



Uludağ Üniversitesi Eğitim Fakültesi Dergisi

<http://kutuphane.uludag.edu.tr/Univder/uufader.htm>

Bireyselleştirilmiş Bilgisayarlı Test Uygulamalarında Farklı Sonlandırma Kurallarının Ölçme Kesinliği ve Test Uzunluğu Açısından Karşılaştırılması¹

Melek Gülşah EROĞLU¹, Hülya KELECİOĞLU²

¹Gazi Üniversitesi Gazi Eğitim Fakültesi E.B.B. Eğitimde Ölçme ve Değerlendirme
A.B.D., mgulsah@gazi.edu.tr

²Hacettepe Üniversitesi Eğitim Fakültesi E.B.B. Eğitimde Ölçme ve Değerlendirme
A.B.D., hulyaebb@hacettepe.edu.tr

ÖZET

Bireyselleştirilmiş testlerde, geleneksel testlerden farklı olarak test algoritması söz konusudur. Test algoritması; teste başlama, devam etme ve testi sonlandırma olmak üzere üç bölümden oluşmaktadır. Bu çalışmanın amacı, bireyselleştirilmiş bilgisayarlı test (BBT) uygulamalarında farklı sonlandırma kurallarının kullanılmasının ölçme kesinliğine ve test uzunluğuna etkisini incelemek ve birbirleri ile karşılaştırmaktır. Araştırma simülasyon çalışması olarak yürütülmüştür. Araştırma kapsamında sabit uzunluk, standart hata, standart hata-en az madde, theta yakınsama ve theta yakınsama-en az madde olmak üzere beş farklı sonlandırma kuralı kullanılmıştır. Her bir sonlandırma kuralında farklı koşullar söz konusu olup toplam 12 koşul birbiriyle karşılaştırılmıştır. Ayrıca sonlandırma kurallarının karşılaştırılmasında BBT’de test algoritmasında önemli yere sahip olan farklı madde havuzu büyüklükleri (250 ve 500 madde) ve yetenek kestirim yöntemleri (Maksimum Olabilirlik Kestirimi ve Beklenen Sonsal Dağılım) seçilmiştir. Her bir BBT uygulamasında ölçme kesinliği için RMSE, yanlışlık ve

¹ Bu makale aynı başlıklı doktora tezinden üretilmiştir.

uyum deđerleri hesaplanmıř ve test uzunlukları elde edilip, birbirleriyle karřılařtırılmıřtır. Arařtırmanın sonucunda, genel olarak 20 madde sabit uzunluk, 0,220 standart hata ve 0,02 theta yakınsama sonlandırma kořullarında RMSE, yanlılık deđerlerinin dūřuk elde edildiđi ancak uyum katsayılarının önemli oranda etkilenmediđi belirlenmiřtir. Ayrıca en az madde kořulunun eklenmesi ile bazı sonlandırma kořulları ölçme kesinliđi açasından daha iyi sonuçlar vermiřtir. Ortalama test uzunluk deđerlerinin RMSE deđerleri ile ters yönde deđiřtiđi bulunmuřtur. Aynı sonlandırma kořullarında madde havuzu büyüklüđünün artması ile ölçme kesinliđi için elde edilen RMSE ve yanlılık deđerlerinin genel olarak daha dūřuk elde edilmiřtir. Yetenek kestirim yöntemi olarak Beklenen Sonsal Dađılım yönteminin kullanılmasının RMSE ve yanlılık deđerlerinde dūřmeye neden olduđu belirlenmiřtir.

Anahtar Sözcükler: Bireyselleřtirilmıř Bilgisayarlı Test (BBT), Sonlandırma Kuralları, Ölçme Kesinliđi, Test Uzunluđu.

Comparison Of Different Test Termination Rules in Terms of Measurement Precision and Test Length in Computerized Adaptive Testing

ABSTRACT

In adaptive testing, there exists a test algorithm different than the classical tests. The test algorithm consists of three parts which are starting, resuming and termination. The aim of this study is to measure the effect of different termination rules on measurement precision and test length in computer adaptive testing.

The research was implemented as a simulation study. Five different termination rules have been used for the study which are: fixed length, standard error, standard error-least item, theta convergence and theta convergence-least item. Different conditions are in place in each termination rule and a total of 12 conditions are compared. Additionally, in comparing termination rules, different item pools (250 and 500) and ability estimation methods (Maksimum Likelihood Estimation and Expected a Posteriori) have been selected since these are critical in the algorithms of Computer Adaptive Testing. RMSE, bias and fidelity values were calculated for the measurement precision and test lengths were obtained and compared for each of the CAT implementation.

As a result, for the 20 item fixed length, 0,220 standard error and 0,02 theta convergence termination conditions RMSE and bias values are small but fidelity factors are not significantly affected. And with the addition of the least item factor, better results were achieved in some of the termination conditions in terms of measurement precision. The test length is observed to be negatively correlated with the RMSE values. In the same termination conditions, with the increases in item pool generally smaller RMSE and bias values were for measurement precision were

achieved. Not a significant change was observed in the evaluation of the effect of the starting rules. The preference of Expected a Posteriori method for the ability estimation is observed to cause a drop in values for RMSE and bias values.

Key Words: Computer Adaptive Testing (CAT), Termination Rules, Measurement Precision, Test Length.

GİRİŞ

1980'lerin sonlarına dođru bireysel bilgisayarların eđitimde giderek yaygınlaşması ile ölçme ve deđerlendirme amacıyla kullanılan test formatları bilgisayar ile uyumlu formatlara dođru hızlıca geçmeye başlamıştır (Linden ve Glas, 2002). Bir testin bilgisayar aracılığıyla sunulması iki şekilde olabilir: Birincisi kađıt kalem testlerinin bilgisayar aracılığıyla dođrusal yolla (maddelerin tüm bireylere önceden belirlenmiş belirli bir sıra ile) sunulması ve izlenmesinin mümkün olduđu Bilgisayar Temelli Test (Computer Based Test)tir. Bu test kađıt-kalem testi ile madde, puanlama yöntemleri bağlamında benzer olup tek farkı testin uygulanma biçimidir. İkinci ise Bireyselleştirilmiş Bilgisayarlı Test (Computer Adaptive Test)tir. BBT esas olarak TOEFL (Test of English as a Foreign Language), GRE (Graduate Record Examination), GMAT (Graduate Management Admission Test) gibi dünya çapında yapılan testlerde ve birçok başka alanda kullanılmaktadır. Bireyselleştirilmiş Bilgisayarlı Test (BBT)'lerde kađıt-kalem cinsinden karşılaştığımız şekilde soru formu yoktur. BBT'de testin uygulama şekli test maddelerinin sınavı alan herkesin yeteneklerine göre biçimlendirilmesine olanak sağlar. Ayrıca testin bireylere uygulanma şekli bilgisayar temelli testlerden farklı olarak dođrusaldan bireyselleştirmeye dönüşmektedir (Pearson Assessment, 2009). Bir başka deyişle; bilgisayar, önceden belirlenmiş bir soru grubunu sıra ile uygulamak yerine, soruları testin gidişine (testi cevaplayanların verdiđi cevaba) göre seçerek sormaktadır (Kalender, 2004).

