

Original article

İngilizce Yeterlik Sınavı İçinde Seviye Tespit Sınavı: Yeni Bir Yaklaşım ve Sonuçları

Placement Exam in English Proficiency Exam: A New Approach and Its Results

Alper Şahin *

Department of Basic English, School of Foreign Languages, Atılım University, Ankara, Turkey

Özet

Kovid-19 pandemisi eğitim kurumlarının işleyişlerinde ve rutin uygulamalarında önemli değişikliklere sebep olmuştur. Çünkü pandemi süreci öncelikleri ve ihtiyaçları değiştirmiştir. Bu değişiklik üniversitelerin ölçme ve değerlendirme uygulamalarına da yansımıştır. Bir çok sınav uzaktan çevrimiçi olarak yapılmıştır. Ancak amaçları açısından iki farklı sınav olan İngilizce Yeterlik Sınavı ve Seviye Tespit Sınavları "önemli sonuçları olan sınavlar" (high-stakes exams) olduğundan dolayı bazı üniversitelerde gerekli hijyen, maske ve mesafe önlemleri alınarak mecburen tek bir sınav şeklinde uygulanmak zorunda kalmıştır. Bu araştırmada böyle bir mecburiyet sonucu İngilizce Yeterlik ve Seviye Tespit Sınavlarının tek sınav halinde yapılması için psikometrik olarak akla yatkın bir yaklaşım geliştirilerek teorik bir çerçeveye oturtulmuştur. Bu yaklaşımla hazırlanan sınav 965 öğrenciye uygulanmıştır. Uygulama sonucunda bu yeni yaklaşımın madde parametrelerine ve test/alt test bilgi fonksiyonlarına etkisi incelenmiştir. Elde edilen bulgular kullanılan yaklaşımın önemli ölçüde başarıya ulaştığı ve madde parametrelerinde ve test/alt test bilgi fonksiyonlarında istendik değişikliklere sebep olduğunu göstermiştir.

Anahtar Kelimeler: İngilizce yeterlik sınavı, İngilizce seviye tespit sınavı, İngilizce Dil sınavları, Kovid-19.

Abstract

The Covid-19 pandemic has caused significant changes in the functioning and routine practices of educational institutions. Because of the pandemic, the priorities and the needs have changed. This change was also reflected in the assessment and evaluation practices of the universities. Many exams have been conducted online. However, as the English Proficiency Exam and the Placement Exam, which are two different exams for their purposes, are "high-stakes exams", some universities had to administer them as a single face-to-face exam by taking the necessary hygiene, mask, and distance measures. In this study, as a result of such a necessity, a psychometrically sound approach was developed for the English Proficiency and Placement Exams to be held as a single exam and it was put into a theoretical framework. The single exam prepared was administered to 965 students. Then, the effects of the new approach on the item parameters and test/subtest information functions were examined. The findings obtained showed that the approach developed was highly successful and caused desired changes in the item parameters and test/subtest information functions.

Keywords: English proficiency exam, English placement exam, English language tests, Covid-19.

Received: 19 June 2021 * Accepted: 27 June 2021 * DOI: <https://doi.org/10.29329/dmer.2021.285.4>

* Corresponding author:

Alper Şahin is an assistant professor in the Department of Basic English at Atılım University in Ankara, Turkey. His research interests include computerized adaptive tests, test anxiety in language exams, performance assessment in English Language Teaching, rater performance assessment, and item response theory in English Language Teaching.
Email: alpersahin2@yahoo.com

GİRİŞ

Üniversitelerin İngilizce öğretim yapan bölümlerini kazanan öğrenciler İngilizce düzeylerinin bu bölümlerin derslerini takip edebilecek düzeyde olup olmadığının tespiti için üniversite rektörlükleri tarafından uygulanan İngilizce Yeterlik Sınavı'na tabi tutulurlar (Yükseköğretim Kurumlarında Yabancı Dil Öğretimi ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik, 2013). Bu İngilizce Yeterlik Sınavı'ndan üniversite senatolarınca belirlenen düzeyde puanı alamayan öğrenciler en az bir, en çok dört yarıyıl olmak üzere İngilizce Hazırlık Programı'nda eğitim görerek İngilizcelelerini geliştirirler. İngilizce Hazırlık Programı sonunda tekrar İngilizce Yeterlik Sınavı yapılır. Öğrencilerin bu sınavda başarılı olmaları beklenmektedir. Bu sınavda başarılı olan öğrenciler bölümlerine devam etmeye hak kazanırlar. Kendilerine tanınan süre içerisinde İngilizce Yeterlik Sınavı'nı geçemeyen öğrenciler üniversiteye giriş puanının ilgili bölümün taban puanından az olmaması koşulu ile öğretim dili Türkçe olan kayıt oldukları programa eşdeğer bir programa yerleştirilebilmektedirler (Yükseköğretim Kurumlarında Yabancı Dil Öğretimi ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik, 2013).

Üniversiteye kayıt sonrasında yapılan İngilizce Yeterlik Sınavı'nda başarılı olamayıp İngilizce Hazırlık Programı'na kayıt yaptıran öğrenciler İngilizce seviyelerinin tespiti için bir seviye tespit sınavına tabi tutulmaktadır. Bazı üniversiteler önce seviye tespit sınavını yapıp bu sınavdan belli bir notun üstünde alan öğrencileri İngilizce Yeterlik Sınavına da alabilmektedir. Seviye Tespit Sınavı adından da anlaşılacağı üzere öğrencilerin İngilizce seviyelerini tespit ederek İngilizce Hazırlık Programı'na doğru seviyeden başlamasını sağlama amacıyla yapılmaktadır. Bu sınavda öğrencilerin doğru seviyeye yerleştirilmesi kritik öneme sahiptir. Çünkü İngilizce düzeyinden daha yukarıda bir seviyeye yerleşen bir öğrenci o seviye sınıflarında derse gireceğinden dolayı dersi takip etmekte güçlük çekebilir ve o seviyenin derslerini tamamlayamayabilir. Aynı şekilde İngilizce seviyesinden daha düşük bir seviyeye yerleşen bir öğrenci o seviyenin derslerinde öğrenme motivasyonunu kaybedebilir (Manning, 2019b).

Görüldüğü üzere İngilizce Yeterlik Sınavı ve İngilizce Seviye Tespit Sınavları farklı amaçlara hitap eden iki farklı sınavdır. Bu durum sınavların madde seçimlerine de yansımaktadır. Şöyle ki İngilizce Yeterlik Sınavı herhangi bir eğitime bağlı olmadan öğrencilerin istedik bir İngilizce seviyesinin üzerinden olup olmadıklarını belirlemek amacıyla yapıldıkları (Alderson, Clapham ve Wall, 1995) için sadece kurumca belirlenen İngilizce düzeyinin üstündeki öğrencilerin doğru cevaplayabilecekleri maddeleri içermesi beklenirken, İngilizce Seviye Tespit Sınavı'nın ise İngilizce Hazırlık Programı'nda doğru seviye grubunda ders almasını amaçlamasından ötürü (Alderson, Clapham ve Wall, 1995) sadece belirli bir düzeyin üstündeki İngilizce becerilerini değil daha geniş bir aralıkta İngilizce becerilerini ölçen maddeler içermesi beklenmektedir. Bu nedenle bu iki sınavın tek sınavda birleştirilmesi beklenen bir durum değildir ve sorunlara sebep olabilir. Bu iki sınavın tek sınav şeklinde

yapılması durumunda psikometrik olarak akılcı ve mantıklı bir madde yönetimi ve sınav tasarımı ile bilinçli bir şekilde bir uzman yardımı ile yapılması gerekmektedir. Çünkü bu sınav hatalı ölçme yapması durumunda hem bir kişinin yanlışlıkla en az bir yarı yıl İngilizce Hazırlık Programı'na devam etmesine sebep olabilecek hem de bir öğrenciyi yanlış seviyeye yerleştirerek zaman kaybetmesine ya da öğrenme motivasyonunun düşmesine sebep olabilecek potansiyele sahiptir. Bu özellikleri ile bu iki sınav alanyazında “Kader Sınavı” (Baştürk, 2007), “Yüksek Çıtalı Sınavlar” (Başaran, 2005), ya da “Yüksek Risk içeren Sınavlar” (Kumandas & Kutlu, 2010) diye anılmakla beraber “önemli sonuçları olan sınavlar” (high-stakes exams) şeklinde adlandırılabilirler.

Problem

2020 yılı başlarında ortaya çıkan Kovid-19 salgını tüm dünyada eğitim ve öğretim faaliyetlerini etkilemiştir. Öyle ki eğitim faaliyetlerinin bir kısmı 2020 yılı boyunca uzaktan çevrimiçi olarak gerçekleştirilmek zorunda kalmıştır. Bu dönemde birçok kurum rutin işleyişlerini terkedip yeni yöntem ve teknikler ile eğitim-öğretim faaliyetlerine devam etmişlerdir. Özellikle pandemi koşulları nedeniyle sınavların uzaktan çevrimiçi platformlar üzerinden yapılması gerekliliği doğmuştur. Fakat bazı üniversiteler önemli sınavlarını gerekli hijyen, maske ve mesafe önlemlerini alarak yüzyüze gerçekleştirmek durumunda kalmıştır. Özellikle yılda binlerce yeni öğrenci kaydı alan üniversitelerin öğretim dili İngilizce olan programlarına kayıt olan öğrencileri için en önemli sınavlardan olan hem İngilizce Yeterlik Sınavı hem de Seviye Tespit Sınavlarını ayrı günlerde yapma gibi bir şans pandemi koşulları göz önüne alındığında mümkün olmamıştır. Bu nedenle farklı amaçlara hitap eden bu iki sınavın tek sınav içinde gerçekleştirilmesi ihtiyacı doğmuştur. Ancak bu durum Kovid-19 pandemisi ile mecburiyet üzerine meydana gelen bir durum olduğu için bundan önce bu iki sınavı tek sınav halinde yapmaya ihtiyaç olmamış bu nedenle de bu konuda araştırmacılara ve geliştiricilere destek olacak çalışmalar bulunmamaktadır.

