif yanlış yani doğru sebepli yanlış olması için öğrencinin sorunun I. aşamasına yanlış, III. aşamasına doğru ve II. ile IV. aşamaya "eminim" cevabını vermiş olması gerekmektedir. Negatif yanlış puanlaması "0-1-1-1" şeklinde hesaplanmalıdır (Hestenes & Halloun; 1995) Bir sorunun kavram yanılgıları olarak adlandırılması için öğrencinin sorunun I. ve III. aşamasına yanlış cevap verip verdiği cevaplardan "emin" olması yani II. ve IV. aşamaya "eminim" cevabını vermesi gerekmektedir. Kavram yanılgıları puanlanırken "0-1-0-1" şeklinde puanlanmaktadır. Bilimsel bilgi, pozitif yanlış, negatif yanlış için cevaplar "1" ile kodlanmıştır. Diğer tüm olasılıklar "0" ile kodlanmıştır. Alanyazın incelendiğinde kavram yanılgılarının ele alındığı çalışmalardaki gözlemlerde %10 ve daha fazla tespit edilen kavram yanılgıları dikkate alınmıştır (Caleon & Subramaniam, 2010). Bu çalışmada öğrencilerin %10 ve üzerinde sahip olduğu kavram yanılgıları dikkatle incelenmiştir. Kodlama sonucu veriler SPSS programı ve Excell kullanılarak bilimsel bilgi, kavram yanılgısı ve bilgi eksikliğine göre analizler yapılmıştır.

Verilerin Analiz Süreci

İlk olarak çalışmanın verileri MS Office Excel'de düzenlenmiştir. DADKT ve ÇSKT'nin güvenirliği, MS Office Excel'de KR-20 formülü uygulanarak hesaplanmıştır. DADKT ve ÇSKT'nin yapı geçerliği SPSS istatistik programında açıklayıcı faktör analizi yapılarak test edilmiştir. Yine SPSS istatistik programında DADKT'nin aşamaları arasında korelasyon değerlerine bakılarak aşamalar ve güven düzeyleri arasındaki güvenirlik oranları tespit edilmiştir. DADKT ve ÇSKT uzman görüşlerine sunulmuş ve uzman görüşlerinin uyum oranları Lawshe'nin (1975) formülleri MS Office Excel'de uygulanarak hesaplanmıştır. Yine benzer şekilde DADKT'nin iki aşama arasındaki yanıt uyumları Hestenes ve Halloun'un (1995) "Bilimsel Bilgi", "Yanlış Pozitif", Yanlış Negatif" ve "Kavram Yanılgısı" kodlamasına göre MS Office Excel'de hesaplanarak dış geçerliği test edilmiştir. Son olarak AMOS programı yardımıyla açıklayıcı faktör analizi ile ortaya çıkan DADKT ve ÇSKT'nin yapısı doğrulanmıştır. Çalışmanın birinci alt probleminde hem DADKT hem de ÇSKT'nin yanıtlama performansları ve süreleri için 1-30 saniye arasında 30 farklı eşik değer noktası belirlenerek; YS_i<ES ise "i" test maddesinin "1" olarak kodlanan doğru yanıtı "0", diğer durumlar için "verilen yanıt (0 ya da 1 olabilir)" şeklinde kabul edilmiştir. Her eşik değer için DADKT ve ÇSKT'nin TYP değerleri belirlendikten sonra MS Office Excel'de grafiğe geçirilmiş ve kırılma noktaları incelenmiştir. Çalışmanın ikinci alt probleminde DADKT'nin "Bilimsel Bilgi", "Yanlış Pozitif", "Yanlış Negatif" ve "Kavram Yanılgısı" olarak adlandırılan iç geçerlik oranları için 1-30 saniye arasında 30 farklı eşik değer noktasında YS_i<ES ise DADKT'nin "i" maddesinin birinci ve ikinci aşamasının "11", "10" ve "01" şeklinde ortak olan yanıtları "00"; diğer durumlar için "verilen yanıt ("00", "10", "01", "11" olabilir) kabul edilmiştir. Her eşik değer için DADKT'nin iç geçerlik oranları belirlendikten sonra MS Office Excel'de grafiğe geçirilmiş ve kırılma noktaları incelenmiştir. Çalışmanın üçüncü alt probleminde DADKT'nin ve ÇSKT'nin 1-30 saniye arasındaki her eşik değer için TYP değerleri SPSS istatistik programında Cox-Hazard analizine tabi tutulmuştur. Çalışmanın dördüncü alt

probleminde DADKT'nin 1-30 saniye arasındaki "Bilimsel Bilgi", "Yanlış Pozitif", "Yanlış Negatif" ve "Kavram Yanılgısı" olarak adlandırılan iç geçerlik oranları Cox-Hazard analizi ile test edilmiştir.

BULGULAR

Bu çalışmada bulgular alt problemlerin sırasına göre sunulmuştur. İlk alt problem "DADKT ile ÇSKT'nin eşik değerleri TYP grafiğine göre nasıldır?" şeklinde tanımlanmıştır. Testlerin son aşaması olan teyit aşamasındaki yanıtlara göre test katılımcılarının tepki süreleri de dikkate alınarak 1 ile 30 saniye arasında 30 farklı eşik değer sınaması yapılmıştır. 30 farklı eşik değer noktasında TYP değerleri bulunmuş ve zamana dayalı olarak grafiğe aktarılmıştır.

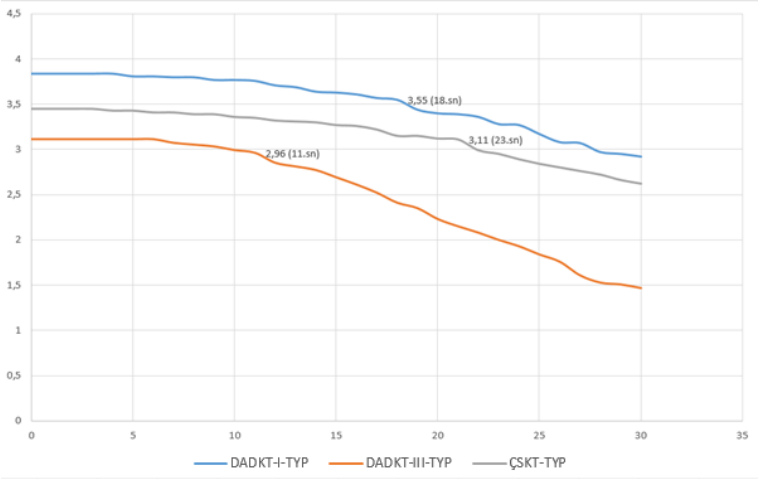
Tablo 3

"1-30 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT ve ÇSKT'nin TYP değerleri

