

## Bireyselleřtirilmiř Testler Üzerine Bir Çalıřma<sup>1</sup>

Metin YAŐAR<sup>2</sup>

### ÖZET

Bu çalıřmada, Türkçe okuduđunu anlamaya yönelik olarak, Klasik Test Kuramı dayalı olarak kâđıt-kalem testi ve Madde Tepki Kuramına (iki-parametrelili lojistik model) dayalı olarak geliřtirilen bireyselleřtirilmiř teste ait KR-20 güvenilirlik düzeylerinin ve bir grup öđrencinin bu iki yöntemle elde edilen yetenek ölçüleri arasında iliřki düzeyleri karřılařtırılması amaçlanmıřtır. İki-parametrelili lojistik modele göre geliřtirilen bireyselleřtirilmiř test, on üç adımda cevaplandırılabilir çoktan seçmeli 61 maddelik bir havuz oluřturularak uygulama gerçekteřtirilmiřtir. Buna karřılık Klasik Test Kuramına göre 47 çoktan seçmeli maddelik kâđıt-kalem testi örnekleme yer alan öđrencilere uygulanmıřtır. İki yöntemle geliřtirilen test aynı gruba uygulandıktan sonra, her iki teste ait KR-20 güvenilirlik katsayısı bireyselleřtirilmiř test için 0,67 ve Klasik Test Kuramına göre hazırlanan kâđıt-kalem testinin güvenilirlik katsayısı ise 0,75 olarak hesaplanmıřtır. Arařtırmada elde edilen her iki yöntemle iliřkin KR-20 güvenilirlik katsayısı Fisher'in Z istatistiđine dönüřtürülerek 0,05 düzeyinde test edilmiřtir. Her iki yöntemle iliřkin hesaplanan KR-20 güvenilirlik katsayıları arasında 0,05 manidarlık düzeyinde anlamlı bir farkın olmadığı tespit edilmiřtir. Bireyselleřtirilmiř test puanları ile kâđıt-kalem testinden elde edilen ölçme sonuçları arasında Pearson momentler çarpımı korelasyon katsayısı 0,36 olarak hesaplanmıřtır. Her iki yöntemle geliřtirilen testten elde edilen ölçme sonuçları arasında pozitif yönde ancak çok düşük düzeyde bir iliřkinin olduđu görülmektedir. Sonuç olarak, her iki yöntemle dayalı olarak geliřtirilen testlerin KR-20 güvenilirlik katsayıları arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir farkın olmadığı gibi her iki teste iliřkin öđrencilerin yetenek ölçüleri arasında düşük düzeyde bir iliřkinin olduđu fakat her iki testten elde edilen yetenek ölçüleri arasında 0,05 düzeyinde anlamlı bir iliřki olmadığı görülmüřtür.

*Anahtar Kelimeler:* Klasik test kuramı, Bireyselleřtirilmiř testler, İki parametrelili lojistik model, Güvenirlik

<sup>1</sup> Bu çalıřma Prof. Dr. Yařar BAYKUL danıřmanlıđında yapılan Doktora tezinden üretilmiřtir.

<sup>2</sup> Yrd. Doç. Dr. - Pamukkale Üniversitesi, Eđitim Fakültesi - myasar@pau.edu.tr

## GİRİŞ

Bloom (1956) taksonomisi, bireylerin davranışlarının benzerliklerini/farklılıklarını dikkate alarak bilişsel, duyuşsal ve devinişsel olmak üzere üç alanda sınıflandırmaktadır. Genellikle eğitim kurumlarında bu üç alana ilişkin özellikler kazandırılmaya çalışılmaktadır. Özellikle bilişsel ve duyuşsal alana ilişkin özelliklerimiz Psikolojik yapı olarak adlandırılabilir. Genellikle okul ortamında bireylere kazandırmaya çalıştığımız psikolojik yapıların gerçekten gerçekleşip gerçekleşmediğinin ya da bireylerin söz konusu psikolojik özelliklere sahip olma düzeylerinin belirlenmesi için ölçme işleminin kaçınılmazlığı ortaya çıkmaktadır.

Söz konusu psikolojik yapıların doğrudan ölçülebilir olmamasından dolayı söz konusu özellikleri ölçebilmek için uygun özelliklere sahip ölçme araçlarının geliştirilmesi zorunluluğu kendiliğinden ortaya çıkmaktadır. Bu doğrultuda ölçme araçlarının geliştirilmesinde dikkate alınan iki temel test kuramının kullanıldığı görülmektedir. Bunlardan birincisi, Klasik Test Kuramı (KTK) ve ikincisi ise Örtük Özellikler Teorisi (ÖÖT) veya Madde Tepki (cevap) Kuramı (MTK) olarak bilinmektedir.

Psikolojik ölçme tarihinin başlangıcından bu güne kadar test geliştirmede, test sonuçlarının analizinde ve psikolojik ölçüklerin puanlanmasında en yaygın olarak kullanılan kuram KTK olurken, KTK'nın bazı sınırlılıklarından dolayı ona alternatif olarak 20. Yüzyılın ortalarında Örtük Özellikler Teorisi (ÖÖT) veya Madde Tepki Kuramı (MTK) olarak bilinen kuramın ortaya çıkışı ve giderek daha yaygın olarak kullanılmaya başlandığı görülmektedir (Baykul, 1979; Crocker, & Algina, 1986, Gelbal, 1994; Hambleton, 1994; Kaptan, 1994; Erkuş, 2003; Reise, Ainsworth, & Haviland, 2005; Kan, 2006; Çelen, & Aybek, 2013).

Bireylerin başarılarının ölçülmesinde başarı puanı KTK'ya göre bireylerin test içinde yer alan maddelerden almış oldukları puanların toplanmasıyla elde edilmektedir. Test ve madde istatistikleri bireylerin test puanlarına veya madde puanına bağlı olarak hesaplanmaktadır. Buradan hareketle bireylerin başarı düzeylerini gösteren puanlar, test içinde yer alan maddelerin güçlük düzeylerine göre değişiklik göstermektedir (Lord, & Novick, 1968). KTK'na göre geliştirilip uygulanan test içindeki maddelerin güçlük düzeyleri dikkate alınarak madde puanlarının ağırlıklandırılmaları yapılmıyorsa, farklı güçlük düzeyine sahip olan maddeler sanki aynı güçlük düzeyine sahipmiş gibi düşünüldüğü ve de her bir maddenin KTK'na göre elde edilen toplam puana katkısının aynı düzeyde olduğu kabul edilmektedir.

Bireylerin sahip olduğunu kabul ettiğimiz bazı psikolojik yapıların ölçülmesi ve bu psikolojik yapıların düzeyinin belirlenmesi için kullanılan ölçme araçlarından elde edilen ölçme sonuçlarının puanlanmasında yaygın olarak kullanılan olarak kullanılan kuram KTK olarak bilinmektedir. Bununla birlikte ikinci bir kuram ise Madde tepki Kuramı (MTK) olarak bilinmektedir. MTK, teknolojiye inanamaz dramatik gelişmelere paralel olarak bilgisayar teknolojisindeki gelişmelere bağlı kullanımı her geçen gün artmakla beraber KTK'na göre daha popüler olmaya ve de daha yaygın olarak kullanılmaya başlanan bir kuram olarak görülmektedir (Çıkrıkçı, 1999; Reise, Ainsworth, & Haviland, 2005; Demirtaşlı, & Arıkan, 2009; Bulut, & Kan, 2012).

Test geliştirme ile ilgili olarak 1970'li yılların başlarından itibaren belirli bir farklı yaklaşımın ortaya çıktığı görülmektedir. Bu farklı yaklaşımdaki düşünce, ölçme işlemine tabi tutulan bireylerin, yetenek ve yeterlilik düzeylerinin ölçülmesinde daha az sayıda test maddesinin kullanılmasıdır. Bu uygulamayla bireylerin yetenek ve yeterlilik düzeyleri daha kesin ve daha güvenilir bir şekilde kestirilmeye çalışılmıştır.