Araştırmanın Önemi

Pandemi koşulları nedeniyle tek bir sınav içinde hem İngilizce yeterlik hem de yerleştirme sınavının yapılması dikkatle ele alınması ve bir uzman yardımı ile yapılması gereken bir işlemdir. Çünkü bu sınavlar “önemli sonuçları olan sınavlar” kategorisindedir. Bu araştırmanın, okuyucularına, iki amaca hitap eden bu iki sınavın tek sınavda birleştirilmesi aşamasında gerekli olan uzman desteğinin en azından bir kısmını sağlayacağı öngörülmektedir. Bu çalışma bu konuda alanyazında ilk araştırmalardan olacağı için önemli bir boşluğu dolduracağı düşünülmektedir. Ayrıca bu araştırmada benimsenen yeni yaklaşımın ve sonuçlarının paylaşıldığı bu çalışmanın benzer ihtiyaçlara sahip İngilizce Hazırlık Programları'nın zamandan, teknik uzmanlık giderlerinden ve emekten tasarruf ederek işlerini daha hızlı gerçekleştirmesine yardımcı olması ve bundan sonra gerçekleştirilecek benzer çalışmalara bir temel oluşturması beklenmektedir.

Araştırma Soruları

Bu araştırmada aşağıdaki iki araştırma soruna yanıt aranacaktır.

1. İngilizce Yeterlik ve Seviye Tespit sınavlarının tek sınavda birleştirilmesi için benimsenen yaklaşımdan madde parametreleri nasıl etkilenmiştir?
2. İngilizce Yeterlik ve Seviye Tespit sınavlarının tek sınavda birleştirilmesi için benimsenen yaklaşımdan tüm test ve alt test Test Bilgi Fonksiyonları nasıl etkilenmiştir?

YÖNTEM

Bu bölümde yöntem “teorik ve kavramsal çerçeve”, “sınav maddelerinin alt testlere dağılımı ve puan hesaplaması”, “araştırmanın katılımcıları”, “madde tepki kuramı varsayımlarının test edilmesi”, ve “madde parametrelerinin kestirimi ve model veri uyumu” başlıkları altında açıklanacaktır.

Teorik ve Kavramsal Çerçeve

Türkiye’de faaliyet gösteren bir üniversitenin pandemi döneminde sınavlarını yüzyüze yapma kararı sonrasında aynı üniversitenin İngilizce Yeterlik sınavı sonuçlarının aynı zamanda İngilizce Seviye Tespit sınavı amacıyla kullanılabilmesi için bazı önlemlerin alınması ve akılcı bir yaklaşımın benimsenmesi gerekmiştir. Çünkü daha önce de belirtildiği gibi İngilizce Yeterlik sınavı ile Seviye Tespit Sınavı amaçları farklı iki ayrı sınav türüdür (Bachman, 1990). Şöyle ki, normal şartlar altında bir İngilizce Yeterlik Sınavında Ortak Avrupa Dil Referans Çerçevesine (CEFR) göre Orta (B1+) ve İleri Orta (B2) seviyelerinde İngilizce yeterlik düzeyini ölçen maddelerin dışında maddeleri kullanmak gerekli olmayabilir. Çünkü bu sınavın amacı daha önce de belirtildiği gibi bölümlerinde derslerini İngilizce olarak yürütebilecek İngilizce Seviyesi belirli bir yetenek düzeyinin üzerinde olan öğrencileri diğer öğrencilerden ayırt ederek onların İngilizce Hazırlık Programı’ndan muaf tutulmalarını sağlamaktır. Türkiye’deki üniversitelerce B2 düzeyi bu amaç için yeterli kabul edilebilir. Çünkü bu seviye için CEFR’de yapılan tanımlamalar tüm beceriler için öğrencilere üniversite öğrenimleri boyunca yeterli olabilecek düzeyde İngilizce düzeyini ifade etmektedir. (bkz. Council of Europe, 2017). Kurumun kriterlerine bağlı olarak bu sınavda C1 (İleri) düzeyinde sorulara gerek olabilir ya da olmayabilir. Eğer daha yüksek bir İngilizce düzeyinde ölçme keskinliği hedefleniyorsa B1+ düzeyindeki maddelerin azaltılması ve yerlerine C1 düzeyinde maddelerin eklenmesi daha uygun olabilir. Çünkü aksi takdirde sınavdaki madde sayısı artabilir ve dolayısıyla sınav süresi uzayabilir.

Normal şartlarda bir İngilizce Seviye Tespit sınavında ise İngilizce Yeterlik Sınavı’nın aksine mümkün olduğunca fazla seviyede yetenek düzeyini ölçecek maddelere yer verilmesi daha uygun olacaktır. Çünkü İngilizce öğretiminde öğrencilerin doğru seviyelere yerleştirilmesi önemlidir (Manning, 2019a). Bu nedenle yerleştirme sınavının daha geniş aralıktaki yetenek düzeylerinde kaliteli ölçme yapması istendik bir durumdur. Ancak eğer bir seviye tespit sınavı hazırlık okulunda yeterlik

sınavını geçemeyen öğrencilerin İngilizce seviyelerini tespit amacıyla kullanılıyorsa zaten C1 ve B2 düzeyindeki öğrenciler yeterlik sınavında başarılı olacağından dolayı bu sınavda bu düzeyde maddelere aslında gerek olmayabilir. Ayrıca A1 ve A2 düzeyindeki maddelere de soru sayısının makul seviyede tutulması için yer verilmeyebilir. Çünkü B1 ve B1+ düzeyindeki soruları doğru yanıtlama derecelerine göre öğrenciler zaten A1 ve A2 seviyelerine yerleştirilebilecektir. Ayrıca B2 düzeyinin altındaki tüm maddeleri doğru cevaplayan bir öğrencinin de B1+ ya da B2 seviyesine yerleştirilmesi uygun olacaktır. Şu da belirtilmelidir ki eğer İngilizce seviye tespit sınavı yeterlik sınavını geçemeyen öğrencilerin seviyelerini tespit için değil de genel amaçlı olarak katılımcılarının İngilizce düzeylerinin tespiti için hazırlanıyorsa A2, B2 ve C1 seviyelerinden de maddeler içermesi daha uygun olacaktır.

Hem Yeterlik hem de Seviye Tespit amaçları için kullanılacak birleşik sınavın tasarımının sınav süresini çok uzatmayacak ve her iki sınavın da ayrı ayrı karakteristik özelliklerini bozmayacak şekilde yapılmasının doğru olacağı değerlendirilmiştir. Çünkü yeni sınavın her iki amaçla da kullanılması ancak bu şekilde mümkün olabilecektir. Bu amaca yönelik olarak benimsenen yaklaşım Tablo 1’de özetlenmiştir.

Tablo 1. Sınavların seviyeleri açısından karşılaştırılması

	Başlangıç (A1)	İleri Başlangıç (A2)	Orta Öncesi (B1)	Orta (B1+)	Orta İleri (B2)	İleri (C1)
İngilizce Yeterlik Sınavı	Gerekli Değil	Gerekli Değil	Gerekli değil	Gerekli	Gerekli	İsteğe bağlı
Seviye Tespit Sınavı	İsteğe bağlı	Gerekli	Gerekli	Gerekli	İsteğe bağlı	İsteğe bağlı
Birleşik Sınav	-	Eklendi	Eklendi	Eklendi	Eklendi	-

Tablo 1’de görüldüğü gibi İngilizce Yeterlik Sınavının madde içeriğine alt yetenek düzeylerindeki öğrencilerin seviyelerinin tespitini mümkün hale getirebilmek amacıyla A2-B1 seviyeleri arasındaki yetenek düzeylerini ölçmeye yönelik maddeler eklenmiştir. Bu amaçla “Dil kullanımı” alt testindeki B1+ ve B2 düzeylerini ölçen 40 maddenin 20 tanesi, A2-B1 seviyeleri arasında dil becerilerini ölçen 20 madde ile değiştirilmiştir. Bu şekilde sınavın A2 yetenek düzeyindeki öğrencileri B1 düzeyindeki öğrencilerden ayırabilmesi amaçlanmıştır. Çünkü sınavın bu iki seviye öğrencileri arasında da ayırım yapabilmesi demek İngilizce Yeterlik Sınavı’nın öğrencileri seviyelere yerleştirme açısından Seviye Tespit Sınavı yetkinliğine yakın bir yetkinliğe ulaşması anlamına gelebilecektir.

Sınav Maddelerinin Alt-testlere Dağılımı ve Puan Hesaplaması

Hem “Yeterlik” hem de “Seviye Tespit” amacıyla kullanılacak birleşik sınavda “Dinleme” alt testinde 20 madde, “Dil Kullanımı” alt testinde 40 madde (ek 20 madde ile birlikte), “Okuma” alt testinde 20 madde olmak üzere toplam dört seçenekli (10 Dinleme maddesi 3 seçenekli olmak üzere) çoktan seçmeli 80 madde bulunmaktadır. Bu alt testlere ek olarak “Yazılı anlatım” alt testinde de öğrencilerden 1 adet İngilizce deneme yazmaları istenmektedir. Sınavın hazırlandığı kurumun test

hazırlama biriminden alınan bilgiye göre sınav maddeleri her bir beceride sınavın ölçmesi beklenen hedeflerin belirlenmesinin ardından ilgili hedeflere yönelik olarak maddelerin yazılmasından sonra her bir madde için üç uzmandan görüş alınmıştır. Sonrasında sınav maddelerinde uzman görüşleri de dikkate alınarak gerekli değişiklikler yapılarak sınava nihai formu verilmiştir.

Yeni eklenen maddeler yeterlik sınavı puanının hesaplanması açısından önemli bir sorunu beraberinde getirmektedir. Eklenen maddelerin bir yeterlik sınavına göre daha kolay maddeler olmasından dolayı yeterlik puanını yapay bir şekilde artırmaması gerekliliği doğmuştur. Bu nedenle A2-B1 düzeyi arasında becerileri ölçmek üzere sınava eklenen maddelerin puanları 0.25 puan ile sınırlı tutulmuştur. Bu sayede eklenen 20 maddenin bu maddelerin tümünü doğru cevaplayan öğrencilerin yeterlik sınavı puanına katkısı toplamda en fazla 5 puanla sınırlı kalmıştır. Bu nedenle her bir öğrenci için Yeterlik Puanı ve Seviye Puanı olmak üzere iki ayrı puan hesaplanmıştır. İki puanın hesaplanma yöntemi Tablo 2’de özetlenmiştir.