Zaman	DADKT-I-TYP	DADKT-III-TYP	ÇSKT-TYP
0	3.84	3.11	3.45
1	3.84	3.11	3.45
2	3.84	3.11	3.45
3	3.84	3.11	3.45
4	3.84	3.11	3.43
5	3.81	3.11	3.43
6	3.81	3.11	3.41
7	3.80	3.07	3.41
8	3.80	3.05	3.39
9	3.77	3.03	3.39
10	3.77	2.99	3.36
11	3.76	2.96	3.35
12	3.71	2.85	3.32
13	3.69	2.81	3.31
14	3.64	2.77	3.30
15	3.63	2.69	3.27
16	3.61	2.61	3.26
17	3.57	2.52	3.22
18	3.55	2.41	3.15
19	3.44	2.35	3.15
20	3.40	2.23	3.12
21	3.39	2.15	3.11
22	3.36	2.08	2.99
23	3.28	2.00	2.95
24	3.27	1.93	2.89
25	3.17	1.84	2.84
26	3.08	1.76	2.80
27	3.07	1.61	2.76
28	2.97	1.53	2.72
29	2.95	1.51	2.66
30	2.92	1.47	2.62

Şekil 3

"1-30 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT ve ÇSKT'nin TYP Değerlerinin Ortalama Değerlerine İlişkin Grafik



Tablo 3'de "1-30 saniye" arasındaki farklı eşik değerlerine göre DADKT ve ÇSKT'nin TYP değerlerinin zaman grafiği Şekil 3'de çizildiğinde ani değişim noktaları görülmüştür. Bu ani değişim noktaları DADKT'nin ve ÇSKT'nin çözümüyle ilgili olarak katılımcıların çözüm davranışı ve tahmin davranışı hakkında bilgi verebilir. Bu noktalar çözüm ve tahmin davranışı için eşik değer noktalarıdır. Bu durumda Tablo 3'den DADKT'nin I. aşamasının eşik değeri 18. saniye (M_{TYP}=3,55); III. aşamasının eşik değeri 11. saniye (M_{TYP}=2,96) ve ÇSKT'nın eşik değeri 23. saniye (M_{TYP}=3,11) olduğu saptanmıştır.

Bu çalışmanın ikinci alt problemi "DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için eşik değerleri zamana dayalı grafiğe göre nasıldır?" şeklinde tanımlanmış olup DADKT'nin I.ve III. aşamasına göre puanlanan iç geçerlik oranlarına göre eşik değer belirlenmiştir. 1-30 saniye arasındaki zamanlarda iç geçerlik oranlarına göre eşik değer Tablo 4'de gösterilmiştir.

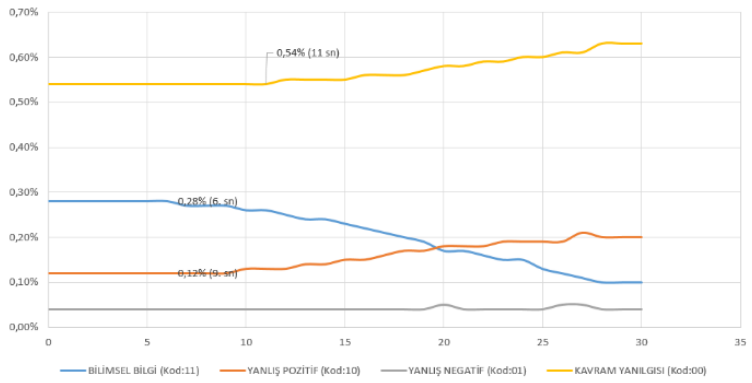
Tablo 4

"1-30 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT I. ve III. Aşamalarının Birlikte Puanlandığı İç Geçerlik Değerlerinin Frekans ve Yüzdeleri

Zaman	**BİLİMSEL BİLGİ (Kod:11)**	**YANLIŞ POZİTİF (Kod:10)**	**YANLIŞ NEGATİF (Kod:01)**	**KAVRAM YANILGISI (Kod:00)**
	f (%)	f (%)	f (%)	f (%)
0	188 (%0.28)	79 (%0.12)	24 (%0.04)	363 (%0.54)
1	188 (%0.28)	79 (%0.12)	24 (%0.04)	363 (%0.54)
2	188 (%0.28)	79 (%0.12)	24 (%0.04)	363 (%0.54)
3	188 (%0.28)	79 (%0.12)	24 (%0.04)	363 (%0.54)
4	188 (%0.28)	79 (%0.12)	24 (%0.04)	363 (%0.54)
5	187 (%0.28)	78 (%0.12)	25 (%0.04)	364 (%0.54)
6	187 (%0.28)	78 (%0.12)	25 (%0.04)	364 (%0.54)
7	184 (%0.27)	80 (%0.12)	25 (%0.04)	365 (%0.54)
8	183 (%0.27)	81 (%0.12)	25 (%0.04)	365 (%0.54)
9	180 (%0.27)	82 (%0.12)	26 (%0.04)	366 (%0.54)
10	177 (%0.26)	85 (%0.13)	26 (%0.04)	366 (%0.54)
11	174 (%0.26)	87 (%0.13)	27 (%0.04)	366 (%0.54)
12	167 (%0.25)	90 (%0.13)	26 (%0.04)	371 (%0.55)
13	164 (%0.24)	92 (%0.14)	26 (%0.04)	372 (%0.55)
14	159 (%0.24)	94 (%0.14)	29 (%0.04)	373 (%0.55)
15	153 (%0.23)	99 (%0.15)	29 (%0.04)	374 (%0.55)
16	148 (%0.22)	103 (%0.15)	28 (%0.04)	376 (%0.56)
17	141 (%0.21)	107 (%0.16)	28 (%0.04)	379 (%0.56)
18	133 (%0.2)	113 (%0.17)	28 (%0.04)	381 (%0.56)
19	126 (%0.19)	112 (%0.17)	30 (%0.04)	387 (%0.57)
20	117 (%0.17)	119 (%0.18)	31 (%0.05)	389 (%0.58)
21	114 (%0.17)	121 (%0.18)	28 (%0.04)	393 (%0.58)
22	110 (%0.16)	124 (%0.18)	28 (%0.04)	395 (%0.59)
23	103 (%0.15)	125 (%0.19)	29 (%0.04)	400 (%0.59)
24	99 (%0.15)	128 (%0.19)	28 (%0.04)	402 (%0.6)
25	91 (%0.13)	129 (%0.19)	29 (%0.04)	408 (%0.6)
26	83 (%0.12)	131 (%0.19)	32 (%0.05)	412 (%0.61)
27	73 (%0.11)	140 (%0.21)	31 (%0.05)	414 (%0.61)
28	69 (%0.1)	138 (%0.2)	30 (%0.04)	422 (%0.63)
29	69 (%0.1)	136 (%0.2)	28 (%0.04)	426 (%0.63)
30	65 (%0.1)	138 (%0.2)	29 (%0.04)	427 (%0.63)

Şekil 4

DADKT'nin I. ve III. Aşamasının Birlikte Puanlandığı İç Geçerlik Değerlerinin Zamana Dayalı Grafiği



Tablo 4'de "1-30 saniye" arasındaki farklı eşik değerlerine göre DADKT I. ve III. aşamasına göre puanlanan iç geçerlik değerlerinin zaman grafiği Şekil 4'de çizildiğinde ani değişim noktaları görülmüştür. Bu ani değişim noktaları DADKT'nin I. ve III. aşamalarına göre puanlanan iç geçerlik değerleri katılımcıların çözüm davranışı ve tahmin davranışı hakkında bilgi verebilir. Bu noktalar çözüm ve tahmin davranışı için eşik değer noktalarıdır. Bu durumda Tablo 4'den DADKT'nin "Bilimsel Bilgi" eşik değeri 6. saniye (%28); "Yanlış Pozitif" eşik değeri 9. saniye (%12); "Kavram Yanılgısı" eşik değeri 11. saniye (%54) olarak belirlenmiştir. Ancak "Yanlış Negatif" eşik değeri grafik yoluyla belirlenememiştir.

