



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

FARKLI SEÇKİSİZ MADDE ÖRNEKLEME YAKLAŞIMLARINA GÖRE ANG OFF
VE EVET/HAYIR YÖNTEMLERİ İLE BELİRLENEN GEÇME PUANLARININ
KARŞILAŞTIRILMASI

Hakan KARA

Yüksek Lisans Tezi

Ankara, 2018



Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

FARKLI SEÇKİSİZ MADDE ÖRNEKLEME YAKLAŞIMLARINA GÖRE ANG OFF
VE EVET/HAYIR YÖNTEMLERİ İLE BELİRLENEN GEÇME PUANLARININ
KARŞILAŞTIRILMASI

COMPARISON OF PASSING SCORES SET BY ANG OFF AND YES/NO
METHODS WITH RESPECT TO DIFFERENT RANDOM ITEM SAMPLING
APPROACHES

Hakan KARA

Yüksek Lisans Tezi

Ankara, 2018

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Hakan KARA'nın hazırladığı "Farklı Seçkisiz Madde Örneklemeye Yaklaşımlarına Göre Angoff ve Evet/Hayır Yöntemleri İle Belirlenen Geçme Puanlarının Karşılaştırılması" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans Tezi** olarak kabul edilmiştir.

Jüri Başkanı

Dr. Öğr. Üyesi, Kübra ATALAY
KABASAKAL



Jüri Üyesi (Danışman)

Dr. Öğr. Üyesi, Sevda ÇETİN



Jüri Üyesi

Dr. Öğr. Üyesi, Funda
NALBANTOĞLU YILMAZ



Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 18 / 06 / 2018 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu araştırmada, bir standart belirleme çalışmasında uzmanların değerlendireceği madde sayısını azaltmada çeşitli seçkisiz örnekleme yöntemlerinin etkililiği incelenmiş ve bu yöntemler birbirleriyle karşılaştırılmıştır. İlgili karşılaştırmalar Angoff ve Evet-Hayır yöntemlerine göre ayrı ayrı yapılmıştır. Araştırmada öncelikle SBS 2012 ve 2013 matematik alt testleri birleştirilerek tüm test oluşturulmuş, sonrasında tüm testten farklı büyüklükte alt testler (%30, %40, %50 ve %70) oluşturmak için basit (BSÖ), içerik tabakalı (İ-TSÖ), madde güçlüğü tabakalı (G-TSÖ) ve içerik-madde güçlüğü tabakalı (İG-TSÖ) seçkisiz örnekleme yöntemleri kullanılmıştır. Toplam 16 çalışma durumu (4 yöntem x 4 alt test) değerlendirilmiştir. Verilerin analizinde, alt testlere ilişkin geçme puanının tüm test geçme puanından manidar düzeyde farklılaşıp farklılaşmadığı ANOVA analiziyle incelenmiş, her bir durum için RMSE ve TSH (tahmini standart hata) değerleri hesaplanmıştır. Elde edilen bulgulara göre, yalnızca %30-İ-TSÖ, %40-İ-TSÖ, %30-G-TSÖ ve %30-İG-TSÖ çalışma durumlarında kestirilen ortalama Angoff geçme puanları tüm testin geçme puanından (45.12) manidar düzeyde farklılaşmıştır. Benzer şekilde, yukarıdaki çalışma durumlarında ve %50-İG-TSÖ çalışma durumunda kestirilen ortalama Evet-Hayır geçme puanları tüm testin geçme puanından (56.25) manidar düzeyde farklılaşmıştır. RMSE hata değerlerine göre en küçük hata İ-TSÖ yöntemi, en büyük hata ise BSÖ yöntemi kullanıldığında hesaplanmıştır. Ayrıca, TSH incelemeleri sonucunda elde edilen bulgularda, tüm testin Angoff geçme puanına benzer (1 TSH içerisinde) kestirimler elde etmek için İ-TSÖ yöntemiyle maddelerin %50'sinin seçilmesinin yeterli olduğu görülmüştür. Öte yandan 16 çalışma durumunun hiçbirisiyle tüm testin Evet-Hayır geçme puanına benzer kestirimler elde edilemediği saptanmıştır. Sonuç olarak, Angoff ve Evet/Hayır yöntemleriyle yapılan MGP belirleme çalışmalarında uzmanların değerlendireceği madde sayısının azaltılmasında İ-TSÖ yöntemi diğer yöntemlere kıyasla daha etkili olmuştur.

Anahtar sözcükler: standart belirleme, angoff, evet-hayır, seçkisiz örnekleme yöntemleri, minimum geçme puanı

Abstract

In this study, the efficiency of various random sampling methods in order to reduce the number of items rated by judges in a standard setting study were examined and those methods were compared with each other. Those comparisons were made with respect to Angoff and Yes-No methods separately. Firstly, full-length test was formed by combining SBS 2012 and 2013 mathematics subsets. After then, simple random sampling (SRS), content stratified (C-SRS), item-difficulty stratified (D-SRS) and content-by-difficulty random sampling (CD-SRS) methods were used in order to constitute different length of subsets (30%, 40%, 50%, 70%) from the full-test. In total, 16 different study conditions (4 methods x 4 subsets) were investigated. In data analysis part, ANOVA analysis were conducted in order to examine whether minimum passing scores (MPSS) for the subsets were significantly different from the MPSSs of the full-length test. As a follow up analysis, RMSE and SEE (Standard Error of Estimation) values were calculated for each study condition. Results indicated that only for 30% C-SRS, 40% C-SRS, 30% D-SRS and 30% CD-SRS study conditions, estimated Angoff MPSSs were significantly different from the full-test Angoff MPS (45.12). In addition to the conditions above, 50% CD-SRS were also had an estimated Yes-No MPS significantly different from the full-test Yes-No (56.25). According to RMSE values, C-SRS method had smallest error while SRS method had biggest one for both Angoff and Yes-No cases. Moreover, SEE examinations revealed that in order to achieve estimations similar to the full-test Angoff MPS (within one SEE), it is sufficient to get 50% of items with C-SRS method. On the other side, similar estimations to the full-test Yes-No MPS could not been obtained in any of 16 study conditions. In conclusion, C-SRS method were more effective one compared to the others in reducing the number of items rated by judges in MPS setting studies conducted with Angoff and Yes-No methods.

Keywords: standard setting, angoff, yes-no, random sampling methods, minimum passing scores

Teşekkür

Yüksek lisans tez çalışmam boyunca bana her konuda destek veren, beni yönlendiren ve başarabileceğime inanan danışmanım, değerli hocam Dr. Öğr. Üyesi Sevda ÇETİN'e, dönütleriyle tezime katkı sağlayan değerli jüri üyelerim Dr. Öğr. Üyesi Kübra ATALAY KABASAKAL'a ve Dr. Öğr. Üyesi Funda NALBANTOĞLU YILMAZ'a,

Tez çalışmama başladığım ilk günden bu güne kadar süren bu uzun yolculuğun her ânında eşsiz destek ve katkılarıyla hep yanımda olan, bütün sıkıntıları dinleyen ve çözüm bulmaya çalışan, varlığı ile kendimi iyi hissetmemi sağlayan ve başarmamda büyük role sahip olan canım eşim Başak ERDEM KARA'ya,

Her zaman yanımda olduklarını hissettiğim, her konuda mutluluğumu ve heyecanımı paylaştığım canım annem, babam ve kardeşlerime,

Keyifli vakit geçirerek tez stresinden uzaklaşmamı sağlayan arkadaşlarım Ferhat TUŞTAŞ'a, Ahmet SAKA'ya, Başak ÇİĞDEMTEKİN'e, Bulut YILDIZTEKİN'e ve Haydar KARAMAN'a,

Çalışmaya katılarak verilerin toplanması konusunda yardımlarını esirgemeyen değerli öğretmen ve akademisyen arkadaşlarıma,

Son olarak bugünlere gelmemde emeği geçen herkese

Çok teşekkür ederim...

İçindekiler

Öz.....	ii
Abstract	iii
Teşekkür.....	iv
Tablolar Dizini.....	vii
Simgeler ve Kısaltmalar Dizini	ix
Bölüm 1 Giriş.....	1
Problem Durumu	1
Araştırmanın Amacı ve Önemi	4
Araştırma Problemi	5
Sayıtlılar	6
Sınırlılıklar	6
Tanımlar	6
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	8
Standart Belirleme.....	8
Standart Belirleme Yöntemleri.....	9
Standart Belirleme Sürecindeki Meseleler	11
Seçkisiz Madde Örnekleme Yöntemleri	13
Madde Sayısının Azaltılması ile İlgili Çalışmalar	14
Seçkisiz Örnekleme Yöntemlerinin Karşılaştırılması ile İlgili Çalışmalar	16
İlgili Araştırmalar Özet.....	18
Bölüm 3 Yöntem	20
Araştırmanın Yöntemi	20
Araştırmanın Evreni ve Örnekleme	20
Veri Toplama Süreci.....	21
Veri Toplama Araçları	21
Verilerin Analizi	22

Bölüm 4 Bulgular ve Yorumlar	31
Alt Problemlere İlişkin Bulgular ve Yorumlar	31
Bölüm 5 Sonuç, Tartışma ve Öneriler.....	52
Sonuç ve Tartışma	52
Öneriler	54
Kaynaklar	56
EK-A: Uzman Görüşleri için Gönüllü Katılım Formu	63
EK-B: Angoff Yöntemine göre Uzmanların Güçlük Tahminleri	82
EK-C: Evet/Hayır Yöntemine göre Uzmanların Kararlarına İlişkin Puanlar.....	84
EK-Ç: Angoff ve Evet/Hayır Yöntemine göre Uzmanların Bireysel Geçme Puanları	86
EK-D: Alt Testlerin Oluşturulmasında Kullanılan Bir R Kodu Örneği	87
EK-E: ANOVA Analizi Varsayım Kontrolleri (Normallik).....	88
EK-F: ANOVA Analizi Varsayım Kontrolleri (Varyansların Homojenliği)	89
EK-G: İçerik Kategorilerine Göre Alt Testlere Seçilecek Madde Sayıları.....	89
EK-Ğ: Güçlük Gruplarına Göre Alt Testlere Seçilecek Madde Sayıları	89
EK-H: İçerik-Güçlük Kategorilerine Göre Alt Testlere Seçilecek Madde Sayıları..	90
EK-I: Etik Komisyonu Onay Bildirimi.....	91
EK-İ: Etik Beyanı	92
EK-J: Yüksek Lisans Tez Çalışması Orijinallik Raporu.....	93
EK-K: Thesis Originality Report.....	94
EK-L: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	95

Tablolar Dizini

Tablo 1 <i>Matematik Alt Testlerine İlişkin Betimsel İstatistikler</i>	24
Tablo 2 <i>Tüm Teste İlişkin Madde Güçlük İndeksleri</i>	25
Tablo 3 <i>Tüm Testin İçerik Kategorilerindeki Madde Sayısı ve Numaraları</i>	26
Tablo 4 <i>Tüm Testin Güçlük Kategorilerindeki Madde Sayısı ve Numaraları</i>	27
Tablo 5 <i>Tüm Testin İçerik-Güçlük Tabakalarındaki Madde Sayısı ve Numaraları</i>	28
Tablo 6 <i>BSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine İlişkin Betimsel İstatistikler</i>	32
Tablo 7 <i>BSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine ait RMSE ve Alt Test Yüzdeleri</i>	33
Tablo 8 <i>İ-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'sine İlişkin Betimsel İstatistikler</i>	34
Tablo 9 <i>İ-TSÖ ile Oluşturulan Alt Testlerin Angof MGP'sine ait RMSE ve Alt Test Yüzdeleri</i>	35
Tablo 10 <i>G-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine İlişkin Betimsel İstatistikler</i>	36
Tablo 11 <i>G-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine Ait RMSE ve Alt Test Yüzdeleri</i>	37
Tablo 12 <i>İG-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'sine İlişkin Betimsel İstatistikler</i>	38
Tablo 13 <i>İG-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'sine ait RMSE ve Alt Test Yüzdeleri</i>	39
Tablo 14 <i>Angoff MGP'lerin Örnekleme Durumlarına Ait RMSE ve Alt Test Yüzdesi Değerleri</i>	40
Tablo 15 <i>BSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler</i>	41
Tablo 16 <i>BSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine ait RMSE ve Alt Test Yüzdeleri</i>	43
Tablo 17 <i>İ-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler</i>	44
Tablo 18 <i>İ-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine ait RMSE ve Alt Test Yüzdesi Değerleri</i>	45

Tablo 19 <i>G-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler</i>	46
Tablo 20 <i>G-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine Ait RMSE ve Alt Test Yüzdesi Değerleri</i>	47
Tablo 21 <i>İG-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler</i>	48
Tablo 22 <i>İG-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine Ait RMSE ve Alt Test Yüzdesi Değerleri</i>	49
Tablo 23 <i>Evet/Hayır MGP'lerin Örnekleme Durumlarına Ait RMSE Değerleri</i>	50
Tablo 24 <i>Evet/Hayır MGP'lerin Örnekleme Durumlarına Ait Alt Test Yüzde Değerleri</i>	51



Simgeler ve Kısaltmalar Dizini

MGP: Minimum Geme Puanı

BSÖ: Basit Sekisiz Örnekleme

İ-TSÖ: İerik Tabakalı Sekisiz Örnekleme

G-TSÖ: Gülük Tabakalı Sekisiz Örnekleme

İG-TSÖ: İerik-Gülük Tabakalı Sekisiz Örnekleme

TSH: Tahminin Standart Hatası



Bölüm 1

Giriş

Bu bölümde; problem durumu, araştırmanın amacı ve önemi, araştırma problemi, sayıtlar, sınırlılıklar ve tanımlar yer almaktadır.

Problem Durumu

Günlük hayatın neredeyse her anında insanlar karar verme süreciyle karşı karşıya kalmaktadır. Doğru kararlar verebilmek için bilgiye ihtiyaç duyulmaktadır ve karar alan kişilere doğru ve uygun bilgiyi sağlamak, ölçme ve değerlendirmenin amacıdır (Mehrens & Lehmann, 1991). Ölçme, birey ya da nesnelerin niteliklerinin uygun araçlar kullanarak gözlenip, gözlem sonuçlarının sembollerle gösterilmesi olarak tanımlanırken (Büyüköztürk, Çakmak, Akgün, Karadeniz & Demirel, 2016); değerlendirme, ölçme sonuçlarının bir ölçütle kıyaslanarak birey ya da nesnelerin ölçülen nitelikleri hakkında bir karara varılması olarak tanımlanabilir (Turgut & Baykul, 2012). Karar verme sürecinde bir ölçüt ya da standardın belirlenmesi daha adil, mantıklı, geçerli ve savunulabilir kararlar alınmasına imkân tanımaktadır. Ayrıca, standardın açıkça ifade edilmesi hem başarıya olan ilgiyi artırmakta hem de alınan kararların daha güvenilir ve anlaşılabilir olmasını sağlamaktadır (Cizek & Bunch, 2007).

Standart, “Ne kadar yeterlidir?” sorusunun cevabı olarak düşünülebilir. Eğitimde farklı amaçlar için kullanılan çeşitli standartlar vardır. İçerik standartları, eğitimsel sınavlar bağlamında testin amaçları, kazanımları ve çıktılarını ifade etmede; performans standartları ise, testin geçme puanını belirlemede kullanılmaktadır. Standartlar, mutlak ya da bağıl olabilir. Bağıl standartlarda, bireylerin birbirlerine göre başarıları önemli iken; mutlak standartlarda önemli değildir. Bağıl standartlarda, testi alan her bir bireyin puanı, testi alan diğer bireylerin puanlarıyla karşılaştırıldığı için bağıl standartlar gruba bağlıdır ve norm temelli olarak da adlandırılmaktadır. Testi alan bireylerin elde ettikleri puanlar yükseldikçe bağıl standart da o kadar yükselir. Mutlak standartlar ise, testi alan bireylere bağlı değildir ve sınavdan önce belirlenen ölçüt temelli standartlar olarak da bilinmektedirler. Bu standartlarda, testi alan bir birey için diğer kişilerin puanları önemli değildir; çünkü onların puanları standardı etkilememektedir (Livingston & Zieky, 1982). Örneğin; “Soruların % 60’ını doğru cevaplayanlar geçecek” ifadesi

mutlak bir standart oluřtururken; “Sınava girenlerin % 60’ı geecek” ifadesi baėıl bir standart oluřturmaktadır.

Bir testte bir ya da birden fazla geme puanı belirleme sreci olarak tanımlanan standart belirleme, son zamanlarda birçok alanda nemli kararlar alabilmek iin gerekli hale gelmiřtir. Bu kararlar, bařta eėitim olmak zere saėlık, hukuk vb. alanlarındaki seme, sınıflama, lisans veya sertifika verme kararlarıdır. Bu kararların doėruluėu, ltn (standartın) doėru belirlenmiř olmasına baėlıdır. Standartın doėru belirlenmiř olması da uygun standart belirleme yntemlerinin seilmesine ve kullanılmasına yani srecin etkili bir řekilde izlenmesine baėlıdır (Downing, 2006; Kane, 2001).

