

BÖLÜM 1

TIMSS SORU KİTAPÇIKLARI ÜZERİNE ÖRNEK BİR UYGULAMA¹

Arş. Gör. Bedirhan TEKE²

Aybige ARABACI³

¹ Bu çalışmanın bir bölümü 27-29 Ocak 2023 tarihleri arasında Ankara'da düzenlenen V. Uluslararası Ankara Multidisiplinler Çalışmalar Kongresi'nde sözlü olarak sunulmuştur.

² Kilis 7 Aralık Üniversitesi, Kilisli Muallim Rıfat Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Kilis, Türkiye. bedirhan.teke@kilis.edu.tr, Orcid ID: 0000-0002-8565-215X

³ Bolu Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Bolu, Türkiye. aybige9045@gmail.com, Orcid ID: 0000-0002-4912-5782

GİRİŞ

Test kuramı, test verileri arasındaki ampirik ilişkiyi açıklayan sistematik bir teoridir. Test teorisyenleri test kuramını genellikle iki ekole ayırır: Biri temel olarak gerçek puan modeline dayanan Klasik Test Kuramı (KTK) (Lord & Novick, 1968); diğeri ise model temelli ölçmelere dayanan Madde Tepki Kuramı'dır (MTK) (Embretson & Reise, 2000). KTK'nın bazı sınırlılıklarının üstesinden gelmek için MTK 20. yüzyılın ortalarında ortaya çıkmıştır. KTK'na göre hesaplanan güçlük, ayırt edicilik ve güvenilirliğin örnekleme bağılı olması, aynı puana sahip her katılımcının aynı yetenek ve standart hataya sahip olması gibi özellikleri sınırlılıkları arasındadır (Crocker & Algina, 1986; Embretson & Reise, 2000).

KTK'da temel kavram belirli bir testteki gerçek puandır (Verhelst, 2004). KTK, gerçek puanın gözlenen puanlardan yola çıkarak kestirilebileceği varsayımına dayanır. Gözlenen puan ile gerçek puan arasındaki doğrusal bir ilişkiyi baz alarak test puanları üzerinden çıkarımda bulunur (Hambleton & Swaminathan, 1985). Tek bir kişiye ait olası test puanlarının dağılımının ortalaması o kişinin "gerçek puanıdır" (Verhelst, 2004), yani gerçek puan sonsuz çoklukta tekrar edilmiş ölçmelerin ortalaması olarak ele alınır (Hambleton & Swaminathan, 1985). Bir uygulama sonunda erişilen gözlenen puan bu ortalamaya eşit çıkmayacaktır. Bu ortalama (gerçek puan) ile gözlenen puan arasındaki bu fark ölçmenin hatasıdır (de Ayala, 2009; Verhelst, 2004). KTK'da kişinin gerçek puanını elde etmenin yolu gözlenen puanlara karışan hatayı en az seviyeye indirmektir (Crocker & Algina, 1985c). KTK ile elde edilen madde güçlük ve ayırt edicilik indekleri gibi test madde istatistikleri örnekleme bağılı olarak hesaplanır. Yani katılımcıların yetenekleri kendilerine uygulanan test maddelerinin ayırt ediciliğini ve güçlüğüne bağlıdır (Hambleton & Swaminathan, 1985). Dolayısıyla, test ve madde istatistikleri sadece katılımcılar açısından tanımlanabilir (Hambleton vd., 1991).

MTK ise katılımcıların yeteneklerinin, belirli test maddelerinden bağımsız olarak tahmin edilebileceği varsayımı üzerine yapılandırılmıştır (Hambleton & Swaminathan, 1985). Yani, MTK'da madde parametreleri test grubunun yeteneğine göre değişmez ve bireyin yeteneği maddeyi doğru yanıtlama olasılığı ile ilişkilidir (Wu & Adams, 2007). MTK, katılımcıların yeteneğini yani gizil özelliği üzerindeki konumu ile altında yatan faktörü belirlemek için oluşturulmuş maddelere verilen yanıtlar arasındaki ilişkiyi anlamlandırmak için oluşturulmuş genel bir çerçevedir (Verhelst, 2004; Wu &

Adams, 2007). MTK, bu ilişkiyi ortaya koymak için pek çok model tanımlamış (Hambleton & Jones, 1993) ve bu modelleri madde parametrelerinin ele alınmasına göre sınıflandırmıştır. Bunlar; madde ayırt edicilik parametresi(a), madde güçlük (b) ve şans (c) parametreleridir. Sadece a parametresini kullanan model bir parametrelili lojistik model, a ve b parametresini kullanan model iki parametrelili lojistik model ve a, b ve c parametrelerinin hepsini kullanan model üç parametrelili lojistik model olarak isimlendirilir (Crocker & Algina, 1986; de Ayala, 2009; Embretson & Reise, 2000). Bir parametrelili lojistik model olan Rasch modelinde, maddelerin aynı ayırt ediciliğe sahip ve şans başarısının sıfır olduğu varsayılır (Wu & Adams, 2007).

Bu çalışmada, TIMSS 2015 Kitapçık 3'teki başarı testinde yer alan 28 maddenin incelenmesi ve analiz sonucunda elde edilen KTK ve MTK çıktılarının yorumlanması amaçlanmıştır. Uluslararası Eğitim Başarısını Değerlendirme Derneği'nin (IEA) 1995'ten beri uyguladığı dört yılda bir yapılan TIMSS, dördüncü ve sekizinci sınıf öğrencilerinin matematik ve fen becerilerini değerlendirir. Bu tür kaynaklardan elde edilen bilgiler, uluslararası bir bağlamda öğrencilerin başarılarına ilişkin gelecekte eğitim gerektiren alanlardaki gelişimleri için geri bildirim sağlayabilmesi ve öğrenci ilerlemesindeki eğilimleri analiz etme olanağı sağlaması açısından önemlidir (Lee vd., 2011). Bu sebeple, çalışma gerçek veri seti üzerinden hem MTK hem de KTK ile madde analizlerini içermesi bakımından önem taşımaktadır.

1. ÖRNEK UYGULAMA

Bu çalışmada, internet erişimli Trends in International Mathematics and Science Study (TIMSS) 2015 Türkiye sekizinci sınıf öğrencilerine yönelik uygulanan kitapçıklardan biri olan Kitapçık 3'ün içerik, betimsel ve istatistiksel olarak analiz edilmesi amaçlandığından doküman analizi yöntemi kullanılmıştır. Bu yöntemin seçilmesindeki bir diğer gerekçe ise, elektronik olarak paylaşılan TIMSS veri setinin gözden geçirilerek analiz edilmesi sonucunda değerlendirilmesine fırsat tanınmasıdır (Bowen, 2009). Çalışmanın evrenini ise TIMSS 2015 Türkiye örneklemini içerisinde yer alan sekizinci sınıf düzeyinde öğrenim gören toplam 6079 öğrenci oluşturmakta iken örneklemini TIMSS 2015 veri setinde Kitapçık 3'ü cevaplayan toplam 440 öğrenci oluşturmaktadır.

1.1. Kitapçığın Betimlenmesi ve Hazırlığı Süreci

Bu çalışmada incelenen kitapçık, TIMSS 2011’de uygulanan M04 matematik blok soruları kaynak gösterilerek oluşturulmuş M03 matematik bloğu ile TIMSS 2015 kapsamında yeni soru maddeleriyle hazırlanan M04 matematik bloğunun bir araya getirilmesiyle oluşturulmuş ve Booklet 3 ismini almıştır. M03 ve M04 blok soru paketleri içerisinde yer alan maddelerin içerik alanlarına, konu alanlarına ve bilişsel etki alanlarına göre birleştirilerek oluşturulan Booklet 3’e ait betimsel analizler Tablo 1’de verilmiştir.

Tablo 1: Booklet 3’te Yer Alan Maddelere Ait İçerik Değerlendirmesi

Soru	İçerik Alanı	Konu Alanı	Bilişsel Etki Alanı	Maddelerin Kodlanması
M052209	Sayılar	Kesirler, Ondalık Sayılar ve Tam Sayılar	Bilme	0-1
M052035	Sayılar	Kesirler, Ondalık Sayılar ve Tam Sayılar	Bilme	0-1
M062212	Sayılar	Kesirler, Ondalık Sayılar ve Tam Sayılar	Uygulama	0-1
M052142	Sayılar	Tam Sayılar	Uygulama	0-1
M052006	Sayılar	Tam Sayılar	Muhakeme	0-1
M062329	Sayılar	Tam Sayılar	Bilme	0-1
M052016	Sayılar	Oran, Orantı ve Yüzde	Uygulama	0-1
M062151	Sayılar	Oran, Orantı ve Yüzde	Muhakeme	0-1
M062346	Sayılar	Oran, Orantı ve Yüzde	Uygulama	0-1
M052064	Cebir	İfadeler ve İşlemler	Bilme	0-1
M052126	Cebir	İfadeler ve İşlemler	Uygulama	0-1
M052103	Cebir	İfadeler ve İşlemler	Bilme	0-1
M062056	Cebir	İfadeler ve İşlemler	Bilme	0-1
M052066	Cebir	Denklemler ve Eşitsizlikler	Uygulama	0-1
M062078	Cebir	Denklemler ve Eşitsizlikler	Muhakeme	0-1
M062317	Cebir	İlişkiler ve Fonksiyonlar	Bilme	0-1
M062350	Cebir	İlişkiler ve Fonksiyonlar	Uygulama	0-1
M052041	Geometri	Geometrik şekiller	Muhakeme	0-1
M052057	Geometri	Geometrik şekiller	Muhakeme	0-1
M052417	Geometri	Geometrik şekiller	Uygulama	0-1
M062284	Geometri	Geometrik şekiller	Bilme	0-1
M062245	Geometri	Konum ve Hareket	Uygulama	0-1
M062287	Geometri	Geometrik Ölçüm	Muhakeme	0-1