Bu çalışmanın üçüncü alt problemi "DADKT ile ÇSKT'nin yanıtlama eşik değerleri TYP'leri Cox- Hazard modeline göre göre nasıldır?" şeklinde tanımlanmıştır. Bu bölümde DADKT ve ÇSKT'nin 1-30 saniye arasındaki farklı eşik değerlerine göre DADKT'nin I. aşamasının TYP değerleri Cox-Hazard analiziyle belirlenmiş ve örnek olması için Tablo 5'de gösterilmiştir. Aynı şekilde DADKT'nin III. aşamasının TYP değerleride ve ÇSKT'nin TYPdeğerleri Cox-Hazard analizi kullanılarak belirlenmiş ancak genel bulguları tablolar halinde sunulmamıştır.

Tablo 5

"1-13 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT'nin I. Aşamasının TYP değerlerini Cox-Hazard Analiz Sonuçları

	B	SE	Wald	SD	p	Exp(B)	Exp(B) için%95 CI	
							En Düşük	En Yüksek
Zaman			59,704	30	0,001			
1.	0.000	0.083	0.000	1	1.000	1.000	0.849	1.177
2.	-0.018	0.084	0.044	1	0.834	0.983	0.834	1.158
3.	-0.021	0.084	0.063	1	0.802	0.979	0.831	1.154
4.	-0.035	0.084	0.177	1	0.674	0.965	0.819	1.138
5.	-0.039	0.084	0.214	1	0.644	0.962	0.816	1.134
6.	-0.053	0.084	0.401	1	0.527	0.948	0.803	1.119
7.	-0.057	0.085	0.457	1	0.499	0.944	0.800	1.115
8.	-0.061	0.085	0.517	1	0.472	0.941	0.797	1.111
9.	-0.072	0.085	0.719	1	0.396	0.931	0.788	1.099
10.	-0.079	0.085	0.873	1	0.350	0.924	0.782	1.091
11.	-0.110	0.086	1.647	1	0.199	0.896	0.757	1.060
12.	0.000	0.083	0.000	1	1.000	1.000	0.849	1.177
13.	-0.122	0.086	2.003	1	0.157	0.885	0.748	1.048

Tablo 5 (Devamı)

"14-30 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT'nin I. Aşamasının TYP değerlerini Cox-Hazard Analiz Sonuçları

	B	SE	Wald	SD	p	Exp(B)	Exp(B) için%95 CI	
							En Düşük	En Yüksek
Zaman			59.704	30	0.001			
14.	-0.126	0.086	2.130	1	0.144	0.882	0.745	1.044
15.	-0.134	0.086	2.396	1	0.122	0.875	0.739	1.036
16.	-0.158	0.087	3.297	1	0.069	0.854	0.721	1.013
17.	-0.162	0.087	3.461	1	0.063	0.851	0.717	1.009
18.	-0.191	0.088	4.738	1	0.029	0.826	0.696	0.981
19.	-0.221	0.088	6.235	1	0.013	0.802	0.675	0.954
20.	-0.225	0.088	6.467	1	0.011	0.799	0.672	0.950
21.	-0.256	0.089	8.223	1	0.004	0.774	0.650	0.922
22.	-0.265	0.089	8.768	1	0.003	0.767	0.644	0.914
23.	0.000	0.083	0.000	1	1.000	1.000	0.849	1.177
24.	-0.274	0.090	9.332	1	0.002	0.760	0.638	0.907
25.	0.000	0.083	0.000	1	1.000	1.000	0.849	1.177
26.	-0.007	0.083	0.007	1	0.933	0.993	0.843	1.170
27.	-0.007	0.083	0.007	1	0.933	0.993	0.843	1.170
28.	-0.010	0.084	0.016	1	0.900	0.990	0.840	1.166
29.	-0.010	0.084	0.016	1	0.900	0.990	0.840	1.166
30.	-0.018	0.084	0.044	1	0.834	0.983	0.834	1.158

Tablo 5'de "1-30 saniye" arasındaki farklı eşik değerlerine göre DADKT'nin I. aşamasının TYP değerlerinin Cox-Hazard analiz sonuçları incelendiğinde katılımcıların DADKT'nin I. aşamasındaki 18. saniye öncesindeki doğru yanıt performanslarının "0" kabul edilmesi durumuna göre genel test puan değerlendirmelerinde önemli bir risk oluşturmazken, 18. saniyden sonra Cox-Hazard analizine göre bir risk oluşturduğu görülmüş ve bu nokta eşik değer olarak kabul edilmiştir (Wald=4,738; Exp(B)=0,826; $p<0,05$). Benzer şekilde eşik değer belirlemede aynı analizler DADKT'nin III. Aşaması ve ÇSKT için gerçekleştirilmiştir. DADKT'nin III. Aşaması için 17. saniye (Wald=4,571; Exp(B)=0,811; $p<0,05$) ve ÇSKT için 18. saniye (Wald=4,341; Exp(B)=0,824; $p<0,05$) olarak belirlenmiştir.

Bu çalışmanın dördüncü alt problemi "DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için eşik değerleri Cox-Hazard analizine göre nasıldır?" şeklinde tanımlanmıştır. Bu aşamada DADKT'nin I.ve III. aşamalarının yanıtlama durumlarına göre belirlenen "Bilimsel Bilgi", "Yanlış Pozitif", "Yanlış Negatif" ve "Kavram Yanılgısı" zihinsel modellerine ilişkin iç geçerlik oranlarının eşik değerleri Cox-Hazard analiziyle belirlenmiştir.