Standart belirleme iin literatrde 50’den fazla yntem bulunmaktadır (Smith, 2011). Farklı yntemlerin aynı sınav iin benzer standartlar verip vermediėi pek ok alıřmada incelenmiř ve yntem seiminin geme puanları zerinde etkili olduėu, farklı yntemlerin aynı sınav zerinde farklı geme puanları ortaya ıkarabileceėi sonucuna varılmıřtır (Berk, 1996; etin, 2011; Irwin, 2007; Jaeger, 1989; Kane, 1998; Mehrens, 1995). Bu nedenle, standart belirleme alıřmalarında birden fazla yntemin bir arada kullanımı nerilmiřtir. Sonuların benzer ıkması elde edilen geme puanını desteklerken, farklı ıkması ise sonuların tekrar gzden geirilmesine dair bir neri sunmaktadır (Cizek, 2001; Irwin, 2007).

Benzer řekilde; Behuniak, Archambault ve Gable (1982), Angoff ve Nedelsky yntemlerini karřılařtırdıkları alıřmalarında standartların sadece yntemler arasında deėil, aynı yntemi kullanan uzmanlar arasında da deėiřtiėini saptamıřlardır. Bu sonu, uzman kanısına dayalı standart belirleme yntemleri kullanıldıėı durumlarda, aynı yntemin aynı sınav zerinde farklı sonular verebileceėi řeklinde yorumlanabilir. Ayrıca, Smith (2011) uzmanların hipotetik bir birey hakkında yargıda bulunmalarını gerektiren standart belirleme yntemlerinin, bu kiřiler iin biliřsel olarak olduka yorucu olabildiėini belirtmiřtir. Uzmanlardan beklenen biliřsel aba zellikle Angoff yntemi iin nemli bir eleřtiri kaynaėı oluřturmuřtur (Lewis, Green, Mitzel, Baum & Patz, 1998). Uzmanların her bir madde iin minimum yeterlik dzeyindeki bireyin performansına iliřkin kestirimde bulunması ve bunu her bir performans dzeyi iin yapması beklenmektedir. Bu nedenle, uzmanlardan beklenen iřlemler zaman alıcı, yorucu ve biliřsel olarak zorlayıcı hale gelebilmektedir. Sonu olarak, eėer her bir uzmanın deėerlendireceėi

soru sayısı azaltılabilirse, uzmanlar daha az soru değerlendirecekleri için daha az zaman alıcı ve yorucu işlemler sonucunda daha doğru değerlendirme yapabilirler (Ferdous & Plake, 2007; Smith, 2011).

Uzmanların değerlendireceği soru sayısının azaltılması iki farklı yolla mümkündür. Bunlardan ilki, bir standart belirleme çalışmasındaki madde sayısını, tüm testi temsil eden bir alt test oluşturacak şekilde azaltmak, böylelikle uzmanların gözden geçireceği toplam madde sayısını düşürmektir (Buckendahl, Ferdous & Gerrow, 2010; Ferdous & Plake, 2005, 2007). İkincisi ise testi, tüm testi temsil eden daha küçük alt testlere bölerek bu alt testleri değerlendirmek üzere eşit sayıda uzman alt grupları tahsis etmektir (Norcini, Shea & Ping, 1988; Plake & Impara, 2001; Sireci, Patelis, Rizavi, Dillingham & Rodriguez, 2000). Çoklu matris örnekleme (Norcini vd., 1988) ve dengeli tamamlanmamış blok tasarımı (Sireci vd., 2000) yöntemleri ikinci yolun çeşitli seçenekleri arasındadır. Bu yolla yapılan çalışmalarda, göz önünde bulundurulan toplam madde sayısı değişmezken, her bir uzmanın göz önünde bulunduracağı madde sayısı azalmaktadır.

Yukarıdaki bilgiler ışığında, standart belirlemenin eğitimde alınan kararlara temel oluşturması açısından önemli olduğu, verilen kararların doğruluğunun doğru standardı belirlemeye bağlı olduğu görülmektedir. Doğru standart belirleme noktasında birden fazla standart belirleme yönteminin bir arada kullanımı önerilmektedir. Ayrıca, özellikle Angoff ve Evet-Hayır yöntemlerinin kullanıldığı standart belirleme süreçlerinin çok zaman alması, çok yorucu olması ve fazla bilişsel çaba gerektirmesi göz önünde bulundurulduğunda, madde azaltma yoluna gidilmesi tavsiye edilmiştir. İlgili literatür incelendiğinde, Angoff yönteminin kullanıldığı standart belirleme çalışmalarında madde sayısının azaltılmasına ilişkin çalışmaların var olduğu (Ferdous & Plake, 2005; Ferdous & Plake, 2007; Kannan, Katz, Sgammato & Tannenbaum Katz, 2015; Plake & Impara, 2001; Smith, 2011) görülmüş; ancak madde sayısının azaltılması noktasında tabakalı seçkisiz örnekleme yöntemlerinin (içerik tabakalı [İ-TSÖ], güçlük tabakalı [G-TSÖ], içerik-güçlük tabakalı [İG-TSÖ], içerik-güçlük-ayrıt edicilik tabakalı [İGA-TSÖ] vb.) etkililiğinin incelendiği çalışmaların sınırlı kaldığı belirlenmiştir. Araştırma kapsamında Seviye Belirleme Sınavı (SBS) Matematik alt testine ilişkin geçme puanları Angoff ve Evet-Hayır yöntemleri kullanılarak belirlenmeye çalışılmıştır. Madde sayısının az olması nedeniyle, bu çalışmada sadece madde güçlük

indeksleri ve içerik alanları dikkate alınarak alt testler oluşturulmuş ve tüm testin geçme puanına benzer kestirimler elde edebilmek için test maddelerinin yüzde kaçının yeterli olabileceği incelenmiştir.

Araştırmanın Amacı ve Önemi

Bu çalışmada, Angoff ve Evet/Hayır yöntemlerinin kullanıldığı standart belirleme çalışmalarında değerlendirilen toplam madde sayısını azaltmada kullanılabilecek seçkisiz örnekleme yöntemlerinin etkililiğini incelemek ve farklı seçkisiz örnekleme yöntemlerini (basit seçkisiz örnekleme [BSÖ], içerik tabakalı [İ-TSÖ], güçlük tabakalı [G-TSÖ], içerik-güçlük tabakalı [İG-TSÖ]) birbirleriyle karşılaştırmak amaçlanmıştır. İ-TSÖ yönteminde sadece içerik kategorileri; G-TSÖ yönteminde sadece güçlük kategorileri; İG-TSÖ yönteminde ise hem içerik hem de güçlük kategorileri dikkate alınarak alt testler oluşturulmuştur.

Genellikle, standart belirleme çalışmaları sürecinde tavsiye edilen minimum geçme puanları iki turda elde edilmektedir (Hambleton, 1998; Reckase, 2001). Uzmanların her turda, minimum yeterlik düzeyindeki öğrenciyi düşünmesi ve her bir madde için ayrı ayrı tahminler yapması çok zaman alabilmektedir. Bu zorlu ve uzun süreç, uzmanların kararlarını sağlıklı bir şekilde almalarına engel olabilmektedir. Bu nedenle, her bir uzmanın değerlendireceği madde sayısının azaltılmasının, uzmanların puanlama kalitesini yükselteceği düşünülmektedir (Ferdous & Plake, 2005). Madde sayısını azaltmanın yöntemlerinden birisi, seçkisiz madde örnekleme yöntemleri kullanılarak tüm testi temsil edecek şekilde alt testlerin oluşturulmasıdır. Bu çalışma kapsamında, SBS 2012 ve SBS 2013 matematik testleri birleştirilerek 40 madde içeren tüm test oluşturulmuş ve bu tüm testten, farklı seçkisiz örnekleme yöntemleriyle alt testler türetilmiştir. Oluşturulan tüm test ve alt testler için minimum geçme puanları (MGP), Angoff ve Evet-Hayır yöntemlerine göre belirlenmiş ve karşılaştırılmıştır. Bu şekilde, farklı seçkisiz örnekleme yöntemlerinin etkililiği de incelenmiştir. Bu açılarından bakıldığında çalışmanın, standart belirleyen kurum ve kişilere, özellikle geniş ölçekli sınavlarda hangi yöntemin nasıl kullanılabileceği hakkında önemli bilgiler sunacağı düşünülmektedir. Bunun yanı sıra, çalışmanın, bir testin geçme puanına benzer kestirimler elde edebilmek için test maddelerinin yüzde kaçının yeterli olabileceği noktasında fikir verebileceği; bu sayede, soru sayısının azaltılmasıyla birlikte uzmanların iş yüklerinin hafifletilerek daha sağlıklı kararlar almalarına yardımcı olacağı düşünülmektedir.

İlgili literatür incelendiğinde, standart belirleme yöntemlerinin karşılaştırılmasına ilişkin çalışmaların çok fazla olduğu; fakat, her bir uzmanın dikkate alacağı madde sayısının azaltılmasına ilişkin tabakalı seçkisiz madde örnekleme yöntemlerinin etkililiğinin incelendiği çalışmaların ise sınırlı sayıda olduğu görülmüştür (Ferdous & Plake, 2005, 2007; Ferdous, 2007; Kannan, Sgammato, Tannenbaum & Katz, 2015; Smith & Ferdous, 2007; Smith, 2011). Bu çalışma kapsamında, Angoff ve Evet/Hayır Angoff yöntemleriyle yapılan standart belirleme çalışmalarında değerlendirilen toplam madde sayısını azaltmada kullanılabilecek seçkisiz örnekleme yöntemlerinin etkililiğini incelemek ve farklı seçkisiz örnekleme yöntemlerini (BSÖ, İ-TSÖ, G-TSÖ, İG-TSÖ) birbirleriyle karşılaştırmak amaçlanmıştır. Madde sayısının azaltılması noktasında, tüm testi temsil edebilecek alt testler oluşturacak şekilde azaltma yoluna gidilmiş ve işlemler sırasında farklı seçkisiz örnekleme yöntemleri kullanılmıştır. Bu yöntemler araştırmanın kuramsal temeli bölümünde detaylı şekilde açıklanmıştır.

Araştırma Problemi

Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin geçme puanları, seçkisiz madde örnekleme yöntemlerine göre farklılaşmakta mıdır?

Alt problemler. Bu çalışma kapsamında aşağıdaki alt problemlere yanıt aranmıştır.

1. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Angoff geçme puanları arasında;

- a. Basit seçkisiz örnekleme (BSÖ),
- b. İçerik tabakalı seçkisiz örnekleme (İ-TSÖ),
- c. Madde güçlüğü tabakalı seçkisiz örnekleme (G-TSÖ) ve
- d. İçerik ve madde güçlüğü tabakalı seçkisiz örnekleme (İG-TSÖ) yöntemine göre manidar düzeyde fark var mıdır?

2. Farklı seçkisiz madde örnekleme yöntemleriyle oluşturulan alt testlere ilişkin ortalama Angoff geçme puanları birbirlerinden nasıl farklılaşmaktadır?

3. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Evet/Hayır geçme puanları arasında;

- a. Basit seçkisiz örnekleme (BSÖ),
- b. İçerik tabakalı seçkisiz örnekleme (İ-TSÖ),
- c. Madde güçlüğü tabakalı seçkisiz örnekleme (G-TSÖ) ve
- d. İçerik ve madde güçlüğü tabakalı seçkisiz örnekleme (İG-TSÖ) yöntemine göre manidar düzeyde fark var mıdır?

4. Farklı seçkisiz madde örnekleme yöntemleriyle oluşturulan alt testlere ilişkin ortalama Evet/Hayır geçme puanları birbirlerinden nasıl farklılaşmaktadır?

Sayıtlılar

- Uzmanların test maddelerine ilişkin kararlarını verirken minimum yeterlik düzeyi tanımını göz önünde bulundurdıkları varsayılmıştır.
- Uzmanların veri toplama formundaki minimum yeterlik düzeyi tanımını aynı şekilde algılayarak puanlama yaptıkları varsayılmıştır.
- Uzmanların test maddelerine ilişkin yargıda bulunurken, Angoff ve Evet/Hayır standart belirleme yöntemlerine uygun biçimde puanlama yaptıkları varsayılmıştır.

Sınırlılıklar

- Araştırma, SBS 2012 ve 2013 Matematik testindeki toplam 40 madde ile sınırlıdır.
- Geçme puanı belirleme süreci, standart belirleme yöntemlerinden Angoff ve Evet/Hayır yöntemi ile sınırlıdır.
- Araştırma, test maddelerini değerlendiren 28 uzmanın yargıları ile sınırlıdır.
- Araştırma, madde örnekleme yöntemlerinden BSÖ, İ-TSÖ, G-TSÖ ve İG-TSÖ yöntemleri ile sınırlıdır.

Tanımlar

Minimum Yeterlik Düzeyi: Uluslararası Öğrenci Değerlendirme Programı (PISA) matematik yeterlik düzeylerinde yer alan düzey 2, minimum matematik yeterlik düzeyi olarak tanımlanmıştır. Tanım daha detaylı olarak EK-A'deki Uzman

Görüşleri için Gönüllü Katılım Formu'nda verilmiştir. Uzmanlar maddelere ilişkin yargılarını bu düzeyi esas alarak vermişlerdir.

Minimum Geçme Puanı (MGP): Minimum yeterlik düzeyinde olan öğrenci ile bu düzeye ulaşamamış öğrenciyi ayırt edebilen uygun performans noktası.

Uzman: Matematik bilgisine sahip, alanında uzmanlaşmış, maddelere ilişkin yargılarda bulunarak geçme puanının belirlenmesinde rol oynayan puanlayıcılar.



Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde, standart belirleme konusuna ilişkin kavramsal çerçeve çizilmiş ve literatürde yer alan ilgili araştırmalara yer verilmiştir.

Standart Belirleme

Standart belirleme, bir teste bir ya da daha fazla geçme puanı belirleme sürecidir (Cizek & Bunch, 2007). Geçme puanı ise yeterli düzeyde olan öğrenci ile yeterli düzeyde olmayan öğrenciyi ayırt eden performans noktasıdır. Cizek (1993)'e göre standart belirleme, iki veya daha fazla performans derecesini ayıran bir sayının belirlenmesinde takip edilen kurallar veya süreçler bütünüdür. Cizek (1993), geçme puanı ve performans standardının eş anlamlı olarak kullanılabileceğini belirtirken; Kane (1994), geçme puanı ile performans standardı arasında bir ayırma gitmenin önemli olabileceğini savunmuştur. Ona göre performans standardı, yeterli performans düzeyindeki kişinin sahip olması gereken bilgi ve beceriyi; geçme puanı ise puanlama ölçeğindeki bir noktayı göstermektedir. Performans standardı istenilen yeterlik düzeyinin kavramsal tanımı iken; geçme puanı, istenilen yeterlik düzeyinin işlevsel tanımıdır. Böyle bir ayırma gitmek, geçme puanları ile ilgili yorumların geçerli olup olmadığını belirlemekte kolaylık sağlamaktadır (Kane, 1994).

Bir test için geçme puanı belirlemeden önce performans düzeyleri açık bir şekilde tanımlanmalıdır. Performans düzeyleri ile bireylerin belli bir performans düzeyinde sahip olması gereken bilgi, beceri ve yetenekleri tanımlanmaktadır. Bireyler, test performanslarına göre geçti-kaldı; temel, yeterli veya ileri düzey gibi performans düzeylerine sınıflandırılabilir. Genellikle, kullanılacak olan performans düzeylerini karar alıcılar belirlemektedir (Zieky, Perie & Livingston, 2008). Özellikle Amerika Birleşik Devletleri gibi eyaletlerden oluşan ülkelerde okulların çoğu kendi standartlarını belirlemekte ve öğrenciler de temel, yeterli ve ileri düzey gibi performans düzeylerine ayrılabilir (Cizek & Bunch, 2007). Performans düzey sayısı arttıkça, düzeyler arasındaki anlamlı farklılıkları ölçmek zorlaşabilmektedir. Bu nedenle, mümkün olduğunca az sayıda düzey kullanılması gerektiği tavsiye edilmektedir (Zieky vd., 2008). Performans düzey tanımları, performans standartlarının ve standart belirleme sürecinin geçerliği açısından önemlidir (Hambleton, 2001). Bu nedenle, geçme puanı belirleme sürecinden önce

performans düzey tanımları açıkça yapılmalı ve tüm uzmanlar tarafından kabul edilmelidir.

Standart Belirleme Yöntemleri

Standart belirlemede kullanılmak üzere Wang, Pan ve Austin (2003), 50'den daha fazla yöntem tanımlamıştır. Fakat, farklı yöntemlerin farklı sonuçlar ürettiği görülmüştür (Berk, 1996; Jaeger, 1989; Mehrens, 1995).