Klasik Test Kuramı ile Madde Tepki Kuramının karşılaştırılması istendiğinde şunların görülmesi mümkündür; KTK'na göre geliştirilen testlerde yer alan maddelerin, madde güçlük ve madde ayırıcılık indeksleri testin uygulandığı grubun yapısına bağlı olduğundan grup değiştikçe madde güçlük ve madde ayırıcılık düzeyi de değişmektedir. Buna karşılık MTK gruptan bağımsız olduğundan dolayı madde parameterlerinin değişmezliği söz konusudur. Bireyler değişse bile madde parametrelerinin değişmezliği KTK'na göre bir üstünlük olarak görülebilir.

MTK'na göre madde parametrelerinin değişmezliğinin sağlanabilmesi için de Madde parametrelerinin elde edilmesinde yapılan ön deneme uygulamasının ve uygulamanın yapıldığı grubun bazı özellikleri taşınmasına (Hambleton, & Swaminathan, 1985; Hambleton, 1990; Hambleton vd., 1991; Kelecioğlu, 2001; Çelen, 2008) bağlı olduğu söylenebilir.

Klasik Test Kuramında ölçme hataları tüm grup için hesaplanırken ve elde edilen ölçmenin standart hatası bütün cevaplayıcılar için aynı düzeydeymiş gibi kabul edilirken, Madde Tepki Kuramında her bir birey için ayrı ayrı hesaplanması söz konusu olmaktadır. Dolayısıyla da bireylere ait hata düzeyi sabit değildir. Bireylerin kestirilen yetenek düzeylerine göre farklılaşmaktadır. Bu özellik de KTK'na göre bir başka üstün özellik olarak alınabilir.

Madde Tepki Kuramı, madde düzeyindeki bilgileri, Klasik Test Kuramı'na göre daha fazla temel almaktadır. Buna göre, her bir madde için kestirilen madde parametreleri ile hangi maddelerin hangi yetenek düzeylerinde daha ayırıcı olduğu, ölçme aracının hangi yetenek düzeyindeki bireyler için daha tutarlı ölçme sonuçları verdiği, farklı güçlük düzeyindeki maddelerin doğru yanıtlandırılabilmesi için gerekli olan yetenek düzeyleri Madde Tepki Kuramı'nda açık bir biçimde kestirilebilmektedir (Nartgün, 2002 Akt: Çelen, 2008).

Klasik Test Kuramı'na göre test geliştirilirken deneme uygulaması sonrasında ayırıcılık gücü yüksek olan ve madde varyansını en yüksek yapabilmek için orta güçlükteki maddeler nihai teste alınır. Madde Tepki Kuramı'nda ise madde parametrelerinin yanı sıra asıl ölçüt modele uyumluluktur. Bunun yanı sıra her bir maddenin teste olan katkısını her yetenek düzeyi için gösteren madde bilgi fonksiyonu madde seçiminde kriter olarak alınabilmektedir (Linden, & Hambleton, 1997 Akt: Çelen, 2008).

Yukarıda bahsedilen iki test kuramından birinin diğerine üstün olup olmadığı hakkında yapılan birçok çalışma vardır. Türkiye'de yapılan çalışmalar incelendiğinde bu çalışmaların genellikle geniş ölçekli bir sınav verisi kullanılarak o sınavın MTK modellerine uyumunun incelendiği veya bir veri seti üzerinden iki kurama göre elde edilen parametrelerin karşılaştırıldığı çalışmalar (Kılıç, 1999; Yalçın, 1999; Çalışkan, 2000; Çelik, 2001; Karataş, 2001; Özkurt, 2002; Can, 2003; Yapar, 2003; Yeğin, 2003; Çelen, 2008 Akt: Çelen, & Aybek, 2013) olduğu görülmektedir.

Madde Tepki Kuramının zayıf yanları arasında büyük örneklem gruplara ihtiyaç duyması, kuramın karmaşık olması ve sonuçların yorumlanmasının zor olması, özel yazılımlar gerektirmesi, sağlanmalarının klasik test kuramına göre daha zor olduğu bir takım varsayımlar gerektirmesi sayılabilir (Hambleton, & Swaminathan, 1984; Hambleton, Swaminathan, & Rogers, 1991).

İki kuramın ortak yönleri vardır: Her iki kuramda da test maddelerine verilen tepkilerin doğruluğu veya yanlışlığı üzerine odaklanılmaktadır. Normallik varsayımı her iki kuram için de söz konusudur. DeMars (2010), normallik varsayımının MTK modelleri için gerekli olmadığını belirtse de, normal dağılım sağlanamamasının yetenek kestiriminde sorunlara yol açabileceğini de ifade etmiştir. Bunun yanında, Linden ve Hambleton (1997),

modern MTK'nın normallik varsayımını gerektiren faktör analizi tekniğinden çokça etkilendiğini belirtmiştir. Bununla birlikte test puanlarının normalliği sağlandığında, iki kurama göre elde edilen ayırt edicilik güçleri ve madde güçlük indeksleri arasında geçiş sağlamak mümkün olabilmektedir (Lord, & Novick, 1968; Crocker, & Algina, 1986 Akt: Çelen, & Aybek, 2013). Belirli bir özelliği ölçmek için geliştirilmiş testin sadece o özelliği ölçmesi ve testteki bir maddeye verilen yanıtın diğerini etkilememesi gerekliliği de her iki kuram için karşılanması gereken varsayımlardır (Demirtaşlı, 1998; Linn, 1998 Akt: Çelen, & Aybek, 2013).

### Klasik Test Kuramında Güvenirlik

Herhangi bir ölçme işlemi gerçekleştirildiğinde, ölçme sonuçlarına, kesin olarak kaynağı belli olmayan hatalar karışmaktadır. Bu tür hatalara tesadüfi hata denmektedir. Hata kavramı ile ölçme aracının veya ondan elde edilen ölçme sonuçlarının güvenirlilik düzeyi arasında bir ilişki söz konusudur. Ölçme sonuçlarına karıştığı varsayılan tesadüfi hata düzeyi azaldıkça testin ya da ondan elde edilen ölçme sonuçlarının güvenirliliği yükselmektedir. Turgut (1986) güvenirliliği, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak ifade ederken, Baykul (1979) ölçme aracının tesadüfi hatalara kapalı olma derecesi ve bu derecelerin görünümü olarak da ölçme aracının duyarlılığı, tutarlılığı ve kararlılığı olarak ifade etmektedir. Ölçme sonuçlarının güvenirlilik düzeylerinin yüksek olması, elde edildikleri ölçme araçlarının hatalara kapalı olma düzeyleriyle ilgili olduğu söylenebilir.