Tablo 2. Yeterlik ve Seviye Puanı Hesaplaması

		Dinleme	Dil Kullanımı	Dil Kullanımı (Ek maddeler)	Okuma	Yazılı Anlatım	Toplam
Yeterlik Puanı	Madde Sayısı	20	20	20	20	1	80+1
	Puanlama	20p	15p	5p	30p	30p	100p
Seviye Puanı	Madde Sayısı	20	20	20	20	-	80
	Puanlama	20p	20p	20p	20p	-	80p

Tablo 2’de görüldüğü gibi sınav sonuçlarının Seviye Tespit Sınavı olarak sorunsuz kullanılması için sonradan eklenen maddelerin Seviye Puanına katkısının diğer maddeler ile aynı düzeyde olması gerektiği değerlendirilmiştir. Bu nedenle Seviye Puanı hesaplanırken sonradan eklenen 20 madde de dahil olmak üzere tüm maddeler 1 tam puan olarak puanlanarak her bir öğrenci için Seviye puanı hesaplanmıştır. Ayrıca Tablo 2’de de görüldüğü gibi Yazılı anlatım puanı hem deneme yazma becerisinin çok ileri düzey bir yazılı anlatım becerisi olması hem de seviye puanına öznel değerlendirme puanı katılmamasının daha uygun olacağı düşünüldüğünden Seviye Puanı hesaplamasına dahil edilmemiş ve Seviye Puanı en yüksek 80 puan olarak hesaplanmıştır.

Araştırmanın Katılımcıları

Araştırmanın katılımcılarını sınavın uygulandığı üniversitenin İngilizce eğitim veren bölümlerini 2020-2021 akademik yılında kazanarak üniversiteye kayıt yaptıran ve İngilizce yeterlik ve yerleştirme amacıyla hazırlanan sınava giren toplam 965 öğrenci oluşturmaktadır. Bu öğrencilerden birisi sadece bir soruyu cevaplayıp diğer soruları boş bırakmasından dolayı veri setinden çıkarılmıştır. Sonuç olarak tüm hesaplamalar 964 öğrencilik veri üzerinden yapılmıştır.

Madde Tepki Kuramı Varsayımlarının Test Edilmesi

Tek boyutlu madde tepki kuramı modelleri ile madde ve yetenek düzeyi parametrelerinin sağlıklı kestirilebilmesi için normal dağılım (Embretson & Reise, 2000), tek boyutluluk ve yerel bağımsızlık varsayımlarının (de ayala, 2009) karşılanması gerekmektedir.

Normal dağılım varsayımının test edilmesi için 20 madde eklenmeden önce (60 madde) ve sonra (80 madde) oluşan iki formun klasik test teorisine ve modern test teorisine göre elde edilmiş ortalama ve standart sapma değerleri ile çarpıklık ve basıklık değerlerinin yanısıra her bir formun KR20 iç tutarlık katsayısı değerleri Tablo 3'te gösterilmiştir.

Tablo 3. Klasik ve Modern test teorilerine göre betimsel istatistikler

Test	KR20	\bar{X}	s	Çarpıklık	Basıklık	$\bar{\theta}$	s
60 Madde	0.94	21.52	12.98	0.86	0.06	-0.10	1.24
80 Madde	0.95	31.37	17.22	0.72	-0.19	-0.11	1.24

Tablo 3'te görüldüğü gibi 60 ve 80 maddelik iki formun KR20 iç tutarlık katsayıları sırasıyla 0.94 ve 0.95 gibi oldukça yüksek iki değer almıştır. Sınav 20 ek madde eklenmeden 60 madde olarak kalsaydı 964 öğrencinin klasik ortalama puanının $\bar{X}_{60 \text{ madde}}=21.52$ ve Modern Test Teorisi'ne göre yetenek parametreleri ortalamasının ise $\bar{\theta}_{60 \text{ madde}} = -0.10$ şeklinde olacağı görülmektedir. Ayrıca 20 madde eklendikten sonra oluşan 80 maddelik formun ortalama puanlarının $\bar{X}_{80 \text{ madde}}=31.37$ ve $\bar{\theta}_{80 \text{ madde}} = -0.11$ şeklinde olduğu görülmektedir. Tablo 4'te görülen çarpıklık ve basıklık katsayıları incelendiğinde hem 20 madde eklenmeden önce hem de eklendikten sonra 60 ve 80 maddelik her iki formun da hem basıklık hem de çarpıklık katsayılarının -1.00 ve +1.00 değerleri aralığında olduğu gözlemlenmektedir. Bu da her iki testin verisinde de normal dağılım varsayımını karşılandığının kanıtı olarak kabul edilmiştir (Tabachnick ve Fidell, 2007).

Demars (2010) tek boyutluluk sayılısını, kullanılan modelde bir öğrencinin maddelere verdiği tepki üzerinde sadece bir baskın faktörün etkili olması ve diğer faktörlerin tesadüfi hatadan ibaret olması şeklinde açıklamıştır. Hambleton, Swaminathan ve Rogers (1991) ise tek boyutluluğu, testi alan kişilerin performanslarının açıklanmasına imkan verecek şekilde bir baskın (dominant) örtük yeteneğin etkili olması şeklinde açıklamaktadır. Bu varsayımın sosyal bilimlerde tam olarak karşılanması oldukça zordur. Çünkü kişilik, bilişsel özellikler ve test uygulaması ile ilintili farklı faktörler az da olsa test performansını etkilemektedir (Hambleton, Swaminathan ve Rogers, 1991). Ayrıca tek boyutluluk varsayımının ihlal edilmesinin madde parametreleri üzerindeki etkisi sınırlı olduğu bazı araştırmalarda tespit edilmiştir (Dorans ve Kingston, 1985; Cuesta ve Muniz, 1999). Bir diğer varsayım olan yerel bağımsızlık bir bireyin her bir madde çiftine verdiği yanıtın birbiri ile ilişkili olmaması durumudur (Hambleton, Swaminathan ve Roger, 1991). Tek boyutluluk varsayımı ile yerel bağımsızlık varsayımlarının birbiri ile ilintili olduğu söylenebilir. Hambleton ve Swaminathan (1985)'a göre tek

boyutluluk varsayımının karşılandığı durumlarda yerel bağımsızlık varsayımının da karşılandığı kabul edilebilir.

Tek boyutluluk varsayımının test edilmesi için 60 ve 80 maddelik test formlarının ikisine de temel bileşenler analizi uygulanmış ve faktörlerin toplam varyansı açıklama oranlarına SPSS 23 (IBM Corp, 2015) yardımı ile bakılmıştır. 60 maddelik formda ilk dört faktöre ait özdeğerlerin sırasıyla 14.29, 2.43, 1.34, 1.25, bu faktörlere ait varyans açıklama oranlarının ise sırasıyla %23.81, %4.05, %2.24, %2.07 olduğu tespit edilmiştir. 80 maddelik formda ilk dört faktöre ait özdeğerlerin sırasıyla 18.27, 2.90, 1.48, 1.32 ve ilk dört faktöre ait varyans açıklama oranları ise sırasıyla %22.83, %2.62, %1.85, %1.65 olduğu tespit edilmiştir. Bu veriler incelendiğinde her iki formda da tek boyutluluk varsayımı ile beklendiği üzere öğrenci performansı üzerinde bir baskın (dominant) faktörün etkili olması, ilk faktöre ait özdeğerin diğer faktörlerin özdeğerlerden oldukça büyük olması, ilk faktörden sonraki faktörlerin özdeğerleri arasında çok küçük farklar olması göz önüne alınarak 60 ve 80 maddelik formların her ikisinde de tek boyutluluk varsayımının karşılandığının kanıtı olarak kabul edilmiştir (Şahhüseyinoğlu, 2006).

Madde Parametrelerinin Kestirimi ve Model Veri Uyumu

Madde parameterlerinin kestirimi Madde Tepki Kuramı'nın üç parametrelili lojistik modelinde Xcalibre 4.2 (Guyer & Thompson, 2014) kullanılarak gerçekleştirilmiştir. Parametre kestirim işlemi en yüksek olasılık (Maximum Likelihood) yöntemi kullanılarak yapılmıştır. Model sabiti (D) 1.7 olarak alınmıştır. Madde kestirimi yapılırken hesaplamaların tamamlanması kriteri olarak iki kestirim arası farkın 0.001'den küçük olması kriteri kullanılmıştır.

Üç parametrelili modelde madde parametreleri olan "a", "b" ve "c" parametreleri üzerinden kestirim yapılmaktadır. "a" parametresi ayırt edicilik parametresidir (Hambleton & Swaminathan, 1985) ve bir maddenin alt yetenek düzeylerindeki öğrencileri üst yetenek düzeylerindeki öğrencilerden ne kadar ayırt edebildiğinin göstergesidir. "a" parametresi teorik olarak $-\infty$ ve $+\infty$ arasında değerler almaktadır. Ancak pratikte çoğunlukla 0.00 ile 2.00 değerleri arasında alır (Hambleton & Swaminathan, 1985). "b" parametresi ise madde güçlüğü parametresidir ve bir yetenek düzeyinde bireylerin bir maddeye doğru cevap verme olasılığının %50 olduğu noktayı göstermektedir (Hambleton, Swaminathan & Rogers, 1991). "b" parametresi ise pratikte -2.00 ile +2.00 arasında değerler almaktadır (Hambleton & Swaminathan, 1985). "c" parametresi ise şans başarısı parametresidir ve en düşük yetenek düzeyindeki bireylerin maddelere doğru cevap verme olasılığını göstermektedir (Hambleton, Swaminathan & Rogers, 1991). Bu değerler dört seçenekli bir çoktan seçmeli testte 0.25, bu çalışmada kullanılan dört seçenekli bir çoktan seçmeli maddelerin bulunduğu bir sınavda 0.25 civarı olması beklenmektedir. Ancak "c" parametresi teorik olarak 0 ile 1 arası değerler almaktadır.

Araştırma sorularının cevaplanması ve eklenen 20 maddenin testin madde parametrelerine ve test bilgi fonksiyonlarına etkisinin karşılaştırılabilmesi için 20 madde eklenmeden yani 60 maddelik formun

veri seti ile ve 20 madde eklendikten sonra 80 maddelik formun veri seti üzerinden ayrı ayrı madde parametresi kestirimi yapılmıştır.