Tablo 6

"1-30 saniye" Arasındaki Farklı Eşik Değerlerine Göre DADKT'nin I. Aşamanın Doğru (1) ve III. Aşamanın Doğru (1) Olduğu Bilimsel Bilgi Olarak Modellenen İç Geçerlik Puanının Cox-Hazard Analiz Sonuçları

	B	SE	Wald	SD	p	Exp(B)	Exp(B) için%95 CI	
							En Düşük	En Düşük
Zaman			908.465	40	0.000			
1.	0.000	0.098	0.000	1	1.000	1.000	0.826	1.211
2.	0.000	0.098	0.000	1	1.000	1.000	0.826	1.211
3.	0.000	0.098	0.000	1	1.000	1.000	0.826	1.211
4.	0.000	0.098	0.000	1	1.000	1.000	0.826	1.211
5.	-0.005	0.098	0.002	1	0.961	0.995	0.821	1.206
6.	-0.005	0.098	0.002	1	0.961	0.995	0.821	1.206
7.	-0.019	0.098	0.039	1	0.844	0.981	0.809	1.189
8.	-0.024	0.098	0.061	1	0.806	0.976	0.805	1.184
9.	-0.039	0.099	0.156	1	0.693	0.962	0.792	1.167
10.	-0.054	0.099	0.297	1	0.586	0.947	0.780	1.151
11.	-0.069	0.100	0.485	1	0.486	0.933	0.768	1.134
12.	-0.106	0.101	1.110	1	0.292	0.900	0.739	1.095
13.	-0.122	0.101	1.460	1	0.227	0.885	0.726	1.079
14.	-0.155	0.102	2.315	1	0.128	0.856	0.701	1.046
15.	-0.189	0.103	3.383	1	0.066	0.828	0.677	1.012
16.	-0.218	0.104	4.441	1	0.035	0.804	0.656	0.985
17.	-0.261	0.105	6.192	1	0.013	0.770	0.627	0.946
18.	-0.312	0.106	8.593	1	0.003	0.732	0.594	0.902
19.	-0.359	0.108	11.061	1	0.001	0.699	0.565	0.863
20.	-0.430	0.110	15.211	1	0.000	0.651	0.524	0.808
21.	-0.452	0.111	16.604	1	0.000	0.636	0.512	0.791
22.	-0.490	0.112	19.083	1	0.000	0.612	0.492	0.763
23.	-0.547	0.114	22.891	1	0.000	0.579	0.463	0.724
24.	-0.580	0.115	25.247	1	0.000	0.560	0.446	0.702
25.	-0.651	0.118	30.359	1	0.000	0.522	0.414	0.657
26.	-0.737	0.122	36.755	1	0.000	0.478	0.377	0.607
27.	-0.843	0.126	44.656	1	0.000	0.431	0.336	0.551
28.	-0.900	0.129	48.910	1	0.000	0.407	0.316	0.523
29.	-0.900	0.129	48.910	1	0.000	0.407	0.316	0.523
30.	-0.948	0.131	52.450	1	0.000	0.388	0.300	0.501

Tablo 6'da "1-30 saniye" arasındaki farklı eşik değerlerine göre DADKT'nin I. aşamasının doğru (1 kodlu) ve III. aşamasının doğru (1 kodlu) kabul edildiği "Bilimsel Bilgi" olarak modellenen iç geçerlik oranına göre Cox-Hazard analiz sonuçları incelendiğinde katılımcıların "Bilimsel Bilgi" modeli için DADKT'nin I. ve III. aşaması için 16. saniye öncesindeki doğru yanıt performanslarının "0" kabul edilmesi durumuna göre genel test puan değerlendirmelerinde önemli bir risk oluşturmazken, 16. saniyeden sonra Cox-Hazard analizine göre bir risk oluşturduğu görülmüştür ve bu an eşik değer olarak kabul edilmiştir (Wald=4,441; Exp(B)=0,804; p<0,05). Benzer şekilde eşik değer belirlemede aynı analizler "Yanlış Pozitif", "Yanlış Negatif" ve "Kavram Yanılgısı" olarak modellenen puanlama türleri içinde gerçekleştirilmiştir. "Yanlış Pozitif" için 17. saniye ((Wald=4,183; Exp(B)=1,354; p<0,05) ve "Kavram Yanılgısı" için 28. saniye (Wald=4,426; Exp(B)=1,163; p<0,05) olarak eşik değerler belirlenmiştir. "Yanlış Negatif" için eşik değer belirlenememiştir.

TARTIŞMA, SONUÇ VE ÖNERİLER

5.1. Tartışma

Wise ve Kong (2005), bilgisayar tabanlı bir çoktan seçmeli test sırasında öğrenci katılımının, madde yanıt süresinden (yani, bir maddenin görüntülendiği zaman ve bir yanıt verildiği zaman arasında geçen süre) çıkarılabileceğini araştırmıştır. Wise ve Kong'un araştırmasının gerekçesi Schnipke ve Scrams'ın (1997, 2002), düşük ve yüksek riskli çoktan seçmeli testler sırasında zaman zaman ölçmeciler tarafından test edilen öğrencilerin hızlı cevaplarıyla ilgili çalışmalara dayanıyordu. İki tür test alma davranışı vardır: test katılımcılarının test maddelerine ve doğru tahminlere göre doğru cevabı belirlemek için efor sarf ettikleri çözüm davranışları ve maddelere hızlı bir şekilde cevap verilen hızlı tahmin davranışları. Bu test bağlamlarında, hızlı tahmin davranışının, öğrencinin testten ayrıldığını ve artık iyi bir efor sarf etmediğini ortaya koyduğunu varsaymışlardır. Çözüm davranışı gösterilmeyen testler, öğrencinin başarı düzeyi hakkında bilgilendirici olmayan, temel olarak maddelere rastgele yanıtlar verdiğinden, bu tür tepkilerin varlığı, çeşitli şekillerde ölçümün kalitesini düşürmektedir.

Bu çalışmanın birinci alt probleminde DADKT ile ÇSKT'nin yanıtlama sürelerine bağlı olarak doğru yanıtlama eforlarının grafiğiyle yanıtlama eşik değerleri incelenmiştir. Testlerin yanıtlama süresine göre çözüm davranışı gösterme, test maddelerine ikili modda (0-1) yanıt verme ile madde yanıtlama süresi şartlı bağımlılık ilkesine göre belirlenebilmektedir (Meyer, 2010). Bu çalışmada DADKT ile ÇSKT'nin yanıtlama sürelerine dayalı olarak doğru yanıtlama eforlarının grafiğiyle eşik değer belirlenmesinde 1-30 saniye arasındaki test katılımcılarının madde yanıt eforları hesaplanmış ve grafiğe geçirilerek değişim noktaları belirlenmiştir. Çalışma bulgularına göre aşamalı kavramsal anlama testinde test katılımcılarının I. aşamada yaklaşık 18. saniyeden sonra çözüm davranışı gösterdikleri, III. aşamada 6. saniyeden sonra çözüm davranışı gösterdikleri görülmektedir. Ayrıca DADKT'nin I. aşamasında da aynı test sorularının kullanıldığı çoktan seçmeli testte ise yaklaşık olarak 23. saniyeden sonra çözüm davranışı gösterdikleri görülmüştür.