Tablo 1 devamı

M052501	Veri ve Olasılık	Veri Yorumlama	Muhakeme	0-1
M052410	Veri ve Olasılık	Veri Yorumlama	Uygulama	0-1
M052170	Veri ve Olasılık	Veri setinin özellikleri	Uygulama	0-1
M062345A	Veri ve Olasılık	Veri setinin özellikleri	Uygulama	0-1-2
M062115	Veri ve Olasılık	Şans/Olasılık	Uygulama	0-1

Tablo 1 incelendiğinde, Booklet 3'ün 28 maddeden oluştuğu görülmektedir. Bu maddeler içerisinde yer alan içerik alanları bazında incelendiğinde, Sayılar içerik alanının (%32,14) diğer içerik alanlarına kıyasla daha fazla oranda bulunduğu tespit edilmiştir. Sayılar içerik alanını sırasıyla Cebir (%28,57), Geometri (%21,42) ve Veri ve Olasılık (%17,85) içerik alanları takip etmektedir. Booklet 3 bilişsel etki alanları bazında değerlendirildiğinde ise, uygulama (%46,42) bilişsel etki alanının bilme (%28,57) ve muhakeme (%25) bilişsel etki alanlarına kıyasla daha fazla oranda bulunduğu saptanmıştır. Ek olarak Booklet 3 içerisinde yer alan maddelere ait cevapların genellikle 0-1'li (M062345A maddesi hariç) kodlama sisteminde olduğu gözlemlenmiştir.

1.2. MTK Varsayımlarının Test Edilmesi

Bu bölümde, MTK varsayımlarının (tek boyutluluk, model uyumu, madde karakteristik eğrisinin (MKE) monotonik olarak artması ve yerel bağımsızlık) test edilme süreci verilmektedir.

1.2.1. Tek Boyutluluk

Seçilen kitapçığın analizinde kullanılacak olan modelin tek boyutlu olması nedeniyle kitapçıktan elde edilen verilerin de tek boyutlu olması gerekmektedir. Bu kontrolü sağlarken ise faktör analizi yöntemleri kullanılmaktadır.

Öncelikle TIMSS Türkiye datasından Kitapçık 3'te yer alan maddelerin ve bu maddelere verilen cevapların ayrıştırılması gerekmektedir. Bu sebeple <https://timssandpirls.bc.edu/timss2015/international-database/> web adresinden gerekli SPSS veri setleri elde edilmiş ve Kitapçık 3 veri dosyası ayrıştırılmıştır. Bu kitapçığın puan kodlama sisteminde ise analizlerde kullanılacak olan Rasch Model kapsamında 0-1'li puanlama yapılmıştır.

Ancak puanlama aşamasında veri seti içerisinde işaretlenmemiş bazı maddelerin bulunduğu saptanmıştır (M062345B, M062345BA, M062345BB, M062345BC ve M062345BD). Bu sebeple analize devam edilebilmesi için bu maddeler analiz kapsamının dışında tutulmuştur. Ayrıca bu aşamada bir maddenin (M062345A) 0-1-2'li kodlandığı görülmüştür. Rasch modele uymayan bu puanlama sistemi, model uyumsuzluğuna sebebiyet verme olasılığı göz önünde bulundurularak gerekli revizeler (0-1 puanlı maddeler 0 puan, 2 puanı 1 puan olarak tekrardan kodlanmıştır) yapılmıştır. Yapılan revizeler sonucunda 440 öğrencinin cevap verdiği ve 28 soru maddesinden oluşan Kitapçık 3 elde edilmiştir.

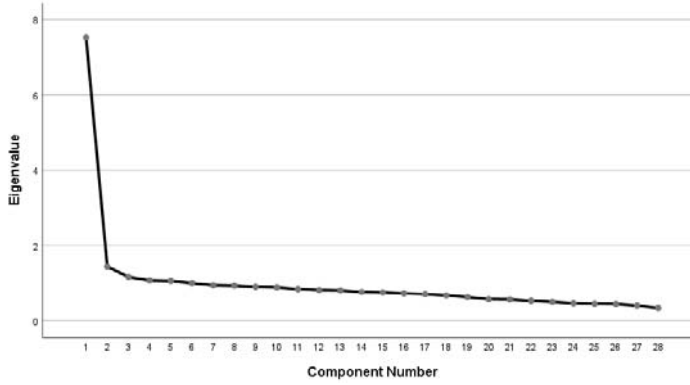
Faktör analizi sürecinde dikkat edilmesi gereken adım, seçilen kitapçığın kullanılacak modelle aynı boyutta olup olmadığının tespitidir. Bu kapsamda Kitapçık 3 SPSS dosyasının faktör analizine uygun olup olmadığını incelemek amacıyla Kaiser-Meyer-Olkin (KMO) katsayısı ve Bartlett küresellik testi sonuçları tespit edilmiştir. Büyüköztürk (2002) çalışmasında, KMO değerinin 0,6'dan yüksek çıkması gerektiğini ve bu değer 0,8'in üstünde olmasının ise mükemmel sayılacağını belirtmektedir. Ayrıca, Bartlett testinin korelasyon matrisini istatistiki olarak anlamlı bulunmasının ($p < 0,05$) çalışılan örneklemin açımlayıcı faktör analizi için uygun/yeterli olduğuna işaret ettiğini söylemektedir. Yapılan analiz sonucunda Kitapçık 3 içerisinde yer alan maddelere ait KMO değerinin 0,937 olduğu ve Bartlett küresellik testinin ki-kare değeri anlamlı bulunmuştur ($p < 0,05$). Böylelikle veri setinin faktör analizine uygun olduğu kararlaştırılmıştır. AFA için veri setinin uygunluğu tespit edildikten sonra metot olarak, yüksek boyutlu verilerde en yüksek varyans ile veri setini tutmak ancak bunu yaparken de boyut indirgemeyi sağlamak amacıyla principal component seçilmiştir. Daha sonra veri seti için rotasyon yönteminin seçimine geçilmiştir. Rotasyon yöntemleri *orthogonal* ve *oblique* yöntem olmak üzere ikiye ayrılmaktadır. Ortogonal yöntemler faktörler arası korelasyona izin vermezken, oblique yöntemler faktörler arası korelasyona izin vermektedir (Yaşlıoğlu, 2017). Veri setinde bulunan maddelerin birbirleriyle korelasyonları incelendiğinde rotasyon yönteminin direk oblimin olmasına karar verilmiştir. Faktör yükü, maddenin faktörlerle arasındaki ilişkiyi açıklayan bir katsayıdır. Bir maddenin bir yapıya ya da faktörü iyi ölçtüğünü söyleyebilmek için faktör yükünün değerinin minimum 0,30 olması gerekmektedir (Kline, 1994; Karaman, 2015). Bu sebeple Kitapçık 3'te yer alan maddelerin oluşturduğu yapıyı test etmek amacıyla faktör yükü 0,32 olarak seçilmiştir. Tablo 2'de soru maddelerinin

oluşturduğu faktör sayısı ile bu faktörlere ait açıklanan toplam varyans değerleri verilmektedir.

Tablo 2: Booklet 3 Varyans Çıktısı

Boyutlar	Varyans	Birikimli Varyans
1	26,892	26,892
2	5,142	32,034
3	4,151	36,185
4	3,835	40,021
5	3,789	43,809
6	3,576	47,385

Tablo 2 incelendiğinde, 1. faktör yapısı altında toplanan maddelerin açıkladığı varyans değerinin 26,892 olduğu görülmektedir. Bu faktör yapısının devamında ise diğer faktör yapılarının açıkladığı varyans değerlerinin düşük olması, Kitapçık 3'ün altı faktörlü bir yapıya sahip olduğunu gösterse de 1. faktör yapısının yoğunluğu tek faktörlü yapıyı desteklemektedir. Boyutların dağılımına ilişkin öz değer dağılım grafiği Şekil 1'de gösterilmiştir.



Şekil 1: Boyutlara İlişkin Öz Değer Dağılım Grafiği

Şekil 1 incelendiğinde, 1. faktör yapısının Kitapçık 3 genelinde dağılımının baskın olduğu ve analizlerde kullanılacak olan modelin tek boyutlu olması nedeniyle kitapçıktan elde edilen verilerin de tek boyutlu olması gerektiği sonucu doğrulanmış olmaktadır.