Norcini ve Guille (2002), standart belirleme yöntemlerini mutlak ve bağıl yöntemler olarak iki gruba ayırmıştır. Bağıl standart belirleme yöntemlerinde, bireylerin önceden belirlenmiş belli bir yüzdesine göre geçti-kaldı kararları alınır. Birey puanları, bireylerin yeterlik seviyelerine göre değil, diğer puanlara kıyasla daha iyi veya daha kötü olarak yorumlanır (Hambleton & Pitoniak, 2006). Downing (2006), geçme puanlarının bireylerin hangi yeterlikte olduğunu yansıtamamasının, bağıl standart belirleme yöntemlerinin bir dezavantajı olduğunu söylemiştir. Mutlak standart belirleme yöntemlerinde ise, uzman kişiler test maddeleri ve öğrenciler hakkında bilgi sahibi olurlar ve karar alırlar. Bu yöntemlerde amaç, bireylerin performans seviyelerini diğer bireylerle karşılaştırmadan belirlemektir (McCauley, 1996). Bağıl yöntemler seçme ve yerleştirme sürecinde kullanılırken; mutlak yöntemler ise, bireylerin yeteneklerini ve ihtiyaçlarını belirlemede kullanılır (Freeman & Miller, 2001; McCauley & Swisher, 1984). Mutlak standart belirleme yöntemleri klasik rasyonel, Madde Tepki Kuramı'na (MTK)-dayalı rasyonel ve ampirik yöntemler olarak 3 temel gruba ayrılır. Test içeriğinin değerlendirilmesine dayalı klasik rasyonel yöntemler Nedelsky (1954), Angoff (1971) ve Angoff varyasyonlarını içerirken; MTK-dayalı rasyonel yöntemler İşaretleme (Lewis, Mitzel & Green, 1996), Madde Eşleştirme (Wang, Wiser & Newman, 2001) ve Mapmark (Schulltz & Mitzel, 2005) yöntemlerini içermektedir. Uygun bir performans standardı belirlemek için dış ölçüt değişkenindeki birey dağılımını kullanan ampirik yöntemler ise Benzer Gruplar ve Sınır Grup (Livingston & Zieky, 1982) yöntemlerinden oluşmaktadır. Bu araştırmada ise klasik rasyonel yöntemlerden Evet/Hayır Angoff (Angoff, 1971; Impara & Plake, 1997) ve Angoff (Angoff, 1971) yöntemleri kullanılmıştır.

Angoff yöntemi. 1971 yılında William H. Angoff tarafından önerilen Angoff yöntemi, özellikle çoktan seçmeli sınavlar için en sık kullanılan standart belirleme

yöntemlerinden birisidir (Impara, 1995; Plake, 1998). Angoff yönteminin farklı çeşitleri olmasına rağmen, en sık kullanılanı Değiştirilmiş Angoff yöntemidir (Smith, 2011). Angoff (1971) Değiştirilmiş Angoff yöntemini aşağıdaki gibi tanımlamıştır:

“Bu yöntemde, her bir uzmandan minimum yeterlik düzeyindeki bir öğrencinin her bir maddeyi doğru cevaplayabilme olasılıklarını belirtmeleri istenir. Aslında, uzmanlar sadece bir öğrenciyi değil, minimum yeterlik düzeyindeki öğrenci sayısını düşünür ve her bir maddeyi doğru cevaplandırarak minimum yeterlik düzeyindeki öğrencilerin yüzdesini tahmin eder. Bu olasılıkların veya yüzdelerin toplamı geçme puanını ifade eder.”

Bu yöntemde, ilk olarak “minimum yeterlik düzeyindeki öğrenci” tanımı uzmanlar için açık hale getirilir ve sonra madde güçlüğüne dayalı olarak uzmanlar tek tek her madde için maddeyi doğru cevaplandırabilecek minimum yeterlik düzeyindeki öğrenci yüzdelerini belirler (Reckase, 2000). Daha sonra, testteki bütün maddelerin tahmin yüzdeleri toplanarak her bir uzmanın geçme puanı hesaplanır ve uzmanların geçme puanlarının ortalamaları alınarak tüm testin geçme puanı elde edilir.

Angoff yönteminin uygulanması sürecinde geri bildirim, uzmanların her bir madde için belirledikleri yüzdelerin tartışılması veya madde ile ilgili bazı istatistiklerin sunulması şeklinde verilebilir. Grup tartışmalarını ve uzman kararlarını içeren çok türlü standart belirleme süreci; uzmanların, geçme puanını genellikle daha iyi tahmin etmesine olanak sağlamıştır (Smith, 2011).

Evet/Hayır Angoff yöntemi. Evet/Hayır Angoff yöntemi, (Angoff, 1971; Impara & Plake, 1997) Angoff yönteminin orijinalidir. Evet/Hayır yönteminde de uzmanlar öncelikle minimum yeterlik düzeyindeki bireyi açıkça tanımlamaya çalışırlar ve sonra minimum yeterlik düzeyindeki öğrencinin doğru cevaplayabileceğini düşündükleri maddeler için “Evet” doğru cevaplayamayacağını düşündükleri maddeler için ise “Hayır” kararını verirler (Smith, 2011). “Evet” yanıtları için 1, “Hayır” yanıtları için 0 puan verilir ve uzmanların testteki her bir maddeye ilişkin puanları toplanarak her bir uzman için geçme puanı hesaplanır. Uzmanların geçme puanlarının ortalamaları alınarak da tüm testin geçme puanı elde edilir (Angoff, 1971).

Standart Belirleme Sürecindeki Meseleler

Kane (1994), standart belirleme kararlarının geçerliliğini iki varsayım üzerinden değerlendirmiştir. Birinci varsayıma göre, betimsel varsayım, geçme puanının üzerinde puan alan bireyler belirli bir performans standardını karşılayabilecekken; geçme puanının altında puan alan bireyler o standardı muhtemelen karşılayamayacaktır. İkinci varsayıma göre ise, karar varsayımı, önceden belirlenmiş performans standardı, geçme puanı kararının amacına uygun olmalıdır. Örneğin, bir lisans sınavı için düşük standartlar belirlenebilirken; çok kişinin katıldığı işe alım sınavlarında daha yüksek standartlar belirlenebilir (Kane, 2001).

Geçerlik, önerilenin açık bir ifadesini gerektirir (Cronbach, 1988). Bu nedenle, geçme puanından sonuçlara kadar elde edilen çıkarımlar ve öne sürülen varsayımlar akla yatkın olmalıdır (Crooks, Kane & Kohen, 1996). Kane (1994), performans standart kararlarını değerlendirmek için üç farklı geçerlik kanıtı tanımlamıştır: iç, dış ve süreç.

İç geçerlik kanıtı, aynı standart belirleme çalışmasından elde edilen farklı sonuç kümelerinin tutarlılığıdır (Kane, 1994). Bu tutarlılık, geçme puanının performans standardını ne kadar iyi yansıttığını gösterir. Ayrıca, Hambleton ve Pitoniak (2006), aynı tekniklerin bir kaç kez tekrarlanmasıyla elde edilen sonuçların tutarlılığı ile de iç geçerlik kanıtının sağlanabileceğini ifade etmiştir. Geçme puanı standart hatası ise, bir standart belirleme çalışmasının tekrarlanması durumunda ne derecede aynı sonuçlar alınacağını ölçüsüdür (Kane, 1994). Bu hata, farklı zamanlarda farklı uzman grupların veya aynı zamanda farklı grupların toplanması ve sonuçların karşılaştırılmasıyla kestirilebilir (ACT, 1993; Norcini & Shea, 1992). Bir diğer iç geçerlik kanıtı da, uzmanların standart belirleme sürecinde farklı turlardaki kendi iç tutarlılığıyla ve her turdaki uzmanlar arası tutarlılıkla elde edilebilir. Katılımcılar, performans standartları hakkında farklı düşüncelere sahip olabileceğinden katılımcılar arasında yüksek düzeyli tutarlılık beklenmemektedir. Bunun yanı sıra, her katılımcı ampirik verileri ve her turdaki tartışmaları dikkate alabileceğinden, katılımcıların kendi iç puanlamaları değişkenlik gösterebilir. Bundan dolayı turlar arasında katılımcıların kendi içindeki yüksek tutarlılığı, iç geçerliğin zayıf olabileceğinin bir kanıtı olabilir (Smith, 2011).

Bir diğer geçerlik kanıtı olan dış geçerlik kanıtı ise geçme puanlarının dış bilgi kaynaklarıyla karşılaştırılması neticesinde elde edilir. Standart belirlemenin amacına uygun olacak şekilde bu bilgi kaynakları; bir ölçüt, farklı bir standart belirleme yöntemi, farklı bir test, farklı bir değerlendirme yöntemi veya geçme oranları olabilmektedir. Fakat tek bir karşılaştırma yerine geçme puanını destekleyen tutarlı birçok sonuç, en uygun standart için daha ikna edici kanıt sunabilmektedir. (Kane,1994).

Süreç geçerlik kanıtı ise standart belirleme sürecinde izlenecek adımların uygunluğu ve bu adımların uygulanmasındaki nitelik ile ilgilidir. Kane (1994), standart belirleme amacını bilen, süreci anlayan ve tarafsız olan kişiler tarafından standartlar belirlendiğinde, standartlara olan güvenin artacağını belirtmiştir. Bu geçerlik kanıtı, performans standartlarının geçerliğini değerlendirmede iki sebepten dolayı önemlidir. Bunlardan ilki, yukarıda bahsedilen diğer geçerlik türlerinin göreceli olarak eksik olmasıdır (Pant, Rupp, Tiffin-Richards & Koller, 2009). İkincisi ise, süreç geçerlik kanıtının politika kararlarını değerlendirmek için yaygın bir şekilde kullanılmasıdır (Sireci, 2007).

Kane (2001), standart belirleme sürecini beş temel unsura ayırmıştır. Bu unsurlar; standart belirleme amaçlarının tanımı, uzmanların seçimi, uzmanların eğitimi, performans standardının tanımı ve veri toplama süreçleridir.

Standart belirleme amaçlarının tanımı. Standart belirleme süreci başlamadan önce geçme puanının genel kullanım amacının tanımlanması gerekir. Tavsiye amaçlı bir test belirli bir alandaki düşük yeterliği tanımlamaya amaçlarken; bir lisans testi ise sorumluluk bilincini vurgulamaya çalışabilir. Bu nedenle, standart belirlemenin amacı uzmanlara açık ve kesin bir dille ifade edilmelidir ve özgün olmalıdır.

Uzmanların seçimi. Nitelikli uzmanların seçimi performans standartlarının geçerliği açısından önemlidir. Aynı zamanda, bu uzmanların seçimi temsil edilebilirlik ve güvenilirlik algısını da etkilemektedir (Messick, 1995). Standart belirleyen kişilerin ilgili alanda gelişmiş teknik bilgiye sahip olmaları ve hedef kitleyi iyi tanımları, standardın daha gerçekçi olmasını sağlayabilir (Kane, 1994). Ayrıca, çalışmaya katılacak uzman sayıları hakkında farklı tavsiyeler bulunmaktadır; bu

tavsiyeler en az 5 uzman olması gerektiğinden (Livingston & Zieky, 1982) en çok 25 uzman olması gerektiğine kadar (Mehrens & Popham, 1992) değişmektedir.

Uzmanların eğitimi. Uzmanların geçerli bir performans standardı belirlemeleri için, standart belirleme sürecinin amacı ve yöntemi konusunda eğitim almaları gerekmektedir (Mills, Melican & Ahluwalia, 1991; Norcini, Lipner, Langdon & Strecker, 1987). Verilecek olan eğitim çalışmanın amaç ve içeriğine göre planlanmalıdır.

Performans standardının tanımı. Bir test için geçme puanı belirlemeden önce performans standardı açık bir şekilde tanımlanmalıdır. Bunun için, performans düzeylerinin belirlenmesi ve tanımlanması gerekir. Performans düzey tanımları, standardın kullanım amacına göre belirlenmelidir.

Veri toplama süreçleri. Tavsiye edilecek standartların uygun olabilmesi için, uzmanlara tipik bir bireyin performansı hakkında normatif bilginin sağlanması gerekmektedir. Bunun yanında, uzmanların kararlarını gözden geçirebilmeleri için tekrarlamalı süreçlerin kullanımı tavsiye edilmektedir (Jaeger, 1989; Shepard, 1980).

Seçkisiz Madde Örneklem Yöntemleri

Ferdous ve Plake (2005), performans standartlarının genellenebilirliği için iki yöntemin varlığından söz etmiştir. Bunlardan ilki, test maddelerini daha küçük alt testlere ayırmak ve bu alt testleri değerlendirmek için eşit sayıda uzman alt grupları oluşturmaktır. Bu yöntemde, toplam değerlendirilen madde sayısı değişmezken; her bir uzmanın göz önünde bulunduracağı madde sayısı azalmaktadır. Fakat mevcut yöntemle aynı seviyede güvenilirlik elde etmek için bu yöntemde çok sayıda uzmana ihtiyaç duyulmaktadır. İkinci yöntem ise, psikometrik özellikler bakımından tüm testi temsil eden bir alt test oluşturarak uzmanların gözden geçireceği madde sayısını azaltmaktır. Bu yöntemde, madde sayısını azaltmak için farklı seçkisiz madde örneklem yöntemleri kullanılabilir. Bu çalışmada, basit seçkisiz madde örneklem (BSÖ) ve tabakalı seçkisiz madde örneklem (TSÖ) yöntemleri kullanılmıştır.

Basit seçkisiz madde örneklem. Basit seçkisiz örneklemde, evrendeki her bir bireyin eşit ve birbirinden bağımsız seçilme hakkı bulunmaktadır (Fraenkel, Wallen & Hyun, 2012). Seçkisiz örneklem büyüklüğü arttıkça, evrenin temsil

edilebilirliği de artmaktadır. Evrendeki her bir bireyin eşit seçilme hakkına sahip olması için rastgele sayılar tablosu kullanılmaktadır. Bu yöntemin en büyük dezavantajı yeterince büyük ve temsil edebilen bir örnekleme sahip olmanın kolay olmamasıdır. Ayrıca, araştırmacılar, belli alt grupların veya tabakaların evrendeki ağırlıkları oranında örnekleme var olmasını isteyebilmektedir. Bu durumda, tabakalı seçkisiz örnekleme kullanılmaktadır.

Tabakalı seçkisiz madde örnekleme. Tabakalı seçkisiz örnekleme, belli alt grupların veya tabakaların, evrendeki ağırlıkları oranında örnekleme temsil edilmelerini sağlayan bir süreçtir (Fraenkel, Wallen & Hyun, 2012). Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel (2016), bu süreçte ilk olarak araştırma problemini etkileyebileceği düşünülen bir faktöre göre evren içinde homojen alt grupların (tabakaların) oluşturulması gerektiğini belirtmiştir. Sonrasında, her bir alt grup için belirlenen örneklem büyüklüğü kadar birimin ayrı ayrı seçimi basit seçkisiz örnekleme ile yapılmaktadır. İki faktörün olması durumunda, örneklem oluşturma işlemi iki aşamada gerçekleşmektedir. Birinci faktöre göre oluşturulan alt grupların her biri, ikinci faktöre göre evren içinde homojen alt kategorilere ayrılabilir. Sonrasında, her bir alt kategori için belirlenen örneklem büyüklüğü kadar birimin ayrı ayrı seçimi yine basit seçkisiz örnekleme ile yapılabilir. Özellikle, örneklemin yeterince büyük olmadığı durumlarda, tabakalı seçkisiz örnekleme yöntemi ile evren daha iyi temsil edilebilmektedir.

Bu çalışma kapsamında, tabakalı seçkisiz örnekleme yöntemi ile madde seçim aşamasında sadece içerik, sadece madde gücü ve hem içerik hem de madde gücü kategorileri göz önüne alınarak farklı tabakalandırmalar yapılmış ve her bir durum ayrı ayrı incelenmiştir.

Madde Sayısının Azaltılması ile İlgili Çalışmalar

Ulusal Değerlendirme Yönetim Kurulu (NAGP, National Assessment Governing Board, 1994), Ulusal Eğitim Süreçlerini Değerlendirme (NAEP, The National Assessment of Educational Progress) Coğrafya başarı seviyeleri belirleme çalışmalarında uzman alt grup tekniği kullanarak uzman kararlarının tutarlılık gösterip göstermediğini incelemişlerdir. Testi, madde gücüne dayalı olarak iki eşit büyüklükte gruba ayırmış ve her bir uzmandan testteki maddelerin en az bir

yarısı için kararlar almasını istemişlerdir. Araştırma sonucunda, uzmanların puanlamaları tutarlılık göstermiştir.

Harvey ve Way (1999), internete dayalı standart belirleme sistemi geliştirdikleri çalışmalarında, bu sisteme ilişkin geçme puanlarının klasik yöntemle ilişkin geçme puanlarından farklılık gösterip göstermediğini incelemişlerdir. Analiz sonucuna göre, internete dayalı çalışmadan elde edilen geçme puanları klasik standart belirleme çalışmasındaki geçme puanlarıyla benzerlik göstermiştir.

Sireci, Patelis, Rizavi, Dillingham ve Rodriguez (2000) ise, çalışmalarında, Bilgisayar Uyarlamalı Test (BUT) maddeleri için standart belirlemek üzere üç Angoff yöntemini (Angoff, Sıralama (lojistik regresyon) ve Sıralama (medyan)) karşılaştırmışlar ve her biri tüm testin üçte biri büyüklüğünde olan alt testlerle puanlama tutarlılığını incelemişlerdir. Çalışmanın sonunda, toplam 112 maddeden oluşan tüm teste "göreceli olarak" benzer geçme puanı elde etmek için maddelerin üçte ikisinin (iki alt test) yeterli olabileceği sonucuna ulaşmışlardır.

Plake ve Impara (2001), bir standart belirleme çalışmasında uzmanların değerlendireceği madde sayısını azaltma yöntemlerinden birini incelediği araştırmalarında, 230 maddeden oluşan sertifika sınavını, içeriklerine ve psikometrik özelliklerine göre denk olmak şartıyla eşit sayıda madde içeren 2 gruba ayırmışlardır. Benzer şekilde 29 kişiden oluşan uzman grubunu da benzer sayıda 2 gruba seçkisiz olarak tahsis etmişler ve her bir grupta bulunan 115 maddeye 35 madde daha ekleyerek iki grup arasında 70 ortak madde olmasını sağlamışlardır. Araştırma sonucunda, iki alt gruba ilişkin geçme puanlarının benzer olduğu sonucuna ulaşmışlardır.