Bir testte motive olan katılımcılar, test maddelerinin ortaya koydukları zorlukları çözebilmek için tüm bilgi, beceri, yeteneklerini kullanma eğilimindedirler. Katılımcılar yeterli eforu göstermezlerse yetersiz test performansı, katılımcının bilgi eksikliği, motivasyon eksikliği ya da her ikisinden de kaynaklanma derecesini ayırt etmemizi zorlaştırır. Efor gösterilmeyen testler, başarı testi puanları üzerinde olumsuz yanlılığa neden olma eğiliminde, kişiye özgü, yapı ile alakasız davranışları oluşturur (Haladyna & Downing, 2004). Birinci alt problemin bulgularına göre DADKT I. ve III. aşama ve ÇSKT'nin çözüm davranışındaki farklılıkların gerekçesi olarak DADKT'nin üçüncü aşamasının öğrencileri motive etmesidir. Çünkü üçüncü aşama I. aşama için ipucu bilgiler verebilmektedir. Bu DADKT'nin III. Aşamasının kısa sürede yanıtlanması ÇSKT'nin daha uzun sürede yanıtlanması sonucundan çıkarılmaktadır.

Çoktan seçmeli bir testte, öğrenciler bir sonraki maddeye geçmeden önce bir cevap vermek zorunda kaldıklarında (örneğin, test bilgisayar tabanlıyken), motivasyonu olmayan öğrenciler hızlı, rastgele cevaplar verebilirler (Baştürk & Türkoguz, 2024; Wise & Kong, 2005). Bu davranışlar, testten ayrılan ve test etkinliklerini sona erdirmeye çalışan bireylerinkilerle tutarlıdır. Davranışların her birinin etkisi, test performansında aşağı yönlü bir yanlılığa yol açmaktır, bu da öğrencilerin bildiklerini ve yapabileceklerini yeterince tahmin etmeyen test puanlarıyla sonuçlanır. Birinci alt problemin sonuçlarına göre DADKT'nin I. ve III. Aşamasında çözüm davranışına başladıkları toplam süreye bakıldığında 24. Saniye gibi bir değere ulaşılmaktadır. Öğrencilerin DADKT'nin cevaplanmasında karşılaştığı kelime yoğunluğu ÇSKT'ye göre daha fazladır, ancak aynı sürelerde çözüm davranışı göstermektedir. Bu sonuç öğrencilerin yoğun kelime içeren soru maddelerinde hızlı tahmin davranışı gösterdiği anlaşılmaktadır. DADKT tarzı testlerde öğrencilerin motivasyonlarının düştüğü çıkarımı yapılabilir.

Bu çalışmanın ikinci alt probleminde DADKT ile ÇSKT'nin eşik değerlerine göre iç geçerlik oranları gözlemlenmeye çalışılmıştır. Katılımcılardan toplanan verilere bakıldığında 1-1 olarak kodlanan bilimsel bilgi 6. saniyede, 1-0 olarak kodlanan yanlış pozitif 9. saniyede, 0-0 olarak kodlanan kavram yanılgısı 11. saniyede bulunmuştur. 0-1 olarak kodlanan yanlış negatif düzeyi için veri bulunmamıştır. Bu durum literatürü destekler niteliğindedir. Litaratürde birçok çalışmada hız azaldıkça madde doğruluğunun düştüğü gözlemlenmiştir (Bugbee, 1996; Kong vd., 2007; Setzer vd., 2013; Wise vd., 2006). Çalışma sonuçları incelendiğinde benzer sonuçlara ulaşılmıştır. Her iki aşamanın da doğru olduğu bilimsel bilgi düzeyinde katılımcılar 6. saniyeden itibaren çözüm davranışı göstermişlerdir. Bu durum her iki aşamanın da yanlış olduğu kavram yanılgısı düzeyinde 11. saniye çıkmıştır. Katılımcılar doğruluğundan emin oldukları maddeleri cevaplarken daha hızlı davranmışlardır. Ancak kavram yanılgısı yaşadıklarında çözüme daha geç başlamışlardır.

Bu çalışmanın üçüncü alt probleminde DADKT ile ÇSKT'nin yanıtlama eşik değerleri madde yanıtlama eforunun Cox-Hazard modeline göre incelemesi yapılmıştır. DADKT'nin I. aşamasını Cox-Hazard modeline göre incelediğimizde 18. saniyeden itibaren çözüm davranışı gösterdiğini III. aşamasında 17. saniyeden itibaren çözüm davranışı gösterdiği görülmüştür. Çoktan seçmeli testte ise 18. saniyeden itibaren çözüm davranışı gösterdiği belirlenmiştir.

Bu çalışmanın dördüncü alt probleminde DADKT ile ÇSKT'nin yanıtlama eşik değerleri Cox-Hazard modeline göre testlerin iç geçerlik oranları incelenmiştir. Ortaya çıkan bulguların ışığında Cox-Hazard modeline göre iç geçerlik oranının eşik değeri (1-1) kodlanan bilimsel bilgi düzeyinde 16. saniyede olduğu görülmüştür. Aynı şekilde yapılan kodlamada (1-0) olarak kodlanan yanlış pozitif düzeyinde 17. saniyede ve (0-0) olarak kodlanan gerçek negatif düzeyinde 28. saniyede olduğu ortaya çıkmıştır. Ancak (0-1) kodlama yapılan yanlış negatif düzeyinde bir eşik değer belirlenememiştir.

Tahmin ya da çözüm davranışı, uluslararası büyük ölçekli değerlendirmelerde bilinen bir sorundur ve rastgele hata ya da önyargı getirerek değerlendirmenin güvenilirliğini ve geçerliliğini etkileyebilir (Pokropek & Khorramdel, 2024). Bununla birlikte öğrenciler bazen bir testin bazı bölümlerinde yeterli çözüm davranışı gösteremezler ve böylelikle testin geçerliliği etkilenebilir. Bu nedenle testin tüm aşamalarında ya da belirli bölümlerinde öğrencilerin test çözüm davranışları madde bazlı incelenmesi fayda sağlayabilir. Bu çalışmanın üçüncü ve dördüncü alt probleminden ulaşılan bu sonuçlara göre yanlış cevaplar doğru cevaplara göre daha uzun süreye sahiptir (Lasry vd., 2013). Bu çalışmada doğru cevap verenlerin ve yanlış cevap verenlerin yanıtlama süreleri karşılaştırılmıştır. Bu çalışmada doğru cevap veren katılımcılar yanlış cevap verenlere göre daha hızlı davranmıştır. Bu çalışmada van der Linden ve Glas (2010) tarafından geliştirilen hız-yeterlilik kombinasyonuna uygun bir sonuç çıkmıştır. Doğru yanıt verenlerle yanlış cevap verenler arasındaki yanıtlama süresi farkı testteki iç geçerlik oranlarını etkilemiştir.

5.2. Sonuç ve Öneriler

Bu çalışmada DADKT ve ÇSKT için öğretmen adaylarının yanıtlama eforları ve iç geçerlik oranlarının yanıtlama süresine göre zaman eşik değerleri incelenmiştir. Bu çalışmada birinci alt problem olarak, DADKT ve ÇSKT'nin yanıtlama sürelerine ve yanıtlama eforlarına birlikte bakılarak iki testin de her aşaması için eşik değer belirlenmeye çalışılmıştır. Eşik değer belirlenmesinde 1-30 saniye arasındaki öğretmen adaylarının madde yanıt eforları hesaplanmış ve grafiğe geçirilerek değişim noktaları belirlenmiştir. Çalışma bulgularına göre öğretmen adayları DADKT'nin I. aşamasında yaklaşık 18. saniyeden sonra çözüm davranışı göstermiş, III. aşamasında ise 6. saniyeden sonra çözüm davranışı göstermişlerdir. ÇSKT'de ise yaklaşık olarak 23. saniyeden sonra çözüm davranışı göstermişlerdir.