1.2.2. Model Uyumu Varsayımı

Bu varsayımın test edilmesindeki gerekçe, modelin kullanışlı oluşunun model-veri uyumuna bağlı olmasıdır. Model-veri uyumunu değerlendirmede uyum iyiliği indeksleri kullanılmakta olup bu çalışmada, model-veri uyumu maksimum olabilirlik fonksiyonu olan -2loglikelihood değeri ve Rasch modelin kişi homojenliğini değerlendiren Andersen's LR-testi ile test edilmiştir. Bu analizlere ait bulgular incelendiğinde;

-2loglikelihood değeri
-4712.114 (df=28)

Andersen LR-test
LR-value: 274,704
Chi-square df: 26
p-value: 0

şeklinde verilebilir.

Bu çıktılar yorumlandığında -2loglikelihood değeri ne kadar yüksek olursa model ile veri seti arasındaki uyumun o kadar iyi olduğu söylenmektedir (Andersen, 1972; de Ayala, 2009). Veri setine ait -2loglikelihood değeri göz önüne alındığında model ile veri setinin uyumlu olduğu söylenebilir. Ancak Andersen LR-testi sonuçları incelendiğinde ise, hesaplanan LR değerinin (274,704) anlamlı ($p>0,05$) olmadığı sonucuna ulaşılmıştır. Böyle bir sonucun ortaya çıkmasındaki gerekçe ise, veri seti içerisinde farklılaşan madde güçlüklerinin bulunmasıdır (Andersen, 1973). Bu sebeple veri setinde yer alan maddelerin uyumunun incelenmesi gerekmektedir.

1.2.3. MKE'nin Monotonik Artış Göstermesi

Bu varsayımla, madde karakteristik fonksiyonunun örtük değişkenler ile madde yanıtları arasındaki doğru ilişkiyi yansıtma durumu incelenmektedir. Varsayım gereği MKE'ler monotonik olarak artmakta ise bireylerin yetenek düzeyi arttıkça doğru yanıt verme olasılığının arttığını ifade eden bir eğri oluşmaktadır (Hambleton vd., 1991). Booklet 3 kapsamında her bir madde için örtük değişken ile madde yanıtları arasında doğrusal bir ilişki olduğu ve oluşan MKE eğrilerinin eğimleri bakımından farklılaştıkları görülmektedir.

1.2.4. Yerel Bağımsızlık

Alan yazın incelendiğinde (Embretson & Reise, 2000; Hambleton vd., 1991), tek boyutluluk varsayımının sağlanması yerel bağımsızlık varsayımının da sağlanması için yeterli olduğu belirtilmektedir. Bu nedenle, Booklet 3 veri

setinin tek boyutluluk varsayımını sađlamasından tr, yerel bađımsızlık varsayımını da sađladıđı kabul edilmiřtir.

1.3. Verilerin Analizi

KTK analizlerinde ilk olarak, veri setine ait yzdelik sıra ve z-puan deđiřkenleri ayrı ayrı Microsoft Excel yardımıyla hesaplanmıřtır. İkinci olarak, veri setine ait gvenirlik kat sayıları hem TAP programı hem de test yarılama yntemi kullanılarak hesaplanmış ve karřılařtırılmıřtır. nc olarak, veri setinde yer alan maddelerin madde glk ve ayırt edicilik indeksleri (Adj Pt Bis⁴) hesaplanmıřtır. Son ařamada ise, veri setinde yer alan maddelerin ierik alanları bazında eldirici analizleri yapılmıřtır. MTK analizlerinde ise R studio programının kullanılmasıyla veri setine ait; eta (madde glđ), thesholds eřik deđeri (bir sorunun yapılma olasılıđının %50 olduđu deđeri), p-value (anlamlılık deđeri), outfit MSQ (ađırlařtırılmamıř)-infit MSQ (ađırlařtırılmıř) deđerleri ile outfit t (ađırlařtırılmıř) deđerleri sırasıyla hesaplanmış ve yorumlanmıřtır.

2. UYGULAMA IKTILARI

Bu blmde ilk olarak KTK kapsamında veri setine ait yzdelik sıra, z-puan deđiřkeni, gvenirlik kat sayısı, madde glk ve ayırt edicilik indekslerinin yanı sıra Booklet 3'te yer alan maddelere ait eldirici analizleri daha sonra ise MTK kapsamında veri setine ait eta, thesholds eřik deđeri, p-value, outfit MSQ-infit MSQ deđerleri ile outfit t-infit t deđerleri sırasıyla hesaplanmıřtır.

2.1. Yzdelik Sıra

Tablo 3'te Booklet 3'te yer alan maddelere ait frekans deđerleri, kmlatif yzde (%) deđerleri, birikimli frekans ve yzdelik sıra deđiřkenlerine ynelik analiz ıktıları verilmektedir.

⁴ Adj Pt Bis deđeri, ilgili soru maddesinin dahil edilmeden hesaplanan toplam test puanı zerinden yapılan iřlemdir.

Tablo 3: Booklet 3 Yüzdeler Sıra Analizi

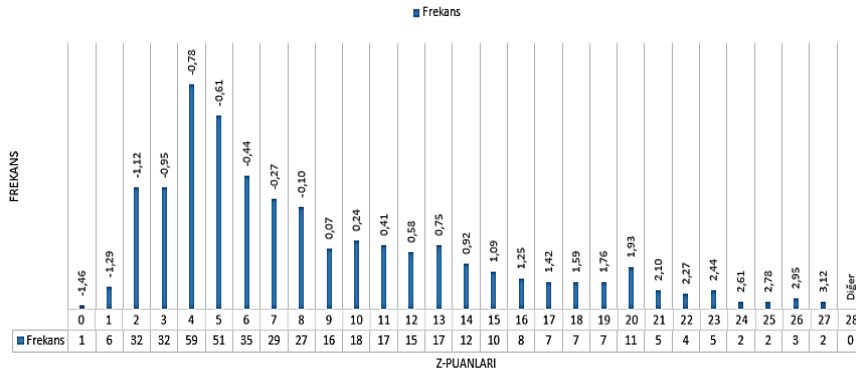
Alnabilecek Puanlar	Frekans	Kümülatif %	Birikimli frekans	Yüzdeler Sıra
0	1	0,23%	1	0%
1	6	1,59%	7	1%
2	32	8,86%	39	5%
3	32	16,14%	71	13%
4	59	29,55%	130	23%
5	51	41,14%	181	35%
6	35	49,09%	216	45%
7	29	55,68%	245	52%
8	27	61,82%	272	59%
9	16	65,45%	288	64%
10	18	69,55%	306	68%
11	17	73,41%	323	71%
12	15	76,82%	338	75%
13	17	80,68%	355	79%
14	12	83,41%	367	82%
15	10	85,68%	377	85%
16	8	87,50%	385	87%
17	7	89,09%	392	88%
18	7	90,68%	399	90%
19	7	92,27%	406	91%
20	11	94,77%	417	94%
21	5	95,91%	422	95%
22	4	96,82%	426	96%
23	5	97,95%	431	97%
24	2	98,41%	433	98%
25	2	98,86%	435	99%
26	3	99,55%	438	99%
27	2	100,00%	440	100%
28	0	100,00%	440	100%

Tablo 3'te yer alan yüzdeler sıra değerleri, yüzdeler sıra satırındaki "birikimli frekans" değeri ile bir sonraki satırdaki "frekans" değerinin yarısı alınarak toplanır ve çıkan değeri birey sayısına bölünür. Böylelikle elde edilen değeri, yüzdeler sırayı temsil eder. Örneğin yüzdeler sıra değerleri, Booklet 3 veri setinde 4 puanını alan bir bireyin 440 kişinin %23'ünü, 7 puan alan bir bireyin 440 kişinin %52'ni ve 18 puanını alan bir bireyin ise 440 kişinin %90'ını geride bıraktığını ifade etmektedir. Ancak, yüzdeler sıra değişkeninin

istatistiksel bir hatası bulunmaktadır. Örneğin 4 puanını alan bir birey diğer bireylerin %23'ünü, 18 puanını alan bir birey ise diğer bireylerin %90'ını geride bırakmaktadır. Bu puanlara sahip iki öğrenci de sınav sürecinde bir tane daha doğru cevap işaretleselerdi sırasıyla puanları 5 ve 19 olacaktı. Bu öğrencilerin yeni puanlarına ait yüzdelik sıra değerleri incelendiğinde 5 puanını alan bir birey diğer öğrencilerin %35'ini, 19 puanını alan bir birey ise diğer öğrencilerin %91'ini geride bırakacaktı. Ancak, bu öğrencilerin bir soru daha yapmaları sonucunda geride bıraktıkları birey sayısı yüzde olarak farklı olacak (%12-%1) ve böylelikle bir sorunun anlamı farklılaşacaktır.

2.2. Z-puanları

Şekil 2'de öğrencilerin Booklet 3'ten aldıkları toplam test puanlarına ilişkin hesaplanan z-puanları verilmektedir. Bu hesaplama sürecine geçilmeden önce testten elde edilen toplam puanların ortalaması (8,59) ve standart sapması (5,90) hesaplanmıştır.