Klasik Test Kuramında, güvenirlilik katsayısı, *gerçek puan varyansının, gözlenen puan varyansına oranı* olarak tanımlanmakta olup aşağıdaki eşitlik ile gösterilmektedir (Lord, 1968).

$$\rho_X = \rho_{XT}^2 = \frac{\rho_T^2}{\rho_X^2} \quad (1.1)$$

Eşitlik (1.1) de verilen ve araştırma evreninin parametresi olan  $\rho_X$  güvenirlilik katsayısı, pratik uygulamalarda mevcut özelliğinden dolayı normal koşullarda güvenirlilik hesaplamalarında kullanılamamaktadır. Bunun sebebi, eşitlik (1.1) de yer alan yalnız gözlenen puanlar varyansının ( $\rho_X^2$ ) hesaplanabilir olmasına karşılık, gerçek puan varyansının ( $\rho_T^2$ ) ve hata puanı varyansının ( $\rho_E^2$ ) hesaplanamamasıdır. Söz konusu gerçek puan varyansı ( $\rho_T^2$ ) ve hata puanı varyansının ( $\rho_E^2$ ) hesaplanmadığı için, ölçme sonuçlarının ve dolayısıyla da ölçme araçlarının güvenirlilik katsayılarının hesaplanması için farklı güvenirlilik yöntemlerinin geliştirilip kullanılması kaçınılmaz olmuştur. Klasik Test Kuramına dayalı olarak geliştirilen güvenirlilik hesaplama yöntemlerine bakıldığında; aynı özelliği ölçmek amacıyla farklı ifadeler kullanılarak oluşturulan eş değer iki test formunu esas alan “*eşdeğer formlar yöntemi*”, bir testin farklı zamanlarda aynı guruba iki defa uygulanması esasına dayalı olarak geliştirilen “*test tekrar-test yöntemi*”, bir guruba uygulanan ve iki eşit yarıya bölünebilirlik özelliğine sahip olan “*iki yarı yöntemi*” ve bir testin uygulanmasına bağlı olarak elde edilen madde puanları arasındaki tutarlılığı ve ilişkileri inceleyen Kuder-Richardson'un KR-20 ve KR-21 numaralı eşitlikleri olarak bilinen güvenirlilik hesaplama yöntemleri geliştirilmiştir.

### Madde Tepki Kuramında Güvenirlilik

Özellikle MTK'nın getirilerinden biri olan madde karakteristik eğrisi doğrultusunda kullanılan soruların niteliğine ilişkin elde edilen detaylı bilgi ile bireyin yetenek düzeyinin belirlenmesinde maddeler daha verimli kullanılabilmektedir. Mevcut ölçme ve değerlendirme yaklaşımları göz önüne alındığında gruba bağlı madde parametreleri ve maddelere bağlı yetenek parametreleri kestirimi sınırlılığını ortadan kaldıran değişmezlik özelliği ile de

bireylerin yeteneği hakkında daha geçerli ve güvenilir bir kestirim imkânı sağlamaktadır (Crocker, & Algina, 1986; Hambleton, Swaminathan, & Rogers 1991; Kezer, & Koç, 2014).

Madde Tepki Kuramında, Klasik Test Kuramındaki güvenilirlik kavramının yerini madde ve test bilgi fonksiyonları kavramı almaktadır.

Madde bilgi fonksiyonu;

$$I(\theta, U_i) = \frac{[P'_i(\theta)]^2}{P_i(\theta) \cdot Q_i(\theta)} \quad (1.2)$$

ve test bilgi fonksiyonu da;

$$I(\theta) = \sum_{i=1}^K \frac{[P'_i(\theta)]^2}{P_i(\theta) \cdot Q_i(\theta)} \quad (1.3)$$

eşitliği ile tanımlanmaktadır. Buradaki (1.2) ve (1.3) nolu eşitlikteki  $P_i(\theta)$  seçilen MTK olasılık fonksiyonunu,  $P'_i(\theta)$  ise bunun türevini göstermekte olup,  $Q_i(\theta)$  ise  $Q_i(\theta) = 1 - P_i(\theta)$  dır.

(1.2) eşitliği ile tanımlanan madde bilgi fonksiyonunun ifadesinde,  $P_i(\theta)$  olasılık fonksiyonu, bunun varyansı ve türevinde yer almaktadır. Türev değerleri büyüdükçe ve varyans değerleri küçüldükçe kesir değeri büyüyeceğinden test maddelerinden elde edilecek bilgi de artacaktır. Ayrıca, eşitlik 1.3'e göre test bilgi fonksiyonunun madde bilgi fonksiyonlarının toplamına eşit olduğu görülmektedir. Bu özellik, madde bilgi fonksiyonunun teste olan katkısının madde dışındaki faktörlere bağlı olmadan katkı sağladığını göstermektedir. Aslında bu özellik, MTK'nın KTK'na karşı olan bir üstünlüğünü göstermektedir.

MTK'da ölçmenin Standart Hatası, bilgi fonksiyonuna bağlı olarak tanımlanmakta olup madde ve test için aşağıdaki eşitliklerle gösterilmektedir.

$$SE_i(\theta) = \frac{1}{\sqrt{I(\theta, U_i)}} \quad (1.4)$$

$$SE_T(\theta) = \frac{1}{\sqrt{\sum I(\theta, U_i)}} \quad (1.5)$$

(1.4) ve (1.5) eşitliklerinde görüldüğü gibi, ölçmenin standart hatası bilgi fonksiyonlarıyla ters orantılıdır. Madde veya testin verdiği bilgi arttıkça standart hata azalmaktadır. Bunun aksi söz konusu olduğunda, yani, madde ve testin verdiği bilgi azaldıkça standart hata da artmaktadır. Test bilgi fonksiyonunun hesaplanması buna bağlı olarak da testin standart

hatasının hesaplanabilir olması bireyin yeterliliğinin ya da yeteneğinin kestirilmesinde, yeteneğin kestirilme işinin, belli bir hata sınırına ininceye kadar devam etmesine imkân sağlamaktadır. Bu özellik, yetenek kestiriminde hatanın sabit tutulması bakımından önemlidir.