Coraggio (2008), yaptığı simulasyon çalışmasında, madde kümelerinin özelliklerinin (madde güçlük dağılımları, doğru performans standardının yeri, madde alt kümelerinin büyüklüğü gibi) ve standart belirleme sürecinin (uzman sayısı, güvenilir olmayan uzmanların yüzdesi gibi) simule edilmiş "doğru" performans standardını etkileme kapsamını incelemiştir. Analiz sonuçlarına göre, yapılan 1000 tekrarlı ölçüm neticesinde, mantıklı bir geçme puanı elde etmek için testteki maddelerin %50-66'sını kullanmanın yeterli olacağı belirtilmiştir. Bunun yanı sıra, madde dağılımları daha küçük bir varyansa sahip olduğunda, bir alt kümeden "doğru" performans standardını elde etmek daha kolay olmuştur. Ayrıca, güvenilir

olmayan uzmanların yüzdesi gibi bu çalışmada bulunan değişkenlerin çoğunun “doğru” performans standardı bulmaya önemli bir etkisi olmadığı anlaşılmıştır.

Buckendahl, Ferdous ve Gerrow (2010) tarafından yapılan araştırmada, operasyonel bir standart belirleme çalışması bağlamında madde alt testlerinin kullanımı incelenmiştir. 300 maddelik sertifika sınavından elde edilen geçme puanı ile yine aynı testten seçkisiz olarak seçilen 150 madde ile elde edilen geçme puanı Angoff ve İşaretleme yöntemleriyle hesaplanmış ve alt testler ile tüm testten elde edilen puanlar karşılaştırılmıştır. Analiz sonuçlarına göre, çalışmada tüm testin ve alt testlerin geçme puanları Angoff yöntemine göre farklılık gösterirken, İşaretleme yöntemine göre farklılık bulunamamıştır.

Seçkisiz Örnekleme Yöntemlerinin Karşılaştırılması ile İlgili Çalışmalar

Ferdous ve Plake (2005), araştırmalarında, standart belirleme çalışmalarında her bir uzmanın gözden geçireceği madde sayısını azaltmada çeşitli stratejilerin etkililiğini incelemişlerdir. Araştırmada, Sireci vd. (2000)'nin çalışmasındaki yöntemlere ek olarak tabakalı seçkisiz örnekleme yönteminden de faydalanılmıştır. Çalışmanın verilerini, sertifikasyon sınavları için yapılan iki standart belirleme çalışmasındaki veriler oluşturmuştur. İlk standart belirleme çalışmasında alt testler tüm testi sadece madde gücü açısından temsil ederken; ikinci çalışmada hem madde gücü hem de içerik ağırlıkları açısından temsil etmiştir. Her iki testten de, toplam madde sayısının %5 ile %70 i arasında değişen 8 farklı büyüklükte alt test oluşturulmuştur. Analiz sonrasında, toplam madde sayısının %50'sini içeren alt test ile belirlenen geçme puanının, tüm testin geçme puanının 1 puan içerisinde yer aldığı görülmüştür. Ayrıca uzmanların kendi içindeki tutarlılık kestirimleri ortalaması, %50'lik alt test ve tüm test için neredeyse aynı değeri almıştır. Araştırma sonucunda, toplam madde sayısının %50'sini içeren tabakalı alt test ile tüm testin geçme puanına eşit bir kestirim elde edilebileceği belirtilmiştir.

Ferdous ve Plake (2007), araştırmalarında, standart belirleme çalışmalarındaki madde sayısını azaltmada üç aşamalı-tabakalı seçkisiz örnekleme yönteminin etkililiğini incelemişlerdir. Tabakalı seçkisiz örnekleme yöntemiyle alt testler oluşturulurken madde gücü ve içerik ağırlıklarına ek olarak madde ayırt edicilik değerleri de dikkate alınmıştır. Maddeler, tüm testin her bir güçlük, içerik ve ayırt edicilik tabakasındaki madde sayısı ile orantılı olacak şekilde seçkisiz olarak

seçilmiş ve tüm testin %10, %20, %30, %40 ve %50'si büyüklüğünde alt testler oluşturulmuştur. Analiz sonrasında, tüm testin geçme puanının 1 tahmini standart hatası (TSH) içerisinde kestirimler elde etmek için maddelerin % 40 ile % 50 arasında bir oranda seçilmesinin yeterli olacağı kararına varılmıştır. Ayrıca, üç aşamalı tabakalı madde örnekleme yönteminin iki aşamalıya göre daha kararlı sonuçlar verdiği belirlenmiştir.

Smith ve Ferdous (2007) ise, çalışmalarında, standart belirleme çalışmalarındaki madde sayısını azaltmada madde gücünü farklı yollarla tabakalandırmanın etkililiğini incelemişlerdir. Çalışmada, madde gücünün iki farklı yolla tabakalandırılmasının yanı sıra üç farklı eşitleme yöntemi kullanılmıştır. Farklı büyüklükteki alt testleri (30, 50 ve 70 madde) oluşturmak için tamamen rastgele, tabakalı p değeri aralıklarıyla orantılı ve p değerlerinin yoğunluğuyla orantılı olan üç örnekleme yöntemi kullanılmıştır. Oluşan 9 durum için üç eşitleme yöntemine (tüm teste ölçekleme, ortalama eşitleme, doğrusal eşitleme) başvurulmuştur. Çalışmanın analiz sonuçlarının, Ferdous ve Plake (2005)'in araştırmasının sonuçlarını desteklediği görülmüştür. Maddeleri tamamen rastgele seçmek yerine güçlük tabakası kullanarak seçmenin, geçme puanlarının standart hatasını daha fazla azalttığı sonucuna ulaşılmıştır. Ayrıca, çalışma sonucunda, kestirilen geçme puanı standart hatası açısından ortalama ve doğrusal eşitleme yöntemlerinin, tüm teste ölçekleme yöntemine göre daha küçük hatalar verdiği görülmüştür.

Ferdous (2007), araştırmasında, üç aşamalı-tabakalı seçkisiz örnekleme yöntemi kullanımının bir standart belirleme çalışmasında yargıda bulunulan madde sayısını azaltmaya etkisini incelemiştir. Çalışmadaki alt testleri oluşturmak için maddeler sırasıyla içeriklerine, madde tipine ve MTK-dayalı madde güçlüklerine (b değerleri) göre tabakalandırılmıştır. Tüm testin içerik, madde tipi ve güçlük kategorilerindeki madde sayılarıyla orantılı olacak şekilde tüm testin %50, %60 ve %70'i büyüklüğündeki alt testler oluşturulmuştur. Araştırma sonucunda, test maddelerinin en az %70'i ile tüm testin geçme puanına benzer bir kestirim elde edilebileceği belirtilmiştir.

Smith (2011), lisans ve sertifika ile ilgili 30 bağımsız standart belirleme çalışmasından aldığı verilerle tabakalı madde örneklemenin, standart belirleme çalışmalarındaki madde sayısını azaltmadaki etkililiğini incelemiştir. Alt testler

sırasıyla içerik, madde güçlüğü ve madde ayırt ediciliği tabakaları kullanılarak oluşturulmuştur. Çalışmanın sonunda tüm testin geçme puanının 1 TSH içerisinde bir kestirim elde etmek için tabakalı seçkisiz örnekleme yöntemi kullanılarak test maddelerin %35'inin, basit seçkisiz örnekleme yöntemi kullanılarak ise %40'ının seçilmesinin yeterli olabileceği kanısına varılmıştır. Ayrıca, tabakalı ve basit seçkisiz örnekleme yöntemleri ile tahminin yüzde bir puan aralığı içinde bir geçme puanı elde etmek için sırasıyla tüm testin %65'inin ve %71'inin yeterli olabileceği görülmüştür. Çalışmanın bir diğer önemli bulgusu ise, madde güçlüğü tüm testlerin %86.7'sinde, madde ayırt ediciliği tüm testlerin %16.7'sinde ve içerik tabakası da tüm testlerin %40'ında önemli bir yordayıcı olmuştur.

Kannan, Sgammato, Tannenbaum ve Katz (2015) Genellenebilirlik Teorisi (G- Teori) ve Karar Çalışması çerçevesi kullanarak genellenebilir geçme puanı kestirimleri elde etmek amacıyla madde alt kümelerini oluşturmuşlardır. Çalışmada, madde sayısı 30, 45, 60, 75 ve 90 olan alt testlerin her birinin oluşturulmasında basit seçkisiz örnekleme (BSÖ), içerik tabakalı (İ-TSÖ), madde güçlüğü tabakalı (G-TSÖ) ve içerik ve madde güçlüğü tabakalı (İG-TSÖ) 4 farklı seçkisiz örnekleme yöntemi kullanılmıştır. Çalışmanın sonunda, yaklaşık 100 maddeyi içeren tüm teste ait genellenebilir geçme puanları üretmek için İ-TSÖ yöntemiyle yaklaşık 45 maddenin seçilmesinin yeterli olabileceği bulgusuna ulaşılmıştır.

Kannon, Sgammato ve Tannenbaum (2015), Kannan vd. (2015)'nin çalışmasında ulaşılan bulgudan yola çıkarak 45 maddeden oluşan alt testin, yaklaşık 100 maddeden oluşan tüm testin Angoff geçme puanına benzer kestirim (1 TSH içerisi) verebilme durumunu incelemişlerdir. 45 maddeden oluşan alt testten elde edilen geçme puanı ile tüm testin geçme puanı, bağımsız ve bağımlı örneklem t-testleri kullanılarak karşılaştırılmıştır. Analizlerin sonucunda, 45 madde içeren içeriğe göre tabakalandırılmış alt testin, güvenilir minimum geçme puanları elde etmek için yeterli olacağı bulgusu desteklenmiştir.

İlgili Araştırmalar Özet

Araştırmalardan elde edilen sonuçlar incelendiğinde, farklı seçkisiz örnekleme yöntemleriyle tüm testin geçme puanına benzer bir kestirim elde edebilmek için tüm testin %40-70 arasında değişen yüzdeliklerinin yeterli olabileceği sonucu çıkmıştır. Ayrıca, farklı seçkisiz örnekleme yöntemleri karşılaştırıldığında,

tabaka sayısı arttıkça geme puanını yordama gcnn de arttıęı grlmřtr. Son yıllarda yapılan G-alıřmasına gre ise, yaklaşık 100 madde ieren tm teste genellenebilir kestirimler elde edebilmek iin 45 maddenin yeterli olabileceęi bulgusuna varılmıřtır.

Bu arařtırmada ise, SBS 2012 ve SBS 2013 matematik testlerinin birleřtirilmesiyle oluřturulan tm test iin Angoff ve Evet/Hayır yntemlerine gre ayrı ayrı geme puanı belirlenmiř ve madde sayısını azaltmakta kullanılan farklı sekisiz rnekleme yntemleri her iki yntem baęlamında da karřılařtırılmıřtır. Ayrıca alan yazına bakıldığında lkemizde bu konuyla ilgili benzer alıřmaların olmadığı grlmřtr. Bu arařtırma, lkemizde bir standart belirleme alıřmasındaki madde sayısının azaltılmasıyla ilgili yrtlen ilk arařtırma olması bakımından nemlidir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın yöntemi, araştırmanın evreni ve örnekleme, veri toplama süreci, veri toplama araçları ve verilerin analizi açıklanmıştır.

Araştırmanın Yöntemi

Değişkenleri etkileyecek herhangi bir müdahalenin olmadığı ve iki ya da daha fazla değişken arasındaki ilişkinin incelendiği araştırma modelleri ilişkisel türden araştırma modelleridir (Fraenkel, Wallen & Hyun, 2012). Bu çalışmada, bir standart belirleme çalışmasındaki madde sayısını azaltmak için kullanılabilecek çeşitli seçkisiz örnekleme yöntemleri karşılaştırılmıştır. Araştırma bu açıdan ilişkisel türden araştırma modellerindendir. Aynı zamanda, bu araştırma, Angoff ve Evet/Hayır yöntemlerine ilişkin betimleyici istatistikleri elde etme bakımından betimsel bir çalışmadır.

Araştırmanın Evreni ve Örnekleme

Araştırmanın çalışma evreni, 2011-2012 ve 2012-2013 öğretim yıllarında Seviye Belirleme Sınavı'na girmiş olan sırasıyla 1.075.533 (MEB, 2012) ve 1.112.604 (MEB, 2013) 8. sınıf öğrencisinden oluşmaktadır. Araştırmanın örneklemini ise, öğrenciler ve uzmanlar olmak üzere iki farklı grup oluşturmuştur. Öğrenci grubunda, SBS 2012 ve SBS 2013'e girmiş olan öğrenciler arasından ayrı ayrı seçkisiz örnekleme yöntemiyle seçilmiş olan sırasıyla 10187 ve 10424 öğrenci olmak üzere toplam 20611 öğrenci; uzman grubunda ise, 12 akademisyen ve 16 ortaokul matematik öğretmenini olmak üzere toplam 28 uzman yer almıştır. Bu çalışmada, uzmanların belirlenmesinde amaca uygun örnekleme yöntemi kullanılmış ve uzman seçiminde gönüllülük durumu esas alınmıştır. Amaca uygun örnekleme yönteminde, araştırmacılar, çalışma grubunun önceki bilgilerine ve araştırmanın amacına göre, örneklem oluşturmak için kişisel değerlendirmelerini kullanabilmektedir (Fraenkel, Wallen & Hyun, 2012). Uzman grubundaki akademisyenler, İlköğretim Matematik Öğretmenliği lisans programından mezun olmuş, Eğitim Bilimleri veya Matematik Eğitimi alanlarında lisansüstü öğrenim gören akademisyenler arasından; uzman grubundaki öğretmenler ise mesleğinde en az beş yıllık tecrübeye sahip ortaokul matematik öğretmenleri arasından seçilmiştir. Çalışma kapsamında her bir uzman, 40 madde için Angoff ve Evet/Hayır

yöntemlerine göre iki ayrı değerlendirmede bulunmuş ve bu değerlendirmelerden yola çıkılarak Angoff ve Evet/Hayır geçme puanları hesaplanmıştır.

Veri Toplama Süreci

Araştırmada kullanılan SBS 2012 ve SBS 2013 matematik alt testlerine ilişkin öğrenci cevapları Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'nden temin edilmiştir. Uzmanların maddelere ilişkin değerlendirmeleri ise, araştırmacı tarafından hazırlanan veri toplama formu ile toplanmıştır. Uzmanlardan veri toplama işlemleri, araştırmacının kendisi tarafından yürütülmüştür.

Veri Toplama Araçları

Bu araştırmada iki farklı veri kullanılmıştır. Araştırmada kullanılan öğrenci verileri, Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından uygulanmış sınavların (SBS 2012 ve SBS 2013) sonuçlarına ilişkin veriler olup; diğer veriler ise, araştırmacı tarafından hazırlanmış olan “Uzman Görüşleri için Gönüllü Katılım Formu” (EK-A) ile 28 uzmandan elde edilmiştir.

Seviye belirleme sınavı (SBS). Milli Eğitim Bakanlığı tarafından ortaöğretime geçiş için yapılan sınavlar yıllar içinde yöntem ve içerik olarak farklılık göstermiştir. 2008 yılına kadar farklı isimlerle yapılan bu sınavlar, 2008-2009 eğitim-öğretim yılından 2013-2014 eğitim-öğretim yılına kadar 6., 7. ve 8. sınıflar için Seviye Belirleme Sınavı (SBS) adı altında yapılmıştır. Bu yıllar arasında kademeli olarak tüm sınıflarda uygulamaya geçilmiş olup, ilerleyen yıllarda da kademeli olarak kaldırılmıştır. Bu çalışmada, 2011-2012 ve 2012-2013 öğretim yıllarında yapılmış olan Seviye Belirleme Sınavı 8. sınıf matematik alt testlerine (her test için 20 soru) ilişkin öğrenci cevapları kullanılmıştır.

Uzman görüşleri için gönüllü katılım formu. Geçme puanı belirlemek amacıyla uzman görüşleri için gönüllü katılım formu hazırlanmıştır. Bu formda öncelikle, Uluslararası Öğrenci Değerlendirme Programı matematik yeterlik düzeylerinde yer alan düzey 2 (PISA, 2007) esas alınarak “minimum yeterlik düzeyi” tanımı açık bir şekilde ifade edilmiştir. Uzmanlardan, değerlendirmeye başlamadan önce çoktan seçmeli test sorularını dikkatle incelemeleri ve SBS 2012 ve 2013 matematik alt testlerinde bulunan toplam 40 soru için önce Angoff sonra da

Evet/Hayır yöntemlerine göre değerlendirmeler yapmaları istenmiştir. Uzmanlardan karar alınan bu iki yöntemde sürecin nasıl işlediği aşağıda açıklanmıştır:

- Angoff Yöntemi için uzmanlardan, minimum yeterlik düzeyindeki öğrencilerin yüzde kaçının soruyu doğru cevaplayabileceğini düşünmeleri ve her bir madde için ayrı ayrı bir yüzdelik değeri tahmininde bulunmaları istenmiştir. Uzmanların her bir madde için belirledikleri güçlük tahminlerinin toplanmasıyla elde edilen puanların 100 puanlık ölçeğe dönüştürülmesiyle her bir uzmanın bireysel geçme puanı hesaplanmış ve bireysel geçme puanlarının ortalaması alınarak da testin nihai geçme puanı belirlenmiştir.