Çalışmanın ikinci alt probleminde DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için eşik değerleri yanıtlama zamanına dayalı grafiğe göre nasıldır? sorusuna cevap aranmaya çalışılmıştır. Bu aşamadan sonra veri setinde kodlamalar

yapılmıştır. Her iki aşamanın da doğru cevaplandığı "bilimsel bilgi düzeyi" (11), I. aşamanın doğru, III. aşamanın yanlış cevaplandığı "yanlış pozitif" (10), I. aşamanın yanlış, III. aşamanın doğru olduğu "yanlış negatif" (01), ve her iki aşamanın da yanlış olduğu "kavram yanılgısı" (00) olarak kodlanmıştır. "11" kodlamasına sahip olan "bilimsel bilgi" düzeyinin eşik değeri 6. saniye, "10" kodlamasına sahip "yanlış pozitif" düzeyinin eşik değeri 9. saniye, "00" kodlamasına sahip "kavram yanılgısı" düzeyinin eşik değeri 11. saniye olarak bulunmuştur. Ancak "01" kodlamasına sahip "yanlış negatif" düzeyine ait eşik değer bulunamamıştır. Çalışmanın sonuçlarından görüldüğü üzere öğretmen adaylarının doğru cevap verme performansı azaldıkça cevaplama sürelerinin arttığı anlaşılmıştır. Bunun sebebi cevabı bilinen ya da konuya aşina olunan test sorularının daha kısa sürede yanıtlanabilmeleridir. Diğer bir ifadeyle cevabı bilinmeyen ya da konuya tanıdık olunmayan test sorularında daha fazla süre harcanmaktadır.

Çalışmanın üçüncü alt probleminde DADKT ve ÇSKT'nin yanıtlama performansına ilişkin zaman eşik değerleri madde yanıtlama eforunun Cox-Hazard analizine göre nasıl olduğu belirlenmiştir. Cox-Hazard analizine göre DADKT'nin I. aşamasına bakıldığında öğretmen adaylarının 18. saniyeden sonra çözüm davranışı gösterdikleri ortaya konmuştur. DADKT'nin III. aşamasında ise öğretmen adayları 17. saniyeden sonra çözüm davranışı göstermişlerdir. ÇSKT'de ise öğretmen adayları DADKT'nin I. aşaması gibi yine 18. saniyeden itibaren çözüm davranışı göstermişlerdir. Bu eşik değerlerden önceki zamanlarda verilen cevaplar hızlı tahmin davranışı olarak belirlenmiştir.

Bu çalışmanın son alt probleminde DADKT'nin I. ve III. aşamasının birlikte değerlendirildiği iç geçerlik oranları için zaman eşik değerleri Cox-Hazard modeline göre belirlenmiştir. Cox-Hazard modeline göre iç geçerlik eşik değerleri "11" kodlaması yapılan "bilimsel bilgi" düzeyi için 16. saniye, "10" kodlaması yapılan "yanlış pozitif" düzeyi için 17. saniye, 00 kodlaması yapılan "kavram yanılgısı" için ise eşik değer 28. saniyede bulunmuştur. Ancak "01" kodlamasına sahip olan "yanlış negatif" düzeyi için eşik değer noktası bulunamamıştır. Sonuç olarak doğru cevap verenler yanlış cevap verenlere göre daha hızlı cevap vermişlerdir. Yanıtlama sürelerindeki bu fark iç geçerlik oranlarını da etkilemiştir.

Bu çalışmada kullanılan testler öğretmen adayları için düşük riskliydi. Bu durumun öğretmen adaylarının motivasyonlarını etkilediği düşünülmektedir. İleriki çalışmaların yüksek riskli testler için yapılması önerilmektedir. Bu çalışma bilgisayar ortamında online test olarak yapılmıştır. Bu yüzden öğretmen adaylarının testi nerede, nasıl, ne şekilde cevaplandığının gözlemlenebilme imkanı olmamıştır. İleriki çalışmalarda yanıtlama süresi ve yanıtlama eforunu gözlemlemek için akıllı telefon, tablet, akıllı kalemler gibi daha teknolojik cihazlar kullanılabilir (Edgecomb vd., 2014; Mehlhorn vd., 2011; Moharkan vd., 2017). Bu çalışmada öğretmen adaylarına soru maddelerine tekrar dönme ve cevaplama şansı verilmemiştir. Ancak yapılan çalışmalarda tekrar cevaplama hakkı verildiğinde iç geçerlik oranlarının arttığı kanıtlanmıştır (Turkoguz, 2019). İleriki çalışmalar için tekrar cevaplama hakkının verilmesi kavram yanılgıları için farklı sonuçlar ortaya çıkarabileceğinden önerilmektedir.

Kaynakça

Arbuckle, J. (2008). *AMOS 17.0 user's guide.* SPSS Inc.

Arı, A., & Önder, H. (2013). Farklı veri yapılarında kullanılabilecek regresyon yöntemleri. *Anadolu J. Agr. Sci.,28*(3):168-174. https://doi.org/10.7161/anajas.2013.28.3.168.

Ata, N., Karasoy, D.S., & Sözer, M.T. (2007) Orantılı Tehlike Varsayımının İncelenmesinde Kullanılan Yöntemler ve Bir Uygulama. *Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 20*(1), 57-80.

Bal, C. (1997). Tedavi sonrası izlem verilerinin cox regresyon aracılığı ile incelenmesi. (Yayımlanmış Yüksek Lisans Tezi). Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Türkiye.

Baştürk, C., & Türkoguz, S. (2024). Dört aşamalı kimya tanı testinin yanıtlama süresi ve yanıtlama performanslarının incelenmesi. *International Journal of New Trends in Arts, Sports &Science Education(IJTASE), 13*(1), 31-43

Bilge, F. (2006). Examining the burnout of academics in relation to job satisfaction and other factors. *Social Behavior and Personality: an international journal, 34*(9), 1151-1160.

Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika, 82*(4), 1126-1148. https://doi.org/10.1007/s11336-016-9537-6

Bugbee A.C. (1996) The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education, 28*(3), 282- 299, https://doi.org/10.1080/08886504.1996.10782166.

Bulut, O. (2015). An empirical analysis of gender-based DIF due to test booklet effect. *European Journal of Research on Education, 3*(1), 7-16.

Caleon, I.S., & Subramaniam, R. (2010). Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education, 40*(3), 313-337.

Cox, D.R., & Oakes, D. (2018). *Analysis of survival data.* Chapman and Hall/CRC.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları (2. baskı).* Pegem.