Weiss (1998)'a göre psikometrik bakış açısından, BBT uygulamasının iki önemli avantajı vardır (Hambleton, Swaminathan ve Rogers, 1991): Biri, ölçmenin kesinliğini/dođruluđunu (precision/accuracy of measurement) arttırması; diđeri ise daha kontrollü ve güvenli bir test ortamı sağlamasıdır. BBT uygulaması beş temel bileşenden oluşmaktadır. Bunlar (a) madde havuzunun oluşturulması, (b) ilk maddenin seçimine ilişkin kural, (c) bir sonraki maddenin seçimine karar verecek madde seçim algoritması, (d) bireyin önceki cevaplarını temel alan yetenek kestirim süreci, (e) testin sonlandırılmasına ilişkin kuraldır (Weiss ve Kingsbury, 1984). Tüm bu bileşenlerin birbiriyle uyumlu olarak işleminde bir

matematiksel modelleme söz konusudur. En uygun matematiksel model, Madde Tepki Kuramı (MTK)'dır. MTK bireyin yeteneđinden yararlanarak, bireyin madde üzerinde göstereceđi performansın olasılıđı üzerinde yoğunlaşır. Bu kuramın en önemli noktası bireylerin yeteneklerini, aldıkları maddelerden bađımsız olarak kestirmektedir. Bireyler farklı maddeler içeren iki farklı test olsa bile kestirilen yetenek düzeyleri farklı olmamaktadır.

MTK'ye dayalı BBT'de verilen bir özelliđe ilişkin madde havuzunun oluşturulması aşamasında, öncelikle görelilik olarak geniş bir madde havuzu geliştirilir ve maddelerin madde bilgi fonksiyonu belirlenir. İyi bir BBT bankası, bireylerin yetenek düzeyi θ 'nın bütün aralıkları boyunca bilgi sunan maddelere sahip olmalıdır (Weiss, 2004).

İlk maddenin seçimine ilişkin kural teste başlama test algoritmasının altında incelenebilir. BBT'de uygulanacak ilk maddenin güçlüđu orta düzeyde ise psikometrik açıdan daha etkin olmaktadır. Çünkü bireyin yeteneđi hakkında ön bilğimiz yoksa en iyi tahmin orta düzeyde yeteneđe sahip olduğunu kabul etmektir. (Mills ve Stocking, 1996). Teste başlamada kullanılan farklı bir yöntem ise bireyin yeteneđine en uygun soru ile başlanmasıdır. Fakat daha önceki cevaplara dayandırılmış bir yetenek düzeyi olmadığı için testin hangi madde ile başlayacağı belirsizdir. Bu nedenle, benzer testleri almış deneklerin yetenek puanlarının ortalaması ilk sorunun güçlük düzeyini belirlemek için kullanılabilir. Ayrıca deneklerin yaş, aldıkları kurslar gibi demografik bilgileri bilindiđi durumlarda, bu özellikleri yansıtan daha küçük farklı bir grubun yetenek ortalaması ilk tahmin olarak kullanılabilir. Bireylerin daha önce aldığı teste benzer başka bir test alması durumunda ise bu iki test arasındaki ilişkinin çıkarılması da yetenek düzeyi tahmininde kullanılabilir.

Bir sonraki maddenin seçimine karar verecek madde seçim algoritması ve bireyin önceki cevaplarını temel alan yetenek kestirim süreci teste devam etme test algoritmasının altında incelenebilir. Kullanılan yetenek kestirimi yöntemine göre ilk maddenin seçiminin ardından takip edilecek farklı kestirim yöntemleri mevcuttur. Bireyin yeteneđini tahmin etmede kullanılan iki temel yöntem; Maksimum Olabilirlik Kestirimi (Maximum Likelihood Estimation - MLE) ve Bayes yöntemidir. Bayes temelli yetenek kestirim yöntemlerinden BBT uygulamalarında en çok kullanılan Beklenen Sonsal Dađılım (Expected a Posteriori - EAP) ve En Çok Sonsal Dađılım (Maksimum a Posteriori - MAP) dır. Bireyin yeteneđini tahmin etmede bu kestirim yöntemlerinden Maksimum Olabilirlik Kestirimi kullanıldığında kestirim için en az bir doğru, bir de yanlış yanıt verilmesi gerekir. Bu nedenle, bu kestirim yöntemi seçildiğinde bilgisayarın sıradaki maddeyi

bireyin bir dođru veya bir yanlış cevap verecek şekilde seçmesi gerekmektedir. Bir başka deyişle, bireyin ilk maddeye verdiđi cevap dođru ise bilgisayar daha zor bir madde sorarak bireyin yanlış yanıt verme olasılıđını artırmaktadır. Eđer birey tüm maddeleri dođru veya yanlış cevaplarlsa; MLE yöntemleri yeteneđin dođru tahminde yetersiz kalabilir ve yetenek düzeyini negatif veya pozitif sonsuzlukta tahmin edebilir. Bu sınırlılık bireyin yeteneđinin kestirilmesi için MLE de olduđu gibi bir dođru veya bir yanlış yanıt verme zorunluluđu olmayan Bayes kestirim yöntemleri kullanılarak giderilebilir. Bu yöntemlerde bireylerin yetenek tahminleri hakkındaki ön bilgileri olabilirlik fonksiyonu ile ilişkilendirilir. (Hambleton, Swaminathan ve Rogers, 1991). Hangi yetenek kestirim yönteminin daha iyi olduđunu belirlemede, BBT uygulaması diđer bileşenleri ile birlikte düşünölmelidir. Madde seçiminde hangi yöntemin kullanılacağına karar verme bu aşamada kritik bir rol oynamaktadır. Madde havuzu, bireyin demografik bilgilerinin tahmin sürecinde kullanılıp kullanılmayacağı, madde kullanım sıklılıđının belirlenmesindeki yöntem seçimi, madde seçiminde içerik dengesinin göz önüne alınması gibi deđişkenler ise yetenek kestirim yöntemi seçiminde etkili olan boyutlardır (Linden ve Glas, 2002).

Bireylerin ilk yetenek düzeyi kestirilmesinden sonraki aşama bundan sonra uygulanacak maddelerin seçilmesidir. BBT uygulamasındaki madde seçiminde Maksimum Bilgi Yöntemi (Maximum Information Method) ve Bayes Yöntemi yaygın olarak kullanılmaktadır (Çıkrıkçı Demirtaşlı, 1999). Maksimum Bilgi Yönteminde, belirli bir yetenek düzeyi için en yüksek bilgiyi sağlayarak yetenek düzeyinin standart hatasına beklenen katkısı minimum olan maddeler seçilir. Bayes yöntemlerde, yetenek dağılımının varyansını minimize edebilecek maddeler seçilmektedir.

BBT uygulamalarında kullanılabilcek çok sayıda testi sonlandırma kuralı bulunmaktadır. Bu kuralların her biri testin standart hatası ve yetenek kestirimlerinin yanlışlıđı üzerinde farklı etkilerde bulunmaktadır (Blais ve Raiche, 2000; Yi, Wang ve Ban, 2001; Hart, Cook, Mioduski, Teal ve Crane, 2006; Choi, Grady ve Dodd, 2011; Babcock ve Weiss, 2012). Testin sonlandırılmasında genel olarak, deđişen uzunluk ve sabit uzunluk sonlandırma kuralları kullanılmaktadır. Test geliřtiricileri, sonlandırma kuralının seçiminin genel olarak; testin amacına, madde havuzunun özelliđine ve işlevsel kısıtlılıklara dayalı olarak yapıldıđını bulmuşlardır (Segall, 2004).