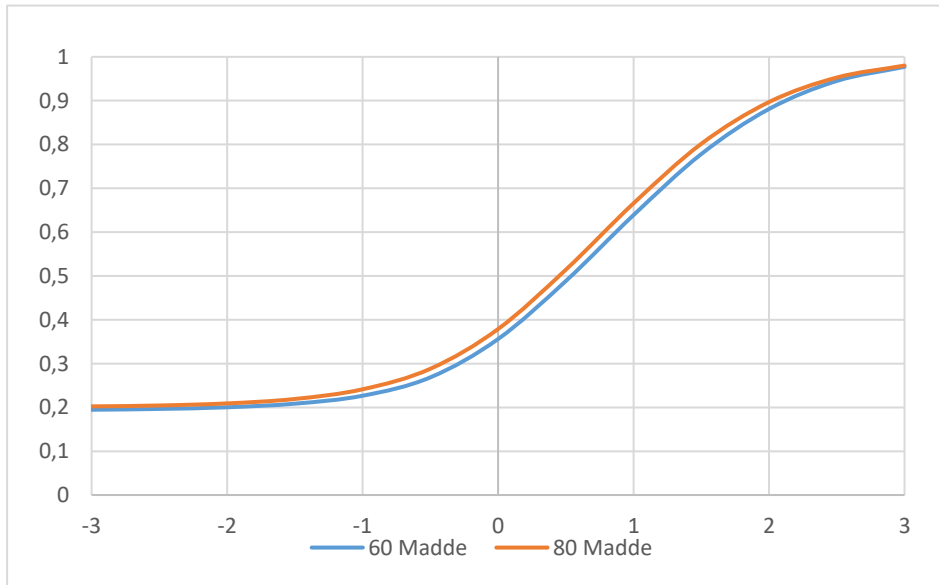
Model Veri Uyumu

Her iki test için de model veri uyumu için Ki-kare (X^2) değeri ve serbestlik derecesinin (sd) oranına bakılmıştır (Drasgow, Levine, Tsien, Williams, ve Mead, 1995). Elde edilen X^2/sd değerleri Tablo 4'te gösterilmiştir.

Tablo 4. 60 ve 80 maddelik formlarda X^2/sd değerleri

Test	X^2 değeri	sd	X^2/sd
60 Madde	1362.65	720	1.89
80 Madde	1752.58	960	1.83

Chernyshenko, Stark, Chan, Drasgow, ve Williams (2001) X^2/sd oranının 3'ten küçük olmasını model veri uyumunun bir kanıtı olarak almışlardır. Tablo 4 incelendiğinde 20 madde eklenmeden (60 madde) ve eklendikten sonra (80 madde) elde edilen X^2/sd oranının 60 ve 80 maddelik test formları için sırasıyla $X^2/Sd_{60\text{ Madde}}=1.89$ ve $X^2/Sd_{80\text{ Madde}}=1.83$ olduğu ve her ikisinin de 3'ten küçük olması nedeniyle bu durumu her iki test formu için de model veri uyumunun sağlandığının bir kanıtı olarak kabul edilmiştir. Ayrıca Hambleton ve Swaminathan (1985) model veri uyumunun bir göstergesi olarak madde ve yetenek parametrelerinin değişmezliğini göstermiştir. Bu nedenle her ikisinin de bir yansıması olan test karakteristik eğrileri görsel incelemeye tabi tutulmuştur (Şahin, 2012). Elde edilen Test Karakteristik Eğrileri Şekil 1'de gösterilmiştir.



Şekil 1. 60 ve 80 maddelik testlere ait test karakteristik eğrileri

Şekil 1 incelendiğinde 60 ve 80 maddelik formlara ait test karakteristik eğrilerinin birbirine çok benzer olduğu belirlenmiştir. Bu durum her iki formda da madde ve yetenek parametrelerinin değişmezliğinin ve dolayısıyla model veri uyumunun kanıtı olarak kabul edilmiştir.

BULGULAR

Bu bölümde araştırma sorularının cevaplanmasına yönelik yapılan analizlerin bulgularına “Eklenen maddelerin madde parametrelerine etkisi”, ”Eklenen maddelerin test bilgi fonksiyonlarına etkisi” başlıkları altında yer verilmiştir.

Eklenen Maddelerin Madde Parametrelerine Etkisi

Sınava eklenen 20 maddenin sınavın madde parametrelerine yaptığı etkiyi görmek amacıyla bu 20 madde eklenmeden ve eklendikten sonra yani 60 ve 80 madde üzerinden tüm test ve alt testlerde “a” (madde ayırt ediciliği), “b” (madde güçlüğü) ve “c” (şans başarısı) parametreleri kestirilerek ortalamaları Tablo 5’te gösterilmiştir.

Tablo 5. Tüm test ve alt testlerde ortalama madde parametrelerinin karşılaştırması

Test	Madde Para.	Dinleme	Dil Kullanımı	Dil Kullanımı (Ek Maddeler)	Dil Kullanımı (Birleşik)	Okuma	Tüm Test
60 Madde	a	1.48	1.40			1.42	1.43
	b	0.79	1.14			0.68	0.87
	c	0.20	0.19			0.19	0.19
80 Madde	a	1.55	1.46	1.15	1.30	1.48	1.41
	b	0.83	1.16	0.41	0.78	0.73	0.78
	c	0.19	0.19	0.23	0.21	0.19	0.20

Tablo 5’te görüldüğü üzere 20 madde eklenmeden Dinleme, Dil kullanımı, okuma alt testlerine ve tüm teste ait “a” yani ayırt edicilik parametreleri ortalamaları sırasıyla 1.48, 1.40, 1.42 ve 1.43 şeklindedir. 20 madde eklendikten sonra dinleme, dil kullanımı, okuma alt testleri ve tüm test için elde edilen “a” parametreleri ortalamaları sırasıyla 1.55, 1.46, 1.48, 1.41 olmuştur.

Ayrıca 20 madde eklenmeden dinleme, dil kullanımı, okuma alt testlerin ve tüm teste ait “b” yani madde güçlüğü parametreleri ortalamaları sırasıyla 0.79, 1.14, 0.68 ve 0.87 şeklindedir. 20 madde eklendikten sonra dinleme, dil kullanımı, okuma alt testleri ve tüm test için elde edilen “b” parametreleri ortalamaları ise sırasıyla 0.83, 0.78, 0.73, 0.78 olmuştur. Tablo 5’ten görüleceği üzere eklenen 20 madde’nin ortalama “b” parametresi değeri 0.41’dir ve testin geneline göre çok daha kolay sorular oldukları bu ortalamadan anlaşılmaktadır. Eklenen maddeler dil kullanımı alt testi için normalde 1.16 olan ortalama “b” parametresi değerini 0.78’e çekmiştir. Bu da bu maddelerin eklenmesi ile istenilen amaca ulaşıldığını göstermektedir.

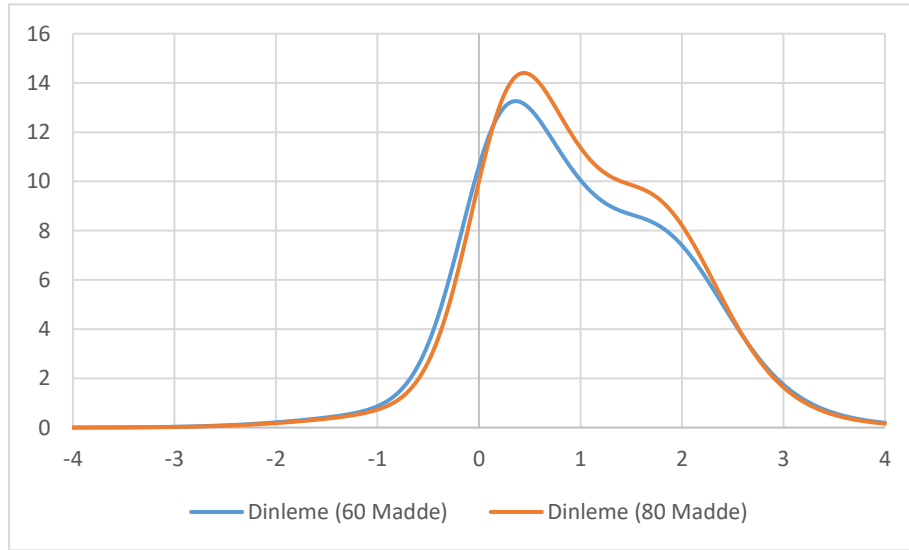
Son olarak 20 madde eklenmeden önce dinleme, dil kullanımı, okuma alt testleri ve tüm teste ait “c” yani şans başarısı parametresi ortalamaları sırasıyla 0.20, 0.19, 0.19 ve 0.19 şeklindedir. 20 madde

eklendikten sonra dinleme, dil kullanımı, okuma alt testleri ve tüm test için elde edilen “c” parametreleri ortalamaları ise sırasıyla 0.19, 0.19, 0.19 ve 0.20 olmuştur.

Her bir alt testte hem ek maddeler eklenmeden hem de eklendikten sonra elde edilen “a”, “b” ve “c” parametreleri incelendiğinde 60 ve 80 maddelik formların her ikisinin de her alt boyutta benzer düzeyde ayırt edicilik ve güçlük değerlerine sahip olması her iki formda da benzer düzeyde ölçme keskinliğine ulaşıldığının bir göstergesi olarak alınmıştır.