- Evet/Hayır yöntemi için ise uzmanlardan, “minimum yeterlik düzeyinde olan bir öğrenci soruyu doğru cevaplayabilir mi?” sorusunu “doğru cevaplayabilir” ya da “doğru cevaplayamaz” seçeneklerinden uygun olanı işaretleyerek cevaplandırmaları istenmiştir. Uzmanların yanıtlarına göre, öğrencinin doğru yanıtlayabileceği düşünülen maddelere “1” puan; doğru yanıtlayamayacağı düşünülen maddelere ise “0” (sıfır) puan verilmiştir. Uzmanların her bir madde için aldıkları kararlara ilişkin puanların toplanmasıyla elde edilen puanların 100 puanlık ölçeğe dönüştürülmesiyle her bir uzmanın bireysel geçme puanı hesaplanmış ve bireysel geçme puanlarının ortalaması alınarak da testin nihai geçme puanı belirlenmiştir.

- Angoff ve Evet/Hayır yöntemlerinde 28 uzmanın testteki her bir madde için değerlendirmelerine ilişkin puanlar sırasıyla EK-B ve EK-C’de, bireysel geçme puanları ise EK-Ç’de verilmiştir.

Verilerin Analizi

Verilerin analizi aşamasında, ilk olarak iki farklı testin (SBS 2012 ve 2013) maddelerine ilişkin öğrenci cevapları, doğru yanıtlar için “1”, yanlış ve boş bırakılan yanıtlar için “0” kodlanarak 1-0 verisine dönüştürülmüştür. Sonrasında, aşağıdaki basamaklar izlenmiştir.

- Test ve madde istatistikleri hesaplamaları için 1-0 şeklinde kodlanan iki farklı veri TAP 14.7.4 (Test Analysis Program) aktarılmış ve her bir test için test ve madde istatistikleri ayrı ayrı hesaplanmıştır.

- SBS 2012 ve SBS 2013 Matematik alt testlerinin birleştirilmesiyle 40 madde içeren tüm test oluşturulmuştur. Her iki testteki maddelerin aynı kazanımları ölçüyor olması ve iki yılda da sınava giren grubun yetenek düzeyleri bakımından eşdeğer olduğu düşünülerek, iki testin birleştirilmesinde sakınca görülmemiştir.

- Angoff ve Evet/Hayır yöntemlerine göre tüm teste ait geçme puanı ayrı ayrı hesaplanmıştır.

- Geçme puanı açısından tüm testi temsil etmesi düşünülen dört farklı alt test (toplam madde sayısının %30, %40, %50 ve %70'u) dört farklı seçkisiz örnekleme yöntemi (basit seçkisiz örnekleme [BSÖ], içerik tabakalı [İ-TSÖ], madde güçlüğü tabakalı [G-TSÖ] ve içerik ve madde güçlüğü tabakalı [İG-TSÖ]) kullanılarak oluşturulmuştur. Yani, dört farklı seçkisiz örnekleme yöntemi ve dört farklı alt test içeren 4x4'lük desen kullanılmış, toplamda 16 çalışma durumu incelenmiştir.

- Alt testlerin oluşturulması sürecinde, her bir çalışma durumu için 1000 replikasyon yapılmıştır. Örneğin, basit seçkisiz madde örnekleme yöntemiyle oluşturulan, 40 maddenin %30'u kadar madde içeren %30'luk alt test için madde seçim işlemleri 1000 defa tekrarlanmış, 1000 farklı alt test ve geçme puanı elde edilmiştir. Böylelikle, elde edilen sonuçların daha tutarlı ve güvenilir olması sağlanmaya çalışılmıştır. Yapılan işlemlerde kullanılan R koduna ilişkin bir örnek EK-D'de sunulmuştur.

- Tüm testin geçme puanı ile oluşturulan alt testlere ilişkin ortalama geçme puanları arasında manidar düzeyde farklılık olup olmadığını incelemek için tek-yönlü varyans analizi (ANOVA) yapılmıştır.

- Her bir ANOVA analizine ilişkin normallik ve varyansların homojenliği varsayımları kontrol edilmiştir. Her bir durum için normallik varsayımının sağlandığı görülmüştür. Tüm durumlara ilişkin basıklık ve çarpıklık değerleri (-2, 2) aralığında ve Shapiro-Wilk p değeri .05'ten büyük bulunmuştur (EK-E). Varyansların homojenliği varsayımı ise Levene testi ile kontrol edilmiş (EK-F) ve hiçbir durum için varsayımın sağlanmadığı görülmüştür ($p < .05$). Bu durumda kullanılması önerilen Welch testi (Pallant, 2005) sonuçlarına başvurulmuştur.

- Daha detaylı incelemeler yapmak adına hata kareleri ortalamasının karekökü (Root Mean Square Error-RMSE) ve TSH değerlerinden yararlanılmıştır. TSH değerleri yorumlanırken, oluşturulan 1000 alt testin geçme puanlarından yüzde

kaçının tüm test geçme puanı ± 1 TSH içerisinde kaldığı incelenmiştir. Bu noktada benzerlik kriteri olarak %95 alınmış olup, 1000 geçme puanının %5'ten fazlasının aralık dışına çıktığı durumda tüm testin geçme puanına benzer sonuç elde edilmediği yorumuna gidilmiştir.

- Geçme puanlarının hesaplanması işlemi Microsoft Excel 2010, varyans analizleri SPSS 23, alt testlere ilişkin replikasyon ve geçme puanı hesapları ise R yazılımında yapılmıştır.

Test ve madde istatistikleri. SBS 2012 ve SBS 2013 Matematik alt testlerine ilişkin betimsel istatistikler Tablo 1'de sunulmuştur. İstatistikler, 2011-2012 ve 2012-2013 öğretim yıllarında sınava giren sırasıyla 10187 ve 10424 öğrenciye ait cevaplar göz önünde bulundurularak hesaplanmıştır.

Tablo 1

Matematik Alt Testlerine İlişkin Betimsel İstatistikler

	SBS 2012 Matematik Testi	SBS 2013 Matematik Testi
Madde Sayısı	20	20
Öğrenci Sayısı	10187	10424
Ortalama	6.41	4.97
Varyans	19.35	16.89
Standart Sapma	4.40	4.11
Güvenirlilik (KR20)	.84	.83
Ortalama Güçlülük	.32	.25

Tablodaki değerler, SBS 2012 Matematik alt testi ($\bar{X}=6.41$) ve SBS 2013 Matematik alt testinin ($\bar{X}=4.97$) zor olduğunu göstermektedir. Matematik alt testlerinin ortalama güçlükleri sırasıyla .32 ve .25 olarak hesaplanmıştır.

Sonuçlarına dayanılarak önemli kararlar alınacak testlerin güvenilirliğinin, madde sayısının az olduğu durumlarda, .80 ve üzerinde olması gerekmektedir (Özçelik, 2013). Matematik alt testlerinin güvenirlik katsayıları ise .84 ve .83 değerlerini almıştır. Buna göre, bu testlere ilişkin puanların güvenilir olduğu söylenebilir.

SBS 2012 Matematik alt testi 1-20 arası maddeler ve SBS 2013 Matematik alt testi 21-40 arası maddeler olmak üzere toplam 40 maddeye ilişkin madde güçlük değerleri (p) Tablo 2’de verilmiştir.

Tablo 2

Tüm Teste İlişkin Madde Güçlük İndeksleri

Madde No	Madde Güçlüğü (p)	Madde No	Madde Güçlüğü (p)	Madde No	Madde Güçlüğü (p)	Madde No	Madde Güçlüğü (p)
1	.28	11	.24	21	.20	31	.20
2	.39	12	.30	22	.43	32	.22
3	.41	13	.23	23	.22	33	.19
4	.63	14	.17	24	.22	34	.41
5	.18	15	.53	25	.32	35	.21
6	.25	16	.53	26	.13	36	.20
7	.40	17	.34	27	.27	37	.23
8	.34	18	.20	28	.30	38	.14
9	.30	19	.43	29	.18	39	.49
10	.13	20	.15	30	.25	40	.19
Ortalama		.29					

SBS 2012 ve SBS 2013 Matematik alt testlerinin birleştirilmesiyle oluşturulan tüm testin 40 maddesine ilişkin madde güçlük değerleri .13 ile .63 arasında değişmektedir. Buna göre, testin zor ve orta güçlükte maddeler içerdiği fakat kolay madde içermediği görülmektedir. Testin en zor maddeleri (p=.13) 10. ve 26. maddeler iken; en kolay maddesi 4. maddedir (p=.63). Tüm teste ait genel ortalama güçlük ise .29 olarak hesaplanmış olup zor bir test olduğu söylenebilir.

Alt testlerin oluşturulması. Geçme puanı açısından tüm testi temsil etmesi düşünülen alt testlerin oluşturulması için basit seçkisiz ve tabakalı seçkisiz örnekleme yöntemleri kullanılmıştır. Her bir alt problem için toplam 40 maddenin %30, %40, %50 ve %70’i seçilerek sırasıyla 12, 16, 20, 28 madde içeren dört ayrı alt test 1000 replikasyonla oluşturulmuştur.

Basit seçkisiz madde örnekleme yöntemi. Bu yöntem kullanılarak oluşturulan alt testlerde, maddeler 40 madde içerisinde seçkisiz olarak seçilmiştir.

Tabakalı seçkisiz madde örnekleme. Tabakalı seçkisiz madde örnekleme yönteminde, alt testler oluşturulurken, içerik ve madde güçlük kategorilerine göre maddeler seçilmiştir. Bu bağlamda, içerik tabakalı seçkisiz örnekleme (İ-TSÖ), madde güçlüğü tabakalı seçkisiz örnekleme (G-TSÖ) ve içerik ve madde güçlük tabakalı seçkisiz örnekleme (İG-TSÖ) yöntemleri kullanılmıştır.

İ-TSÖ yöntemiyle alt testlere madde seçiminden önce alan uzmanı tarafından tüm test maddeleri içeriklerine göre 5 kategoriye ayrılmış; kategoriler belirlenirken ortaokul matematik müfredatındaki öğrenme alanları (sayılar, geometri, ölçme, olasılık ve istatistik, cebir) dikkate alınmıştır (MEB, 2009). Alt testlere madde seçiminde içerik kategorileri tabaka olarak kullanılmış ve alt testlere seçilen maddeler, tüm testin her bir içerik kategorisindeki toplam madde sayısı ile orantılı olacak şekilde seçkisiz olarak her bir kategoriden seçilmiştir. Tüm test maddelerinin içeriklerine göre oluşturulan kategorileri ve her bir kategorideki madde sayısı ve numaraları Tablo 3'te sunulmuştur.

Tablo 3

Tüm Testin İçerik Kategorilerindeki Madde Sayısı ve Numaraları

İçerik Kategorisi	Madde Sayısı	Madde Numarası
Sayılar	9 (%22.5)	10, 18, 26, 33, 1, 2, 25, 3, 22
Geometri	10 (%25)	21, 36, 6, 7, 9, 23, 28, 35, 4, 39
Ölçme	7 (%17.5)	14, 38, 11, 12, 13, 30, 32
Olasılık ve İstatistik	5 (%12.5)	17, 24, 15, 16, 34
Cebir	9 (22.5)	5, 20, 29, 31, 40, 8, 27, 37, 19

Tablo 3'e göre, tüm testteki maddelerin %22.5'i sayılar, %25'i geometri, %17.5'i ölçme, %12.5'i olasılık ve istatistik ve %22.5'i de cebir kategorisinde yer almaktadır. Bu bilgiler ışığında, alt testlere seçilen maddelerin de yine aynı oranlarda her bir kategoride olması sağlanmıştır. Örneğin, 20 maddelik bir alt test oluşturmak için 5 madde sayılar, 5 madde geometri, 3 madde ölçme, 2 madde olasılık ve istatistik ve 5 madde ise cebir kategorilerindeki maddeler arasından seçkisiz olarak seçilmiştir. İçerik kategorilerine göre alt testlere seçilmesi beklenen madde sayısı EK-G'de verilmiştir.

Madde güçlüğü tabakalı seçkisiz örnekleme (G-TSÖ) yönteminde ise öncelikle, tüm test maddeleri güçlük değerlerine göre 3 kategoriye ayrılmış; .00-.20 aralığında güçlük parametresine sahip maddeler Kategori 1, .20-.40 aralığındaki maddeler Kategori 2, .40-.63 aralığındaki maddeler Kategori 3 içerisinde alınmıştır. Alt testlere madde seçiminde güçlük kategorileri tabaka olarak kullanılmış ve alt testlere seçilen maddeler, tüm testin her bir güçlük kategorisindeki toplam madde sayısı ile orantılı olacak şekilde seçkisiz olarak her bir kategoriden seçilmiştir. Kategorilerin madde parametre değer aralıkları ve tüm testin her bir güçlük kategorisindeki madde sayısı ve numaraları Tablo 4'te verilmiştir.

Tablo 4

Tüm Testin Güçlük Kategorilerindeki Madde Sayısı ve Numaraları

Güçlük Kategorisi	Madde Sayısı	Madde Numarası
Kategori 1 (.00<p≤.20)	13 (%32.5)	5, 10, 14, 18, 20, 21, 26, 29, 31, 33, 36, 38, 40
Kategori 2 (.20<p≤.40)	19 (%47.5)	1, 2, 6, 7, 8, 9, 11, 12, 13, 17, 23, 24, 25, 27, 28, 30, 32, 35, 37
Kategori 3 (.40<p≤.63)	8 (%20)	3, 4, 15, 16, 19, 22, 34, 39

Tablo 4 incelendiğinde, madde güçlük değerlerine göre maddelerin %32.5'inin Kategori 1'de, %47.5'inin Kategori 2'de ve %20'sinin de Kategori 3'te yer aldığı görülmektedir. Bu durumda oluşturulmak istenen bir alt teste seçilen maddelerin güçlük dağılımının da, tüm test ile aynı olması sağlanmıştır. Örneğin, 20 maddelik bir alt test oluşturmak için 6 madde Kategori 1'deki, 10 madde Kategori 2'deki ve 4 madde ise Kategori 3'teki maddeler arasından seçkisiz olarak seçilmiştir. Her bir alt test için 3 ayrı güçlük kategorilerine seçilmesi beklenen madde sayısı EK-Ğ'de verilmiştir.

İçerik ve madde güçlüğü tabakalı örnekleme yönteminde (İG-TSÖ) ise, maddeler hem içerik alanları hem de güçlük değerleri göz önünde bulundurularak seçilmiştir. Maddeler ilk olarak, içeriklerine göre 5 kategoriye ayrılmış, sonrasında her bir içerik kategorisindeki maddeler güçlüklerine göre 3 gruba bölünmüş ve toplamda 15 tabaka oluşturulmuştur. Alt testlere seçilen maddeler, tüm testin her bir tabakasındaki toplam madde sayısı ile orantılı olacak şekilde seçkisiz olarak her bir içerik-güçlük tabakasından seçilmiştir. Tüm testin her bir içerik-güçlük tabakasındaki madde sayısı ve numaraları Tablo 5'te verilmiştir.

Tablo 5

Tüm Testin İçerik-Güçlük Tabakalarındaki Madde Sayısı ve Numaraları

İçerik Kategorisi	Güçlük Grubu	Madde Sayısı	Madde Numarası
Sayılar	Grup 1 ($.00 < p \leq .20$)	4 (%10)	10, 18, 26, 33
	Grup 2 ($.20 < p \leq .40$)	3 (%7.5)	1, 2, 25
	Grup 3 ($.40 < p \leq .63$)	2 (%5)	3, 22
Geometri	Grup 1 ($.00 < p \leq .20$)	2 (%5)	21, 36
	Grup 2 ($.20 < p \leq .40$)	6 (%15)	6, 7, 9, 23, 28, 35
	Grup 3 ($.40 < p \leq .63$)	2 (%5)	4, 39
Ölçme	Grup 1 ($.00 < p \leq .20$)	2 (%5)	14, 38
	Grup 2 ($.20 < p \leq .40$)	5 (%12.5)	11, 12, 13, 30, 32
	Grup 3 ($.40 < p \leq .63$)	0	
Olasılık ve İstatistik	Grup 1 ($.00 < p \leq .20$)	0	
	Grup 2 ($.20 < p \leq .40$)	2 (%5)	17, 24
	Grup 3 ($.40 < p \leq .63$)	3 (%7.5)	15, 16, 34
Cebir	Grup 1 ($.00 < p \leq .20$)	5 (%12.5)	5, 20, 29, 31, 40
	Grup 2 ($.20 < p \leq .40$)	3 (%7.5)	8, 27, 37
	Grup 3 ($.40 < p \leq .63$)	1 (%2.5)	19

Tablo 5'te her bir içerik kategorisi kendi içerisinde güçlük değerlerine göre gruplara ayrılmış, 15 ayrı tabaka oluşturulmuştur. Tablodaki maddelerin dağılımı incelendiğinde, en fazla sayıda maddenin Geometri-Grup2'de (6 madde), en az sayıda maddenin ise Cebir-Grup3'te (1 madde) bulunduğu görülmektedir. Ayrıca, Ölçme-Grup3 ve Olasılık ve İstatistik-Grup1 tabakalarında hiçbir madde yer almamıştır. Alt testlere seçilen maddelerin, tüm testin bu 13 tabakasındaki madde dağılımını temsil edecek şekilde orantılı olması sağlanmaya çalışılmıştır. Örneğin; tüm testteki 40 maddeden 4'ü (%10) Sayılar-Grup1'de yer aldığı için; her bir alt teste seçilecek toplam madde sayısının %10'unun Sayılar-Grup1 tabakasındaki maddeler arasından seçkisiz olarak seçilmesi sağlanmıştır. Her bir alt teste 15 ayrı tabakadan seçilmesi beklenen madde sayısı EK-H'de verilmiştir.