Şekil 2: Z Puanının Betimlenmesi

Şekil 2 incelendiğinde, ortalaması 8,59 olan Booklet 3'e ait soru maddelerinin ortalamasının ne kadar üstünde ya da altında kaldığı görülmektedir. Örneğin, 6 puanına karşılık gelen z standart puanı -0,44 iken 18 puanına karşılık gelen z standart puanı 1,59'dur. Bu değerler yorumlandığında, 6 puanını alan bir öğrenci ortalamasının -0,44 standart sapma altında iken 18 puanını alan bir öğrenci ortalamasının 1,59 standart sapma üzerindedir. Ayrıca, z-puanları grafiğinin sağa çarpık dağılım sergilediği görülmektedir. Yüzdelik sırada ortaya çıkan puan sorununa karşılık z puanları, bu sorunu ortadan kaldırmaktadır. Örneğin, 6 puan alan bir öğrenci 7

puan alırsa z puanı 0,17 artmakta; benzer şekilde 18 puan alan bir öğrenci 19 puan alırsa z puanı 0,17 artmaktadır. Bu sebepten ötürü z puanının, eşit miktarda aralıklarını koruduğu söylenebilir.

2.3. Güvenirlilik Kat Sayısı

Bu bölümde Booklet 3 kapsamında ele alınan soru maddelerine ait güvenirlilik kat sayıları hem TAP programı hem de test yarılama yöntemiyle ayrı ayrı hesaplanmış ve karşılaştırmalı olarak yorumlanmıştır.

2.3.1. TAP Programının Kullanım Süreci

TAP programının kullanımına geçilmeden önce, SPSS'den Microsoft Excel formatına aktarılan veri dosyası metin formatına getirilerek 0-1'li olacak şekilde kodlanmıştır. Daha sonra metin formatına getirilen veri seti TAP programına aktarılmış ve Booklet 3 veri setinin güvenirlilik katsayısı hesaplanması aşamasında kullanılmıştır. Tablo 4'te TAP program çıktıları yer almaktadır.

Tablo 4: Booklet 3 TAP çıktıları

Güvenirlilik Katsayısı	Varyans	Gerçek Puan Varyans	Hata Puan Varyans	Hata Standart Sapma (HSS)
0,88	34,78	30,68	4,10	2,03

Tablo 4 incelendiğinde, Booklet 3 kapsamında yer alan soru maddelerine ait güvenirlilik katsayısının 0,88; varyansın ise 34,78 olduğu görülmektedir. Bu iki değer göz önüne alındığında, soru maddelerine ait gerçek puan varyansının 30,68; hata puan varyansının ise 4,10 olduğu tespit edilmiştir. Ayrıca hata puan varyansının karekökü olan hata standart sapma değeri 2,03 olarak hesaplanmıştır.

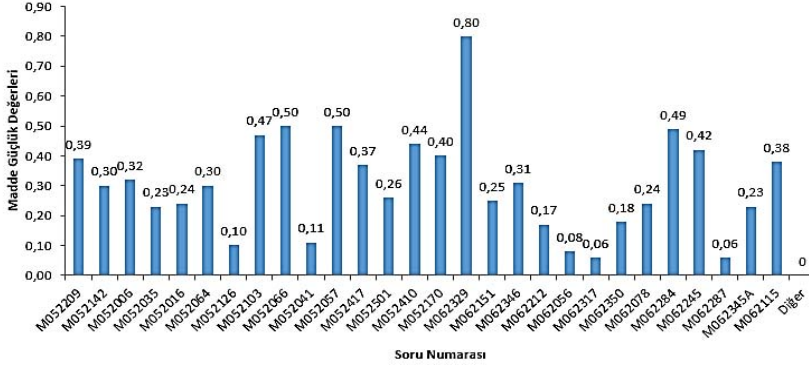
2.3.2. Test Yarılama Tekniğinin Kullanım Süreci

Test yarılama tekniğinde öncelikle Booklet 3'te yer alan soru maddelerinin (28 adet) oluşumunda yer alan M03 (15 soru) ile M04 (13 soru) soru blokları iki farklı test olarak düşünülmüştür. Uygulama adımlarına geçildiğinde ise, öncelikle her iki testin öğrenci bazında toplam puanları hesaplanmıştır. Daha sonra, hesaplanan toplam puanların testler bazında korelasyonları tespit edilmiştir (0,79). Ancak soru bloklarındaki farktan dolayı (madde sayısı) düzeltme formülü kullanılmış ve güvenirlilik kat sayısı 0,88 olarak hesaplanmıştır. Sonuç olarak hesaplanan değerler göz önüne alındığında, TAP programında hesaplanan güvenirlilik katsayısı ile test

yarılama tekniği ile hesaplanan güvenilirlik katsayısı değerinin benzer çıktığı gözlenmiştir.

2.4. Madde Güçlük Değerleri

Şekil 3'te Booklet 3 kitapçığına ait madde güçlük değerlerini betimleyen histogram grafiği yer almaktadır.



Şekil 3: Booklet 3 Kitapçığına Ait Madde Güçlük Değerleri

Madde güçlük indeksine göre maddelerin nasıl değerlendirilmesi gerektiği hakkında bilgi Tablo 5'te verilmiştir.

Tablo 5: Madde Güçlük İndeksi Değerleri ve Yorumu

Madde Güçlük İndeksi	Maddenin Değerlendirilmesi
0.00 – 0.29	Zor
0.30 – 0.49	Orta Güçlükte
0.50 – 0.69	Kolay
0.70 – 1.00	Çok Kolay

(Atılğan vd., 2016)

Tablo 5 incelendiğinde, madde güçlük indeksinin 0'a yaklaşması sonucu soru maddesine ait güçlüğü arttığı, 1'e yaklaştıkça ise güçlüğü azaldığı görülmektedir. Ayrıca, madde güçlük indeksinin 0,50 olması sorunun orta güçlükte olduğunu göstermektedir. Bu açıklama doğrultusunda Booklet 3 kapsamındaki soru maddelerine ait güçlük değerleri yorumlandığında; 13 maddenin zor, 14 maddenin orta güçlükte ve 1 maddenin çok kolay olduğu ve bundan dolayı da ortalama madde güçlük indeksinin orta güçlükte (0,307) olduğu tespit edilmiştir.

2.5. Madde Ayırt Edicilik Değerleri

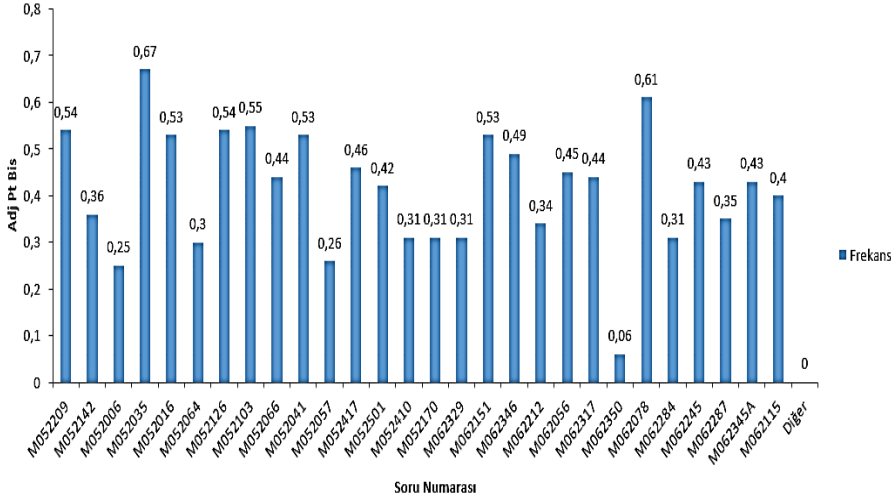
Atılğan vd. (2016) göre, madde ayırt edicilik indeksi değerlerinin yorumlanmasında kullanılması gerekli kriterler Tablo 6’da verilmiştir.

Tablo 6: Madde Ayırt Edicilik İndeksi Değerleri ve Yorumu

Madde Ayırt Edicilik İndeksi	Maddenin Değerlendirilmesi
0.19 ve daha küçük	Kesinlikle teste alınmamalı ya da tamamen düzeltilmelidir.
0.20 ile 0.29 arasında	Sınırdaki maddelerdir ve gerekirse düzeltilerek teste alınabilir.
0.30 ile 0.39 arasında	Düzeltilme yapılmaksızın ya da küçük düzeltmelerle teste alınabilir.
0.40 ve daha yüksek	Çok iyi işleyen maddeler, teste olduğu gibi alınabilir.

(Atılğan vd., 2016)

Tablo 6 incelendiğinde, bir soru maddesine ait madde ayırt edicilik indeksi değerinin 0,19 ve aşağısında olmaması gerektiği, ideal madde ayırt edicilik indeksi değerinin 0,40 ve üzerinde olması gerektiği vurgulanmaktadır. Ancak hesaplanan değerler göz önüne alındığında, madde ayırt edicilik indeksi değerlerinin Adj Pt Bis (nokta çift serili korelasyon katsayısı) değeri olması gerektiği literatürde vurgulanmaktadır (Wu & Adams, 2007). Bu kapsamda Şekil 4’te Booklet 3’e ait Adj Pt Bis değerlerini betimleyen histogram grafiği yer almaktadır.