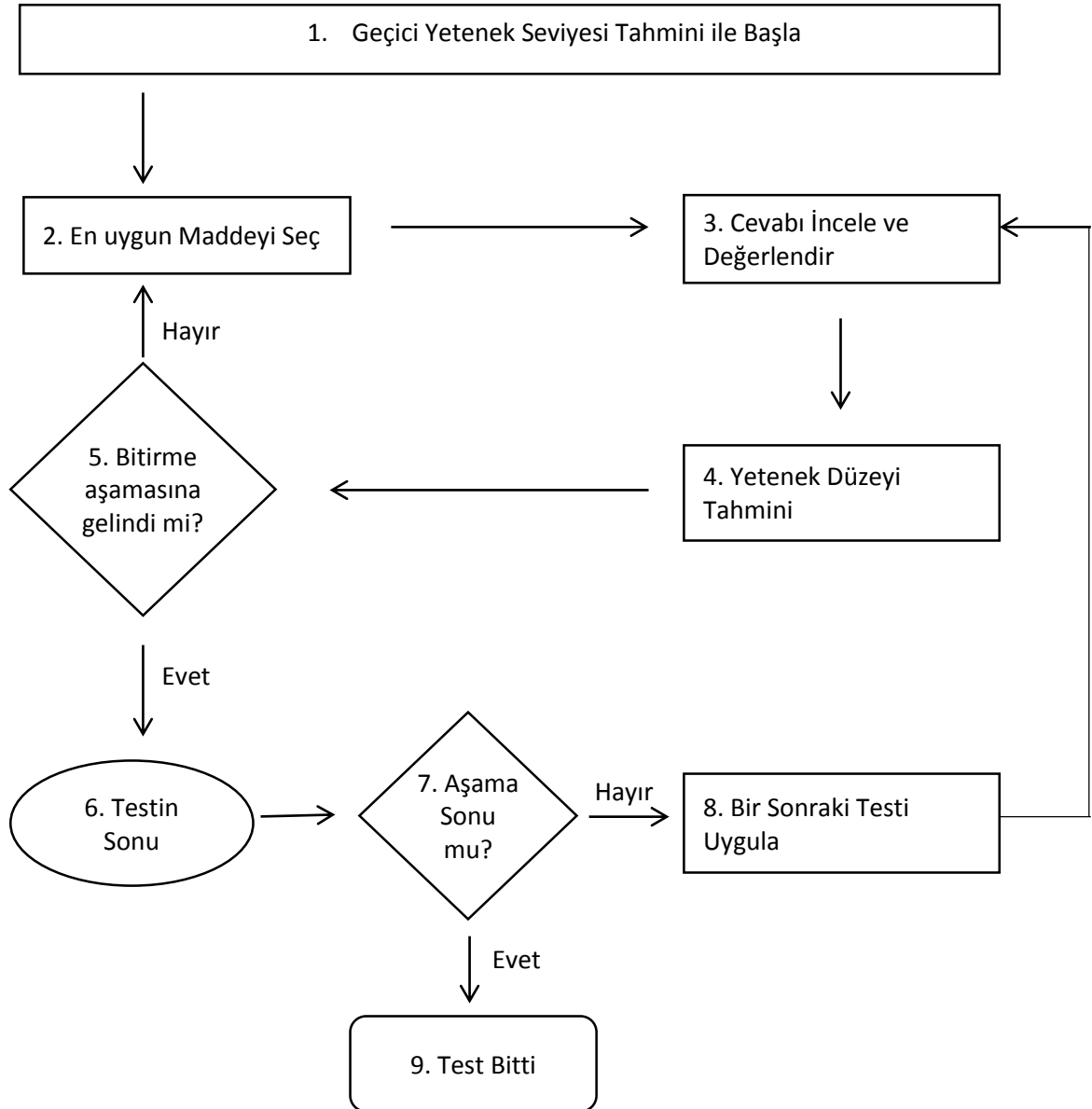
### Bireyselleştirilmiş Testler

Bu noktaya kadar, farklı özelliklere sahip olan KTK'ndan ve MTK ile ilgili olarak kısaca açıklamalarda bulunuldu. Bu iki kuramın dışında, bir yönden test geliştirme, diğer yandan bir test uygulaması olarak algılanabilecek olan bir başka yöntemden de söz edilmektedir. Bu yöntem de bireyselleştirilmiş (veya bilgisayar ortamında test) olarak bilinmektedir. Bireyselleştirilmiş testler, psikometrik özellikleri daha önceden belirlenmiş çok sayıda madde arasından cevaplayıcıların yetenek düzeylerine/özelliklerine uygun maddelerin seçilerek uygulanması esasına dayanan bir test uygulama yöntemidir. Bireyselleştirilmiş test uygulamasına bakıldığında kâğıt-kalem testlerinin aksine her birey için onların yeteneklerine göre ayarlanmış bireylere özgü test olarak bilinmektedir. Bireyselleştirilmiş test uygulamalarında farklı yetenek düzeylerindeki bireyler farklı maddeleri cevaplandırdıklarından dolayı test bireye özgü olmaktadır. Bu özelliği bakımından bireyselleştirilmiş testler kâğıt-kalem testlerinden farklılaşmaktadır. Kâğıt kalem testlerinde bütün bireylerin ölçülen özelliğe yönelik yetenek düzeyleri ne olursa olsun, aynı maddeleri cevaplandırmak zorunda kalmaktadırlar.

Bireyselleştirilmiş testlerde, madde havuzundaki, cevaplayıcılar tarafından verilen cevapların doğru veya yanlış olması durumuna göre madde havuzundan değişik madde güçlük ( $b_g$ ) düzeyindeki maddeler verilerek ölçme işlemi devam ettirilir. Bu mantığa göre belirlenen maddelerin, bireylerin ölçülmeye çalışılan yetenek düzeyini belirlemede etkili daha gerçekçi yetenek ölçü değerlerini üretilmesine imkân verdiği söylenebilir. Cevaplayıcıların hangi maddeyi cevaplandıracağı daha önceki maddeye vermiş olduğu cevabın doğru veya yanlış olmasına bağlı olarak ortaya çıkmaktadır. Eğer cevaplayıcı maddeyi doğru olarak cevaplandırmışsa, bir sonraki madde, doğru olarak cevaplandığı maddeden biraz daha zor madde olarak belirlenir veya madde yanlış cevaplandırılmışsa bir sonraki madde yanlış cevaplandırılan maddeye göre biraz daha kolay maddelerden seçilerek cevaplandırma işlemi devam ettirilmektedir.

Geleneksel testlerle ilgili belirtilen sınırlılıklarından dolayı bireye biçimlendirilmiş (tailored tests) veya bireye uyarlanmış (adaptive testing) testlere ihtiyaç duyulmuştur (Lord, 1970; Wood, 1973; Weiss, 1983; Hambleton, & Swaminathan 1985). Bireyselleştirilmiş testlerde, bireyin ölçülmek istenen yeteneğine uygun olarak test maddelerinin güçlüğünün ayarlanması esastır. Test maddelerinin güçlüğünün bireyin yeteneğine göre ayarlanıyor olması ile bireyin yeteneğinin çok üstünde ya da çok altında soru almasının da önüne geçilebilmektedir (Eroğlu, 2013).

Bireylerin bireyselleştirilmiş testleri cevaplandırmasında, belirli bir test algoritması takip edilir. Test algoritması (I) teste başlama, (II) testi cevaplandırmaya devam etme, (III) testi cevaplandırmayı sonlandırma. Söz konusu test algoritması Şekil 1'de verilmiştir.



Şekil 1. Bireyselleştirilmiş Test Alma Algoritması

Bu çalışmada, Türkçe okuduğunu anlama alanında Klasik Test Kuramına ve Madde Tepki Kuramına göre iki-parametrelili lojistik model kullanılarak geliştirilen bireyselleştirilmiş testin KR-20 güvenirlik düzeyleri ile bu iki yöntemle göre elde edilen yetenek ölçüleri arasında manidar düzeyde bir farklılık var mıdır? sorusuna yanıt aranmıştır. Yine bu problem cümlesine bağlı olarak cevap aranan alt problemler de şöyle ifade edilmiştir:

1. Türkçe okuduğunu anlama alanında Klasik Test Kuramına ve Madde Tepki Kuramına göre iki-parametrelili lojistik model kullanılarak geliştirilen bireyselleştirilmiş testin KR-20 güvenirlik düzeyleri arasında manidar düzeyde bir farklılık var mıdır?
2. Türkçe okuduğunu anlama alanında Klasik Test Kuramına ve Madde Tepki Kuramına göre iki-parametrelili lojistik model kullanılarak geliştirilen bireyselleştirilmiş teste göre elde edilen yetenek ölçüleri arasındaki ilişki hangi düzeydedir?