Geçme puanlarının belirlenmesi ve yorumlanması. SBS 2012 ve SBS 2013 Matematik alt testlerinin birleştirilmesiyle 40 madde içeren tüm test oluşturulmuş ve Angoff ve Evet/Hayır yöntemleriyle tüm teste ilişkin geçme puanı belirleme sürecinde ilk olarak uzmanlar arası uyumun olup olmadığı Kendall'ın W uyum katsayısı ile incelenmiştir. Parametrik olmayan bu teknikte Kendall'ın uyum katsayısı aşağıdaki formülle hesaplanır (Siegel, 1956).

$$W = \frac{12\sum R_i^2 - 3k^2 N (N + 1)^2}{k^2 N (N^2 - 1)}$$

Eşitlikte;

k: Puanlayıcı sayısını, N: Puanlanan madde sayısını, R: Her bir madde için tüm puanlayıcıların verdiği puanların toplamını gösterir.

Puanlayıcı sayısının yediye eşit ve büyük olduğu durumlar için χ^2 kullanılır ve $\chi^2_{(N-1)} = k (N - 1)W$ değeri N-1 serbestlik derecesinde χ^2 dağılımını gösterir (Siegel, 1956).

Sonrasında, Angoff ve Evet/Hayır yöntemlerinin her biriyle tüm test için geçme puanı hesaplanmıştır. Angoff yönteminde, uzmanların her bir madde için belirledikleri güçlük tahminlerinin toplanmasıyla elde edilen puanların 100 puanlık ölçeğe dönüştürülmesiyle her bir uzmanın bireysel geçme puanı hesaplanmış ve bireysel geçme puanlarının ortalaması alınarak da testin nihai geçme puanı elde edilmiştir. Evet/Hayır yönteminde ise, öğrencinin doğru yanıtlayabileceği düşünülen maddelere “1” puan; doğru yanıtlayamayacağı düşünülen maddelere ise “0” (sıfır) puan verilmiş ve uzmanların her bir madde için aldıkları kararlara ilişkin puanların toplanıp 100 puanlık ölçeğe dönüştürülmesiyle her bir uzmanın bireysel geçme puanı hesaplanmıştır. Bireysel geçme puanlarının ortalaması alınarak da testin nihai geçme puanı belirlenmiştir. Her bir alt test için Angoff ve Evet/Hayır yöntemlerine göre ayrı ayrı geçme puanı yine yukarıdaki adımlar takip edilerek hesaplanmış ve tüm testle aynı ölçeğe yerleştirilmiştir. Tüm teste ve oluşturulan alt testlere ilişkin geçme puanları arasında manidar düzeyde farklılık olup olmadığını incelemek için tek-yönlü ANOVA yapılmıştır.

Yukarıdaki analizler ışığında sonuçların genellenebilirliğini incelemek amacıyla RMSE ve TSH incelemeleri yapılmıştır. Bu sayede, madde sayısını

azaltmada farklı madde örnekleme yöntemlerinin etkililiği incelenmiş ve sonuçların genellenebilirliği hakkında daha sistematik ve kararlı bulgular elde edilmeye çalışılmıştır.

Her bir geçme puanı için TSH aşağıdaki formülle hesaplanmıştır.

$$TSH = \frac{SS}{\sqrt{N}}$$

Bu formülde;

SEE: Standart hata

SS: Uzmanların bireysel geçme puanlarının standart sapması

N: Uzman sayısını temsil etmektedir.

TSH'ye ek olarak incelenen RMSE değerleri ise aşağıdaki formül yardımıyla hesaplanmıştır.

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (\hat{X} - X_j)^2}{N}}$$

Bu formülde;

$\hat{X}$: Tüm testin minimum geçme puanı, X_j : j. replikasyona ilişkin geçme puanı, N: Toplam replikasyon sayısı (=1000) anlamına gelmektedir.

Bölüm 4

Bulgular ve Yorumlar

Bu bölümde alt problemlere ilişkin bulgulara ve tartışmalara yer verilmiştir.

Angoff ve Evet/Hayır yöntemlerinde 28 uzmanın testteki her bir madde için değerlendirmelerine ilişkin puanları sırasıyla EK-B ve EK-C'de verilmiştir. Bu yöntemlerin her birine göre tüm testin minimum geçme puanlarını (MGP) hesaplamadan önce, uzmanlar arasındaki tutarlık derecesi Kendall'ın uyum katsayısı ile analiz edilmiştir. Yapılan analizlere göre Angoff yönteminde Kendall'ın uyum katsayısı .30 olarak bulunurken ($\chi^2=322.99$, $sd=39$, $p<.05$); Evet/Hayır yönteminde ise bu katsayı .23 olarak hesaplanmıştır ($\chi^2=254.44$, $sd=39$, $p<.05$). Bu bilgiler ışığında her iki yöntemde de uzmanlar arasında anlamlı bir uyumun olduğu söylenebilir.

Alt Problemlere İlişkin Bulgular ve Yorumlar

Alt problem 1a'ya ilişkin bulgular ve yorumlar. *Alt problem 1a. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Angoff geçme puanları arasında BSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Bu alt probleme ilişkin analizlerde, öncelikle 40 maddeden oluşan tüm teste ait minimum geçme puanı (MGP) hesaplanmıştır. Sonrasında, her bir alt testte (%30, %40, %50 ve %70) yönelik 1000 replikasyon yapılmıştır. Böylelikle, alt testlerden elde edilen MGP değerlerinin tutarlılığının ve güvenilirliğinin artırılması amaçlanmıştır. Basit seçkisiz örnekleme (BSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyonun MGP'lerine ilişkin betimsel istatistikler Tablo 6'da verilmiştir.

Tablo 6

BSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine İlişkin Betimsel İstatistikler

BSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	45.03	2.47	38.38	52.94
%40	16	45.15	1.96	38.94	51.36
%50	20	45.15	1.51	40.76	49.76
%70	28	45.16	1.02	41.82	48.13
Tüm Test	40	45.12			

Tablo 6'da görüldüğü üzere, Angoff yöntemine göre tüm teste(40 madde) ilişkin MGP 45.12 olarak hesaplanmıştır. Alt testlere ilişkin değerler incelendiğinde, maddelerin %30'unun (12 maddenin) basit seçkisiz yöntemle seçilmesiyle oluşturulan 1000 farklı alt testin MGP ortalaması 45.03, standart sapması 2.47 olarak bulunmuştur. Bu testlerin MGP'lerinin 38.38 ile 52.94 puan arasında değiştiği gözlenmiştir. Maddelerin %40'ının seçilmesiyle oluşturulan 16 maddelik alt testlerin MGP'leri (38.94-51.36) aralığında değişmişken, ortalaması 45.15 standart sapması ise 1.96 olarak bulunmuştur. BSÖ yöntemine göre maddelerin %50'si seçkisiz olarak 1000 defa seçildiğinde, oluşan alt testlerin ortalaması 45.15, standart sapması 1.51 olarak hesaplanmıştır. Bu 20 maddelik alt testlerin MGP'leri 40.76 ile 49.76 puan aralığında değerler almıştır. Ayrıca, tüm test maddelerinin %70'i seçildiğinde ise, oluşan alt testlerin ortalaması 45.16, standart sapması ise 1.02 olarak bulunmuştur. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 41.82 ile 48.13 aralığında değişmiştir. Farklı büyüklükteki alt testlerin standart sapması ve minimum-maksimum değerleri göz önünde bulundurulduğunda, madde sayısı arttıkça alt testlerin MGP'leri tüm testin MGP'sine yaklaşmıştır.

Tüm teste ve alt testlere ilişkin MGP ortalamalarının birbirinden manidar düzeyde farklılaşıp farklılaşmadığını anlamak amacıyla ANOVA analizi yapılmadan önce normallik ve varyansların homojenliği varsayımları test edilmiştir. Normallik varsayımının sağlandığı ancak varyansların homojenliği varsayımının ihlal edildiği görülmüş; bu nedenle Welch testi sonuçlarına bakılmıştır. Elde edilen sonuçlar, alt testlere ilişkin MGP'ler arasında manidar derecede fark olmadığını ($F(4,1998) = 0.824, p=.510$) göstermiştir.

Daha ayrıntılı inceleme yapmak ve tüm testi MGP anlamında en iyi şekilde temsil eden alt testi belirlemek amacıyla, TSH ve RMSE değerleri incelenmiştir.

Tablo 7

BSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine ait RMSE ve Alt Test Yüzdeleri

BSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	2.47	%70.1	%96.4
%40	1.96	%81.3	%99.3
%50	1.51	%91.5	%100
%70	1.02	%98.6	%100
Tüm Test	TSH = 2.58 Ortalama $\pm$ 1TSH= (42.54 - 47.70) Ortalama $\pm$ 2TSH = (39.96 – 50.28)		

Tablo 7'de görüldüğü gibi, tüm test için tahminin standart hatası değeri (TSH) 2.58 olarak hesaplanmıştır. Alt test yüzdesi kolonunda verilen değerler, her bir alt test için oluşturulan 1000 farklı alt testten yüzde kaçının MGP'sinin belirlenen sınırlar içerisinde kaldığına dair bilgi vermektedir. Bu bilgilere göre, tüm testin geçme puanının 1 TSH ve 2 TSH değeri içerisinde yer alan MGP yüzdeleri, %30'luk test için en düşük, %70'lik test için en büyük değerleri almışlardır. RMSE değerleri incelendiğinde de beklenildiği gibi, madde sayısının artışıyla birlikte hata değeri azalmıştır. En az hata %70'lik (1.02), en fazla hata ise %30'luk (2.47) alt testte elde edilmiştir. Ek olarak, sadece %70'lik alt testlere ilişkin Angoff MGP'leri, istenen kriterin (en az %95 olasılıkla nihai geçme puanının 1 TSH içerisi) içerisinde yer almıştır. Maddelerin %70'iyle oluşturulan alt testlerin MGP'leri ile tüm testin MGP'si arasındaki farkın mutlak değerleri, 1000 alt testin %98.6'sında 1 TSH'den azdır. Sonuç olarak, tüm testin MGP'sine benzer geçme puanı elde edebilmek için, BSÖ yöntemiyle test maddelerinin en az %70'inin kullanılması önerilebilir.

Alt problem 1b'ye ilişkin bulgular ve yorumlar. *Alt problem 1b. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Angoff geçme puanları arasında İ-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Tablo 8’de içerik tabakalı seçkisiz örnekleme (İ-TSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyona ilişkin betimsel istatistikler verilmiştir.

Tablo 8

İ-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP’sine İlişkin Betimsel İstatistikler

İ-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	44.81	2.05	37.92	50.93
%40	16	45.43	1.66	40.25	49.89
%50	20	45.08	1.37	41.42	49.11
%70	28	45.20	0.86	42.63	48.16
Tüm Test	40	45.12			

Tablo 8 incelendiğinde, maddelerin %30’unun İ-TSÖ yöntemine göre seçilmesiyle oluşturulan 1000 farklı alt testin ortalaması 44.81, standart sapması 2.05 olarak bulunmuştur. Bu testlerin MGP’lerinin 37.92 ile 50.93 puan aralığında değiştiği gözlenmiştir. Maddelerin %40’ının seçilmesiyle oluşturulan 16 maddelik alt testlerin MGP’leri (40.25-49.89) aralığında değişmişken, ortalaması 45.43 standart sapması ise 1.66 olarak bulunmuştur. Test maddelerinin %50’si 1000 defa seçildiğinde, oluşan alt testlerin ortalaması 45.08, standart sapması 1.37 olarak hesaplanmıştır. Bu alt testlerin MGP’leri 41.42 ile 49.11 puan aralığında değerler almıştır. Ayrıca, test maddelerinin %70’i seçildiğinde ise, oluşan alt testlerin ortalaması 45.20, standart sapması ise 0.86 olarak bulunmuştur. Bu alt testlerin MGP’leri ise 42.63 ile 48.16 puan aralığında değişmiştir. Farklı büyüklükteki alt testlerin standart sapması ve minimum-maksimum değerleri göz önünde bulundurulduğunda, madde sayısı arttıkça alt testlerin MGP’leri tüm testin MGP’sine yaklaşmıştır.

Tüm test ve alt testlerin geçme puanı ortalamaları bakımından manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. ANOVA analizi öncesinde yapılan varsayım kontrollerine göre normallik varsayımı sağlanmış ancak varyansların homojenliği ihlal edilmiştir. Bu nedenle Welch testi sonuçları yorumlanmış ve analiz sonuçları, MGP ortalamalarının testler arasında manidar düzeyde farklılaştığını ($F(4,1998) = 16.866, p=.000$) ortaya koymuştur.

Hangi alt testlerin MGP ortalamalarının tüm testin MGP'sinden manidar düzeyde farklılaştığını belirlemek amacıyla çoklu karşılaştırma (Post-Hoc) yapılmıştır. Karşılaştırma sonuçlarına göre, hem %30'luk ($X_{ort}=44.81$) hem de %40'luk ($X_{ort}= 45.43$) alt testlerin MGP ortalamaları, tüm testin MGP'sinden ($X_{ort}=45.12$) manidar derecede farklılık göstermiştir ($p<.05$). Dolayısıyla, İ-TSÖ yöntemiyle oluşturulan %30 ve %40'luk alt testlerin geçme puanı bakımından tüm testi temsil ettiği yorumu yapılamaz.

Ayrıca, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve elde edilen değerler Tablo 9'da sunulmuştur.

Tablo 9

İ-TSÖ ile Oluşturulan Alt Testlerin Angof MGP'sine ait RMSE ve Alt Test Yüzdeleri

İ-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	2.07	%78.9	%99.3
%40	1.69	%86.8	%100
%50	1.37	%95.0	%100
%70	0.86	%99.9	%100
Tüm Test	TSH = 2.58 Ortalama $\pm$ 1TSH= (42.54 - 47.70) Ortalama $\pm$ 2TSH = (39.96 – 50.28)		

Tablo 9'a göre, tüm testin geçme puanının 1 TSH ve 2 TSH değeri içerisinde yer alan MGP yüzdeleri, %30'luk test için en düşük, %70'lik test için en büyük değerleri almışlardır. Bununla birlikte, sadece %50 ve %70'lik alt testlere ilişkin Angoff MGP'leri, istenen kriterin (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) içerisinde yer almıştır. Test maddelerinin %50 ve %70'sini içeren 1000 farklı alt testin sırasıyla %95 ve %99.9'unun MGP'leri, tüm testin MGP'sinin 1 TSH değeri içerisinde yer almıştır. Ayrıca, farklı büyüklükteki alt testlerin neredeyse tamamının MGP'si, tüm testin MGP'sinin 2 TSH değeri içerisinde yer almıştır. RMSE değerleri incelendiğinde ise, beklenildiği gibi madde sayısının artışıyla birlikte hata değeri azalmıştır. Hata değeri; %50'lik test için 1.37, %70'lik alt test için ise 0.86 olarak bulunmuştur. Bu bilgiler ışığında, tüm testin MGP'sine benzer geçme puanı elde

edebilmek için, İ-TSÖ yöntemiyle test maddelerinin en az %50'sinin kullanılması önerilebilir.

Alt problem 1c'ye ilişkin bulgular ve yorumlar. *Alt problem 1c. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Angoff geçme puanları arasında G-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Tablo 10'da tüm teste ve madde güçlüğü tabakalı seçkisiz örnekleme (G-TSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyona ilişkin betimsel istatistikler verilmiştir.

Tablo 10

G-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine İlişkin Betimsel İstatistikler

G-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	44.85	2.20	38.50	51.16
%40	16	45.00	2.19	39.79	50.10
%50	20	45.13	1.43	40.36	49.81
%70	28	45.32	0.94	42.32	48.04
Tüm Test	40	45.12			

Tablo 10 incelendiğinde, maddelerin %30'unun (12 maddenin) G-TSÖ yöntemine göre seçilmesiyle oluşturulan 1000 farklı alt testin ortalaması 44.85, standart sapması 2.20 olarak bulunmuştur. Bu testlerin MGP'lerinin 38.50 ile 51.16 puan aralığında değiştiği gözlenmiştir. Maddelerin %40'ının seçilmesiyle oluşturulan 16 maddelik alt testlerin MGP'leri (39.79-50.10) aralığında değişmişken, ortalaması 45.00 standart sapması ise 2.19 olarak bulunmuştur. G-TSÖ yöntemine göre maddelerin %50'si 1000 defa seçildiğinde, oluşan alt testlerin ortalaması 45.13, standart sapması 1.43 olarak hesaplanmıştır. Bu 20 maddelik alt testlerin MGP'leri 40.36 ile 49.81 puan aralığında değerler almıştır. Ayrıca, tüm test maddelerinin %70'i seçildiğinde ise, oluşan alt testlerin ortalaması 45.32, standart sapması ise 0.94 olarak bulunmuştur. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 42.32 ile 48.04 aralığında değişmiştir. Farklı büyüklükteki alt testlerin standart sapması ve minimum-maksimum değerleri göz önünde bulundurulduğunda, madde sayısı arttıkça alt testlerin MGP'leri tüm testin MGP'sine yaklaşmıştır.