Şekil 4: Booklet 3 Kitapçığına Ait Adj Pt Bis Değerleri

Sonuç olarak, Booklet 3 kapsamında soru maddelerine ait Adj Pt Bis değerleri yorumlandığında; 1 soru maddesinin kesinlikle teste alınmaması gerektiği (kötü), 2 soru maddesinin gerekirse düzeltilerek teste alınması gerektiği (orta), 8 soru maddesinin düzeltme yapılmaksızın ya da küçük düzeltmelerle teste alınması gerektiği (iyi) ve 17 soru maddesinin ise çok iyi işleyen maddeler oldukları için teste olduğu gibi alınması gerektiği sonucuna ulaşılmıştır. Ayrıca ortalama Adj Pt Bis değerinin 0,443 olması, Booklet 3 kapsamında yer alan soru maddelerinin ortalama olarak çok iyi işleyen maddelerden oluştuğunu ifade ettiği söylenebilir.

2.6. Çeldirici Analizleri

Çoktan seçmeli sorularda çeldiricilere ait üst grup ile alt grup arasındaki farkın negatif, doğru cevaba ait farkın pozitif olması beklenir (Hasançebi vd., 2020). Bu kapsamda çeldirici analizlerine geçmeden önce, veri setinin TAP programının algılayacağı bir dilde yazılması gerekmektedir. Bu yüzden ilk olarak, SPSS veri setinde yer alan çoktan seçmeli soru cevapları “1” A şıkkını, “2” B şıkkını, “3” C şıkkını, “4” D şıkkını ve “5” Boşu temsil edecek şekilde kodlanmıştır. Veri setinde yer alan açık uçlu sorular ise bu kısma dahil edilmemiştir. Böylelikle TAP programı için, SPSS verileri kodlanarak metin dosyası formatına getirilmiştir.

2.6.1. Sayılar İçerik Alanı ve Çeldiriciler

Tablo 7’de Sayılar içerik alanına ait konular kapsamındaki çeldiricilerin istatistiksel değerleri betimlenmiştir.

Tablo 7: Sayılar İçerik Alanına Ait Çeldiriciler

Soru	Grup	A	B	C	D	BOŞ
M052209	TOTAL	176 (0,400)	60 (0,136)	171* (0,389)	32 (0,073)	1 (0,002)
	High	14 (0,104)	6 (0,045)	110 (0,821)	4 (0,030)	0 (0,000)
	Low	87 (0,669)	20 (0,154)	12 (0,092)	11 (0,085)	0 (0,000)
	Diff	-73 (-0,565)	-14 (-0,109)	98 (0,729)	-7 (-0,055)	0 (0,000)
M062212		A	B	C	D	BOŞ
	TOTAL	200 (0,455)	75* (0,170)	38 (0,086)	126 (0,286)	1 (0,002)
	High	26 (0,194)	51 (0,381)	24 (0,179)	33 (0,246)	0 (0,000)
	Low	90 (0,692)	8 (0,062)	4 (0,031)	28 (0,215)	0 (0,000)
Diff	-64 (-0,498)	43 (0,319)	20 (0,148)	5 (0,031)	0 (0,000)	
M052142		A	B	C	D	BOŞ
	TOTAL	134* (0,305)	83 (0,189)	73 (0,166)	148 (0,336)	2 (0,005)
	High	72 (0,537)	29 (0,216)	4 (0,030)	28 (0,209)	1 (0,007)
	Low	10 (0,077)	20 (0,154)	42 (0,323)	57 (0,438)	1 (0,008)
Diff	62 (0,460)	9 (0,063)	-38 (-0,293)	-29 (-0,230)	0 (0,000)	
M052006		A	B	C	D	BOŞ
	TOTAL	143 (0,325)	83 (0,189)	67 (0,152)	142* (0,323)	5 (0,011)
	High	25 (0,187)	18 (0,134)	21 (0,157)	67 (0,500)	3 (0,022)
	Low	59 (0,454)	28 (0,215)	20 (0,154)	23 (0,177)	0 (0,000)
Diff	-34 (-0,267)	-10 (-0,081)	1 (0,003)	44 (0,323)	3 (0,022)	
M062329		A	B	C	D	BOŞ
	TOTAL	11 (0,025)	353* (0,802)	15 (0,034)	61 (0,139)	0 (0,000)
	High	0 (0,000)	133 (0,993)	0 (0,000)	1 (0,007)	0 (0,000)
	Low	9 (0,069)	68 (0,523)	6 (0,046)	47 (0,362)	0 (0,000)
Diff	-9 (-0,069)	65 (0,469)	-6 (-0,046)	-46 (-0,354)	0 (0,000)	

*: Doğru cevap

Tablo 7’de verilen M052209 ve M062329 soru maddelerine ait veriler incelendiğinde çeldiricilere ait işaretlemelerde üst grup ile alt grup arasındaki farkın negatif, doğru cevabın işaretlenmesinde ise üst grup ile alt grup arasındaki farkın pozitif olduğu görülmektedir. Ancak M062212, M052142 ve M052006 soru maddeleri incelendiğinde ise çeldiricilerin işaretlenmesinde üst grup ile alt grup arasındaki farkın pozitif olduğu dolayısıyla, çeldiricilerin ters yönde çalıştığı söylenebilir.

2.6.2. Cebir İçerik Alanı ve Çeldiriciler

Tablo 8’de Cebir içerik alanına ait konular kapsamındaki çeldiricilerin istatistiksel değerleri betimlenmiştir.

Tablo 8: Cebir İçerik Alanına Ait Çeldiriciler

Soru	Grup	A	B	C	D	BOŞ
M052064	TOTAL	78 (0,177)	138 (0,314)	131* (0,298)	81 (0,184)	12 (0,027)
	High	23 (0,172)	28 (0,209)	69 (0,515)	7 (0,052)	7 (0,052)
	Low	19 (0,146)	55 (0,423)	22 (0,169)	33 (0,254)	1 (0,008)
	Diff	4 (0,025)	-27 (-0,214)	47 (0,346)	-26 (-0,202)	6 (0,045)
M052103		A	B	C	D	BOŞ
	TOTAL	55 (0,125)	206* (0,468)	41 (0,093)	138 (0,314)	0 (0,000)
	High	2 (0,015)	121 (0,903)	3 (0,022)	8 (0,060)	0 (0,000)
	Low	27 (0,208)	19 (0,146)	20 (0,154)	64 (0,492)	0 (0,000)
	Diff	-25 (-0,193)	102 (0,757)	-17 (-0,131)	-56 (-0,433)	0 (0,000)
M052066		A	B	C	D	BOŞ
	TOTAL	69 (0,157)	82 (0,186)	222* (0,505)	65 (0,148)	2 (0,005)
	High	6 (0,045)	6 (0,045)	115 (0,858)	7 (0,052)	0 (0,000)
	Low	31 (0,238)	44 (0,338)	27 (0,208)	28 (0,215)	0 (0,000)
	Diff	-25 (-0,194)	-38 (-0,294)	88 (0,651)	-21 (-0,163)	0 (0,000)

Tablo 8 devamı

		A	B	C	D	BOŞ
M062350	TOTAL	78 (0,177)	124 (0,282)	149 (0,339)	81* (0,184)	8 (0,018)
	High	20 (0,149)	38 (0,284)	44 (0,328)	29 (0,216)	3 (0,022)
	Low	25 (0,192)	34 (0,262)	55 (0,423)	13 (0,100)	3 (0,023)
	Diff	-5 (-0,043)	4 (0,022)	-11 (-0,095)	16 (0,116)	0 (-0,001)

Tablo 8’de verilen M052103 ve M052066 soru maddelerine ait veriler incelendiğinde çeldiricilere ait işaretlemelerde üst grup ile alt grup arasındaki farkın negatif, doğru cevabın işaretlenmesinde ise üst grup ile alt grup arasındaki farkın pozitif olduğu görülmektedir. Ancak M052064 ve M062350 soru maddeleri incelendiğinde ise çeldiricilerin işaretlenmesinde üst grup ile alt grup arasındaki farkın pozitif olduğu dolayısıyla, çeldiricilerin ters yönde çalıştığı söylenebilir.

2.6.3. Geometri İçerik Alanı ve Çeldiriciler

Tablo 9’da Geometri içerik alanına ait konular kapsamındaki çeldiricilerin istatistiksel değerleri betimlenmiştir.

Tablo 9: Geometri İçerik Alanına Ait Çeldiriciler

Soru	Grup	A	B	C	D	BOŞ
M052057	TOTAL	95 (0,216)	77 (0,175)	49 (0,111)	218* (0,495)	1 (0,002)
	High	11 (0,082)	19 (0,142)	9 (0,067)	95 (0,709)	0 (0,000)
	Low	51 (0,392)	24 (0,185)	24 (0,185)	31 (0,238)	0 (0,000)
	Diff	-40 (-0,310)	-5 (-0,043)	-15 (-0,117)	64 (0,470)	0 (0,000)
M062284		A	B	C	D	BOŞ
	TOTAL	143 (0,325)	216* (0,491)	22 (0,050)	57 (0,130)	2 (0,005)
	High	19 (0,142)	96 (0,716)	5 (0,037)	14 (0,104)	0 (0,000)
	Low	72 (0,554)	29 (0,223)	8 (0,062)	19 (0,146)	2 (0,015)
Diff	-53 (-0,412)	67 (0,493)	-3 (-0,024)	-5 (-0,042)	-2 (-0,015)	

Tablo 9 devamı

		A	B	C	D	BOŞ
M062245	TOTAL	184* (0,418)	76 (0,173)	117 (0,266)	59 (0,134)	4 (0,009)
	High	103 (0,769)	14 (0,104)	11 (0,082)	6 (0,045)	0 (0,000)
	Low	20 (0,154)	30 (0,231)	55 (0,423)	24 (0,185)	1 (0,008)
	Diff	83 (0,615)	-16 (-0,126)	-44 (-0,341)	-18 (-0,140)	-1 (-0,008)

Tablo 9’da verilen M052057, M062284 ve M062245 soru maddelerine ait veriler incelendiğinde çeldiricilere ait işaretlemelerde üst grup ile alt grup arasındaki farkın negatif, doğru cevabın işaretlenmesinde ise üst grup ile alt grup arasındaki farkın pozitif olduğu görülmektedir.