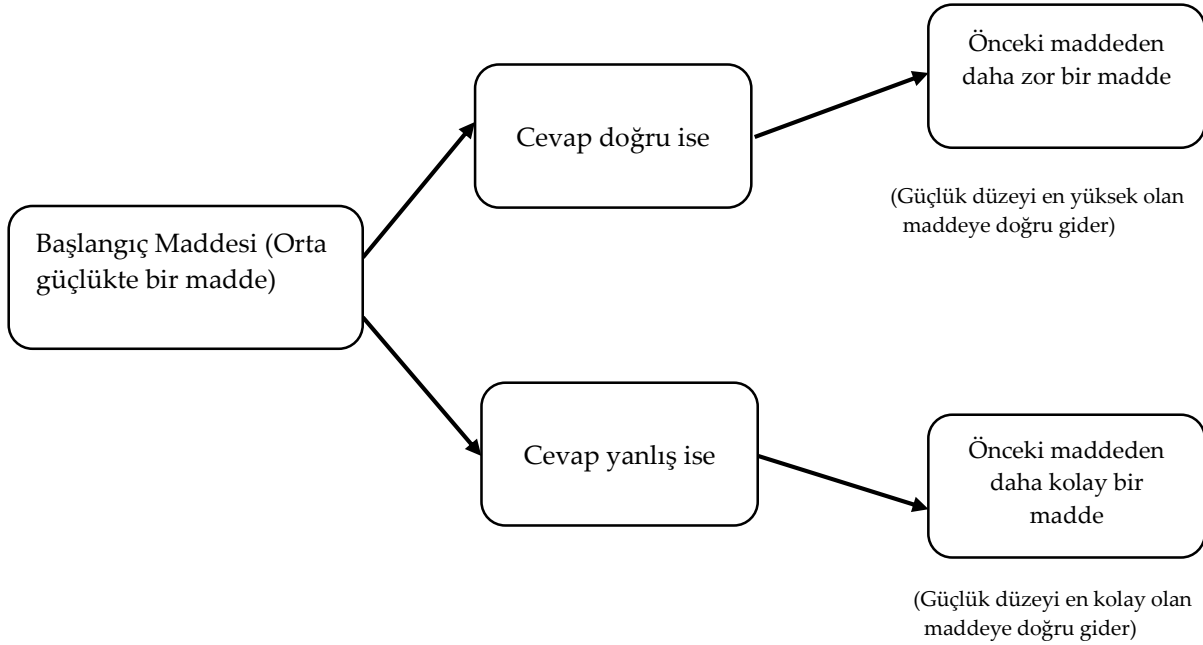
## YÖNTEM

Eldeki çalışma temel araştırma modelinde bir araştırmadır. Temel araştırmalar, alanda var olan teorik bilgilere yeni bilgi eklemeyi amaçlayan araştırma modelleri olarak düşünülebilir. Bu tür araştırmalar, araştırma kavramının en yalın bir biçimde ifade edildiği çalışmalar olarak da düşünülebilir. Bu çalışma temel araştırma modelinde olduğu için araştırma evreni ve örnekleme üzerinde durulmamıştır.

### Veri Toplama Aracı

Veri toplama aracının hazırlanmasında, Türkçe okuduğunu anlama alanında cümle tamamlama, eşanlamlı kelime, deyim ve atasözüne dayalı olarak yazılan çoktan seçmeli test maddeleri alanda görev yapan iki uzman görüşleri de alınarak yazılmıştır. Alana özgü 132 çoktan seçmeli madde oluşturulduktan sonra ön denemesi Denizli İlindeki lise ikinci sınıfta öğrenim görmekte olan 132 lise öğrencisine uygulanmıştır. Deneme amacıyla uygulanan 134 maddeden elde edilen veriler BILOG paket programı kullanılarak bireyselleştirilmiş test havuzu oluşturulmaya çalışılmıştır. Söz konusu test içinde yer alan maddelerin analizi sonucunda istenilen özelliklere sahip olmayan 34 madde test içerisinden çıkarılmıştır. Geriye kalan 100 madde içinden de 61 madde seçilerek bireyselleştirilmiş test temeline dayalı 13 adımda cevaplandırılabilir hale getirilmiştir. Her bir cevaplayıcının, madde havuzunda yer alan kendi yetenek düzeyine uygun olan 13 maddeyi (adım adım) cevaplandırarak test alma işlemini bitirecek formata uygun hale getirilerek uygulanmıştır. Başka bir deyişle, bireyselleştirilmiş test için durdurma kuralı olarak sabit 13 madde kullanılmıştır. Test uygulamasına ilişkin cevaplandırma algoritması şekil 2’de olduğu gibidir. Bireyselleştirilmiş test uygulaması için BASIC programlama dilinde yazılan bir yazılım yardımıyla, her bir cevaplayıcı programa giriş yapabilmek için kendilerine ait şifrelerini girerek soruları cevaplandırmaları sağlanmıştır. Cevaplayıcı şifresini programa girdikten sonra cevaplandıracağı ilk soru bilgisayar ekranına gelmektedir. Cevaplayıcı, cevaplandıracağı maddeye verdiği cevap yanlış ise ilk başlangıç maddesinden daha kolay olan bir madde ekrana gelmekte ve cevaplayıcı ekrana gelen maddeyi cevaplandırmaya çalışmaktadır. Buna karşılık, cevaplayıcı başlangıç maddesini doğru olarak cevaplandırdıysa, başlangıç maddesinden daha zor olan bir başka madde ekrana gelmekte ve cevaplayıcı ekrana gelen yeni maddeyi cevaplandırmaya çalışmaktadır.





Şekil 2. Bireyselleştirilmiş Teste Ait Cevaplandırma Algoritması

Cevaplayıcının bireyselleştirilmiş test içindeki en son maddeyi cevaplandırdıktan sonra test bitirme aşamasına gelmektedir. Test bitirme aşamasında cevaplayıcının vermiş olduğu cevaplar dosyada kayıt altına alınmaktadır. Cevaplayıcının cevaplama işlemi bittikten sonra yetenek ölçüsüne ait değerlerin hesaplanması çok kolaylaşmaktadır.

## BULGULAR

Araştırmanın birinci alt probleminde sorulan *Türkçe okuduğunu anlama alanında Klasik Test Kuramına ve Madde Tepki Kuramına göre iki-parametrelili lojistik model kullanılarak geliştirilen bireyselleştirilmiş testin KR-20 güvenirlik düzeyleri arasında manidar düzeyde bir farklılık var mıdır?* sorucuna cevap bulabilmek amacıyla öncelikle bireyselleştirilmiş testten elde edilen puanlar,

$$X_{\theta} = 9,1 x_{\theta} + 100 \quad (1.6)$$

eşitliği yardımıyla lineer dönüşüme işlemine tabii tutulmuştur. Bireyselleştirilmiş testten elde edilen puanların lineer dönüşüme tabii tutulmasının sebebi, uygulamada yer alan cevaplayıcılardan bir kısmının bireyselleştirilmiş test puanlarının işaretlerinin negatif (-) olması ve de bireyselleştirilmiş testten elde edilen puanlarla kâğıt-kalem testinden elde edilen puanlar arasındaki ilişki (korelasyon) düzeyinin (Pearson momentler çarpımı korelasyon katsayısının) hesaplanabilmesi için negatif (-) olan işaretin hesaplamada karşılaşılabilecek olan sıkıntıdan kurtulmak amacıyla lineer dönüştürme işlemi gerçekleştirilmiştir. Bireyselleştirilmiş testten ve kâğıt-kalem testinden elde edilen ham puanlar tablo 1’de verilmiştir.

Tablo 1. Bireyselleştirilmiş ve Kağıt-Kalem Testine Ait Puanlar

Cevaplayıcılar	Bireyselleştirilmiş Teste Ait Dönüştürülmemiş Puan	Bireyselleştirilmiş Teste Ait Dönüştürülmüş Puan	Kağıt Kalem Test Puanı	Cevaplayıcılar	Bireyselleştirilmiş Teste Ait Dönüştürülmemiş Puan	Bireyselleştirilmiş Teste Ait Dönüştürülmüş Puan	Kağıt Kalem Test Puanı
1	-0,19	98,29	27,48	36	0,48	104,34	28,77
2	-3,01	72,58	22,67	37	0,03	100,28	30,67
3	-0,21	98,06	31,71	38	0,70	106,40	29,40
4	-0,32	97,07	32,17	39	1,54	114,00	36,64
5	-0,36	96,76	32,32	40	1,22	111,12	25,59
6	1,50	113,66	24,41	41	-1,32	87,96	30,27
7	-0,43	103,90	28,97	42	-1,40	87,24	29,40
8	-0,62	94,32	33,46	43	-0,14	98,77	31,38
9	-2,57	76,53	26,57	44	0,16	101,47	30,11
10	1,35	112,24	36,49	45	-0,83	92,41	34,35
11	1,61	114,64	23,45	46	-1,93	82,44	29,42
12	2,02	118,42	42,24	47	-2,17	80,22	26,68
13	-1,24	88,69	36,10	48	0,17	101,58	30,06
14	-2,15	80,47	26,48	49	0,13	101,18	30,25
15	0,05	100,42	32,81	50	1,75	115,97	34,03
16	-2,40	78,20	24,72	51	0,37	103,36	31,27
17	0,47	104,28	34,45	52	-0,10	99,13	23,92
18	-1,40	87,29	24,65	53	0,28	102,53	29,62
19	0,01	100,05	30,77	54	0,54	104,92	28,50
20	0,22	101,99	29,87	55	1,80	116,42	34,74
21	0,94	108,58	33,74	56	0,70	106,34	31,45
22	-1,30	88,17	30,93	57	0,15	101,37	30,16
23	-0,50	95,50	32,91	58	-0,81	92,65	21,87
24	0,94	108,58	31,17	59	1,15	110,48	30,80
25	0,25	102,23	32,34	60	-0,24	97,78	20,27
26	0,50	104,52	31,29	61	-0,22	97,99	31,74
27	-0,76	93,05	27,62	62	-1,95	82,27	19,23
28	-0,26	97,64	31,90	63	-0,37	96,66	27,12
29	0,32	102,91	37,23	64	-0,87	92,12	29,10
30	-0,62	94,37	33,44	65	0,87	107,89	38,94
31	0,08	100,69	30,48	66	-0,57	94,80	30,28
32	-0,38	96,53	32,42	67	0,53	104,82	31,42
33	0,36	103,29	29,26	68	2,20	120,00	31,82
34	1,00	109,10	42,21	69	-2,42	87,02	19,49
35	-0,17	98,44	31,53				
				$\bar{X}$	-0,10	99,08	31,30
				$S_X$	1,15	10,52	10,24