Sabit uzunluk sonlandırma kuralında, önceden belirlenmiş sayıda madde uygulanınca test sonlanmaktadır. Tüm bireylere testin sonunda ulařılan ölçme kesinliđinin derecesi göz ardı edilerek aynı sayıda madde

uygulanmaktadır. En önemli avantajı kolay olmasıdır. Bu kuralın uygulanmasının sonuçlarından biri de tüm bireylerin farklı kesinlik derecelerinde ölçölüyor olması ve genellikle uç yetenek seviyelerinde daha fazla ölçme hatasına neden olmasıdır. Bu duruma ek olarak bireyin yetenek düzeyinde daha az bilgiye sahip maddelerin gereksiz yere uygulanması BBT'nin etkililiđini sınırlayabilir (Choi, Grady ve Dodd, 2010).

Deđişen uzunluk sonlandırma kuralında bireylerin cevap örüntülerine bađlı olarak farklı sayıda maddenin uygulanması söz konusudur. Deđişen uzunluk sonlandırma kuralını uygulamak ölçmenin etkililiđi ve ölçmenin kalitesi bakımından avantaj sađlamaktadır. Ölçmenin etkililiđi ile bireyin yeteneđinin göreceli olarak daha az madde ile kestirilmesi kastedilmektedir. Bu durum, ölçme için gerekli zamanı azaltmakta ve madde havuzunun kullanımını etkili kılmaktadır (Weiss ve Kingsbury, 1984). Deđişen uzunlukta uygulanan BBT, teorik olarak sabit uzunlukta uygulanan BBT ile eşit veya daha üstün ölçme kalitesi sađlamaktadır. Eđer test puanı belli bir ölçme kesinliđine ulařmıřsa, biraz daha fazla maddenin uygulanması puanın kesinliđini önemli ölçüde etkilememektedir. Deđişen uzunluk sonlandırma kuralının sabit uzunluđa kıyasla bir dezavantajı çok az madde ile sonlandırıldıđında ölçme kalitesinin düşük olmasıdır. Madde havuzunda sonlandırma kuralı ile örtüşen bir maddenin kalmaması da başka bir dezavantajdır. Sabit uzunlukta BBT beklendiđi gibi sonlandırılırken; deđişen uzunlukta BBT de çok fazla maddenin uygulanma ihtimali de bulunmaktadır. Ayrıca sabit uzunlukta BBT uygulamalarını sürdürme ve kalite kontrolünü sađlamada deđişen uzunlukta BBT uygulamasına göre daha kolaydır. Testi alan adaylar sabit uzunlukta testi almanın daha adil olduđunu düşünebilirler. (Babcock ve Weiss, 2012). Deđişen uzunlukta sonlandırma kuralları temel aldıđı ölçütlere göre çeřitlilik gösterir. Bunlar Standart Hata (SH), Minimum Bilgi (MB) ve Theta Deđişimi sonlandırma kuralıdır.

Standart hata sonlandırma kuralında bireyin yetenek düzeyi belli bir kesinlik seviyesine ulařıncaya kadar madde uygulanmaktadır. Bu kural, kabul edilebilir standart hata için yeteri kadar bilgi veren maddelerin bulunmasına ve bireye uymayan maddelerin göreceli olarak az olmasına bađlıdır. Weiss ve Kingsbury (1984)'e göre sonlandırma kuralları içinde en yaygın olanı standart hata sonlandırma kuralıdır. Yapılan arařtırmalar, BBT uygulamalarında standart hata sonlandırma kuralı ile her bir yetenek düzeyi kestirildiđinde ölçme kesinliđinin yüksek olduđunu göstermektedir. Ancak burada madde havuzunun, her bir yetenek düzeyinde standart hata deđerinin elde edilmesine imkân vermesi gerekmektedir (Dodd, Koch ve De Ayala, 1993; Wang ve Wang, 2001; Babcock ve Weiss,2012).

Minimum bilgi sonlandırma kuralında, havuzda bireylerin en son yetenek düzeyinde bilgi sađlayabilen madde kalmayınca kadar BBT uygulamasının sürdürölmesi söz konusudur (Gialluca ve Weiss, 1979; Maurelli ve Weiss, 1981; akt: Babcock ve Weiss, 2012). Bazı arařtırmalar minimum bilgi sonlandırma kuralının etkili bir sonlandırma kuralı olduđunu ortaya koymaktadır (Brown ve Weiss, 1979).

Theta deđiřimi sonlandırma kuralı, bireyin yetenek düzeyi theta (Θ) veya theta'ya yakınsayan yetenek düzeyindeki deđiřimi ölçüt almaktadır. BBT'de bireye uygulanan her madde sonrasında bireyin yetenek düzeyi deđiřmektedir. BBT uygulamasının bařlangıcında yetenek düzeyindeki deđiřim fazla iken; genellikle giderek daha azalmaktadır (Weiss ve Kingsbury,1984). Yetenek düzeyine yakınsama da (convergence of Θ estimate) BBT için iyi bir sonlandırma kuralı olabilir (Babcock ve Weiss, 2012).

Bunların dıřında madde bankasının tükenmesi, madde bankalarının küçük olması ve birey tarafından her maddenin alındıđı durumlarda test sonlandırılabilir. Ayrıca BBT yazılımları verilen cevap örüntüsünü sürekli takip eder. Bireyin çok hızlı ya da çok yavař olması durumunda test yöneticisi bireyle görüřüp testi durdurmak veya ertelemek için karar alabilir. Bireye minimum sayıda madde yönlendirilmediyse veya teste dahil olan tüm konulardan soru gelmediyse testin durdurulması kesinlikle önerilmez (Linacre, 2000).

BBT uygulamalarında kullanılabilir çok sayıda testi sonlandırma kuralı bulunmaktadır. Alan yazın incelendiđinde farklı testi sonlandırma kuralları ile birlikte madde havuzu özellikleri, bařlama kuralları ve testi sürdürmede kullanılan yöntemlerin birlikte testin ölçme kesinliđine etkisinin incelendiđi görölmektedir (Wang, Hanson ve Lau, 1999; Yi, Wang ve Ban, 2001; İřeri 2002; Blais ve Raiche, 2002; Simms ve Clark, 2005; Ivie, 2007; Riley, Conrad, Bezruczko ve Dennis, 2007; Wang, 2009; Kalender, 2011; Choi, Grady ve Dodd, 2011; Babcock ve Weiss, 2012).

Amaç ve Önem

Yurtdıřında uzun süredir yapılan BBT uygulamalarına iliřkin çalıřmalar, Türkiye'de sınırlı sayıdadır. Bu arařtırmanın amacı, BBT'nin önemli bir bileřeni olan testi sonlandırma kurallarının farklı deđerler almasının ölçme kesinliđi (RMSE, Bias ve uyum deđerleri) ve test uzunluđunu nasıl etkilediđini incelemektir. Arařtırma kapsamında ele alınan sonlandırma kurallarına ait kořulların, farklı madde havuzu büyüklüğü ve yetenek kestirim yöntemleri bađlamında karşılařtırılması hem ulusal hem de

uluslararası alan yazında katkı sađlayacađı düşünölmektedir. Arařtırmanın ele alınan deđiřkenleri ve kořulları bakımından Türkiye’ de ilk olması, arařtırmanın önemli bir boyutunu oluřturmaktadır.

Arařtırmanın amacı dođrultusunda ařađıdaki sorulara cevap aranmıřtır.

1. BBT uygulamalarında ölçme kesinliđi farklı sonlandırma kurallarına göre

a) Madde havuzu büyüklüğü 250 ve 500 madde seçildiđinde

b) Yetenek kestirim yöntemi Maksimum Olabilirlik Kestirimi (MLE) ve Beklenen Sonsal Dađılım (EAP) seçildiđinde

nasıl deđiřmektedir?