2.6.4. Veri ve Olasılık İçerik Alanı ve Çeldiriciler

Tablo 10’da Veri ve Olasılık içerik alanına ait konular kapsamındaki çeldiricilerin istatistiksel değerleri betimlenmiştir.

Tablo 10: Veri ve Olasılık İçerik Alanına Ait Çeldiriciler

Soru	Grup	A	B	C	D	BOŞ
M052410	TOTAL	127 (0,289)	192* (0,436)	92 (0,209)	26 (0,059)	3 (0,007)
	High	22 (0,164)	89 (0,664)	22 (0,164)	1 (0,007)	0 (0,000)
	Low	49 (0,377)	25 (0,192)	38 (0,292)	16 (0,123)	2 (0,015)
	Diff	-27 (-0,213)	64 (0,472)	-16 (-0,128)	-15 (-0,116)	-2 (-0,015)
M052170		A	B	C	D	BOŞ
	TOTAL	61 (0,139)	174* (0,395)	124 (0,282)	73 (0,166)	8 (0,018)
	High	19 (0,142)	87 (0,649)	14 (0,104)	13 (0,097)	1 (0,007)
	Low	18 (0,138)	28 (0,215)	57 (0,438)	26 (0,200)	1 (0,008)
Diff	1 (0,003)	59 (0,434)	-43 (-0,334)	-13 (-0,103)	0 (0,000)	

Tablo 10 devamı

		A	B	C	D	BOŞ
M062115	TOTAL	96 (0,218)	124 (0,282)	168* (0,382)	40 (0,091)	12 (0,027)
	High	11 (0,082)	17 (0,127)	97 (0,724)	5 (0,037)	4 (0,030)
	Low	35 (0,269)	48 (0,369)	26 (0,200)	16 (0,123)	5 (0,038)
	Diff	-24 (-0,187)	-31 (-0,242)	71 (0,524)	-11 (-0,086)	-1 (-0,009)

Tablo 10’da verilen M052410 ve M062115 soru maddelerine ait veriler incelendiğinde çeldiricilere ait işaretlemelerde üst grup ile alt grup arasındaki farkın negatif, doğru cevabın işaretlenmesinde ise üst grup ile alt grup arasındaki farkın pozitif olduğu görülmektedir. Ancak M052170 soru maddesi incelendiğinde ise çeldiricilerin işaretlenmesinde üst grup ile alt grup arasındaki farkın pozitif olduğu dolayısıyla, çeldiricinin ters yönde çalıştığı söylenebilir.

Sonuç olarak Booklet 3 kapsamında incelenen çoktan seçmeli soru maddelerinden dokuz tanesinde, çeldiricilerin istendiği gibi çalıştığı ancak 6 soru maddesinde ise çeldiricilerin ters yönlü çalıştığı sonucuna ulaşılmıştır.

2.7. MTK’ya Göre İçerik Alanı Analizleri

Bu bölümde MTK’ya göre Booklet 3 veri seti içerik alanları bazında analiz edilmiş ve elde edilen sonuçlar yorumlanmıştır. Ayrıca ilgili soru maddelerine ait madde karakteristik eğrileri betimlenerek her tablodan sonra verilmiştir.

2.7.1. Sayılar İçerik Alanına Ait Sonuçlar

Tablo 11’de Sayılar içerik alanına ait konular kapsamındaki MTK’ya ait istatistiksel değerler betimlenmiştir.

Tablo 11: Sayılar İçerik Alanına Ait MTK Analiz Sonuçları

Soru	Eta	Thold (location)	p- value	O. MSQ	I. MSQ	Infit t	Disc.
M052209	-0.572	-0.571	0.996	0.828	0.838	-3.207	0.600
M052035	0.400	0.399	1.000	0.468	0.659	-5.257	0.753
M062212	0.926	0.925	0.005	1.178	1.035	0.414	0.435
M052142	-0.079	-0.079	0.182	1.058	1.135	2.096	0.388
M052006	-0.192	-0.191	0.000	1.324	1.237	3.683	0.322
M062329	-2.801	-2.800	1.000	0.637	0.862	-2.268	0.354
M052016	0.366	0.366	1.000	0.729	0.849	-2.162	0.614

Tablo 11 devamı

M062151	0.317	0.316	1.000	0.733	0.873	-1.834	0.587
M062346	-0.108	-0.107	0.975	0.870	0.915	-1.387	0.544

Eta: Madde güçlüğü; Thold: Eşik değer (bir sorunun yapılma olasılığının %50 olduğu değer; p-value: Anlamlılık değeri; MSQ: Logitlere ne kadar güvenebileceğimiz hakkında verilen bilgi; Outfit MSQ (O.MSQ): Ağırlaştırılmamış MSQ değeri; Infit MSQ (I.MSQ): Ağırlaştırılmış MSQ değeri; Infit t: Ağırlaştırılmış t değeri; Discrim (Disc.): Madde ile veri setinden elde edilen toplam puan arasındaki ilişki

Tablo 11 incelendiğinde, Sayılar içerik alanında yer alan M052209, M052035, M062212, M052016, M062151 ve M062346 soru maddeleri ile öğrencilerin Booklet 3 veri setinden elde ettikleri toplam puan arasında yüksek bir ilişkinin olduğunu (Disc.>0.40) ancak M052142, M052006 ve M062329 soru maddelerince ise bu ilişkinin zayıf olduğu görülmektedir. Soru maddelerine ait Eta değerleri incelendiğinde Sayı içerik alanına ait en zor maddenin M062212, en kolay maddenin ise M062329 olduğu görülmektedir.

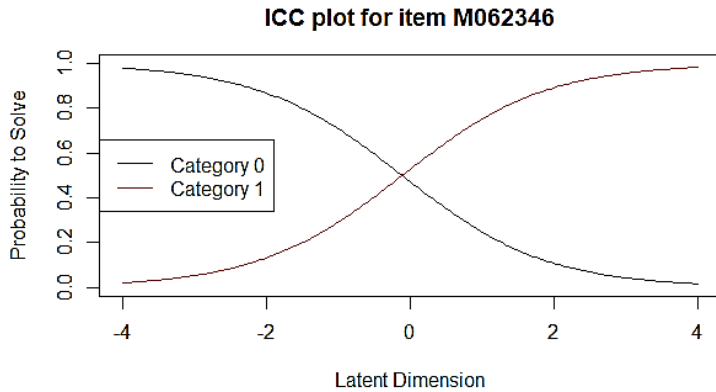
Tablo 11’de yer alan infit MSQ değerlerinin işlevi, kişiler ve öğelerle ilişkili ölçülere (logitlere) ne kadar güvenebileceğimiz hakkında bilgi sağlamaktır. Bu değerlerin +0,5 ile +1,5 arasında olması durumunda maddelerin veya kişilerin Rasch modeli için uygun olduğu, aksi takdirde veriler uyumsuz veri olarak sınıflandırılmakta ve bu da verilerin Rasch modeline uymadığını göstermektedir (Green, 2013, s. 167). Eğer bu değer +0,5’ten küçükse, bir kişinin veya bir öğenin çok öngörülebilir bir şekilde performans gösterdiği; +1,5’ten büyük ise kişi veya öğelerin öngörülemez bir şekilde performans gösterdiği söylenebilir. Ancak Rasch model genellikle +1,5’ten yüksek MSQ değerlerine odaklanmaktadır (Green, 2013, s. 169). Bu ifadeye karşın Wu & Adams (2007) çalışmalarında, infit MSQ değerinin 1’in altında olmasının kabul edilebilir olduğunu ancak üstünde bir değer olması durumunda bu maddenin genellikle modelin öngördüğünden daha az ayırt edici olduğunu ve maddenin modelle uyumunun zayıf olduğunu belirtmektedirler.

Bu kapsamda Green (2013)’in veri uyumu dikkate alındığında Sayılar içerik alanına ait M052035 soru maddesi dışındaki tüm maddelerin modelle uyumlu olduğu; Wu & Adams (2007)’in veri uyumu dikkate alındığında ise M062212, M052142 ve M052006 soru maddeleri dışındaki tüm maddelerin modelle uyumlu olduğu sonucuna varılmıştır (p>0,05).