Bireyselleştirilmiş test içinde yer alan maddelere ait iki-parametrelili lojistik modele göre  $a_j$  ve  $b_j$  parametre değerleri,

$$r_{jx} = \frac{a_j}{\sqrt{1+a_j^2}} \quad (1.8)$$

ve

$$P_j = \frac{1}{\sqrt{2\pi}} \int_{\frac{a_j \cdot b_j}{\sqrt{1+a_j^2}}}^{\pm\infty} e^{-\frac{1}{2}z^2} \cdot dz \quad (1.9)$$

$P_j$ : j maddesinin KTK'na göre madde güçlük indeksi  
 $r_{jx}$ : j maddesinin KTK'na göre madde ayırıcılık indeksi  
 $a_j$ : j maddesinin MTK'na göre ayırıcılık indeksi parametresi

dönüştürme için (1.8) ve (1.9) eşitlikler kullanılarak,  $P_j$  madde güçlük indeksi ve  $r_{jx}$  çift serili korelasyon katsayısı olarak madde ayırıcılık gücü katsayısına dönüştürüldükten sonra, bu değerlerden,

$$r_j = r_{jx} \sqrt{P_j \cdot q_j} \quad (1.10)$$

$P_j$ : j maddesinin KTK'na göre madde güçlük indeksi  
 $r_j$ : j maddesinin KTK'na göre madde güvenilirlik katsayısı  
 $q_j$ :  $1 - P_j$

(1.10) numaralı eşitlik yardımıyla, KTK'na göre madde güvenilirlik katsayıları hesaplanmıştır. Elde edilen madde güvenilirlik katsayıları Tablo 2'de verilmiştir.

Tablo 2. Bireyselleştirilmiş Test Uygulamasındaki Maddelere Ait Parametrelerinin KTK'daki Madde İstatistikleri Karşılıkları

Item No	$a_j$	$b_j$	$p_j$	$r_{jx}$	Item No	$a_j$	$b_j$	$p_j$	$r_{jx}$
1	0,41	0,11	0,52	0,38	32	0,73	-0,58	0,66	0,59
2	0,47	0,13	0,52	0,43	33	0,82	-0,69	0,53	0,63
3	0,65	0,12	0,53	0,55	34	0,69	-1,27	0,61	0,57
4	0,52	-0,20	0,54	0,46	35	0,78	0,56	0,63	0,62
5	0,72	-0,32	0,58	0,58	36	0,57	0,79	0,67	0,50
6	0,49	0,12	0,52	0,44	37	0,94	0,22	0,57	0,69
7	0,69	0,58	0,63	0,57	38	0,93	0,20	0,76	0,68
8	0,82	0,31	0,58	0,63	39	0,86	-0,19	0,63	0,65
9	0,74	-0,13	0,53	0,60	40	0,72	0,18	0,65	0,58
10	1,03	0,46	0,62	0,72	41	1,06	0,28	0,45	0,73
11	0,93	-0,75	0,70	0,68	42	0,94	0,22	0,55	0,68
12	0,83	-0,27	0,57	0,64	43	1,15	-0,16	0,43	0,76
13	0,67	1,046	0,72	0,56	44	1,07	0,86	0,41	0,73
14	1,12	0,25	0,57	0,74	45	0,41	0,99	0,57	0,38
5	0,78	1,58	0,83	0,62	46	0,77	0,43	0,55	0,61
16	0,48	-0,53	0,59	0,43	47	0,92	-0,12	0,42	0,68
17	0,52	-0,68	0,62	0,46	48	1,07	-0,78	0,72	0,73
18	0,55	1,72	0,80	0,48	49	0,86	0,82	0,70	0,65
19	0,66	1,24	0,75	0,55	50	0,53	1,68	0,78	0,47
20	0,59	0,39	0,58	0,51	51	0,61	0,06	0,41	0,52
21	0,47	-0,86	0,64	0,43	52	0,47	2,10	0,81	0,42
22	0,86	0,22	0,55	0,65	53	1,00	0,66	0,57	0,71
23	0,49	0,95	0,63	0,44	54	0,79	-0,74	0,68	0,62
24	0,73	0,19	0,54	0,59	55	0,98	1,03	0,67	0,70
25	1,21	-0,43	0,66	0,77	56	0,67	-0,86	0,76	0,56
26	0,94	-1,27	0,51	0,69	57	1,09	0,27	0,68	0,74
27	0,51	0,19	0,54	0,45	58	0,94	0,12	0,53	0,69
28	0,73	0,15	0,60	0,59	59	0,81	0,09	0,52	0,06
29	0,59	0,83	0,81	0,51	60	1,06	0,11	0,53	0,08
30	0,66	0,17	0,53	0,55	61	0,80	-0,57	0,63	0,36
31	0,51	0,76	0,54	0,45					

Bireyselleştirilmiş test uygulamasında yer alan her bir cevaplayıcının cevaplandığı test maddesine ait  $a_j$  ve  $b_j$  değerleri yerine bunların dönüştürülmüş değerleri kullanılarak KTK'ındaki test standart kayması

$$S_x = \sum_{k=1}^K r_j \quad (1.11)$$

eşitliği yardımıyla kestirilmiştir. Elde edilen dönüştürülmüş madde istatistikleri bireyselleştirilmiş test uygulamasında her öğrenciye uygulanan maddelerin KTK'na ait dönüştürülmüş madde istatistikleri ve test standart sapma değeri kullanılarak her öğrenci için bireyselleştirilmiş teste ait KR-20 güvenilirlik katsayısı hesaplanmıştır. Daha sonra, söz konusu güvenilirlik katsayılarının aritmetik ortalaması alınarak bireyselleştirilmiş teste ait KR-20 güvenilirlik katsayısı hesaplanmıştır. KTK'na göre KR-20 güvenilirlik katsayısı 0,75 olurken bireyselleştirilmiş teste ait güvenilirlik katsayısı da 0,67 olarak hesaplanmıştır.