Tüm test ve alt testlerin geçme puanı ortalamaları bakımından manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. Analiz öncesi yapılan varsayım kontrollerinde varyansların homojenliğinin sağlanmadığı görülmüş, bu nedenle Welch tetsi sonuçları yorumlanmıştır. Analiz sonuçları, MGP ortalamalarının testler arasında manidar düzeyde farklılaştığını ($F(4,1998)=16.110$, $p=.000$) ortaya koymuştur.

Hangi alt testlerin MGP ortalamalarının tüm testin MGP'sinden manidar düzeyde farklılaştığını belirlemek amacıyla çoklu karşılaştırma (Post-Hoc) yapılmıştır. Karşılaştırma sonuçlarına göre, sadece %30'luk ($X_{ort}=44.85$) alt testlerin MGP ortalaması, tüm testin MGP'sinden ($X_{ort}=45.12$) manidar derecede farklılık göstermiştir ($p<.05$). Dolayısıyla, G-TSÖ yöntemi kullanılarak üretilen %30'luk alt test ile tüm testin MGP'sine benzer geçme puanı elde edilebileceği söylenemez.

Yukarıdaki analizler dışında, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve Tablo 11'de sunulmuştur.

Tablo 11

G-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'lerine Ait RMSE ve Alt Test Yüzdeleri

G-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	2.21	%74.5	%98
%40	1.67	%87.8	%99.9
%50	1.43	%93.5	%100
%70	0.96	%99.1	%100
Tüm Test	TSH = 2.58 Ortalama $\pm$ 1TSH= (42.54 - 47.7) Ortalama $\pm$ 2TSH = (39.96 – 50.28)		

Tablo 11'e göre, tüm testin MGP'sinin 1 TSH değeri içerisinde kalan yüzdeleri, madde sayısı arttıkça artmıştır. Ancak, sadece %70'lik alt testlere ilişkin Angoff MGP'leri, istenen kriteri (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) sağlamıştır. Test maddelerinin %70'ini içeren 1000 farklı alt testin MGP'leri

ve tüm testin MGP'si arasındaki farkın mutlak değeri, alt testlerin %99.1'inde 1 TSH'den küçüktür. Bunun yanında, farklı büyüklükteki alt testlerin neredeyse tamamının MGP'leri, tüm testin MGP'sinden en fazla 2 TSH farklılaşmıştır. Beklenildiği gibi, madde sayısı arttıkça RMSE hata değeri azalmıştır. En az hata %70'lik (0.96), en fazla hata ise %30'lık (2.21) alt testte elde edilmiştir. Sonuç olarak, tüm testin MGP'sine benzer geçme puanı elde edebilmek için, toplam madde sayısının en az %70'inin G-TSÖ yöntemi dikkate alınarak seçilmesi tavsiye edilebilir.

Alt problem 1d'ye ilişkin bulgular ve yorumlar. *Alt problem 1d. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Angoff geçme puanları arasında İG-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Tablo 12'de içerik ve madde güclüğü tabakalı seçkisiz örnekleme (İG-TSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyona ilişkin betimsel istatistikler verilmiştir.

Tablo 12

İG-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'sine İlişkin Betimsel İstatistikler

İG-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	44.66	2.14	39.02	51.07
%40	16	45.12	1.64	40.67	50.02
%50	20	45.13	1.44	41.57	49.81
%70	28	45.16	0.93	41.93	48.05
Tüm Test	40	45.12			

Tablo 12'ye göre, maddelerin %30'unun (12 maddenin) İG-TSÖ yöntemine göre seçilmesiyle oluşturulan 1000 farklı alt testin ortalaması 44.66, standart sapması 2.14 olarak bulunmuştur. Bu testlerin MGP'lerinin 39.02 ile 51.07 puan aralığında değiştiği gözlenmiştir. Maddelerin %40'ının seçilmesiyle oluşturulan 16 maddelik alt testlerin MGP'leri (40.67-50.02) aralığında değişmişken, ortalaması 45.12 standart sapması ise 1.64 olarak bulunmuştur. İG-TSÖ yöntemine göre maddelerin %50'si 1000 defa seçildiğinde, oluşan alt testlerin ortalaması 45.13, standart sapması 1.44 olarak hesaplanmıştır. Bu 20 maddelik alt testlerin MGP'leri 41.57 ile 49.81 puan aralığında değerler almıştır. Ayrıca, tüm test maddelerinin %70'i seçildiğinde ise, oluşan alt testlerin ortalaması 45.16, standart sapması ise

0.93 olarak bulunmuştur. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 41.93 ile 48.05 aralığında değişmiştir. Farklı büyüklükteki alt testlerin standart sapması ve minimum-maksimum değerleri göz önünde bulundurulduğunda, madde sayısı arttıkça alt testlerin MGP'leri tüm testin MGP'sine yaklaşmıştır.

Tüm test ve alt testlerin MGP ortalamalarının manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. Varsayımlardan normallik sağlanmış ancak varyansların homojenliği ihlal edilmiştir. Buna göre yorumlanan Welch testi sonuçları, MGP ortalamalarının testler arasında manidar düzeyde farklılaştığını ($F(4,1998)=12.133$ $p=.000$) ortaya koymuştur.

Yapılan çoklu karşılaştırma (Post-Hoc) sonuçlarına göre, sadece %30'luk ($X_{ort}=44.66$) alt testlerin MGP ortalaması, tüm testin MGP'sinden ($X_{ort}=45.12$) manidar derecede farklılık göstermiştir ($p<.05$). Bu nedenle, İG-TSÖ yöntemine göre oluşturulan %30'luk alt test ile tüm testin MGP'sine benzer geçme puanı elde edilebileceği söylenemez.

Daha detaylı inceleme yapmak amacıyla, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış olup Tablo 13'te sunulmuştur.

Tablo 13

İG-TSÖ ile Oluşturulan Alt Testlerin Angoff MGP'sine ait RMSE ve Alt Test Yüzdeleri

İG-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	2.19	%84.9	%98.3
%40	1.64	%88.3	%100
%50	1.44	%93.0	%100
%70	0.93	%99.5	%100
Tüm Test	TSH = 2.58 Ortalama $\pm$ 1TSH= (42.54 - 47.7) Ortalama $\pm$ 2TSH = (39.96 – 50.28)		

Yukarıdaki tabloya göre, sadece %70'lik alt testlere ilişkin Angoff MGP'leri, istenen kriterin içerisinde (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi)

yer almıştır. Her biri toplam madde sayısının %70'i kadar madde içeren 1000 farklı alt testin neredeyse tamamının MGP'leri, tüm testin MGP'sinden en fazla 1 TSH farklılaşmıştır. Ayrıca, farklı büyüklükteki alt testlerin neredeyse tamamının MGP'leri, tüm testin MGP'sinin 2 TSH içerisinde yer almıştır. RMSE hata değerlerine bakıldığında ise, madde sayısı arttıkça RMSE hata değeri azalmıştır. En az hata %70'lik (0.93), en fazla hata ise %30'lık (2.19) alt testte elde edilmiştir. Bu doğrultuda, tüm testin MGP'sine benzer geçme puanı elde edebilmek için, toplam madde sayısının en az %70'inin İG-TSÖ yöntemi dikkate alınarak seçilmesi tavsiye edilebilir.

Alt problem 2'ye ilişkin bulgular ve yorumlar. *Alt problem 2. Farklı seçkisiz madde örnekleme yöntemleriyle oluşturulan alt testlere ilişkin Angoff geçme puanları birbirlerinden nasıl farklılaşmaktadır?*

Farklı seçkisiz örnekleme yöntemlerini birbirleriyle karşılaştırmak amacıyla, tüm testin geçme puanına benzer sonuçlar veren %50 ve %70'lik alt testler seçilmiştir. Farklı örnekleme durumları için ortalamadan 1 TSH fark içerisinde yer alan geçme puanı yüzdeleri ve RMSE değerleri Tablo 14'te gösterilmiştir.

Tablo 14

Angoff MGP'lerin Örnekleme Durumlarına Ait RMSE ve Alt Test Yüzdesi Değerleri

	Alt Test Yüzdesi							
	RMSE				Ort.±1TSH			
	BSÖ	İ-TSÖ	G-TSÖ	İG-TSÖ	BSÖ	İ-TSÖ	G-TSÖ	İG-TSÖ
%50	1.51	1.37	1.43	1.44	%91.5	%95.0	%93.5	%93.0
%70	1.02	0.86	0.96	0.93	%98.6	%99.9	%99.1	%99.5
Tüm Test	TSH = 2.58 Ortalama±1TSH= (42.54 - 47.7)							

Yukarıdaki tablodan anlaşılabacağı üzere, her bir seçkisiz örnekleme yönteminde madde sayısı arttıkça beklenildiği gibi RMSE hata değeri azalmıştır. Bunun yanı sıra, RMSE hata değerlerinin, BSÖ yöntemine kıyasla TSÖ yöntemleri için daha küçük olduğu görülmektedir. Test maddelerinin %50 ve %70'i büyüklüğündeki alt testler için en az hata değerleri (1.37;0.86) İ-TSÖ yöntemiyle elde edilmiştir.

Tüm testin MGP'sinin 1 TSH içerisinde (42.54 - 47.70) yer alan yüzdelik oranları incelendiğinde, en küçük oran BSÖ yöntemiyle en büyük oran ise İ-TSÖ yöntemiyle elde edilmiştir. Dolayısıyla, tüm testin MGP'sine en yakın geçme puanı elde etme hususunda İ-TSÖ yönteminin daha etkili olduğu söylenebilir. Buna ek olarak uygulanan benzerlik kriteri (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) dikkate alındığında da, yine aynı şekilde İ-TSÖ yönteminin daha etkili olduğu görülmektedir. Tüm testin MGP'sine benzer kesme puanı elde etmek için İ-TSÖ yöntemiyle toplam madde sayısının %50'sinin seçilmesi yeterliyken; BSÖ, G-TSÖ ve İG-TSÖ yöntemleriyle en az %70'inin seçilmesi gerekmektedir. Sonuç olarak, MGP açısından tüm testi temsil etmesi düşünülen alt test/testlere madde seçiminde içerik tabakalı seçkisiz örnekleme yöntemi (İ-TSÖ) daha etkili bir yöntem olabilir.

Alt problem 3a'ya ilişkin bulgular ve yorumlar. *Alt problem 3a. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Evet/Hayır geçme puanları arasında BSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Bu alt probleme ilişkin analizlerde, öncelikle 40 maddeden oluşan tüm teste ait Evet/Hayır minimum geçme puanı (MGP) hesaplanmıştır. Sonrasında, sonuçların tutarlığını ve güvenilirliğini arttırmak amacıyla her bir alt teste (%30, %40, %50 ve %70) yönelik 1000 replikasyon yapılmıştır. Basit seçkisiz örnekleme (BSÖ) yöntemiyle her bir alt test için 1000 replikasyon üretilmiş ve bunların geçme puanlarına ilişkin betimsel istatistikler Tablo 15'de verilmiştir.

Tablo 15

BSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler

BSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	56.30	5.52	36.61	72.02
%40	16	56.37	4.42	42.19	71.88
%50	20	56.31	3.52	46.07	66.96
%70	28	56.26	2.31	49.49	64.54
Tüm Test	40	56.25			

Tablo 15'de görüldüğü üzere, Evet/Hayır yöntemine göre tüm teste (40 madde) ilişkin MGP 56.25 olarak hesaplanmıştır. Alt testlere ilişkin değerler incelendiğinde, maddelerin %30'unun (12 maddenin) basit seçkisiz yöntemle

seçilmesiyle oluşturulan 1000 farklı alt testin ortalaması 56.30, standart sapması 5.52 olarak bulunmuştur. Bu testlerin MGP'lerinin 36.61 ile 72.02 puan aralığında değiştiği gözlenmiştir. Maddelerin %40'ının seçilmesiyle oluşturulan 16 maddelik alt testlerin MGP'leri (42.19-71.88) aralığında değişmişken, ortalaması 56.37 standart sapması ise 4.42 olarak bulunmuştur. Maddelerin %50'si seçkisiz olarak 1000 defa seçildiğinde, oluşan alt testlerin ortalaması 56.31, standart sapması 3.52 olarak hesaplanmıştır. Bu 20 maddelik alt testlerin MGP'leri 46.07 ile 66.96 puan aralığında değerler almıştır. Ayrıca, tüm test maddelerinin %70'i seçildiğinde ise, oluşan alt testlerin ortalaması 56.26, standart sapması ise 2.31 olarak bulunmuştur. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 49.49 ile 64.54 aralığında değişmiştir. Farklı büyüklükteki alt testlerin standart sapması ve minimum-maksimum değerleri dikkate alındığında, madde sayısı arttıkça alt testlerin MGP'leri tüm testin MGP'sine yaklaşmıştır.

Tüm teste ve alt testlere ilişkin MGP ortalamalarının birbirinden manidar düzeyde farklılaşıp farklılaşmadığını araştırmak amacıyla ANOVA analizi yapılmıştır. Analiz öncesi kontrol edilen varsayımlardan normalliğin sağlandığı, varyansların homojenliğinin ise ihlal edildiği görülmüştür. Sonrasında Welch istatistiği yorumlanmış ve elde edilen sonuçlar, alt testlere ilişkin geçme puanları ortalamaları arasında manidar derecede fark olmadığını ($F(4,1998) = 0.301$, $p=0.877$) göstermiştir.

Daha ayrıntılı inceleme yapmak amacıyla, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve elde edilen değerler Tablo 16'da verilmiştir.

Tablo 16

BSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine ait RMSE ve Alt Test Yüzdeleri

BSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	5.52	%41.8	%72.6
%40	4.42	%52.1	%82.7
%50	3.52	%62.9	%91.9
%70	2.31	%82.9	%99.3
Tüm Test	TSH = 3.09 Ortalama $\pm$ 1TSH = (53.16- 59.34) Ortalama $\pm$ 2TSH = (50.07 – 62.43)		

Tablo 16'da görüldüğü gibi, tüm test için tahminin standart hatası (TSH) değeri 3.09 olarak bulunmuştur. Tüm testin geçme puanının 1 TSH ve 2 TSH değeri içerisinde yer alan MGP yüzdeleri, %30'luk test için en düşük, %70'lik test için en büyük değerleri almışlardır. Ancak, farklı büyüklükteki bu 4 alt testin hiçbirinin MGP'si belirlenen kriterin (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) içerisinde yer almamıştır. Bununla birlikte, sadece %70'lik alt teste ait MGP'ler, en az %95 olasılıkla nihai geçme puanının 2 TSH içerisinde yer almıştır. RMSE değerleri incelendiğinde ise, beklenildiği gibi madde sayısının artışıyla birlikte hata değeri azalmıştır. En az hata %70'lik (2.31), en fazla hata ise %30'luk (5.52) alt testte elde edilmiştir. Bu bilgiler ışığında, seçkisiz olarak oluşturulan farklı büyüklükteki bu 4 alt testin herhangi biri ile tüm testin MGP'sine benzer kesme puanı(en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) elde etmek mümkün olmayabilir. Ancak, toplam madde sayısının %70'i ile tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde edilebilir.

Alt problem 3b'ye ilişkin bulgular ve yorumlar. *Alt problem 3b. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Evet/Hayır geçme puanları arasında İ-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

İçerik tabakalı seçkisiz örnekleme (İ-TSÖ) yöntemiyle her bir alt test için 1000 replikasyon oluşturulmuş ve bu replikasyonların geçme puanlarına ilişkin betimsel istatistikler Tablo 17'de sunulmuştur.

Tablo 17

İ-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler

İ-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	55.60	4.14	43.75	67.56
%40	16	56.91	3.28	46.88	67.41
%50	20	56.29	2.76	48.93	63.93
%70	28	56.55	1.80	51.15	61.61
Tüm Test	40	56.25			

Tablo 17 incelendiğinde, tüm testin içerik kategorilerinden toplam madde sayısının %30'u (12 madde) kadar maddenin 1000 defa seçilmesiyle oluşan alt testlerin MGP ortalaması 55.60, standart sapması 4.14 olarak bulunmuştur. Bu testlerin MGP'lerinin 43.75 ile 67.56 puan aralığında değiştiği gözlenmiştir. Toplam madde sayısının %40'ının (16 madde) yine 1000 defa seçilmesiyle oluşan alt testlerin MGP'leri (46.88-67.41) aralığında değişmişken, ortalaması 56.91 standart sapması ise 3.28 olarak bulunmuştur. Toplam madde sayısının %50'si (20 madde) seçildiğinde ise, oluşan alt testlerin ortalaması 56.29. standart sapması 2.76 olarak hesaplanmıştır. Bu 20 maddelik alt testlerin MGP'leri 48.93 ile 63.93 puan aralığında değerler almıştır. Ayrıca. 40 maddenin %70'i (28 madde) seçildiğinde ise, oluşan alt testlerin ortalaması 56.55. standart sapması ise 1.80 olarak bulunmuştur. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 51.15 ile 61.61 aralığında değişmiştir. Standart sapma ve minimum-maksimum değerleri karşılaştırıldığında, alt teste seçilen madde sayısı arttıkça alt testlerin MGP'lerinin nihai MGP'ye yaklaştığı görülmektedir.