Uygun verileri uyumsuz verilerden ayırt edebilen bir diğer değer ise Tablo 11’de infit t olarak verilmiştir. Bu değer -2,0 ile +2,0 arasında olması

kabul edilebilirken, bu değerler dışındaki değerler maddenin Rasch model ile daha az uyumlu olduğunu belirtmektedir (Bond & Fox, 2001). Ancak MSQ değeri kabul edilen sınırlar içerisinde ise infit t değerleri göz ardı edilebilmektedir (Linacre, 2002). Bu sebeple Green (2013)'in veri uyumu değerleri dikkate alındığında M052035 soru maddesine ait infit t değerlerinin kabul edilen sınır değerleri dışında olduğu ve böylelikle M052035 maddenin model ile uyumunun zayıf olduğu sonucuna varılmıştır. Buna karşın Wu & Adams (2007)'in veri uyumu dikkate alındığında ise M052142 ve M052006 soru maddelerine ait infit t değerlerinin kabul edilen sınır değerleri dışında olduğu ve maddelerin model ile uyumunun zayıf olduğu sonucuna varılmıştır. Diğer tüm maddelerin MSQ değerleri, kabul edilen sınırlar içerisinde olduğu için infit t değerleri göz ardı edilmiştir.

Thesholds (Thold) temel olarak madde zorluğudur ve İCC grafik çizimlerinde eğrilerin kesiştikleri noktalardır. Bu kesişimlere eşik değerler denmekte ve eşik değerleri bir kişinin %50 ihtimalle ilgili soru maddesini yapma olasılığını ifade etmektedir. Şekil 5'te Sayı içerik alanı kapsamında yer alan soru maddelerinden rastgele seçilmiş bir maddeye ait İCC grafiği ve yorumu verilmiştir.



Şekil 5: ICC grafiği (1)

ICC grafiklerinin yorumlanmasını M062346 soru maddesi üzerinden gerçekleştirebilirsek eğer, eşik değerinin (-0,107) solunda kalan değerler (-∞, -0,107, -∞) için bir öğrencinin bu soru maddesini yanlış yapma olasılığı doğru yapma olasılığından fazladır. Ancak eşik değeri geçildiği zaman (-0,107, +∞) bu sorunun doğru cevaplanma olasılığı yanlış cevaplanma olasılığından fazla

olacaktır. Buradaki eşik değeri ise bu sorunun doğru veya yanlış cevaplanma olasılığının eşit (%50-%50) olduğu değerdir.

2.7.2. Cebir İçerik Alanına Ait Sonuçlar

Tablo 12’de Cebir içerik alanına ait konular kapsamındaki MTK’ya ait istatistiksel değerler betimlenmiştir.

Tablo 12: Cebir İçerik Alanına Ait MTK Analiz Sonuçları

Soru	Eta	Thold (location)	p-value	O. MSQ	I. MSQ	Infit t	Disc.
M052064	-0.036	-0.036	0.000	1.384	1.188	2.821	0.348
M052126	1.721	1.721	1.000	0.343	0.762	-2.107	0.624
M052103	-0.994	-0.993	0.997	0.825	0.799	-4.557	0.594
M062056	2.076	2.076	1.000	0.712	0.767	-1.774	0.529
M052066	-1.179	-1.178	0.878	0.920	0.952	-1.057	0.481
M062078	0.383	0.382	1.000	0.525	0.712	-4.345	0.711
M062317	2.334	2.333	1.000	0.306	0.787	-1.432	0.536
M062350	0.802	0.802	0.000	2.010	1.511	5.217	0.071

Tablo 12 incelendiğinde, Cebir içerik alanında yer alan M052126, M052103, M062056, M052066, M062078 ve M062317 soru maddeleri ile öğrencilerin Booklet 3 veri setinden elde ettikleri toplam puan arasında yüksek bir ilişkinin olduğunu ($Disc.>0.40$) ancak M052064 ve M062350 soru maddelerince ise bu ilişkinin zayıf olduğu görülmektedir. Soru maddelerine ait Eta değerleri incelendiğinde Cebir içerik alanına ait en zor maddenin M062317, en kolay maddenin ise M052126 olduğu görülmektedir.

Tablo 12’de yer alan infit MSQ değerleri Green (2013)’in veri uyumuna göre incelendiğinde, Cebir içerik alanına ait M062350 soru maddesi dışındaki tüm maddelerin modelle uyumlu olduğu; Wu & Adams (2007)’in veri uyumuna göre incelendiğinde ise M052064 ve M062350 soru maddeleri dışındaki tüm maddelerin modelle uyumlu olduğu sonucuna varılmıştır ($p>0,05$).

Veri uyumu infit t değerleri Green (2013)’in veri uyumu değerleri dikkate alınarak incelendiğinde M062350 soru maddesinin; Wu & Adams (2007)’in veri uyumu değerleri dikkate alınarak incelendiğinde ise M052064 ve M062350 soru maddelerinin infit t değerlerinin kabul edilen sınır değerleri dışında olduğu ve böylelikle bu maddelerin model ile uyumunun zayıf olduğu sonucuna varılmıştır. Diğer tüm maddelerin MSQ değerleri, kabul edilen sınırlar içerisinde olduğu için infit t değerleri göz ardı edilmiştir.

2.7.3. Geometri İçerik Alanına Ait Sonuçlar

Tablo 13'te Geometri içerik alanına ait konular kapsamındaki MTK'ya ait istatistiksel değerler betimlenmiştir.

Tablo 13: Geometri İçerik Alanına Ait MTK Analiz Sonuçları

Soru	Eta	Thold (location)	p-value	O. MSQ	I. MSQ	Infit t	Disc.
M052041	1.539	1.538	1.000	0.438	0.724	-2.682	0.640
M052057	-1.133	-1.132	0.000	1.540	1.163	3.403	0.279
M052417	-0.445	-0.444	0.729	0.956	0.980	-0.337	0.497
M062284	-1.110	-1.109	0.113	1.080	1.093	1.976	0.369
M062245	-0.732	-0.732	0.648	0.971	0.985	-0.271	0.486
M062287	2.381	2.380	1.000	0.697	0.879	-0.743	0.434

Tablo 13 incelendiğinde, Geometri içerik alanında yer alan M052041, M052417, M062245 ve M062287 soru maddeleri ile öğrencilerin Booklet 3 veri setinden elde ettikleri toplam puan arasında yüksek bir ilişkinin olduğunu (Disc.>0.40) ancak M052057 ve M062284 soru maddelerince ise bu ilişkinin zayıf olduğu görülmektedir. Soru maddelerine ait Eta değerleri incelendiğinde Geometri içerik alanına ait en zor maddenin M062287, en kolay maddenin ise M052057 olduğu görülmektedir.

Tablo 13'te yer alan infit MSQ değerleri Green (2013)'in veri uyumuna göre incelendiğinde, Geometri içerik alanına ait tüm maddelerin modelle uyumlu olduğu; Wu & Adams (2007)'in veri uyumuna göre incelendiğinde ise M052057 soru maddesi dışındaki tüm maddelerin modelle uyumlu olduğu sonucuna varılmıştır ($p>0,05$).

Veri uyumu infit t değerleri Green (2013)'in veri uyumu değerleri dikkate alınarak incelendiğinde tüm maddelerin MSQ değerleri, kabul edilen sınırlar içerisinde olduğu için infit t değerleri göz ardı edilmiş; Wu & Adams (2007)'in veri uyumu değerleri dikkate alınarak incelendiğinde ise M052057 soru maddesinin infit t değerlerinin kabul edilen sınır değerleri dışında olduğu ve böylelikle bu maddelerin model ile uyumunun zayıf olduğu sonucuna varılmıştır.

2.7.4. Veri ve Olasılık İçerik Alanına Ait Sonuçlar

Tablo 14'te Veri ve Olasılık içerik alanına ait konular kapsamındaki MTK'ya ait istatistiksel değerler betimlenmiştir.

Tablo 14: Veri ve Olasılık İçerik Alanına Ait MTK Analiz Sonuçları

Soru	Eta	Thold (location)	p-value	O. MSQ	I. MSQ	Infit t	Disc.
M052501	0.236	0.236	0.952	0.888	1.057	0.830	0.453
M052410	-0.829	-0.828	0.002	1.207	1.153	2.979	0.333
M052170	-0.609	-0.609	0.000	1.301	1.178	3.212	0.341
M062345A	-1,064	-0.572*	0.494	0.997	0.942	-0.877	0.609
M062115	-0.534	-0.534	0.008	1.167	1.055	1.029	0.441

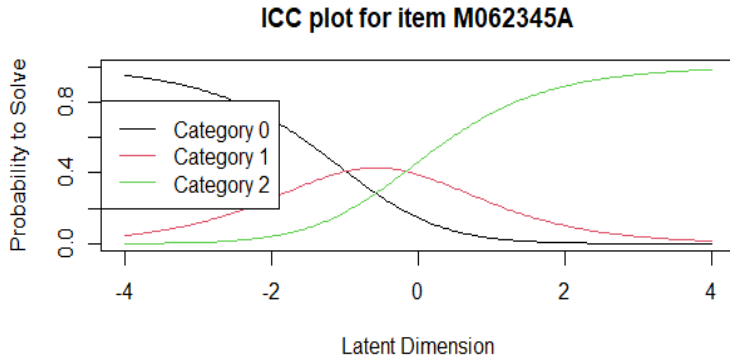
*: Threshold değeri iki tane (Th1:-0,982, Th2: -0,163) olup locatioan değeri yazılmıştır.