2. BBT uygulamalarında test uzunluđu farklı sonlandırma kurallarına göre

a) Madde havuzu büyüklüğü 250 ve 500 madde seçildiđinde

b) Yetenek kestirim yöntemi Maksimum Olabilirlik Kestirimi (MLE) ve Beklenen Sonsal Dađılım (EAP) seçildiđinde

nasıl deđiřmektedir?

YÖNTEM

Arařtırmanın Türü

Bu arařtırmada, BBT’lerde ölçme kesinliđi ve test uzunluđu, farklı testi sonlandırma kurallarına göre kontrollü kořullarda incelenmektedir. Kontrollü kořullar oluřturularak uygun verilerin türetilmesi bakımından arařtırma, bir simölasyon çalıřmasıdır. Arařtırmada simölasyon verileri ile yöntemleri etkileyecek deđiřkenler kontrol edilmekte ve yöntemlerin üstünlükleri/zayıflıkları deđerlendirilmektedir. Arařtırma bu yönüyle de yöntemlerin geliřtirilmesine katkı sađlayacađından temel arařtırma olarak kabul edilebilir (Karasar, 2004).

Verilerin Türetilmesi

Arařtırmada SimulCAT simölasyon veri üretme programından (Han, 2011) yararlanılmıřtır. Veri üretimi için öncelikle bireylere ait yetenek parametre deđerleri, ikinci ařamada ise madde parametre deđerleri elde edilmiřtir.

Testi alan bireylerin yetenek parametresi (θ deđeri) -3 ve +3 arasında tek biđimli dađılacak Őekilde oluŐturulmuŐtur. Cevaplayıcı sayısı 1000 olarak belirlenmiŐtir.

AraŐtırmada kullanılmak üzere oktan Őemeli maddelerden oluŐan iki farklı byklkte madde havuzu oluŐturulmuŐtur. AraŐtırma kapsamında kullanılan veri retim programı, 3 parametrelili lojistik model ile uyumlu veri retmektedir. Programda 250 ve 500 maddeden oluŐan iki farklı byklkteki havuzun parametreleri ayrı ayrı tretilmiŐtir. AraŐtırmada sadece madde havuzu byklđnn etkisi incelenmek istendiđinden her iki byklkteki havuz iin madde parametre aralıkları aynı belirlenmiŐtir. Her iki madde havuzundaki, maddelerin parametreleri tek biđimli olacak Őekilde; a parametresi [0,50;2,00]; b parametresi [-3,00;+3,00] ve c parametresi ise [0,05;0,20] aralıđında belirlenmiŐtir.

BBT KoŐulları

AraŐtırmada verilerin tretilmesinden sonra BBT koŐulları belirlenmiŐtir. AraŐtırma kapsamında madde Őeme yntemi Maksimum Fisher Bilgisi (Maximum Fisher Information -MFI) ve baŐlama kuralı ise $b=0$ olarak sabit ele alınmiŐtir. Yetenek kestirim yntemi MLE ve EAP yntemleri belirlenmiŐtir. Testi sonlandırma sabit uzunluk, yetenek dzeyinin standart hatası ve theta deđerinin yakınsaması kuralları kullanılmıŐtır. Belirlenen sonlandırma kurallarına iliŐkin Őeilen koŐullar aŐađıda verilmiŐtir.

1. Sabit uzunluk: Bu kural iin 15 ve 20 maddelik testler olmak üzere iki koŐul belirlenmiŐtir. Blaise ve Raiche (2002) alıŐmalarında madde sayısının sonlandırma kuralı olarak temel alındıđı alıŐmalarda en az 13 maddenin uygulanmasını nermiŐlerdir.

2. Yetenek dzeyinin standart hatası ($SE(\theta)$): Standart hata sonlandırma kuralı araŐtırmada iki farklı kurala gre ele alınmiŐtir.

a) $SE(\theta)$ deđerine gre sonlandırma: Bu kural iin, BBT'nin tahmin edilen gvenirlik deđeri belirli bir dzeye ulaŐtıđında sonlandırılmasına iliŐkin koŐullar belirlenmiŐtir. BBT uygulamasında kestirilen yetenek dzeyi standart hatası ve gvenirlik deđeri sonlandırma kuralı zerinde aynı etkiye sahiptir (Wang, Hanson ve Lau,1999). Blaise ve Raiche (2002) alıŐmalarında [-3,00;+3,00] aralıđında deđiŐen yetenek dzeylerinde standart hatanın 0,40' a eŐit veya daha dŐk olmasının lme kesinliđi iin uygun olduđunu nermiŐlerdir. Ayrıca; Babcock ve Weiss (2012) $SE(\theta)$ deđerlerinin belirlenmesinde gvenirliđin karesini (r^2) gz nne almıŐlardır.

Bu alıřmada $SE(\theta)$ deđerinin en az 0.40 olması olutü ve r^2 olutü birlikte ele alınmıřtır. Buna gre sırasıyla; $r^2 = 0,85$, $r^2 = 0,90$ ve $r^2 = 0,95$ deđerlerine karřılık gelen $SE(\theta)$ deđerleri 0,385, 0,315 ve 0,220 sonlandırma kuralı olarak alınmıřtır.

b) $SE(\theta)$ ve minimum madde sayısına gre sonlandırma: Minimum madde sayısı kısıtlaması diđer sonlandırma kuralları ile birlikte kullanılması, deđiřen uzunlukta sonlandırma iin uygulanan kullanıřlı bir zumdür (Babcock ve Weiss, 2012). Bu kuralda $SE(\theta)$ deđerinin belirlenen aralıđa dřmesi ve belirlenen minimum sayıda maddenin uygulanması gerekliliđi sz konusudur. Eđer $SE(\theta)$ deđer belirlenen deđerin altına dřse dahi belirlenen maddeden daha az uygulama sz konusu ise test sonlandırılmayacaktır. Bock ve Mislevy (1982) [-3,00;+3,00] aralıđında deđiřen yetenek dzeyinde en az 10 maddenin uygulanması durumunda yanlılıđın sıfıra yaklařacađını belirtmiřlerdir. Bu alıřmada $SE(\theta)$ deđerinin 0,385; 0,315 ve 0,220 deđerlerinin altına dřmesi ve en az madde sayısı 10 olarak belirlenmiřtir.

3. Yetenek kestirimi deđerinde mutlak deđiřim yakınsaması (θ deđerinde yakınsama): θ deđerinde yakınsama sonlandırma kuralı arařtırmada iki farklı kurala gre ele alınmıřtır.

a) θ deđerinde yakınsama deđerine gre sonlandırma: Yetenek kestirimleri arasındaki fark 0.05 ve 0.02 olarak belirlenmiřtir (Babcock ve Weiss, 2012).

b) θ deđerinde yakınsama ve minimum madde sayısına gre sonlandırma: Yetenek kestirimi (θ) deđerleri arasındaki fark 0,05 ve 0,02 ve en az madde sayısı 10 seilmiřtir. Bu kuralda; (θ) deđerleri arasındaki farkın belirlenen aralıđa dřmesi ve belirlenen minimum sayıda maddenin uygulanması gerekliliđi sz konusudur. Eđer (θ) deđerleri arasındaki fark; belirlenen deđerin altına dřse dahi 10 maddeden daha az uygulama sz konusu ise test sonlandırılmayacaktır.

Arařtırmada ele alınacak beř ayrı sonlandırma kuralına gre oluřturulan 12 sonlandırma kořulu Tablo 2.1.'de yer almaktadır.