Her iki yönetime ilişkin hesaplanan güvenilirlik katsayıları (1.12) ve (1.13) eşitlikleri yardımıyla Fisher'in Z istatistiğine dönüştürülerek 0,05 düzeyinde karşılaştırılmış istatistiksel olarak manidar düzeyde anlamlı bir farkın olmadığı görülmüştür.

$$z' = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \quad (1.12)$$

$$t = \frac{z'_1 - z'_2}{\sqrt{\frac{2}{N-3}}} \quad (1.13)$$

t: istatistik

r : korelasyon

Z': r'nin standart normal değeri

N: Örneklemden eleman sayısı

Bu sonuçlara göre, KTK'na göre elde edilen güvenilirlik katsayısı ile bireyselleştirilmiş teste ait elde edilen güvenilirlik katsayısı istatistiksel anlamda manidar fark göstermemesi her iki güvenilirlik düzeyinin aynı düzeyde olduğunu göstermektedir. Bu sonuca göre bireyselleştirilmiş testlerin KTK'na göre ölçme araçlarının ya da ölçme araçlarında yer alan maddelere ait güvenilirlik düzeylerinde istatistiksel olarak anlamlı düzeyde katkı sağladığını iddia etmek mümkün görünmemekle beraber, bireyselleştirilmiş test algoritmasına bağlı olarak her bir öğrenci 13 madde cevaplandırırken KTK'na göre ise kâğıt-kalem testi içinde yer alan 47 maddeyi cevaplandırmıştır. Bireyselleştirilmiş teste göre madde havuzundaki soru sayısı arttırılmış olsaydı, öğrencilerin cevaplandıracağı madde sayısı da doğal olarak artacaktı. Bu durumda bireyselleştirilmiş testlere göre elde edilen güvenilirlik katsayısının değeri KTK'na göre elde edilen güvenilirlik katsayısının değerinden daha büyük olacak ve de istatistiksel olarak manidar çıkma olasılığı da söz konusu olabilecekti. Bu durumda bireyselleştirilmiş testlerde daha az sayıda madde ile ulaşılan güvenilirlik katsayısı daha çok sayıdaki maddeden oluşturulan kâğıt-kalem testinden daha büyük ve anlamlı güvenilirlik katsayısına ulaşılmasına olanak sağlayarak, bireyselleştirilmiş testlerin kâğıt-kalem testine (KTK'na) göre daha kullanışlı olduğunu ve üstünlüğünü gösterecekti.

Araştırmada cevabı aranan ikinci alt problemi olarak sorulan soru ise “KTK'na ve İki-parametrelili Lojistik model kullanılarak geliştirilen “Türkçe okuduğunu anlama” testine ilişkin öğrencilerin yetenek düzeylerine ait ölçüleri arasında hangi düzeyde bir ilişki olduğu şeklinde ifade edilmiş olup, bu sorunun cevabını bulabilmek amacıyla öğrencilerin bireyselleştirilmiş test ile KTK'na göre hazırlanan kâğıt-kalem testinden elde edilen ham puanları arasında Pearson momentler çarpımı korelasyon tekniği kullanılarak korelasyon katsayısı hesaplanmıştır. Her iki yaklaşıma ait ham puanlar kullanılarak korelasyon katsayısı 0,36 olarak hesaplanmıştır.

Bireyselleştirilmiş teste ve KTK'na göre hazırlanan kâğıt kalem testinden elde edilen korelasyon katsayısının bu denli düşük çıkmasının sebebi araştırmanın belki de en büyük sınırlılıklarından biri olarak görülebilir. Her iki yöntemde göre hazırlanan testten elde edilen korelasyon katsayısının düşük çıkmasına sebep olarak bireyselleştirilmiş testin oluşturulduğu madde havuzunda istenilen sayıda içerikle ilgili maddenin olmayışından kaynaklanmış olabileceğini dikkate almakta yarar görülmektedir. Madde havuzunda yeteri sayıda maddenin olmayışı, çalışmayı yapan araştırmacının araştırmaya ilişkin mevcut olanaklarıyla ilişkili olabilir.

## SONUÇ, TARTIŞMA VE ÖNERİLER

Bu araştırmada, Türkçe okuduğunu anlama alanına yönelik olmak üzere, Klasik Test Kuramına dayalı olarak geliştirilen kâğıt-kalem ile Madde Tepki Kuramına (İki parametrelili Lojistik modele) dayalı olarak geliştirilen bireyselleştirilmiş test güvenilirlik düzeyleri karşılaştırılmıştır. Klasik Test Kuramının bazı sınırlılıklarından dolayı 20. Yüzyılın ortalarına doğru Madde Tepki Kuramı (IRT) veya Örtük Özellikler Kuramı olarak adlandırılan kuram ortaya çıkmıştır (Crocer, & Algina, 1986). Bireylere özgü ölçülmek istenen özelliklerin ölçülmesinde kullanılan ölçeklerin puanlanmasında yaygın olarak KTK kullanılmakla beraber Madde Tepki Kuramı (MTK) gittikçe daha fazla kullanılır olmaya başlamıştır (Hambleton, 1994; Reise, Ainsworth, & Haviland, 2005).

Klasik Test Kuramından elde edilen ölçümler gruba bağlı olurken, Madde Tepki Kuramı gruptan bağımsız (sample free) ve değişmez parametre kestirimleri yapma iddiasındadır. Değişmezlik hem aynı özelliği ölçmeye yönelik olarak hazırlanmış olan farklı maddelere verilen tepkilere dayalı olarak kestirilen yetenek parametrelerinin de içermesi (test free) hem de aynı testin farklı bireylere uygulanmasıyla elde edilen madde parametrelerinin değişmezliği olarak ele alınabilir. Bu durum, bir testin Madde Tepki Kuramı'na göre bir kez ölçeklendikten sonra maddelerin özelliği değişmediğinde, pek çok kez kullanılmasına olanak sağlar. Ancak bu değişmezliğin sağlanması, madde parametrelerinin elde edilmesinde yapılan deneme uygulamasının ve bu uygulamanın yapıldığı grubun bazı şartları sağlamasına bağlıdır (Hambleton, & Swaminathan, 1985; Hambleton, 1990; Hambleton vd., 1991; Kelecioğlu, 2001). Klasik Test Kuramı'nda ise bireylerin aldıkları puanlar o testin güçlük düzeyine göre değişmektedir (Lord, & Novic, 1968 Akt: Çelen, 2008).

Yine Klasik test Kuramında ölçmenin standart hatası bütün grup için hesaplanırken, madde Tepki Kuramında her bir birey için ayrı ayrı hesaplanabilmektedir. Klasik Test Kuramı'nda ölçme hataları tüm grup için hesaplanırken, Madde Tepki Kuramı'nda her birey için ayrı ayrı kestirilebilmektedir. Yine güvenilirlik katsayısı Klasik Test Kuramı'nda cevaplayıcı grubun puan dağılımı için tek bir değer olarak hesaplanırken, Madde Tepki Kuramı'nda her bir madde ve yetenek düzeyi için güvenilirlik madde ve test bilgi fonksiyonu şeklinde hesaplanabilmektedir. Klasik Test Kuramı'nda elde edilen tek katsayı, güvenilirliğin farklı yetenek düzeyleri için değişmediği anlamına gelmekte iken, tekrarlı ölçmelerle hesaplanan güvenilirlik katsayılarına bakıldığında, bunların ölçülen özelliğe üst düzeyde sahip bireyler için daha yüksek olduğu görülmektedir. Bu durum farklı yetenek düzeyindeki bireyler için ölçme aracının aynı düzeyde güvenilirliğe sahip olamayacağını göstermektedir (Nartgün, 2002 Akt: Çelen, 2008).

Eldeki Araştırmanın yapıldığı zaman dilimine kadar Türkçe okuduğunu anlama alanına ilişkin, Klasik Test Kuramına bağlı olarak geliştirilen kâğıt-kalem testi ve Madde Tepki Kuramına göre iki-parametrelili lojistik model kullanılarak geliştirilen bireyselleştirilmiş teste ait yetenek düzeylerine ait ölçümlerin güvenilirlik düzeyini karşılaştırmayı amaçlayan bir çalışmaya rastlanılmamıştır.

Bu araştırmada, Türkçe okuduğunu anlama alanına yönelik olmak üzere, Klasik Test Kuramına dayalı olarak geliştirilen kâğıt-kalem ile Madde Tepki Kuramına (İki parametrelili Lojistik modele) dayalı olarak geliştirilen bireyselleştirilmiş test güvenilirlik düzeyleri karşılaştırılmıştır. Klasik Test kuramına göre geliştirilen testin KR-20 güvenilirlik düzeyi 0,75 ve İki-parametrelili lojistik modele göre hesaplanan güvenilirlik düzeyi 0,67 olarak hesaplanmıştır. Her iki yönteme göre hesaplanan güvenilirlik Fisher'in Z istatistiğine

dönüştürülerek 0,05 anlamlılık düzeyinde test edilmiş ve manidar bir fark olmadığı görülmüştür.