Tablo 14 incelendiğinde, Veri ve Olasılık içerik alanında yer alan M052501, M062345A ve M062115 soru maddeleri ile öğrencilerin Booklet 3 veri setinden elde ettikleri toplam puan arasında yüksek bir ilişkinin olduğunu (Disc.>0.40) ancak M052410 ve M052170 soru maddelerince ise bu ilişkinin zayıf olduğu görülmektedir. Soru maddelerine ait Eta değerleri incelendiğinde Veri ve Olasılık içerik alanına ait en zor maddenin M052501, en kolay maddenin ise M062345A olduğu görülmektedir.

Tablo 14’te yer alan infit MSQ değerleri Green (2013)’in veri uyumuna göre incelendiğinde, Veri ve Olasılık içerik alanına ait tüm maddelerin modelle uyumlu olduğu; Wu & Adams (2007)’in veri uyumuna göre incelendiğinde ise M052410, M052170 ve M062115 soru maddeleri dışındaki tüm maddelerin modelle uyumlu olduğu sonucuna varılmıştır ($p>0,05$).

Veri uyumu infit t değerleri Green (2013)’in veri uyumu değerleri dikkate alınarak incelendiğinde tüm maddelerin MSQ değerleri, kabul edilen sınırlar içerisinde olduğu için infit t değerleri göz ardı edilmiş; Wu & Adams (2007)’in veri uyumu değerleri dikkate alınarak incelendiğinde ise M052410 ve M052170 soru maddelerinin infit t değerlerinin kabul edilen sınır değerleri dışında olduğu ve böylelikle bu maddelerin model ile uyumunun zayıf olduğu sonucuna varılmıştır.

Şekil 6’da Veri ve Olasılık içerik alanı kapsamında yer alan M062345A maddesine ait İCC grafiği verilmiştir.



Şekil 6: ICC grafiği (2)

Genel olarak 0-1’li kodlanan Booklet 3 veri setinde yer alan M062345A maddesinin 0-1-2’li kodlanmasından dolayı bu soru maddesine ait ICC grafiğinin yorumlanmasının karşılaştırma yapılabilmesi açısından fayda sağlayacağı düşünülmektedir. M062345A soru maddesine ait iki farklı eşik değeri (Th1: -0,982, Th2: -0,163) olup, -0,982 eşik değerinin solunda kalan değerler (-0,982, $-\infty$) için bir öğrencinin bu soru maddesini yanlış yapma olasılığı doğru yapma olasılığından fazladır. Ancak 1. eşik değeri geçildiği zaman (-0,982 ile -0,163 arası) 2. eşik değerine kadar olan değerlerde (yaklaşık -0,572) öğrencilerin bu soruyu kısmi olarak doğru yapmaları en yüksek olasılık değerine sahiptir. Fakat 2. eşik değeri geçildiğinde (-0,163, $+\infty$) ise bu sorunun tam olarak cevaplanma olasılığı, kısmi olarak cevaplanma olasılığından fazla olacaktır.

SONUÇ

Bu çalışmada Klasik Test Kuramı ve Madde Tepki Kuramı kullanılarak TIMSS 2015 Türkiye sekizinci sınıf öğrencilerine yönelik uygulanan kitapçıklardan biri olan Booklet 3, hem içeriksel hem de betimsel istatistikler kullanılarak analiz edilmiştir. Yapılan analizler sonucunda ilgili kitapçığa yönelik 8 soru maddesinin model ile uyumlu olmadığı tespit edilmiştir. Tablo 15’te bu sorulara ilişkin istatistiksel değerler betimlenmiştir.

Tablo 15: Uyumsuz Verilere Ait MTK Analiz Sonuçları

Sorular	Adj Pt Bis	Thold	p-value	I. MSQ	Infit t	Disc.	Çeldiriciler
M052035 *	0,67	0.399	1.000	0.659	-5.257	0.753	İstendiği gibi
M052142 **	0,36	-0.079	0.182	1.135	2.096	0.388	Ters yönlü
M052006 **	0,25	-0.191	0.000	1.237	3.683	0.322	Ters yönlü
M052064 **	0,30	-0.036	0.000	1.188	2.821	0.348	Ters yönlü
M052057 **	0,26	-1.132	0.000	1.163	3.403	0.279	İstendiği gibi
M052410 **	0,31	-0.828	0.002	1.153	2.979	0.333	İstendiği gibi
M052170 **	0,31	-0.609	0.000	1.178	3.212	0.341	Ters yönlü
M062350 ***	0,06	0.802	0.000	1.511	5.217	0.071	Ters yönlü

*: Green (2013)'e göre uyumsuz maddeler; **: Wu & Adams (2007)'e göre uyumsuz maddeler; ***: Ortak uyumsuz maddeler

Sırasıyla Tablo 15'te verilen soru maddelerinin uyumları incelendiğinde; M052035 soru maddesinin ayırt edici, orta güçlükte ve çeldiricilerinin istendiği gibi çalıştığı bir madde olduğu görülmüştür. Uyum indekslerine bakıldığında p değerinin 0,05'ten büyük, MSQ değerinin 1'in altında olduğu ve toplam test puanı ile arasında yüksek bir ilişkinin olduğunu ancak t değerinin yüksek çıktığı görülmektedir. Eğer MSQ değeri istenen değer kapsamında ise t değeri göz ardı edileceğinden dolayı bu maddenin uyumlu olduğu sonucuna ulaşılmıştır.

M052142, M052006, M052064, M052170 ve M062350 soru maddeleri incelendiğinde ise, ayırt ediciliklerinin ve güçlüklerinin düşük (M062350 zor bir sorudur) ve çeldiricilerinin ters yönlü olduğu görülmüştür. Uyum indekslerine bakıldığında p değerinin 0,05'ten küçük (M052142 hariç), MSQ değerinin 1'in üstünde olduğu, toplam test puanı ile aralarında düşük bir ilişkinin olduğu ve t değerinin yüksek çıktığı görülmektedir. Bu sebeple bu maddelerin model ile uyumsuz olduğu sonucuna ulaşılmıştır.

M052057 ve M052410 soru maddeleri incelendiğinde ise, ayırt ediciliklerinin ve güçlüklerinin düşük ancak çeldiricilerinin istendiği gibi çalıştığı görülmüştür. Uyum indekslerine bakıldığında p değerinin 0,05'ten

küçük, MSQ değerinin 1'in üstünde olduğu, toplam test puanı ile aralarında düşük bir ilişkinin olduğu ve t değerinin yüksek çıktığı görülmektedir. Bu sebeple bu maddelerin model ile uyumsuz olduğu sonucuna ulaşılmıştır.

Sonuç olarak Booklet 3 kitapçığında yer alan 7 soru maddesinin model ile uyumsuz olduğu ve bu sebeple bu soru maddelerinin gözden geçirilmesi gerektiği düşünülmektedir. Ayrıca genel Türkiye TIMSS 2015 sonuçları düşünüldüğünde, Booklet 3 kitapçığının bu sonucu doğrular nitelikte olduğu söylenebilir.

KAYNAKÇA

- Andersen, E. B. (1972). The Numerical Solution of a Set of Conditional Estimation Equations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1), 42–54. <https://doi.org/10.1111/j.2517-6161.1972.tb00887.x>
- Andersen, E. B. (1973). A goodness-of-fit test for the Rasch model. *Psychometrika*, 38(1), 123–140.
- Atılğan, H., Kan, A., & Doğan, N. (2016). *Eğitimde ölçme ve değerlendirme* (9.baskı). Ankara: Anı Yayıncılık.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32(32), 470-483.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Javonich College.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY, Guilford Publications.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory For Psychologists*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Green, R., 2013. *Statistical analysis for language testers*. New York: Palgrave Macmillan.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park CA: Sage Publications.
- Hambleton, R. K., & Swaminathan H. (1985). *Item response theory: principals and applications*. Norwell, MA: Kluwer
- Hambleton, R. K., & Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hasançebi, B., Terzi, Y., & Küçük, Z. (2020). Madde Güçlük İndeksi ve Madde Ayırt Edicilik İndeksine Dayalı Çeldirici Analizi. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 10(1), 224-240.

- Karaman, H. (2015). *Açımlayıcı Faktör Analizinde Kullanılan Faktör Çıkartma Yöntemlerinin Karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Yüksek Öğretim Kurulu tez merkezi veri tabanından erişildi (381434).
- Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge
- Lee, Y.S., Park, Y. S., & Taylan, D. (2011). A Cognitive Diagnostic Modeling of Attribute Mastery in Massachusetts, Minnesota, and the U.S. National Sample Using the TIMSS 2007. *International Journal of Testing, 11(2)*, 144–177. doi:10.1080/15305058.2010.5345
- Linacre, J. M. (2002). What do Infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16(2)*, 878.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley
- Verhelst, N. D. (2004). Section C: Classical Test Theory. Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Section C). Strasbourg: Council of Europe (download from <http://rm.coe.int/0900001680667a1e>)
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions, Melbourne.
- Yaşlıoğlu, M. M. (2017). Sosyal Bilimlerde Faktör Analizi ve Geçerlilik: Keşfedici ve Doğrulayıcı Faktör Analizlerinin Kullanılması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi, 46*, 74-85.