Tablo 2.1. Arařtırma Kapsamında Kullanılan Sonlandırma Kořulları

Kořullar	Sonlandırma Kořulu
Kořul 1	Sabit uzunluk-15 madde
Kořul 2	Sabit uzunluk-20 madde
Kořul 3	SE(θ): 0,385
Kořul 4	SE(θ): 0,315
Kořul 5	SE(θ): 0,220
Kořul 6	SE(θ): 0,385 ve en az 10 madde
Kořul 7	SE(θ): 0,315 ve en az 10 madde
Kořul 8	SE(θ): 0,220 ve en az 10 madde
Kořul 9	(θ) deđeri farkı: 0,05
Kořul 10	(θ) deđeri farkı: 0,02
Kořul 11	(θ) deđeri farkı: 0,05 ve en az 10 madde
Kořul 12	(θ) deđeri farkı: 0,02 ve en az 10 madde

Arařtırma kapsamında BBT uygulaması iin belirlenen deđiřkenler ve deđerleri Tablo 2.2’de toplu olarak sunulmuřtur.

Tablo 2.2. Arařtırmada Kapsamında Ele Alınan Deđiřkenler ve Kořulları

Deđiřkenler	Kořullar
Yetenek kestirimi	ML, EAP
Madde havuzu	250, 500 madde
Bařlama kuralı	b=0
Madde seim yntemi	MFI
Sonlandırma kuralı	
1. Sabit uzunluk	15, 20
2. Standart hata ve r^2	0.385, 0.315 ve 0.222
3. Standart hata ve minimum madde sayısı	Se: 0.385, 0.315, 0.222 Madde sayısı: 10
4. Yetenek kestirimlerinin yakınsaması	0.05 ve 0.02
5. Yetenek kestirimlerinin yakınsaması ve minimum madde sayısı	0.05 ve 0.02 Madde sayısı: 10

Tablo 2.2.’de yer alan deđiřkenler ve kořulları aprazlandıđında her alt problem iin 48 kořul belirlenmektedir. Arařtırmanın genelinde 48 kořul iin program alıřtırılmıřtır. Harwell (1996) rneklem yanlılıđının ortadan kaldırılması iin en az 25 tekrarın kullanılmasını nermektedir (akt: Evans, 2010). Arařtırmada her kořul iin 25 tekrar kullanılmıřtır.

Verilerin Analizi

Arařtırmada her bir kořul iin lme kesinliđi ve test uzunluđu deđerleri hesaplanmış ve birbirleri ile karřılařtırılmıřtır.

lme Kesinliđi: lme kesinliđi, lmenin tutarlıđı ile ilgilidir. lme kesinliđinin artması durumunda hata deđerleri daha dūřuk deđer olacaktır. Arařtırma kapsamında; lme kesinliđinin belirlenmesi iin, hata gstergeleri olan RMSE, yanlılık (bias) ve uyum (fidelity) deđerleri hesaplanmıştır.

- i. RMSE (Root Mean Squared Error): BBT uygulaması ile elde edilen bireye ait tahmin edilen yetenek dūzeyi ile gerek yetenek dūzeyi arasındaki mutlak farklılıđa iliřkin istatistiktir. RMSE deđeri ařađıdaki eřitlik yardımıyla hesaplanır.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}}$$

Burada; n: toplam birey sayısı, θ_i : i.bireyin gerek yetenek dūzeyi deđeri; $\hat{\theta}_i$: i.bireyin kestirilen yetenek dūzeyi deđerini gstermektedir.

- ii. Yanlılık (Bias): BBT uygulaması ile elde edilen bireye ait tahmin edilen yetenek dūzeyi ile gerek yetenek dūzeyi arasındaki ortalama anlamlı farklılık istatistiđidir. Yanlılık deđeri

$$BIAS = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)}{n}$$

ile hesaplanır. RMSE deđeri ile Bias deđeri arasındaki iliřkiyi;

$$RMSE^2 = Bias^2 + SE^2$$

gstermek mmkündür. Burada RMSE deđerinin, yanlılık ve tahminin standart hatasını ierdiđi grlmektedir.

- iii. Uyum (Fidelity): Uyum katsayısı bireylerin gerek yetenek dūzeyi ile tahmin edilen yetenek dūzeyi arasındaki Pearson korelasyon olarak tanımlanır (Weiss,1982). Bireye ait tahmin edilen yetenek dūzeyi ile gerek yetenek dūzeyi arasındaki korelasyon katsayısı;

$$r = \frac{cov(\hat{\theta}, \theta)}{SS(\hat{\theta})SS(\theta)}$$

ile hesaplanır. Bu deđerin yüksek elde edilmesi; tahmin edilen yetenek düzeyi ($\hat{\theta}_i$) ile gerçek yetenek düzeyi (θ_i) arasındaki yüksek uyumu belirtmektedir.

iv. Test Uzunluđu: Çalıřma kapsamında, her bir kořul için bireylerin aldıkları maddelerin ortalama deđeridir. Bu deđiřken sabit uzunluk sonlandırma kuralı için hesaplanmamıřtır.

BULGULAR ve YORUM

BBT uygulamalarında 12 farklı sonlandırma kořulları kullanıldıđında, ölçme kesinliđini belirlemeye yönelik hesaplanan RMSE, yanlılık (bias) ve uyum (fidelity) deđerleri Tablo 3.1’de sunulmuřtur. Yorumlar, madde havuzu büyüklüđu ve yetenek kestirim yöntemine göre çaprazlanan dört farklı duruma göre verilmiřtir. Ölçme kesinliđini yorumlamada tüm hata göstergeleri birlikte deđerlendirilmiřtir.

Tablo 3.1. Her Bir Kořula Ait Elde Edilen Ortalama RMSE, Yanlılık ve Uyum Deđerleri

Durumlar	RMSE				Yanlılık (Bias)				Uyum (Fidelity)			
	250madde MLE	500madde MLE	250madde EAP	500madde EAP	250madde MLE	500madde MLE	250madde EAP	500madde EAP	250madde MLE	500madde MLE	250madde EAP	500madde EAP
15	0,302	0,273	0,269	0,247	0,026	0,021	0,010	0,004	0,984	0,987	0,988	0,990
20	0,237	0,217	0,227	0,208	0,015	0,013	0,007	0,002	0,990	0,992	0,991	0,993
0,385	0,546	0,516	0,463	0,466	0,097	0,079	0,016	0,014	0,949	0,954	0,965	0,964
0,315	0,386	0,382	0,363	0,363	0,044	0,042	0,009	0,009	0,974	0,974	0,978	0,978
0,220	0,222	0,227	0,229	0,235	0,010	0,012	0,002	0,002	0,991	0,991	0,991	0,991
0,385-en az 10 madde	0,397	0,359	0,348	0,324	0,052	0,041	0,013	0,004	0,972	0,977	0,980	0,983
0,315-en az 10 madde	0,356	0,341	0,337	0,313	0,040	0,032	0,011	0,005	0,978	0,979	0,981	0,984
0,220-en az 10 madde	0,222	0,227	0,229	0,235	0,010	0,012	0,002	0,002	0,991	0,991	0,991	0,991
0,05	0,305	0,293	0,250	0,228	0,009	0,010	0,006	0,003	0,984	0,985	0,989	0,991
0,02	0,266	0,262	0,189	0,168	0,006	0,007	0,003	0,001	0,988	0,988	0,994	0,995
0,05-en az 10 madde	0,246	0,228	0,245	0,227	0,012	0,009	0,007	0,003	0,989	0,991	0,990	0,991
0,02-en az 10 madde	0,190	0,170	0,189	0,168	0,002	0,005	0,003	0,000	0,994	0,995	0,994	0,995