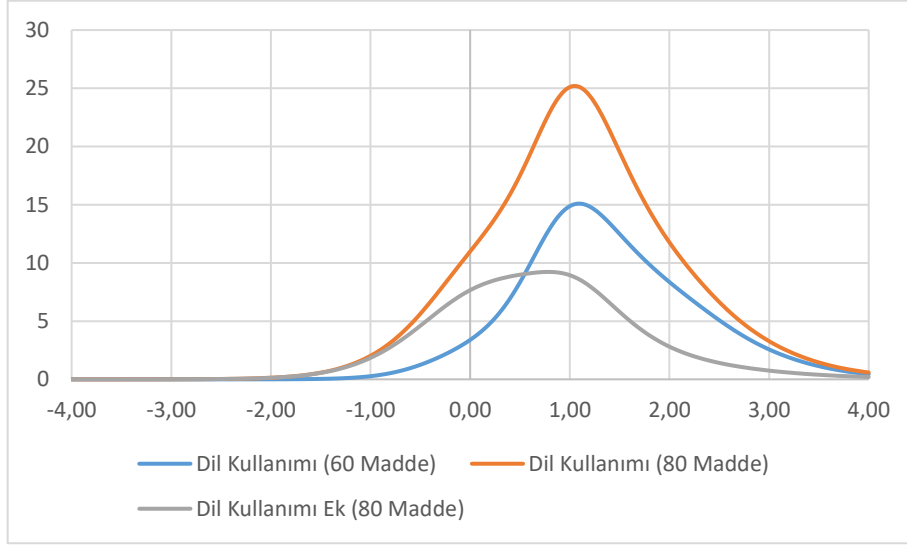
Eklenen Maddelerin Test Bilgi Fonksiyonlarına Etkisi

Yeterlik Sınavının aynı zamanda Seviye Tespit Sınavı olarak kullanılmasını sağlamak amacıyla sınava eklenen 20 maddenin tüm test ve her bir alt testin test bilgi fonksiyonlarına etkisi incelenmiştir. Şekil 2’de 20 madde eklenmeden ve eklendikten sonra elde edilen “Dinleme” alt testine ait test bilgi fonksiyonları görülmektedir.



Şekil 2. Dinleme alt testlerine ait Test Bilgi Fonksiyonları Karşılaştırması

Şekil 2’deki “Dinleme” alt testlerine ait test bilgi fonksiyonları incelendiğinde yeni eklenen maddeler “Dil kullanımı” alt testine eklenmiş olmalarına rağmen “Dinleme” alt testinde de bilgi düzeyinde $\theta=0.2$ ile $\theta=2.00$ yetenek düzeyleri arasındaki bölgede az da olsa artışa neden olduğu görülmektedir. Bu durum madde parametreleri kestirimi işlemi tüm maddelerin hesaba katıldığı “Expectation-Maximization” işlemini de içeren tekrarlı (iterative) bir işlem olduğundan dolayı yeni eklenen maddelerin de modele eklenmesi ile yeni bir kestirim işlemi yapılırken madde parametrelerinde bir kaymadan olmuş olabileceğini düşündürmektedir. Ayrıca madde parametreleri kestirimi işlemi tekrarlı bir şekilde toplu olarak yapıldığından dolayı madde sayısındaki artış sebebiyle belirli yetenek düzeylerinde oluşan bilgi artışı diğer alt testlerin bilgi fonksiyonlarına da göreceli şekilde yükseliş olarak yansımış olabileceği de değerlendirilebilir.

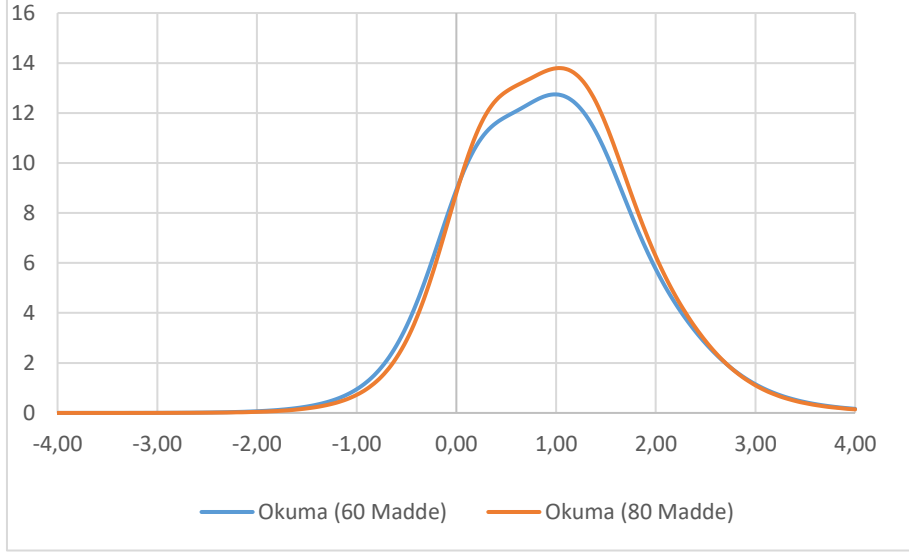


Şekil 3. Dil kullanımı alt testlerine ait Test Bilgi Fonksiyonları karşılaştırması

Maddeler eklenmeden önce “Dil kullanımı” alt testine ait test bilgi fonksiyonu, eklendikten sonra “Dil kullanımı” alt testine testine ait Test Bilgi Fonksiyonu ve sadece eklenen maddelere ait (Dil Kullanımı Ek) test bilgi fonksiyonu Şekil 3’te gösterilmektedir.

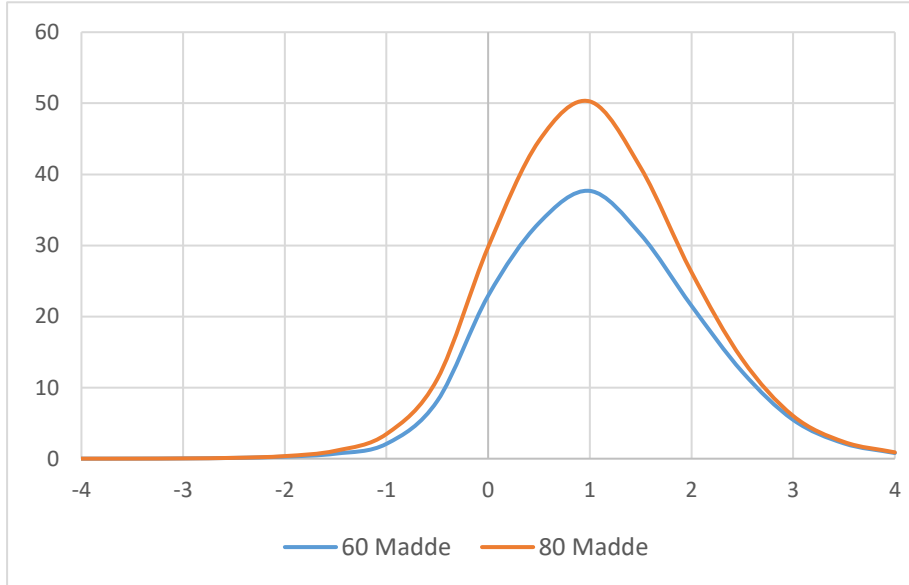
Şekil 3 incelendiğinde 60 maddelik teste eklenen 20 maddenin “Dil kullanımı” alt testinin alt test bilgi fonksiyonu üzerindeki etkisi açıkça görülmektedir. Şekil 3’te görüldüğü gibi test 60 madde olsaydı testte “Dil kullanımı” alt testinde en yüksek bilgi düzeyine yaklaşık 1.00 yetenek düzeyinde ulaşılacaktı ve alt test yaklaşık olarak $\theta=0.00$ ile $\theta=3.00$ yetenek düzeyleri arasında bilgiye sahip olacaktı. 60 maddelik teste eklenen 20 maddenin “Dil kullanımı” alt testine yaptığı katkıyı görmek için hazırlanan sadece bu ek maddelere ait test bilgi fonksiyonu incelendiğinde 80 maddelik testin “Dil kullanımı” alt testine eklenen 20 maddenin bilgi fonksiyonuna -1.00 ve 2.00 yetenek düzeyleri arasında katkı yaptığı açıkça görülmektedir. Sonuç olarak 20 madde eklendikten sonra elde edilen 80 maddelik testte “Dil kullanımı” alt testine ait test bilgi fonksiyonu $\theta=-1.00$ ile $\theta=3.00$ arası yetenek düzeylerini de kapsayacak şekilde genişlediği ve $\theta=-1.00$ ile $\theta=2.00$ düzeylerinde ölçme keskinliğinin hissedilir düzeyde artmasına neden olduğu gözlemlenmiştir.

Teste eklenen 20 madde eklenmeseydi elde edilecek 60 maddelik testin “Okuma” alt testine ait test bilgi fonksiyonu ile eklendikten sonraki “Okuma” alt testine ait test bilgi fonksiyonu Şekil 4’te gösterilmiştir.



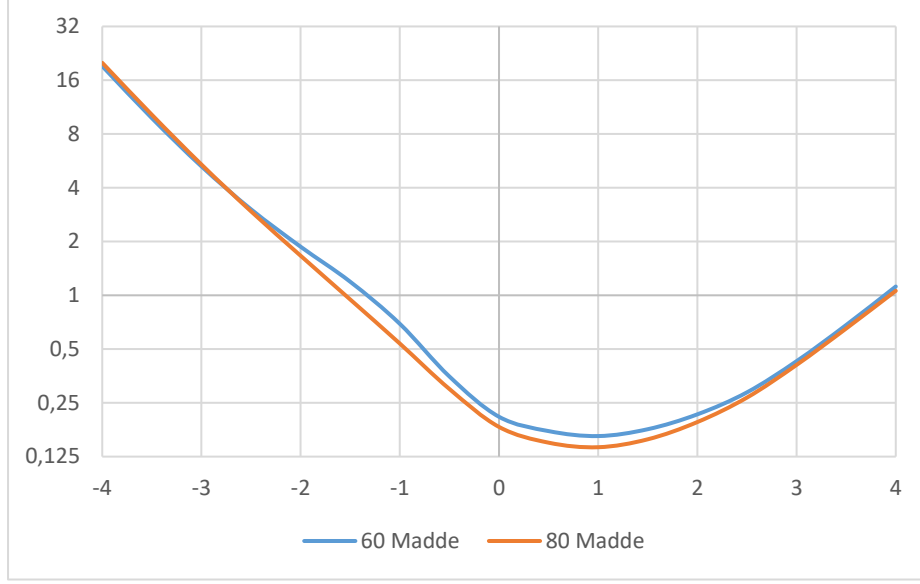
Şekil 4. Okuma alt testlerine ait Test Bilgi Fonksiyonları karşılaştırması

Şekil 4'te de görüldüğü gibi eklenen 20 madde “okuma” alt testine ait test bilgi fonksiyonunun belirli yetenek düzeylerinde yükselmesine neden olmuştur. Bunun neden olabileceği “Dinleme” alt testine ait test bilgi fonksiyonu incelenirken açıklanan sebepler olabilir. Şekilde 4’te görüldüğü gibi teste eklenen maddeler “Okuma” test bilgi fonksiyonun ölçtüğü yetenek düzeyi aralığında “Dil Kullanımı” alt boyutunda gerçekleşen düzeyde bir değişime neden olmamıştır.