Edgecomb, T.L., Van Schaack, A., & Marggraff, J. (2014). U.S. Patent No. 8,638,319. Washington, DC: U.S. Patent and Trademark Office.

Guo, H., Rios, J.A., Haberman, S., Liu, O.L., Wang, J., & Paek, I. (2016). A new procedure for detection of students' rapid guessing responses using response time. *Applied Measurement in Education, 29*(3), 173-183. https://doi.org/10.1080/08957347.2016.1171766.

Gürçay, D., & Gülbaş, E. (2015). Development of three-tier heat, temperature and internal energy diagnostic test. *Research in Science and Technological Education, 33*(2), 197-217. https://doi.org/10.1080/02635143.2015.1018154.

Gürel, D.K., Eryılmaz, A., & McDermott, L.C. (2015). A review and comparison of diagnostic ınstruments to ıdentify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(5), 989-1008.

Haladyna, T.M., & Downing, S.M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23* (1), 17-27

Halkitis, P.N., Jones, J.P., & Pradhan, J. (1996, April). *Estimating testing time: The effects of item characteristics on response latency.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Hanley, C. (1962). The "difficulty" of a personality inventory item. *Educational and Psychological Measurement, 22*(3), 577-584.

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *The Physics Teacher, 33*, 502-506. https://doi.org/10.1119/1.2344278.

İnceoğlu, F. (2013). *Sağkalım analiz yöntemleri ve karaciğer nakli verileri ile bir uygulama (Yüksek lisans tezi)*. İnönü Üniversitesi Sağlık Bilimleri Enstitüsü, Türkiye.

Kaltakçı, D. (2012). *Fizik öğretmen adaylarının geometrik optik ile ilgili kavram yanılgılarını ölçmek amacıyla dört basamaklı bir testin geliştirilmesi ve uygulanması (Yayınlanmamış Doktora Tezi).* Orta Doğu Teknik Üniversitesi, Türkiye.

Karasar, N. (2005). *Bilimsel araştırma yöntemi (17. Baskı).* Nobel yayın dağıtım, 81-83.

Klein, J.P., & Moeschberger, M.L. (2003). *Survival analysis techniques for censored and truncated data.* Springer-Verlang.

Kline, R.B. (2011). *Principles and practice of structural equation modeling (3rd. Edition).* Guilford.

Kong, X.J., Wise, S.L., & Bhola, D.S. (2007). Setting the Response Time Threshold Parameter to Differentiate Solution Behavior From Rapid-Guessing Behavior. *Educational and Psychological Measurement, 67*(4), 606– 619. https://doi.org/10.1177/0013164406294779.

Lasry, N., Watkins, J., Mazur, E., & Ibrahim, A. (2013). Response times to conceptual questions. *Am. J. Phys., 81*, 703.

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel psychology, 28*(4), 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x.

Lee, E.T., & Wang, J. (2003). *Statistical methods for survival data analysis*. John Wiley&Sons.

Lee, Y.H., & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling, 53*(3), 359-379.

Lunz, M.E., Bergstrom, B.A., & Gershon, R.C. (1994). Computer adaptive testing. *International Journal of Educational Research, 21*(6), 623-634.

Ma, L., Wise, S.L., Thum, Y.M., & Kingsbury, G. (2011, April). *Detecting response time threshold under the computer adaptive testing environment.* In annual meeting of the National Council on Measurement in Education, New Orleans.

Mehlhorn, S., Parrott, S.D., Mehlhorn, J., Burcham, T. Roberts, J.,& Smartt, P. (Şubat, 2011). *Using Digital Learning Objects to Improve Student Problem Solving Skills*. Southern Agricultural Economics Association Annual Meeting. Corpus Christi, Texas,United States.

Meyer, J.P. (2010). A Mixture rasch model with ıtem response time components. A*pplied Psychological Measurement, 34*(7), 521–538

Moharkan, Z.A., Choudhury, T., Gupta, S.C., & Raj, G. (2017). *Internet of Things and its applications in E-learning.* In Proceedings of the 3rd International Conference on Computational Intelligence and Communication Technology (CICT). IEEE, Ghaziabad India, 1–5. https://doi.org/10.1109/CIACT. 2017.7977333.

Önsal, G. (2016). *Özel görelilik kuramıyla ilgili kavram yanılgılarını belirlemeye yönelik dört aşamalı bir testin geliştirilmesi ve uygulanması (Yüksek lisans tezi)*. Gazi Üniversitesi, Türkiye.

Özdamar, K. (2003). *SPSS ile biyoistatistik.* 5. Baskı. Kaan Kitapevi Yayınları.

Peşman, H., & Eryılmaz, A. (2010) Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103*(3), 208-222. https://doi.org/10.1080/00220670903383002.

Pokropek, A., & Khorramdel, L. (2024). Analyzing Test-Taking Behaviors Through Process Data Using the IRT Explanatory Model for Guessing. *PsyArXiv.* https://doi.org/ 10.31234/osf.io/3krct.

Ranger, J., & Kuhn, J.T. (2011). A flexible latent trait model for response times in tests. *Psychometrika, 77,* 31–47.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Scheike T., Klein J.P., Van Houwelingen H.C., & Ibrahim J.G. (2014). *Handbook of Survival Analysis.* Chapman & Hall.

Schnipke, D. L., & Scrams, D. J. (2002). *Exploring issues of examinee behavior: Insights gained from response-time analyses* (Eds. C.N. Mills, M.T. Potenza, J.J. Fremer, & W.C. Ward). Computer-based testing: Building the foundation for future assessments (pp. 237–266). Lawrence Erlbaum Associates.

Schnipke, D.L., & Scrams, D.J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement, 34*, 213–232.

Semmes, R., Davison, M.L., & Close, C. (2011). Modeling individual differences in numerical reasoning speed as a random effect of response time limits. *Applied Psychological Measurement, 35*(6), 433-446.

Setzer, J.C., Wise, S.L., Van Den Heuvel, J.R., & Ling, G. (2013) An ınvestigation of examinee test-taking effort on a large-scale assessment. Applied Measurement in Education, 26(1), 34-49. https://doi.org/10.1080/08957347.2013.739453.

Streiner, D.L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of personality assessment, 81*(3), 209-219.

Swanson, D.B., Case, S.M., Ripkey, D.R., Clauser, B.E., & Holtman, M.C. (2001). Relationships among item characteristics, examine characteristics, and response times on USMLE Step 1. *Academic Medicine, 76*(10), S114-S116.

Swerdzewski, P.J., Harmes, J.C., & Finney, S.J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education, 24*(2), 162-188.

Taber K.S. (2017). The Use of Cronbach's Alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.*, 1–24. https://doi.org/10.1007/s11165-016- 9602-2.

Terzi, Y. (2003). *Sansürlü veriler için sağkalım analizi ve gerçek verilere uygulaması (Yayınlanmamıs doktora tezi).* Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Türkiye.

Therneau T.M., & Grambsch P.M. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag.

Türkoguz, S. (2020). Comparison of threshold values of three-tier diagnostic and multiple-choice tests based on response time. *Anatolian Journal of Education, 5*(2), 19-36.