Sabit uzunluk sonlandırma kuralına ait 15 ve 20 maddelik kořullar ölçme kesinliđi bakımından karşılaştırıldıđında madde havuzu ve yetenek kestirim yönteminin deđiřtirildiđi durumlarda, 20 maddelik testte daha düşük RMSE ve yanlılık deđerleri ile daha yüksek uyum elde edilmiřtir. Dolayısıyla 20 maddelik test 15 maddelik teste kıyasla ölçme etkililiđi daha yüksektir. Standart hata sonlandırma kuralına ait 0,385 SE(theta), 0,315 SE(theta) ve 0,220 SE(theta) sonlandırma kořulları tüm farklı durumlarda benzer şekilde sonuç üretmiřtir. Ölçme kesinliđi en yüksek 0,220 SE(theta); en düşük ise 0,385 SE(theta) kořulunda elde edilmiřtir. Standart hata kuralının minimum madde sayısı ile kullanılması durumunda da ölçme kesinliđinin etkililiđi benzerdir. Theta düzeyinde yakınsama sonlandırma kuralında 0,02 ve 0,05 yakınsama kořulları karşılaştırıldıđında, tüm durumlarda 0,02 theta yakınsama kořulu ile ölçme kesinliđi daha yüksek çıkmıřtır. Theta yakınsama kuralının minimum madde sayısı ile kullanılması durumunda da 0,02 theta yakınsama ve minimum 10 madde kořulunda daha düşük RMSE ve yanlılık (uyum deđerleri) elde edilerek ölçme kesinliđi daha yüksek bulunmuřtur.

Dört farklı durumun her birinde, sabit sonlandırma, standart hata ve theta yakınsama sonlandırma kuralları kıyaslandıđında; 0,02 theta yakınsama kořulunun tüm kořullara kıyasla en düşük RMSE ve yanlılık deđerini alırken en yüksek uyum deđerine sahip olduđu belirlenmiřtir. Ayrıca 0.02 theta yakınsama kořulunun minimum 10 madde kořulu ile birlikte kullanılmasında çođu durumda tek başına kullanılmasına kıyasla ölçme kesinliđinde daha iyi sonuç elde edilmiřtir. Standart hata sonlandırma kuralının 0,220 SE(theta) kořulu ile sabit uzunluk 15 madde kořulu ile kıyaslandıđında tüm durumlarda; 20 madde kořulu ile kıyaslandıđında ise çođu durumlarda daha düşük RMSE ve yanlılık deđerleri ve daha yüksek uyum deđerleri elde edilmiřtir. Alan yazında da deđiřen uzunluklu sonlandırma kurallarının sabit uzunluk sonlandırma kuralına kıyasla daha güçlü ölçme sonuçları sunduđuna dair arařtırmalar söz konusudur (Weiss, 1982; Babcock ve Weiss, 2012)

Sabit uzunluk sonlandırma kuralı kořullarının tüm durumlarda ölçme kesinliđi, 0,385 SE(theta) ve 0,315 SE(theta) kořullarına kıyasla daha yüksek elde edilmiřtir. Ayrıca sabit uzunluk sonlandırma kuralının 20 madde kořulunun kullanıldıđı test durumunda, theta yakınsama sonlandırma kuralının 0,05 kořulu ile karşılaştırıldıđında da ölçme etkililiđi daha yüksek elde edilmiřtir. Alan yazında da benzer sonuçları veren çalıřmalar söz konusudur (Stocking,1987; Yi, Wang ve Ban, 2001).

En az 10 madde kořulunun standart hata ve theta yakınsama sonlandırma kuralları ile birlikte uygulanmasında özellikle ortalama 10 maddenin ařađısında testin sonlandırıldıđı kořullara olçme etkililiđi bakımından avantaj sađladıđı grlmektedir. Arařtırmada standart hata ile minimum madde sayısının kombinasyonu olçme kesinliđi aısından standart hatanın tek bařına kullanıldıđı duruma gre avantaj sađlamıřtır.

Deđiřen uzunluk sonlandırma kuralları iin BBT uygulamaların drt farklı duruma iliřkin elde edilen test uzunluđu deđerleri Tablo 3.2’de sunulmuřtur.

Tablo 3.2. Her Bir Kořula Ait Elde Edilen Ortalama Test Uzunluđu Deđerleri

Kořullar	SE (theta)			SE (theta)- en az 10 madde			Theta yakınsama		Theta yakınsama- en az 10 madde	
	0,385	0,315	0,220	0,385	0,315	0,220	0,05	0,02	0,05	0,02
250 madde MLE	7,0	9,6	19,4	10,2	10,6	19,4	15,9	28,2	16,6	29,4
500 madde MLE	6,8	8,9	16,3	10,1	10,4	16,3	15,8	28,6	16,6	29,9
250 madde EAP	6,9	9,3	19,1	10,1	10,5	19,1	16,9	30,0	16,9	30,0
500 madde EAP	6,2	8,4	15,8	10,1	10,3	15,8	16,7	30,5	16,8	30,5

Tm durumlarda sonlandırma kořullarında en dřk test uzunluđunun 0,385 SE(theta) sonlandırma kořulunun kullanılmasında elde edildiđi grlmektedir. En yksek test uzunluđuna ise 0,02 theta yakınsama kořulunda rastlanmaktadır. Yetenek kestirim ynteminin MLE seilmesi durumunda ise en yksek test uzunluđu 0,02 theta yakınsama ve minimum 10 madde kořulunda elde edilmektedir.

Madde havuzu byklđnn deđiřtirilmesi durumunda sonlandırma kurallarının karřılařtırılmasında, Standart hata sonlandırma kurallarında madde havuzu byklđ arttıka ortalama test uzunluđunun da genel olarak azaldıđı gzlenmiřtir. Özellikle farklılık 0,220 SE(theta) kořulunda daha fazladır. Farklı olarak theta yakınsama kořulunda 0,02 deđerinin seilmesi durumunda ise ortalama test uzunluđu artmaktadır.

Yetenek kestirim yntemlerinin etkisinin incelenmesinde, standart hata sonlandırma kurallarında yetenek kestirim yntemi olarak EAP seildiđinde ortalama test uzunluđunun genel olarak azaldıđı gzlenmiřtir. Özellikle farklılık 0,385 SE(theta) ve 0,315 SE(theta) kořullarının tek

bařına kullanılması durumunda en az 10 madde kořulu ile birlikte kullanılmasına kıyasla daha fazladır. Farklı olarak theta yakınsama kořulunda ise yetenek kestirim yónteminin EAP seğıilmesi durumunda ortalama test uzunluđu artmaktadır.

SONUÇ ve ÖNERİLER

Bu arařtırmada beř farklı sonlandırma kuralına göre oluřturulan 12 sonlandırma kořulu, 250 ve 500 maddeden oluřan iki farklı büyüklükteki madde havuzu ile MLE ve EAP yetenek kestirim yóntemlerine göre incelenmiřtir. Sonular, bu deđiřkenler göz önüne alınarak sonlandırma kořulları çerevesinde verilmiřtir.

Madde Havuzu Büyüklüđünün 250 veya 500 Madde Olarak Belirlenmesi

Yetenek kestirimi MLE seğıildiđinde; belirlenen sonlandırma kurallarının çođunluđunda 500 maddelik havuzda RMSE ve yanlılık deđerleri daha düşük elde edilmiřtir. Her iki havuz büyüklüđünde her bir sonlandırma kuralı için uygulanan madde sayılarına bakıldıđında, çođunluđunda çok önemli bir deđiřim olmadıđı görülmüřtür. Ancak 0,220 SE(theta) ve 0,220 SE(theta)-en az 10 madde kořullarının her ikisinde de RMSE ve yanlılık deđerlerinde, 500 maddelik havuzda ortalama 3,1 madde daha az uygulanmasına rađmen çok önemli bir deđiřim olmadıđı belirlenmiřtir.