Öte yandan, ker iki yöntemle dayalı olarak geliştirilen testin uygulandığı gruptaki bireylerin hesaplanan yetenek ölçüleri arasında ilişki düzeyinin belirlenmesi için de Pearson momentler çarpımı korelasyon tekniği kullanılarak korelasyon katsayısı hesaplanmıştır. Korelasyon katsayısının hesaplanabilmesi için cevaplayıcıların bireyselleştirilmiş testten elde ettikleri yetenek ölçüleri dönüştürmeye tabi tutulduktan sonra korelasyon katsayısı 0,36 olarak hesaplanmıştır. Elde edilen 0,36'lık korelasyon katsayısı oldukça düşük bir ilişkiye işaret etmektedir. Muhtemelen bunun sebebi çok geniş bir madde havuzu gerektiren bireyselleştirilmiş testteki madde sayısının azlığından kaynaklanıyor olabilir.

Bu noktadan hareketle, bu konuya dayalı çalışma yapacak olan araştırmacıların özellikle bireyselleştirilmiş teste ait madde havuzundaki madde sayısının yeteri sayıda olmasına dikkat etmesi gerekir. Bununla birlikte, bu araştırmada Türkçe okuduğunu anlamaya yönelik olarak Klasik Test Kuramına dayalı olarak geliştirilen Kağıt-kalem testi ile Madde Tepki Kuramına (iki-parametrelili lojistik modele) göre geliştirilen bireyselleştirilmiş teste ait KR-20 güvenilirlik düzeyleri karşılaştırılırken, bunun dışında yine üç-parametrelili lojistik yöntem kullanılarak farklı disiplinlere yönelik olarak geliştirilecek başarı testlerinin güvenilirlik düzeylerinin karşılaştırılmasını sağlayacak çalışmalar yapılabilir.

#### KAYNAKLAR

- Baykul, Y. (1979). *Örtük özellikler ve klasik test kuramları üzerine bir araştırma* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Bloom, B. (1956). *Taxonomy of educational objectives: Handbook I, The cognitive domain*. New York: David McKay & Co
- Bulut, O., & Kan, A. (2012) Application of computerized adaptive testing to entrance examination for graduate studies in Turkey. *Eğitim Araştırmaları-Eurasian Journal of Educational Research*, 49, 61-80.
- Can, S. (2003). *Ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı sözel bölümünün Madde tepki kuramı modellerine göre analizi* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Crocker, L., & Algina, J. (1986). *Introduction Classical and Modern Test Theory*. USA: CBS College Publishing Company
- Çalışkan, M. (2000). *Madde tepki kuramının (MTK) bir, iki ve üç parametrelili modellerinin Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi'nin (MEB-EARGED) fen bilgisi başarı testi verilerine uygunluğu* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çelen, Ü. (2008) Comparison of validity and reliability of two tests developed by classical test theory and item response theory. *Elementary Education Online*, 7(3), 758-768. [Online]: <http://ilkogretim-online.org.tr>
- Çelen, Ü., & Aybek, E. C (2013) Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- Çelik, D. (2001). *Madde tepki kuramının (MTK) bir-, iki-, ve üç parametrelili modellerinin Milli Eğitim Bakanlığı ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı testi verilerine uygunluğu* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara

- Çıkrıkçı-Demirtaşlı, N. (1998). Test geliştirmede yeni yaklaşımlar: Örtük özellikler kuramı – temel özellikleri, varsayımları, model ve sınırlılıkları. *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 2 (28), 161-173.
- DeMars, C. (2010). *Item Response Theory*. Oxford: Oxford University Press.
- Demirtaşlı, N., & Arıkan, S. (2009). ÖİS'de MTK uygulamaları. I. *Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi Bildiri Kitabı*. Ankara Üniversitesi Yayınları. ss. 225-227.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Eroğlu, M. G. (2013). *Bireyselleştirilmiş bilgisayarlı test uygulamalarında farklı sonlandırma kurallarının ölçme kesinliği ve test uzunluğu açısından karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Gelbal, S. (1994). p madde güçlük indeksi ile Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçülerine üzerine bir karşılaştırma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 10, 85-94.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory (2nd Ed.)* California: Sage Publications, Inc.
- Hambleton, R. K., & Swaminathan, H. (1984). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambelton, R. K. (1994). Item response theory: a broad psychometric framework for measurement advances. *Psicothema*, 6, 3, 535-556.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and application*. Kluwer, Nijhoff Publishing a Member of the Kluwer Academic Publisher Group
- Kan, A. (2006). Klasik test teorisine ve örtük özellikler teorisine göre kestirilen madde parametrelerinin karşılaştırılması üzerine ampirik bir çalışma. *Mersin University Journal of the Faculty of Education* 2 (2), 227-235.
- Kaptan, F. (1993). *Yetenek kestiriminde adaptive (bireyselleştirilmiş) test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Karataş, A. G. (2001). *Madde tepki kuramı (MTK) modellerini kullanarak bir İngilizce yeterlilik sınavının ölçeklendirilmesi* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Kezer, F., & Koç, N. (2014) Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması. *Eğitim Bilimleri Araştırmaları Dergisi*. Cilt 4 Sayı 1
- Kılıç, İ. (1999). *Madde tepki kuramının (MTK) bir, iki ve üç parametrelili modellerinin Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) Öğrenci Seçme Sınavına (ÖSS) uygunluğu* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Larkin, K. C., & Weiss, D. J. (1974) *An empirical investigation of computer-administered pyramidal ability testing (Research Report 74-3)*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Linden, W. J. Van Der, & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Lord, F. M., & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Addison Wesley Publishing Company, Educational Testing Service.
- Lord, F. M. (1970). *Some test theory for tailored testing*. In W. H. Holtzman (Ed.), *Computer assisted instruction, testing, and guidance* (pp. 139-183). New York: Harper & Row



- Lord, F. M. (1971a). Tailored testing, an approximation of stochastic approximation. *Journal of the American Statistical Association*, 66, 707–711.
- Lord, F. M. (1971b). A theoretical study of the measurement effectiveness of flexi level tests. *Educational and Psychological Measurement*, 31, 805–813.
- Özkurt, S. (2002). *Madde tepki kuramının (MTK) bir-, iki-, ve üç- parametrelili modellerinin bir İngilizce yeterlik başarı testi verilerine uygunluğu* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14, 2, 95-101.
- Turgut, M. F. (1984) *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Saydam.
- Weiss, D. J. (1974) *Strategies of adaptive ability measurement (Research Report 74-5)*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1974.
- Weiss, D. J. (Ed.). (1983). *New Horizons in Testing: Latent Test Theory and Computerized Adaptive Testing*. USA: Academic Press.
- Yalçın, M. (1999). *Eğitimi araştırma ve geliştirme dairesi başarı testlerinin madde-tepki kuramının bir, iki, üç parametreye uygunluğu* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yeğın, O. P. (2003). *Başkent Üniversitesi İngilizce yeterlik sınavının (Büiys) madde Madde Tepki Kuramının (MTK) üç parametrelili modelinin kullanımıyla elde edilen yetenek kestirimlerinin yordama geçerliği* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yapar, T. (2003). *İki parametrelili tepki kuramı (MTK) modelinin yetenek kestirimleriyle Başkent Üniversitesi İngilizce yeterlik sınavının yordama geçerliğini inceleme çalışması* (Yayımlanmamış Doktora Tezi). Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

**Atıf için:**

Yaşar, M. (2017). Bireyselleştirilmiş testler üzerine bir çalışma. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research*, 7(2), 285-301. <http://ebad-jesr.com/>