Şekil 5. Tüm test Test Bilgi Fonksiyonları karşılaştırması

20 madde eklenmeden önce ve sonra elde edilen 60 ve 80 maddelik testlere ait tüm test bilgi fonksiyonları Şekil 5’te gösterilmiştir. Şekilde 5’te görüldüğü üzere eklenen 20 maddenin tüm test bilgi fonksiyonuna etkisi testin $\theta = -1.00$ ve $\theta = 2.50$ yetenek düzeyleri arasındaki tüm yetenek düzeylerinde daha fazla bilgi elde edilmesi şeklinde olmuştur. Elde edilen test bilgi fonksiyonlarına ait standart hata değerleri de şekilde 6’da gösterilmiştir.



Şekil 6. Test Bilgi Fonksiyonuna ait Standard Hata değerleri fonksiyonları karşılaştırması

Şekil 6’da da görüldüğü gibi eklenen 20 madde 80 maddelik testin $\theta = -2.00$ ve $\theta = 2.00$ yetenek düzeyleri aralığında daha az hata ile ölçmesine katkı yapmış ve dolayısıyla ölçme keskinliğinin artmasına neden olmuştur. Burada ilginç bir durum dikkat çekmektedir. Sınava eklenen 20 madde test bilgi fonksiyonunda $\theta = -2.00$ ile $\theta = 0.00$ aralığında fazla katkı yapmamış gibi görünse de hata değerlerinin bu aralıkta düşmesine sebep olmuştur.

TARTIŞMA VE SONUÇ

Bu kısımda elde edilen bulgular ışığında araştırma soruları cevaplanacak ve tartışılacaktır. Bu amaçla iki araştırma sorusu ayrı başlıklar altında irdelenecektir.

Araştırma Sorusu 1. İngilizce Yeterlik ve Seviye Tespit sınavlarının tek sınavda birleştirilmesi için benimsenen yaklaşımdan madde parametreleri nasıl etkilenmiştir?

Madde parametresi kestirimlerine göre sınava 20 madde eklenmeden önce yani 60 maddelik formda ortalama “a”, “b” ve “c” parametreleri sırasıyla 1.43, 0.87, 0.19 değerlerini almıştır. 1.43 gibi bir ayırt edicilik ortalamasına ulaşması sınavın kalitesi açısından önemli bir göstergedir. Çünkü daha önce de belirtildiği gibi “a” parametresi sınav maddelerinin üst ve alt yetenek düzeylerindeki öğrencileri ayırt etme düzeyini göstermektedir (Hambleton & Swaminathan, 1985). Baker (2001)’a göre “a” parametresi 1.35 ile 1.69 arasında değerler aldığı yüksek ayırt edicilik düzeyine sahip maddeler

olarak adlandırılabilir. Elde edilen 1.43'lük ortalama "a" parametresi değeri test maddelerinin yüksek derecede ayırt edici olduğunun göstergesi olarak alınabilir. Bu durumda sınav maddelerinin olması gerektiği ve beklendiği gibi üst ve alt yetenek düzeylerindeki öğrencileri birbirinden ayırt ettiği söylenebilir.

Elde edilen ortalama "b" parametresi değeri sınavın güçlüğünün "1" yetenek düzeyine yakın bir yerde 0.87'de konumlandığını göstermektedir. Ortalama "b" parametresi 0.00 konumunda ise sınavın orta güçlükte olduğu düşünülür (Kezer, 2013). İngilizce Yeterlik Sınavının ortlamadan biraz daha yüksek becerileri ölçtüğü düşünüldüğünde tüm testten elde edilen ortalama "b" parametresi (0.87) ideale yakın bir konumda olduğu söylenebilir.

"c" parametresi 0.20'den küçük değerler aldığı o sınavda şans başarısının düşük bir değer aldığı düşünülebilir (Özbaşı, 2014). Testten elde edilen ortalama "c" parametresinin de 0.19 değeri alması sınavın dört seçenekli çoktan seçmeli bir sınav olmasından dolayı şans başarısına düşük düzeyde kaldığını düşündürmektedir. Tüm bu veriler ışığında 60 maddelik formun ayırt edicilik, güçlük ve şans başarısı parametreleri açısından istendik düzeyde olduğu söylenebilir.

Sınava 20 madde eklendikten sonra oluşan 80 maddelik teste ait tüm test ortalama madde parametreleri incelendiğinde "a", "b" ve "c" parametrelerinin ortalamasının sırasıyla 1.41, 0.78, 0.20 değerleri aldığı görülmektedir. Bu değerler de 20 madde eklendikten sonra da yukarıda bahsi geçen sebeplerden dolayı sınavın istendik düzeyde ölçme keskinliğine sahip olduğunu göstermektedir. Eklenen maddeler sınavın ortalama ayırt edicilik değerinde sadece 0.02 düzeyinde bir düşüşe sebep olmuştur. Bu da eklenen maddelerin de yüksek ayırt edicilik değerine sahip kaliteli maddeler olduğunu düşündürmektedir.

Ortalama "b" parametresinde 0.87'den 0.78'e doğru gerçekleşen düşüş de olması gereken yani beklenen bir düşüştür. Bu düşüş çok az bir düşüş gibi görünse de madde sayıları dikkate alındığında iki ortalama "b" parametresi değeri (60 madde ve 80 maddelik formlar için) arasındaki farkın (0.09) madde sayısı olan 80 ile çarpıldığında gerçekleşen düşüşün etkisi daha iyi anlaşılacaktır. Konunun daha derinlemesine irdelenmesi için "Dil kullanımı" alt testine eklenen 20 maddenin ortalama "a" "b" ve "c" değerlerine bakmakta fayda vardır. Bu değerler sırası ile 1.15, 0.41 ve 0.23 olarak elde edilmiştir. Bu değerler incelendiğinde testin geneline göre beklendiği üzere daha düşük bir ortalama ayırt edicilik değerine ulaştığı görülmektedir. Bu da beklendik bir durumdur. Çünkü bu maddeler nispeten daha kolay maddeler olduğu için testin genelinden daha düşük puan alan öğrenciler de bu maddelere doğru yanıt verebilmişlerdir. Ancak elde edilen 1.15 ortalama "a" değeri sınavın genelinin ayırt edicilik düzeyini olumsuz etkileyebilecek kadar düşük düzeyde bir değer değildir ve Baker (2001)'a göre orta düzey ayırt edicilik değerlerinin (0.65-1.35 arası) üst yarısını işaret etmektedir.

Sadece “Dil kullanımı” bölümüne eklenen 20 maddenin ortalama “b” parametresi değerine ($b_{ort}=0.41$) bakıldığında testin genelinden elde edilen 0.87’den oldukça uzak olduğu görülmektedir. Fakat bu uzaklık da testin ortalama güçlük düzeyini 0.5’lere çekmeye yetmemiştir. Mevcut durum testin işlevini görebilmesi için oldukça yeterlidir. Ancak testin son halinin 0.50 düzeyinde ya da bu değere yakın bir ortalama “b” parametresine sahip olması sağlanabilirse bu sayede eklenen 20 maddenin daha alt yetenek düzeylerindeki öğrencileri birbirlerinden ayırt etme konusunda daha başarılı olması sağlanabilir. Ancak bu tür “önemli sonuçlara sahip sınavlarda” genellikle daha önceki sınavlarda kullanılarak denenmiş maddeler kullanılmadığı için madde güçlüklerini ve ayırd ediciliğini sadece uzman görüşü ile kestirmek zorunda kalınmaktadır. Bilindiği üzere uzman görüşü ile madde parametrelerinin tahmin edilmesi doğru sonuçlar vermeyebilir (Bejar, 1983). Bu nedenle elde edilen sonuç içinde bulunulan şartlar göz önüne alındığında oldukça başarılıdır. Ancak benzer şartlarda eğer imkan varsa ve madde parametreleri biliniyorsa ortalama güçlüğü daha düşük olan bir madde setinin teste eklenmesinin uygunluğu değerlendirilebilir. Ancak bu değerlendirme yapılırken bu kararın yeterlik sınavı puanına etkisi de göz önünde bulundurulmalıdır.

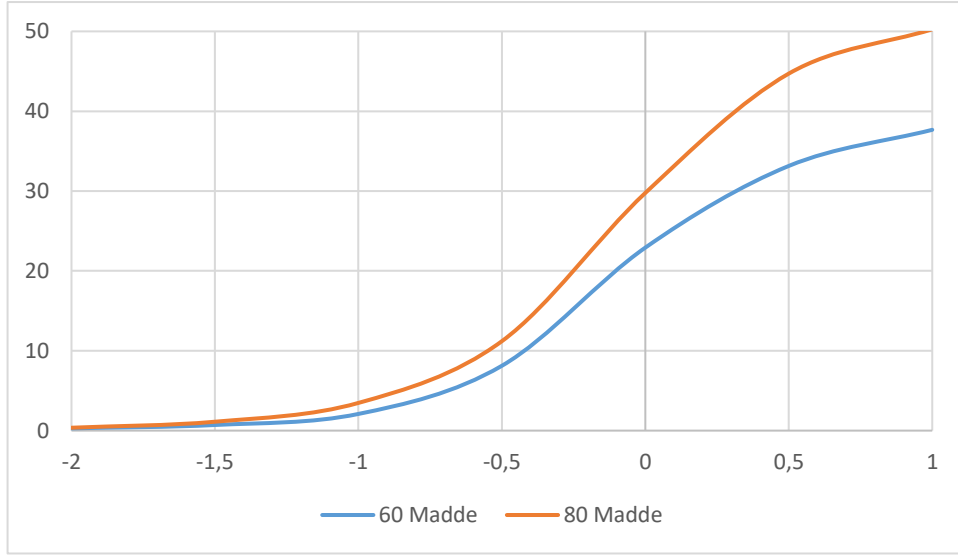
Tek başına madde parametrelerinin değerlendirilmesi de yanıltıcı olabilir. Bu nedenle madde parametrelerine ait bilgileri test/alt test bilgi fonksiyonları ile birlikte de değerlendirilmesi uygun olacaktır. Sonuç olarak madde parametreleri değerlendirildiğinde “Dil kullanımı” alt testine eklenen 20 maddenin testin ortalama ayırt edicilik düzeyinde büyük bir düşüşe neden olmadan testin ortalama güçlüğü üzerinde yeterli düzeyde düşüşe sebep olduğu ve hem bu maddeler eklenmeden elde edilen 60 maddelik formun hem de eklendikten sonra elde edilen 80 maddelik formun madde parametreleri açısından kaliteli bir sınav olduğu söylenebilir.