Yetenek kestirimi EAP seğıildiđinde; belirlenen sonlandırma kurallarının çođunluđunda 500 maddelik havuzda RMSE deđerleri daha düşük elde edilmiřtir. Yanlılık deđerlerine bakıldıđında ise tüm sonlandırma kořullarında 500 maddelik havuzunun lehinedir. Her iki havuz büyüklüđünde her bir sonlandırma kuralı için uygulanan madde sayılarına bakıldıđında, çođunluđunda çok önemli deđiřim olmadıđı görülmüřtür. Geçerli kořullarda; göze arpan en önemli nokta ise 0,220 SE(theta) ve 0,220 SE(theta)-en az 10 madde kořullarında, 500 maddelik havuzda ortalama 3,3 madde daha az uygulanmasına rađmen RMSE ve yanlılık deđerlerinde çok önemli bir deđiřim olmadıđıdır.

Arařtırmada, madde havuzu büyüklüđu ile RMSE deđerleri arasında ters yönlü bir iliřki söz konusudur. Buna dayanarak, madde havuzu büyüklüđu arttıka, ölçmenin kesinliđinin de arttıđı söylenebilir. Choi, Graddy ve Dodd (2011) alıřmalarında iki farklı özelliđe sahip, iki farklı büyüklükte madde havuzlarında üç farklı sonlandırma kuralını oklu

puanlanan maddeler iin karřılařtırmıřlardır. Aynı zelliđe sahip maddelerden oluřan havuzun byklđ arttıķa RMSE deđeri sonlandırma kurallarının hepsinde azalma gstermektedir. Ayrıca Ivie (2007); madde havuzu geniřliđinin arttıķa daha yksek ayırt edicilik ve daha dřk řans parametrelerinin, daha kararlı yetenek dzeyi tahmini sađladıđını belirtmiřtir.

Yetenek Kestirim Yntemi Olarak MLE veya EAP Belirlenmesi

Madde havuzu byklđ 250 seildiđinde; RMSE deđerleri, 0,220 SE(theta) (tek bařına ve en az 10 madde ile kullanımı) dıřındaki tm kořullar iin EAP yetenek kestirim ynteminin kullanıldıđı durumda daha dřk elde edilmiřtir. Yanlılık deđeri ise tm kořullarda EAP yetenek kestirim ynteminin kullanıldıđı durumda daha dřk elde edilmiřtir. Test uzunlukları aısından da ođu sonlandırma kořullarında ok nemli farklılıklar gzlemlenmemiřtir. 0,02 theta yakınsama sonlandırma kořulunun kullanımında EAP kestirim ynteminde zellikle RMSE deđeri, MLE yntemine kıyasla olduka dřktr.

Madde havuzu byklđ 500 seildiđinde; EAP kestirim yntemi ile elde edilen RMSE ve yanlılık deđerleri sonlandırma kořullarının ođunluđunda daha dřk elde edilmiřtir. Test uzunlukları aısından da ok farklılık gzlemlenmemiřtir. EAP ve MLE yntemi ile elde edilen RMSE deđerleri arasındaki fark en yksek deđerini 0,02 theta yakınsama sonlandırma kořulunda almıřtır.

Arařtırmada, EAP kestirim ynteminde MLE'ye kıyasla daha dřk RMSE ve yanlılık deđerleri elde edilmiřtir. Alan yazında da MLE ve Bayes yntemlerinin RMSE ve yanlılık deđerlerinin karřılařtırılması ile ilgili, bu arařtırmanın sonucunu destekleyecek alıřmalar sz konusudur. Wang ve Vispoel (1998) BBT'de simlasyon alıřması yaparak MLE ve Bayes yntemlerini (Oven's method, EAP ve MAP) karřılařtırmıřlardır. MLE yntemi ile daha dřk yanlılık, daha yksek RMSE deđerleri elde edilmiřtir. İřeri (2002) alıřmasında Bayes yntemlerinin lme kesinliđi aısından daha iyi kestirim sađladıđını belirtmiřtir.

Arařtırmanın sonularından yola ıkarak ařađıdaki oneriler sunulabilir.

1. Madde havuzu byklđ 250 ve yetenek kestirim yntemi MLE seildiđinde, hem RMSE ve yanlılık hem de testin kullanıřlılıđı bakımından yeteneđin tahmininde 0,220 SE(theta) ve 0,220 SE(theta)–en az 10 madde sonlandırma kořulları onerilebilir.

2. Madde havuzu byklđ 500 ve yetenek kestirim yntemi MLE seildiđinde, hem RMSE ve yanlılık hem de testin kullanıřlılıđı bakımından 0,220 SE(theta) ve 0,05 theta yakınsama-en az 10 madde sonlandırma kořullarının kullanılması nerilebilir.
3. Madde havuzu byklđ 250 ve yetenek kestirim yntemi EAP seildiđinde, hem RMSE ve yanlılık hem de testin kullanıřlılıđı bakımından 0,220 SE(theta) ve 0,05 theta yakınsama-en az 10 madde sonlandırma kořullarının kullanılması nerilebilir.
4. Madde havuzu byklđ 500 ve yetenek kestirim yntemi EAP seildiđinde; hem RMSE ve yanlılık hem de testin kullanıřlılıđı bakımından 0,220 SE(theta) sonlandırma kořulunun kullanılması nerilebilir.
5. RMSE ve yanlılık deđerlerinin dřk elde edilmesi; bir bařka deyiřle, yksek lme kesinliđi iin byk madde havuzları tercih edilebilir.
6. Hem lme kesinliđi hem de testin ekonomik olması aısından deđiřen uzunluk sonlandırma kurallarından 0,220 SE(theta) byk madde havuzlarında tercih edilebilir.
7. Yetenek kestirim yntemi olarak EAP kullanılması MLE'ye kıyasla daha kullanıřlıdır.
8. zellikle theta yakınsama kuralında yetenek kestirim ynteminin EAP seilmesi daha dřk RMSE ve yanlılık aısından tercih edilebilir.

Benzer alıřmalar iin ařađıdaki neriler sunulabilir.

1. Sonlandırma kořulları ok boyutlu testler iin karřılařtırılabilir.
2. BBT uygulamalarında nemli bir yere sahip olan madde kullanım sıklıđı ve ierik dengesinin incelenmesi bu arařtırma kapsamı dıřında tutulmuřtur. Gelecek arařtırmalarda bu iki durum sonlandırma kořulları ile birlikte alıřılabilir.
3. Minimum Bilgi Sonlandırma kurallarından elde edilen deđerler de diđer kurallardan elde edilenlerle karřılařtırılabilir.
4. Theta yakınsama kuralının standart hata sonlandırma kuralı ile birlikte kullanılmasının sonulara etkisi incelenebilir.
5. Madde havuzunun zellikleri deđiřtirilerek sonlandırma kuralları birbiri ile karřılařtırılabilir.
6. Diđer yetenek kestirim yntemleri kullanılmasının sonlandırma kurallarından elde edilen sonulara etkisi incelenebilir.

7. Arařtırmada madde seęme yntemlerinden Maksimum Fisher Bilgi yntemi kullanılmıřtır. Diđer madde seęme yntemleri de benzer kořullarda karřılařtırılabilir.