Ünsal, A.A. (2019). Fen bilgisi öğretmen adaylarının gaz basıncı konusundaki kavram yanılgılarının belirlenmesi (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Türkiye.

Van Der Linden, W. J., & Glas, C. A. (2010). *Elements of Adaptive Testing.* Springer.

Weeks, J.P., Von Davier, M., & Yamamoto, K. (2016). Using response time data to inform the coding of omitted respo nses. *Psychological Test and Assessment Modeling, 58*(4), 671-701.

Wise, S., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183.

Wise, S.L. Bhola, D., & Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: The effortmonitoring CBT. *Educational Measurement: Issues and Practice, 25*(2), 21–30.

Wise, S.L., & DeMars, C.E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43*, 19–38.

Wise, S.L., & Ma, L. (2012, April). *Setting response time thresholds for a CAT item pool: The normative threshold method.* In annual meeting of the National Council on Measurement in Education, Vancouver, Canada (pp. 163-183).

Wise, S.L., Kingsbury, G.G., Thomason, J., & Kong, X. (2004, April). *An investigation of motivation filtering in a statewide achievement testing program*. In annual meeting of the National Council on Measurement in Education, San Diego, CA.

Yay, M. Çoker, E., & Uysal, Ö. (2007). Yaşam analizinde cox regresyon modeli ve artıkların incelenmesi. *Cerrahpaşa Tıp Dergisi, 38,* 139 – 145

Yetkin, B.B. (2006). *Cox Regresyon analizi ve bir uygulaması (Yayınlanmamış yüksek lisans tezi).* Mimar Sinan Üniversitesi Fen Bilimleri Enstitüsü, Türkiye.

Zacks, S. (1992). *Introduction to reliability analysis. Probability models and statistical methods.* Springer-Verlag.

EXTEND ABSTRACT

Introduction

When students encounter a test item, they can respond to the test item in two ways: rapid guessing behavior or solution behavior. It is important to calculate the threshold value to understand whether students show rapid guessing behavior or solution behavior. This study aimed to examine the threshold values of the four-tier chemistry diagnostic test and the multiple choice chemistry test.

Problem Statement

The problem statement of this study was determined as “What are the threshold values of the Four-Tier Chemistry Diagnostic Test (FTCDT) and Multiple Choice Chemistry Test (MCCT) ?” Four sub-problems were determined regarding the solution of the main problem. These are:

1- How are the threshold values of FTCDT and MCCT according to the Test Respond Performance (TRP) graph?

2- How are the threshold values for the internal validity rates where the I. and III. tiers of FTCDT evaluated together according to the time-based graph?

3- How are the response threshold values of FTCDT and MCCT according to the TRPs according to the Cox-Hazard model?

4- How are the threshold values for the internal validity rates where the I. and III. tiers of FTCDT evaluated together according to the Cox-Hazard analysis?

Method

3.1.Research Model

In this study, the relational survey model, which is one of the general survey models among quantitative research methods, was used (Karasar, 2005).

3.2.Participants

The study was carried out with the participation of 149 pre-service teachers studying in the science teaching department of Dokuz Eylül University, Buca Faculty of Education in the 2020-2021 academic year.

3.3.Data Collection Process and Tools

Four-Tier Diagnostic Chemistry Testing (FTDCT)

In this study, the Four-Tier Diagnostic Chemistry Test (FTDCT) consisting of 9 items on the subject of gas pressure developed by Ünsal (2019) was used. The scientific knowledge reliability of the FTDCT used in this study was calculated as 0.460 (based on the condition of receiving 1 point if all tiers are correct); and the misconception reliability of KR-20 (based on the condition of giving wrong answers in Tiers I and III and being sure in Tiers II and IV) was calculated as 0.570.

Multiple Choice Chemistry Test (MCCT)

In this study, Multiple Choice Chemistry Test on gas pressure was also used. In fact, MCCT is the first tier of the 9-item FTDCT on gas pressure developed by Ünsal (2019). MCCT was used by removing other tiers from FTDCT. For the KR-20 reliability analysis of MCCT, correct answers were given 1 and other answers were given 0. For this study, the KR-20 reliability coefficient of MCCT was found to be 0.520.

3.3.Data Collection and Analysis

Data Collection Process

Data collection was carried out with FTDCT and MCCT applied to students studying in the Department of Science Education at Dokuz Eylül University, Buca Faculty of Education. While the data collection process of the study was being carried out, the COVID-19 pandemic occurred and the courses were conducted online via computer. Therefore, the data of this study were collected in online test environments with computers.

Data Coding and Scoring Process

In this study, scores were calculated based on scientific knowledge, false positives, false negatives, lack of knowledge and misconceptions. In each tier of the questions in the entire test, correct answers were coded as “1” and wrong answers were coded as “0”. Scoring for scientific knowledge was “1-1-1-1”, scoring for positive errors was “1-1-0-1”, scoring for negative errors was “0-1-1-1” and scoring for misconceptions was “0-1-0-1”.

Data Analysis Process

KR-20 formula was used for reliability, explanatory factor analysis and confirmatory factor analysis were used for construct validity. TRP values for threshold value were subjected to Cox-Hazard analysis in SPSS statistics program.

Findings and Conclusion

In this study, the findings are presented in the order of the sub-problems. The first sub-problem is defined as "How are the threshold values of FTDCT and MCCT according to the TRP graph?" According to the study findings, the pre-service teachers showed solution behavior approximately after the 18th second in the first tier of FTDCT, and after the 6th second in the third tier. In MCCT, they showed solution behavior approximately after the 23rd second.

The second sub-problem of this study was defined as “What are the threshold values for internal validity rates when the I. and III. Tiers of FTDCT are evaluated together according to the time-based graph?” The threshold value of the “scientific knowledge” level with the code “11” was found as 6th sec., the threshold value of the “false positive” level with the code “10” was found as 9th sec., and the threshold value of the “misconception” level with the code “00” was found as 11th sec. However, the threshold value for the “false negative” level with the code “01” could not be found.

The third sub-problem of this study was defined as “How are the response threshold values of FTDCT and MCCT compared to the Cox-Hazard model?” According to the Cox-Hazard analysis, when the I. tier of FTDCT was examined, it was revealed that the pre-service teachers showed solution behavior after the 18th second. In the III. tier of FTDCT, the pre-service teachers showed solution behavior after the 17th second. In MCCT, the pre-service teachers showed solution behavior again from the 18th second, like in the I. tier of FTDCT. The answers given before these threshold values were determined as fast guessing behavior.

The fourth sub-problem of this study was defined as “What are the threshold values for internal validity rates when the I. and III. tiers of FTDCT are evaluated together according to the Cox-Hazard analysis?” According to the Cox-Hazard model, the internal validity threshold values were found at the 16th sec for the “scientific knowledge” level coded as “11”, at the 17th sec for the “false positive” level coded as “10”, and at the 28th sec for the “misconception” coded as 00. However, the threshold value point could not be found for the “false negative” level coded as “01”.