Araştırma Sorusu 2. İngilizce Yeterlik ve Seviye Tespit sınavlarının tek sınavda birleştirilmesi için benimsenen yaklaşımdan tüm test ve alt test Test Bilgi Fonksiyonları nasıl etkilenmiştir?

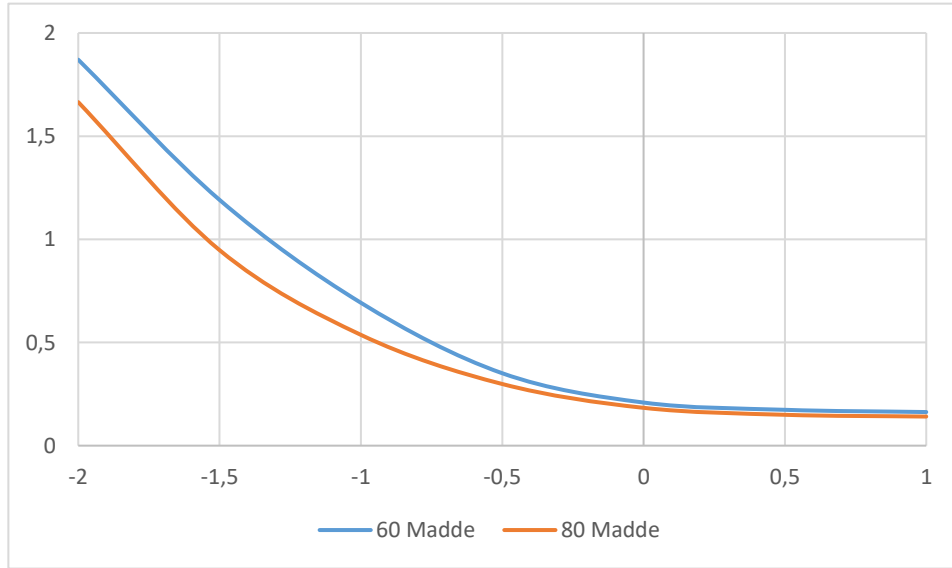
Test Bilgi fonksiyonları bir testin tüm yetenek düzeylerinde ne kadar başarılı kestirim yaptığını gösterirler (Baker, 2001). Test bilgi fonksiyonlarında yükseklik arttıkça varyans azalmakta, bilgi artmakta ve dolayısıyla standart hata azalmaktadır (Hambleton & Swaminathan, 1985). Eklenen 20 maddenin hem alt testlerin hem de tüm testin test bilgi fonksiyonlarına etkisi daha önceki şekillerde gösterilmişti. Söz konusu maddelerin eklendiği “Dil kullanımı” alt testine ait test bilgi fonksiyonu (Şekil 3) eklenen maddelerin aslında doğru şekilde konumlandırıldığını göstermektedir. Çünkü görüldüğü kadarıyla eklenen maddeler testin ortalama güçlüğüde beklenenin biraz altında değişikliğe yol açmış gibi görünse de “Dil Kullanımı” alt testinin test bilgi fonksiyonu incelendiğinde eklenen maddelerin “Dil Kullanımı” alt testi test bilgi fonksiyonuna hem dikey hem de yatay katkı yaptığı görülmektedir. Yatay etki yani alt testin kapsadığı yetenek düzeyleri aralığında yaklaşık 1.00 yetenek düzeyi aşağıya doğru kayma şeklinde gerçekleşmiştir. Bu da 80 maddelik formda 60 maddelik formun aksine bu alt testin $\theta = -1.00$ düzeyini de kapsayacak becerileri ölçmesine sebep olmuştur. Dikey etki ise $\theta = -1.00$ ile

$\theta=+3.00$ yetenek düzeyleri aralığındaki tüm yetenek düzeylerinde bilginin gözle görülür şekilde artması şeklinde gerçekleşmiştir. Şekil 3 incelendiğinde eklenen maddelerin $\theta=0.00$ ile $\theta=1.00$ değerleri arasında da test bilgi fonksiyonuna önemli katkılar yaptığı görülmektedir.

Tüm testin test bilgi fonksiyonu incelendiğinde (Şekil 5) eklenen maddelerin $\theta=-2.00$ ile $\theta=+1.00$ değerleri arasında daha keskin ölçme kabiliyetine sahip olmasını sağladığı görülmektedir. Bu iki aralıkta 60 madde ve 80 maddelik testlerden elde edilen tüm test test bilgi ve hata fonksiyonlarına daha yakından bakmakta fayda vardır. Bu fonksiyonlar Şekil 7 ve Şekil 8’de gösterilmiştir.



Şekil 7. $\theta=-2.00$ ve $\theta=+1.00$ yetenek düzeyleri arasında elde edilen tüm test bilgi fonksiyonları



Şekil 8. $\theta=-2.00$ ve $\theta=+1.00$ yetenek düzeyleri arasında elde edilen hata fonksiyonları

Şekil 7 incelendiğinde yaklaşık olarak $\theta=-1.50$ düzeyinden başlayarak $\theta=+1.00$ düzeyine doğru düzenli bir şekilde bilgi artışı gerçekleştiği görülmektedir. Özellikle $\theta=-1.00$ ve $\theta=0.00$ yetenek düzeyleri arasında bilgi artışının daha fazla olması için bundan sonra yapılacak benzer çalışmalarda eğer

madde parametreleri biliniyorsa bu aralıkta bilgi verecek daha fazla maddenin bilinçli olarak teste eklenmesi bu aralıktaki bilginin artmasını sağlayabilecektir.

Bu aşamada Şekil 8’de $\theta=-2.00$ ile $\theta=+1.00$ değerleri arasında elde edilen hata değerleri incelendiğinde açıkça görüldüğü gibi eklenen 20 maddenin $\theta=-2.00$ ve $\theta=-0.50$ yetenek düzeyleri arasında standart hata değerlerini oldukça pozitif yönde etkilediği açıkça görülmektedir. Bu da eklenen 20 maddenin bilgi düzeyini $\theta=-1.50$ ve $\theta=1.00$ aralığında artırırken standart hata değerlerine $\theta=-2.00$ ve $\theta=0.50$ aralığında olumlu katkı yaptığını net bir şekilde göstermektedir.

SONUÇ

Kovid-19 pandemisinden dolayı kalabalık öğrenci gruplarının kampüse iki ayrı sınav için iki kez gelmesinin önüne geçmek için iki farklı amaca yönelik iki ayrı sınav olan İngilizce Yeterlik Sınavı ve Seviye Tespit Sınavı’nın tek bir sınav şeklinde uygulanması gerekliliği doğmuştur. Bu gereklilik bu iki sınavın işlevini görecektir yeni sınava madde seçiminde ve sınav tasarımında bazı değişiklikleri de beraberinde getirmiştir. Çünkü İngilizce Yeterlik Sınavı’nın sadece yüksek yetenek düzeylerinde daha az hata ile ölçme yapması beklenirken Seviye Tespit Sınavı’nın ise daha alt yetenek düzeylerini de kapsayacak şekilde daha geniş bir yetenek düzeyi aralığında yüksek ölçme keskinliğine ulaşması beklenmektedir. Bu iki bilgiyi göz önünde bulundurarak İngilizce Yeterlik Sınavının madde yapısına müdahalede bulunulmuş ve nispeten daha alt yetenek düzeylerindeki (CEFR’a göre A2-B1 seviyeleri) öğrencilerin de cevaplayabileceği 20 maddelik bir madde seti sınavın “Dil kullanımı” alt testine eklenmiştir. Her bir öğrenci için aynı sınava verilen tepkilerden Yeterlik ve Seviye Puanı olmak üzere iki farklı puan hesaplanması yapılmıştır. Eklenen bu maddeler Yeterlik Puanı hesaplanırken 0.25 puan, Seviye Puanı hesaplanırken bir (1) tam puan şeklinde puanlanmıştır. Böylece sınava eklenen nispeten daha kolay maddelerin yeterlik puanı üzerindeki etkisi en çok 5 puan ile sınırlı tutulurken Seviye Puanı’nda 20 puan etkiye sahip olması sağlanmıştır. Bu işlemin ardından eklenen maddelerin maddeler eklenmeden önceki 60 maddelik test ve eklendikten sonraki 80 maddelik test üzerinden kestirilen madde parametreleri ve test ve alt test bilgi fonksiyonları üzerindeki etkisi incelenmiştir. Elde edilen madde parametreleri, ve test bilgi fonksiyonları incelendiğinde eklenen maddelerin sınavın madde parametreleri açısından kalitesini bozmadan üst yetenek düzeylerinde daha yüksek bilgiye ve seviye tespit sınavı olarak kullanılabilmesi için de daha geniş bir yetenek düzeyi aralığında daha yüksek bilgiye sahip olmasını sağladığı tespit edilmiştir. Bu da benimsenen yaklaşımın başarıya ulaştığının bir kanıtı olarak alınmıştır.