KAYNAKęA

- Babcock, B. and Weiss, D.J., 2012. Termination criteria in Computerized Adaptive Tests: do variable-length CAT's provide efficient and effective measurement? *International Association for Computerized Adaptive Testing*, 1, 1-18.
- Blais, J. and Raiche, G., 2002. *Features of the sampling distribution of the ability estimate in Computerized Adaptive Testing according to two stopping rules*, International Objective Measurement Workshop, New Orleans, April 2002.
- Choi, S. W., Grady, M.W. and Dodd, B.G., 2011. A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71, 37-53.
- ęıkırıęı, Demirtařlı, N., 1999. Psikometride yeni ufuklar: bilgisayar ortamında bireye uyarlanmıř test. *Trk Psikoloji Blteni*. 5(13), 31-36.
- Dodd, B.G., Koch, W.R. and de Ayala, R.J., 1993. Computerized Adaptive Testing using the partial credit model effects of item pool characteristics and different stopping rules. *Educational and Psychological Measurement*, 53, 61-77.
- Embretson, E. and Reise, S. P., 2000. *Item response theory for psychologist principles and application*. London: Lawrence Erlbaum Assc.
- Eggen,T., 2004. *Contributions to the theory and practice of Computerized Adaptive Testing*. Druk: Print Partners Ipskamp B.V., Enschede.
- Evans, J. J., 2010. *Comparability of examinee proficiency scores on Computer Adaptive Tests using real and simulated data*. Doctoral Dissertation. The State University of New Jersey.
- Flaughner, R., 2000. Item pools. In H. Wainer (Eds.), *Computerized Adaptive Testing*, 37-59. London: Lawrence Erlbaum Assc.
- Hambleton, R. K. and Swaminathan, H., 1985. *Item response theory. principles and application*. Boston: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H. and Rogers,H. J., 1991. *Fundamentals of item response theory*. California: Sage Publications.
- Han, K. C., 2011. *User's Manual: SimulCAT*. Graduate Management Admission Council.

- Ivei, J. L., 2007. *Test taking strategies in Computer Adaptive Testing That will improve your score: fact or fiction?* Doctoral Dissertation. University of Michigan.
- İseri, A. I., 2002. *Assessment of students' mathematics achievement through Computer Adaptive Testing procedures.* Middle East Technical University.
- Kalender, İ., 2004. *Bilgisayar ortamında Bireyselleştirilmiş Testlerin eğitimde kullanımı. XIII. Ulusal Eğitim Bilimleri Kurultayı. İnönü Üniversitesi, Eğitim Fakóltesi, Malatya. 6-9 Temmuz 2004.*
- Kalender, İ., 2011. *Effects of different Computerized Adaptive Testing strategies on recovery of ability.* Doctoral Disertation. Middle East Technical University.
- Kaptan, F., 1993. *Yetenek kestiriminde Adaptive (Bireyselleştirilmiş) Test uygulaması ile geleneksel kâğıt-kalem testi uygulamasının karşılaştırılması.* Doktora Tezi. Hacettepe Üniversitesi.
- Karasar, N., 2004. *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler* (13. Baskı). Ankara: Nobel Yayın Dağıtım.
- Köklü, N., 1990. *Klasik test teorisine göre geliştirilen tailored test ile grup testi arasında bir karşılaştırma.* Doktora Tezi. Hacettepe Üniversitesi.
- Linden, W. J and Glas, G. A. W., 2002. *Computerized Adaptive Testing: theory and practice.* USA: Kluwer Academic Publishers.
- Linecra, J. M., 2000. *Computer-Adaptive Testing: a methodology whose time has come,* [Çevrim-içi: <http://www.rasch.org/memo69.htm>], erişim tarihi: 02 Mayıs 2012.
- McBride, J. R. and Martin, J. T., 1983. Reliability and validity of adaptive ability tests in a militarysetting. In D. J. Weiss (Eds.), *New horizons in testing: Latent trait theory and computerized adaptive testing*, 223–226). New York: Academic Press.
- McBride, J. R., Wetzel,C.D. and Hetter, R. D., 2001. Preliminary psychometric research for CAT-ASVAB: selecting an adaptive testing strategy. In W. Sands, B. K. Waters, and J.R.McBride (Eds.). *Computerized Adaptive Testing: from inquiry to operation*, 83–95. Washington, DC: American Psychological Association.
- Mead, A. D. and Drasgow, F., 1993. Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Meijer, R. R. & Nerring, M. L., 2001. New development in the area of Computerized Testing. *Psychologie Francaise*, 46(3), 221-230.
- Mills, C. N. and Stocking, M.L., 1996. Practical issues in large-scale Computerized Adaptive Testing. *Applied Mesurement in Education*, 9(4), 287-304.

- Pearson Assessment., 2012. [Çevrim-içi: www.pearsonassessment.com] Eriřim tarihi: 02.05.12.
- Riley, B. B., Conrad, K. J., Bezruczko, N. and Dennis, M., 2007. Relative precision, efficiency and construct validity of different starting and stopping rules for a Computerized Adaptive Test: the GAIN substance problem scale. *Journal of Applied Measurement*, 8(1).
- Rudner, L. M., 1998. *Interactive Computer Adaptive Testing*, [Çevrim-içi: <http://EdRes.org/scripts/cat>], Eriřim Tarihi: 02.05.12.
- Samejima, F., 1977. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 42(2), 163-191.
- Segall, D. O., 2004. Computerized Adaptive Testing. In Kempf-Leanard (Eds.). *The Encyclopedia of Social Measurement*, 429-438. San Diego, CA: Academic Press.
- Simms, L. J. and Clark, L. A., 2005. Validation of a computerized adaptive version of the schedule for non-adaptive and adaptive personality (SNAP). *Psychological Assessment*, 17, 28-43.
- Stocking, M. L., 1987. Two Simulated feasibility studies in Computerized Adaptive Testing. *Applied Psychology: An International Review*, 36, 263-267.
- Thissen, D. and Mislevy, R. J., 2000. Testing algorithms. In H. Wainer (Eds.). *Computerized Adaptive Testing*, 101-135. London: Lawrence Erlbaum Assc.
- Wainer, H., 2000. *Computerized Adaptive Testing*. London: Lawrence Erlbaum Assc.
- Wainer, H. and Mislevy, R. J., 2000. Item response theory, item calibration and proficiency estimation. In H. Wainer (Eds.). *Computerized Adaptive Testing*. 65-102. London: Lawrence Erlbaum Assc.
- Wang, T., Hanson, B. A. and Lau, C., 1999. Reducing bias in CAT ability estimation: a comparison of approaches. *Applied Psychological Measurement*, 23, 263-278.
- Wang, T. and Vispoel, W. P., 1998. Properties of ability estimation methods Computerized Adaptive Testing. *Journal of Educational Measurement*, 35 (2), 109-135.
- Wang, S. and Wang, T., 2001. Precision of warm's weighted likelihood estimates for a polytomous model in Computerized Adaptive Testing. *Applied Psychological Measurement*, 25(4), 317-331.
- Weiss, D. J., 1982. Improving measurement quality and efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6,473-492.
- Weiss, D. J., 1983. *New horizons in testing*. New York: Academic Press.

- Weiss, D. J., 2004. Computerized Adaptive Testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.
- Weiss, D. J. and Kingsbury, G. G. 1984. Application of Computerized Testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Yi, Q., Wang, T. and Ban, J. C., 2001. Effects of scale transformation and test termination rule on the precision of ability estimation in Computerized Adaptive Testing. *Journal of Educational Measurement*, 38, 267-292.
- Yoo, H., 2011. *Evaluating several Multidimensional Adaptive Testing procedures for diagnostic assessment*. Doctoral Dissertation. University of Massachusetts.

Başvuru: 08.07.2014

Yayına Kabul: 11.02.2015