Bu çalışmada uygulanan yaklaşımın başarısının daha da artırılması için Sınavın Seviye Tespit Sınavı olarak kullanılırken daha keskin ölçmeyi hedeflediği özellikle $\theta=-1.00$ ve $\theta=0.00$ yetenek düzeyleri aralığında daha fazla bilgiye ve dolayısıyla daha az hataya sahip olması için sınava eklenen maddelerin mümkünse bu yetenek düzeyi aralığında daha fazla ya da güçlük düzeyi biraz daha düşük maddeleri içermesi tavsiye edilebilir. Ancak bu tavsiye dikkate alınırken sınavın bu haliyle de başarılı

bir şekilde beklenen işlevlerini yerine getirdiğinin gözlemlendiği göz önünde bulundurulmalıdır. Ayrıca bu tavsiyeye uyulmadan önce yeterlik sınavında güçlük düzeyi daha düşük soru sayısının artırılmasının yeterlik sınavı puanına ve güçlük düzeyi daha düşük soruların eklenmesinin testin madde parametreleri ortalamasına muhtemel etkisi iyice irdelenmeden bu işlem gerçekleştirilmemelidir. Tavsiyelerin testin psikometrik özellikleri üzerindeki muhtemel etkisi göz önüne alındığında benimsenen yaklaşımın ideale yakın bir sonuç verdiği söylenebilir.

Çalışmanın Sınırlılıkları

Bu çalışma Seviye Tespit Sınavı'na giren öğrencilerin puanlarına göre seviye gruplarına ayrılmasına yarayan puan aralıklarını dolayısıyla sınıflama doğruluğunu içermemektedir. Bu aralıklar her bir kurum ve sınava göre değişiklik gösterebileceğinden dolayı bu konu bu çalışmada ele alınmamıştır. Bu çalışma sadece elde edilen iki formun psikometrik özellikleri ile sınırlandırılmıştır.

Sonraki Çalışmalar için Öneriler

Bu araştırma İngilizce Yeterlik Sınavı ve Seviye Tespit Sınavlarının tek bir sınavda yeni bir yaklaşım benimsenerek kullanılabilceğini göstermiştir. Ancak bu çalışma sınavın Seviye Tespit Sınavı olarak kullanıldığında elde edilen öğrenci-seviye eşleştirmelerinin sınıflama doğruluğu analizini içermemektedir. Bundan sonraki çalışmalarda benzer bir yaklaşımla tasarlanan bir İngilizce Yeterlik Sınavı'nın seviye tespit sınavı olarak da kullanılması sonucu elde edilen seviye-öğrenci eşleştirmeleri sınıflama doğruluğu açısından incelenebilir. Ayrıca benzer bir yaklaşımla hazırlanan ve uygulanan bir İngilizce Yeterlik Sınavı'nın "Klasik Test Teorisi" ve "Modern Test Teorisi" temele alınarak elde edilecek madde parametreleri birbiri ile karşılaştırılabilir ve her iki teoride madde parametrelerine etkisi incelenebilir. Son olarak sonraki çalışmalarda bu çalışmada 20 madde ile sınırlı tutulan ek maddelerin sayısının artırılarak 20'den daha fazla sayıda madde eklenmesinin testin madde parametrelerine ve test bilgi fonksiyonlarına etkisi incelenebilir.

KAYNAKÇA

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Newyork, NY: Oxford University Press.
- Baker, F.B. (2001). *The basics of item response theory*. USA: ERIC Clearinghouse on Assessment and Evaluation.
- Başaran, S. (2005). *Diğer ülkelerde lise bitirme sınavları ve Türk eğitim sistemi için lise bitirme sınavı önerisi*. Ankara: MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Baştürk, R. (2007). Kamu personeli seçme sınavına hazırlanan öğretmen adaylarının sınav kaygı düzeylerinin incelenmesi. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 17(2), 163-176.

- Bejar, I. I. (1983). Subject matter experts' assessment of item statistics. *Applied Psychological Measurement*, 3, 303– 310.
- Chernyshenko, O.S., Stark, S., Chan, K., Drasgow, F. & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 36(4), 523-562.
- Council of Europe (2017). *Self-Assessment Grid-Turkish*. Retrieved from <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168045bb69>
- Cuesta, M. & Muniz, J. (1999). Robustness of item response logistic models to violations of the unidimensionality assumption. *Psicothema*, 11 (1), 175-182. <http://www.psicothema.com/pdf/240.pdf>
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. London : Oxford Press.
- Dorans, N. J., & Kingston, N. M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE Verbal Scale. *Journal of Educational Measurement*, 22 (4), 249-262. <https://www.jstor.org/stable/1434701>
- Drasgow, F., Levine, M., Tsien, S., Williams, B., ve Mead, A. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19(2), 143-165.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Psychology Press.
- Guyer, R., & Thompson, N.A. (2014). *User's Manual for Xcalibre item response theory calibration software, version 4.2.2 and later*. Woodbury MN: Assessment Systems Corporation.
- Hambleton, R., K., & Swaminathan H. (1985). *Item response theory: Principals and applications*. Norwell, MA: Kluwer Academic Publishers.
- Hambleton R. K., Swaminathan H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.
- IBM Corp. Released 2015. *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: IBM Corp.
- Kezer, F. (2013). *Bilgisayar ortamında bireye uyarlanmış test stratejilerinin karşılaştırılması*. (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.
- Kumandaş, H., & Kutlu, Ö. (2010). High stakes testing: does secondary education examination involve any risks? *Procedia- Social and Behavioral Sciences*, 9, 758-764.
- Manning, J. (2019a, Mayıs 14). *The high cost of placing university students in the wrong English program*. <https://www.english.com/blog/the-high-cost-of-placing-university-students-in-the-wrong-english-program/>
- Manning, J. (2019b, Temmuz 23). *What should you look for in an English Placement Test*. English.com. <https://www.english.com/blog/what-should-you-look-for-in-an-english-placement-test/>
- Özbaşı, D. (2014). *Bilgisayar okuryazarlığı testinin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliğine ilişkin bir araştırma*. (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara.

- Şahhüseyinoğlu, D. (2006). A comparison of the results of English proficiency exams through three different equation methods. *Hacettepe University Journal of Education*, 31, 115-125.
- Şahin, A. (2012). *Madde Tepki Kuramı'nda Test Uzunluğu ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Tabachnick, G.B. ve Fidell, S.L. (2007). *Using multivariate statistics*. Boston, MA: Pearson International Inc.
- Yükseköğretim Kurumlarında Yabancı Dil Öğretimi ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik. (2013). *T.C. Resmi Gazete* (28725, 1 Ağustos 2013).

EXTENDED SUMMARY

Introduction

Students are placed into the departments of universities in Turkey via a university entrance exam. Among these students, the ones who are placed into the departments that use English as the medium of instruction are subjected to the English Proficiency Exam administered by the university rectorates to determine whether their English level is adequate to follow the courses of those departments (Yükseköğretim Kurumlarında Yabancı Dil Öğretimi ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik, 2013). Students who fail to score at a certain level determined by the university senates in this English Proficiency Exam are expected to improve their English by studying at the English Preparatory Program (Intensive English Program) for one to four semesters. Before they start the English Preparatory Program, they are expected to take an English placement exam in order to determine the level at which they will start their English Preparatory Program courses.

The English Proficiency Exam and the English Placement Exam are two different exams administered for two different purposes. This affects the item selection and design of these two exams. Briefly, since the English Proficiency Exam is conducted to determine whether the students are above a desired level of English without being related to any training (Alderson, Clapham, & Wall, 1995), it is expected that it contains only the items which can be responded correctly by the students above a certain English level determined by the institution, while the English Placement Exam is expected to contain items that measure in a wider range of ability levels since it aims to assign the students to the right level in the English Preparatory Program (Alderson, Clapham, & Wall, 1995). Therefore, combining these two exams in one exam is not expected and may cause serious problems. As these two exams can be categorized as "High-Stakes Exams", if they are administered as a single exam, the design of the new exam should be done consciously with the help of a psychometrician and a psychometrically sound approach should be pursued.

To prevent crowded student groups from coming to the university campuses twice for two separate exams due to the Covid-19 pandemic, for some universities in Turkey, it has become a necessity to administer the English Proficiency Exam and the Placement Exam as a single exam. As the administration of these two exams is not a common practice, a psychometrically sound approach was developed and pursued. The results of the design approach pursued are shared in this study which is one of the first studies on this topic. The approach used in this study has the potential to be used as a base in the upcoming studies in the field.

Method

In order to give the English Proficiency exam the capability to cater for a wider range of ability levels, items to measure the ability levels between A2-B1 levels were added to the content of it. For this purpose, 20 of the 40 items measuring B1+ and B2 levels in the “Use of English” subtest were replaced with 20 items measuring skills between A2-B1 levels. In this way, it was aimed that the exam could distinguish between the students at A2 skill level from students at B1 level. As a result, the new exam to be used for both "Proficiency" and "Placement" purposes had 20 items in the "Listening" subtest, 40 items in the "Use of English" subtest (with additional 20 items), and 20 items in the "Reading" subtest, with a total of four-option 80 multiple-choice items with (except 10 three-option items). Item parameters of both 60 items (before 20 A2-B1 level items were added) and 80 items (after new items were added) were estimated in the three-parameter logistic model of the item response theory and the effects of their newly added items on the average item parameters and test information functions for the whole test and each subtest were examined.

Normal distribution (Embretson & Reise, 2000), unidimensionality, and local independence assumptions (de ayala, 2009) of the item response theory to estimate item and person parameters successfully were examined and found to be met. Moreover, the model-data fit was also examined, and was found that the model fit the data.

Two different scores, namely “Proficiency Score” and “Level Score”, were calculated for the same students from their responses to the same items. The newly added A2-B1 level items were scored as 0.25 points when calculating the “Proficiency Score” which was out of 100 points (including the essay writing section), and as one (1) point when calculating the “Level Score” which was calculated out of 80 points (without essay writing section). Thus, the effect of the relatively easy items added to the students’ “Proficiency Score” was limited to a maximum of 5 points, while it was ensured to have an effect of 20 points on the “Level Score”.

Conclusion

When the average item parameters and test information functions were examined for the whole test and subtests, it was observed that the new added 20 items made the test have higher information at

higher ability levels without affecting the high quality of the exam in terms of item parameters and that they also made it to have higher information in a wider range of ability levels in order it make it possible to be used as a placement exam. These were taken as proof of the success of the theorized approach.

In order to further increase the success of the approach developed and pursued in this study, as the exam is expected to measure more accurately especially in between $\theta=-1.00$ and $\theta=0.00$ ability levels while being used as a Placement Exam, more items with information within this range or slightly easier items could be added to the exam. However, when considering this recommendation, it should be taken into account that the current approach already functioned successfully. In addition, before this recommendation is followed, its effects on the number of questions in the proficiency exam, and the proficiency score should be thoroughly examined.