Tüm test ve alt testlerin geçme puanı ortalamaları bakımından manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. Varyansların homojenliği varsayımının ihlal edildiği görüldüğünden, Welch testi sonuçları incelenmiştir. Analiz sonuçları, MGP ortalamalarının testler arasında manidar düzeyde farklılaştığını ($F(4,1998)=23.280$, $p=.000$) ortaya koymuştur. Çoklu karşılaştırma (Post-Hoc) sonuçlarına göre, hem %30'luk ($X_{ort}=55.60$) hem de %40'lık ($X_{ort}= 56.91$) alt testlerin MGP ortalamaları, tüm testin MGP'sinden

($X_{ort}=56.25$) manidar derecede farklılık göstermiştir ($p<.05$). Dolayısıyla, İ-TSÖ yöntemi kullanılarak üretilen %30 ve %40'lık alt testler ile tüm testin MGP'sine yakın geçme puanı elde edilebileceği söylenemez.

Daha detaylı inceleme yapmak amacıyla, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve bu değerler Tablo 18'de sunulmuştur.

Tablo 18

İ-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine ait RMSE ve Alt Test Yüzdesi Değerleri

İ-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	4.19	%54.1	%84.7
%40	3.34	%63.6	%92.8
%50	2.75	%72.9	%97.4
%70	1.82	%90.9	%100
Tüm Test	TSH = 3.09 Ortalama $\pm$ 1TSH= (53.16- 59.34) Ortalama $\pm$ 2TSH = (50.07 – 62.43)		

Tablo 18'de görüldüğü üzere, tüm testin MGP'sinin 1 TSH ve 2 TSH değeri içerisinde kalan yüzdeleri, madde sayısı arttıkça artmıştır. Buna rağmen, hiçbir alt testin MGP ortalaması, istenen kriterin (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) içerisinde yer almamıştır. Buna ek olarak, sadece %50 ve %70'lik alt testlere ilişkin MGP'ler en az %95 ihtimalle tüm testin MGP'sinin 2 TSH içerisinde yer almıştır. Diğer %30 ve %40'lık alt testlerde bu yüzde daha düşüktür. RMSE değerlerine bakıldığında ise, beklenildiği gibi en az hata değeri %70'lik (1.82), en fazla hata ise %30'luk (4.19) alt testte bulunmuştur. Sonuç olarak, İ-TSÖ yöntemiyle oluşturulan bu 4 alt testin (%30,%40,%50 ve %70) herhangi biri ile nihai geçme puanına benzer geçme puanı (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) elde etmek pek mümkün olmayabilir. Fakat toplam madde sayısının en az %50'i ile tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde edilebilir.

Alt problem 3c'ye ilişkin bulgular ve yorumlar. *Alt problem 3c. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Evet/Hayır geçme puanları arasında G-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

Güçlük tabakalı seçkisiz örnekleme (G-TSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyonun geçme puanlarına ilişkin betimsel istatistikler Tablo 19'da verilmiştir.

Tablo 19

G-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler

G-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	55.54	5.08	41.67	71.13
%40	16	56.03	3.87	43.30	68.30
%50	20	56.57	3.29	46.25	68.57
%70	28	56.67	2.19	50.13	62.88
Tüm Test	40	56.25			

Tablo 19'da görüldüğü gibi, tüm testin güçlük kategorilerinden toplam madde sayısının %30'u (12 madde) kadar maddenin 1000 defa seçilmesiyle oluşan alt testlerin MGP ortalaması 55.54, standart sapması 5.08 olarak bulunmuştur. Bu testlerin MGP'leri 41.67 ile 71.13 puan aralığında değişmiştir. Toplam madde sayısının %40'ının (16 madde) yine 1000 defa seçilmesiyle oluşan alt testlerin MGP'leri (43.30-68.30) aralığında değişmişken, ortalaması 56.03 standart sapması ise 3.87 olarak bulunmuştur. G-TSÖ yöntemine göre tüm testin %50'si (20 madde) seçildiğinde ise oluşan alt testlerin ortalaması 56.57, standart sapması 3.29 olarak hesaplanmıştır. Bu alt testlerin geçme puanları ise 46.25 ile 68.57 puan aralığında değerler almıştır. Son olarak, tüm testin %70'i (28 madde) seçildiğinde ise oluşan alt testlerin ortalaması 56.67, standart sapması ise 2.19 değerlerini almıştır. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 50.13 ile 62.88 aralığında değişmiştir.

Tüm test ve alt testlerin geçme puanı ortalamaları bakımından manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile analiz edilmiştir. ANOVA'nın varsayımlarından olan varyansların homojenliği ihlal edildiği için Welch testi sonuçları ele alınmıştır. Analiz sonuçları, geçme puanları ortalamalarının testler

arasında manidar düzeyde farklılaştığını ($F(4, 1998)=17.167, p=.000$) göstermiştir. Ayrıca, çoklu karşılaştırma (Post-Hoc) analizi yapılmış ve buna göre, sadece %30'luk ($X_{ort}=55.54$) alt testlerin MGP ortalaması, tüm testin MGP'sinden ($X_{ort}=56.25$) manidar derecede farklılık göstermiştir ($p<.05$). Dolayısıyla, G-TSÖ yöntemi kullanılarak üretilen %30'luk alt test ile tüm testin MGP'sine yakın geçme puanı elde edilebileceği yorumu yapılamaz.

Daha ayrıntılı inceleme yapmak amacıyla, her bir alt test için belirlenen sınırlar ($ortalama\pm 1TSH$; $ortalama\pm 2TSH$) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve elde edilen değerler Tablo 20'de sunulmuştur.

Tablo 20

G-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine Ait RMSE ve Alt Test Yüzdesi Değerleri

G-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm 1TSH$	Ortalama $\pm 2TSH$
%30	5.13	%43.6	%75.7
%40	3.87	%53.9	%89.6
%50	3.30	%64.9	%94.3
%70	2.22	%84.2	%99.8
Tüm Test	TSH = 3.09 Ortalama $\pm 1TSH$ = (53.16- 59.34) Ortalama $\pm 2TSH$ = (50.07 – 62.43)		

Tablo 20'de görüldüğü gibi, tüm testin MGP'sinin 1 TSH ve 2 TSH değer aralığının içerisinde kalan yüzdeleri, %30'luk test için en düşük, %70'lik test için en büyük değerleri almışlardır. Diğer bir deyişle, beklenildiği gibi madde sayısı arttıkça bu yüzde değerleri de artmıştır. Buna rağmen, hiçbir alt testin MGP ortalaması, benzerlik kriterinin (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) içerisinde yer almamıştır. Bununla birlikte, sadece %70'lik alt testlere ilişkin MGP'ler tüm testin MGP'sinin 2 TSH içerisinde (en az %95 ihtimalle) yer almıştır. Diğer alt testlerde bu yüzde daha düşük çıkmıştır. RMSE değerlerine bakıldığında ise, beklenildiği gibi en az hata değeri %70'lik (2.22), en fazla hata ise %30'luk (5.13) alt testte hesaplanmıştır. Bu bilgiler ışığında, G-TSÖ yöntemiyle oluşturulan bu 4 alt testin (%30,%40,%50 ve %70) herhangi biri ile nihai Evet/ Hayır MGP'ye benzer

geçme puanı (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içerisi) elde etmek pek olası olmayabilir. Fakat toplam madde sayısının en az %70'i ile tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir MGP elde edilebilir.

Alt problem 3d'ye ilişkin bulgular ve yorumlar. *Alt problem 3d. Tüm teste ve tüm testten oluşturulan farklı alt testlere ilişkin Evet/Hayır geçme puanları arasında İG-TSÖ yöntemine göre manidar düzeyde fark var mıdır?*

İçerik ve madde güçlüğü tabakalı seçkisiz örnekleme (İG-TSÖ) yöntemiyle her bir alt test için oluşturulan 1000 replikasyonun geçme puanlarına ilişkin betimsel istatistikler Tablo 21'de sunulmuştur.

Tablo 21

İG-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine İlişkin Betimsel İstatistikler

İG-TSÖ	Madde Sayısı	MGP Ortalama	Std. Sapma	Minimum	Maksimum
%30	12	54.88	4.45	42.56	66.96
%40	16	56.56	3.18	47.77	65.63
%50	20	56.90	2.83	48.21	65.54
%70	28	56.58	1.94	49.74	62.12
Tüm Test	40	56.25			

Tablo 21'de görüldüğü gibi, İG-TSÖ yöntemi dikkate alınarak toplam madde sayısının %30'u (12 madde) kadar maddenin 1000 defa seçilmesiyle oluşan alt testlerin MGP ortalaması 54.88, standart sapması 4.45 olarak hesaplanmıştır. Bu testlerin MGP'leri 42.56 ile 66.96 puan aralığında değerler almıştır. Toplam madde sayısının %40'ının (16 madde) seçilmesiyle oluşan alt testlerin MGP'leri (47.77-65.63) aralığında değişmişken, ortalaması 56.56 standart sapması ise 3.18 olarak bulunmuştur. İG-TSÖ yöntemine göre tüm testin %50'si (20 madde) seçildiğinde ise, oluşan alt testlerin ortalaması 56.90, standart sapması 2.83 olarak hesaplanmıştır. Bu alt testlerin MGP'leri ise 48.21 ile 65.54 puan aralığında değerler almıştır. Son olarak, tüm testin %70'i (28 madde) seçildiğinde ise, oluşan alt testlerin ortalaması 56.58, standart sapması ise 1.94 değerlerini almıştır. Bu 28 maddeden oluşan alt testlerin MGP'leri ise 49.74 ile 62.12 puan aralığında değişmiştir. Yukarıdaki tablodaki standart sapma ve minimum-maksimum değerlerinden anlaşılabacağı üzere,

madde sayısı arttıkça alt testlerin MGP'lerinin nihai MGP'ye yaklaştığı görülmektedir.

Tüm test ve alt testlerin geçme puanı ortalamaları bakımından manidar düzeyde farklılaşıp farklılaşmadığı tek yönlü ANOVA ile incelenmiştir. Varsayımlardan normallik sağlanmış ancak varyansların homojenliği sağlanmamıştır. Bu durumda Welch testi sonuçları yorumlanmış; analiz sonuçları, geçme puanları ortalamalarının testler arasında manidar düzeyde farklılaştığını ($F(4,1998)=46.575$, $p=.000$) ortaya koymuştur. Ayrıca, çoklu karşılaştırma (Post-Hoc) analizi yapılmış ve buna göre, %30'luk ($X_{ort}=54.88$) ve %50'lik ($X_{ort}=56.90$) alt testlerin geçme puanları ortalamaları, tüm testin geçme puanından ($X_{ort}=56.25$) manidar derecede farklılık göstermiştir ($p<.05$). Bundan dolayı, İG-TSÖ yöntemi kullanılarak üretilen %30 ve %50'lik alt testler ile tüm testin MGP'sine benzer geçme puanı elde edilebileceği söylenemez.

Daha detaylı inceleme yapmak amacıyla, her bir alt test için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri ve RMSE değerleri hesaplanmış ve bu değerler Tablo 22'de sunulmuştur.

Tablo 22

İG-TSÖ ile Oluşturulan Alt Testlerin Evet/Hayır MGP'lerine Ait RMSE ve Alt Test Yüzdesi Değerleri

İG-TSÖ	RMSE	Alt Test Yüzdesi	
		Ortalama $\pm$ 1TSH	Ortalama $\pm$ 2TSH
%30	4.65	%46.9	%79.7
%40	3.19	%66.3	%94.2
%50	2.90	%71.4	%97.1
%70	1.96	%89.5	%99.9
Tüm Test	TSH = 3.09 Ortalama $\pm$ 1TSH= (53.16- 59.34) Ortalama $\pm$ 2TSH = (50.07 – 62.43)		

Tablo 22'de görüldüğü gibi, tüm testin MGP'sinin 1 TSH ve 2 TSH değer aralığının içerisinde kalan yüzdeleri, beklenildiği gibi madde sayısı arttıkça artmıştır. Ancak, alt testteki madde sayısı 16'dan (%40'lık test) 20'ye (%50'lik test)

yükselmesine rağmen, tüm testin MGP'sinin 1 TSH içerisinde kalan yüzde değeri diğerlerine kıyasla daha az artmıştır. Bunun aksine, hiçbir alt testin MGP ortalaması, benzerlik kriterinin (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içeri) içerisinde yer almamıştır. Ancak, sadece %50 ve %70'lik alt testlere ilişkin MGP'ler tüm testin MGP'sinin 2 TSH içerisinde (en az %95 ihtimalle) yer almıştır. Diğer alt testlerde bu yüzde daha düşük çıkmıştır. RMSE değerlerine bakıldığında ise, beklenildiği gibi en az hata değeri %70'lik (1.96), en fazla hata ise %30'luk (4.65) alt testte hesaplanmıştır. Bu bilgilerden yola çıkarak, İG-TSÖ yöntemiyle oluşturulan bu 4 alt testin (%30,%40,%50 ve %70) herhangi biri ile nihai Evet/ Hayır geçme puanına benzer bir puan (en az %95 olasılıkla tüm testin MGP'sinin 1 TSH içeri) elde etmek pek olası olmayabilir. Fakat toplam madde sayısının en az %70'i ile tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde edilebilir.

Alt problem 4'e ilişkin bulgular ve yorumlar. *Alt problem 4. Farklı seçkisiz madde örnekleme yöntemleriyle oluşturulan alt testlere ilişkin ortalama Evet/Hayır geçme puanları birbirlerinden nasıl farklılaşmaktadır?*

Farklı madde örnekleme yöntemleriyle oluşturulan alt testlerin (%30,%40,%50,%70) hiçbirisiyle, tüm testin Evet/Hayır geçme puanına çok benzer bir puan elde edilememiştir. Ancak, toplam madde sayısının %50'i ile %70'i, tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde etmek için yeterli olmuştur. Bu nedenle, farklı seçkisiz örnekleme yöntemlerini birbirleriyle karşılaştırmak amacıyla, %50 ve %70'lik alt testler seçilmiştir. Seçilen alt testler için RMSE değerleri Tablo 23'te gösterilmiştir.

Tablo 23

Evet/Hayır MGP'lerin Örnekleme Durumlarına Ait RMSE Değerleri

Alt Test	RMSE			
	BSÖ	İ-TSÖ	G-TSÖ	İG-TSÖ
%50	3,52	2,75	3,30	2,90
%70	2,31	1,82	2,22	1,96

Tablo 23'te görüldüğü üzere, her bir seçkisiz örnekleme yönteminde madde sayısı arttıkça beklenildiği gibi RMSE hata değeri azalmıştır. Bunun yanı sıra, RMSE

hata değerleri arasındaki fark ölçülebilir olmamasına rağmen, bu değerlerin İ-TSÖ yöntemi için en küçük, BSÖ yöntemi için en büyük olduğu gözükmemektedir. Dolayısıyla, RMSE değerleri dikkate alındığında İ-TSÖ'nün daha etkili bir yöntem olduğu söylenebilir.

Farklı örnekleme durumları için belirlenen sınırlar (ortalama $\pm$ 1TSH; ortalama $\pm$ 2TSH) içerisinde kalan alt test yüzde değerleri Tablo 24'te verilmiştir.

Tablo 24

Evet/Hayır MGP'lerin Örnekleme Durumlarına Ait Alt Test Yüzde Değerleri

MGP Ort. $\pm$ 1TSH					MGP Ort. $\pm$ 2TSH			
Alt test	BSÖ	İ-TSÖ	G-TSÖ	İG-TSÖ	BSÖ	İ-TSÖ	G-TSÖ	İG-TSÖ
%50	%62.9	%72.9	%64.9	%71.4	%91.9	%97.4	%94.3	%97.1
%70	%82.9	%90.9	%84.2	%89.5	%99.3	%100	%99.8	%99.9

Tüm Test TSH = 3.09
Ortalama $\pm$ 1TSH= (53.16- 59.34)
Ortalama $\pm$ 2TSH = (50.07 – 62.43)

Tablo 24'te görüldüğü gibi, yukarıda belirtilen puan aralıklarının içerisinde kalan alt test yüzdeler oranları, BSÖ yöntemiyle en küçük, İ-TSÖ yöntemiyle ise en büyük olarak bulunmuştur. Bundan dolayı, İ-TSÖ yöntemiyle tüm testin MGP'sine daha yakın geçme puanı elde edilebileceği yorumu yapılabilir. Bunun yanı sıra, tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde etmek için İ-TSÖ ve İG-TSÖ yöntemlerine göre maddelerin en az %50'sinin, BSÖ ve G-TSÖ yöntemlerine göre ise maddelerin en az %70'inin seçilmesi gerekmektedir. Sonuç olarak, her ne kadar 16 durumun hiçbirisiyle (4 yöntem x 4 alt test) tüm testin Evet/Hayır geçme puanına çok benzer bir puan elde edilemese de, İ-TSÖ ve İG-TSÖ yöntemleriyle daha ümit verici sonuçlar elde edildiği açıktır.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Bu bölümde araştırmanın bulgularına dayalı olarak elde edilen sonuçlar özetlenmiş ve literatürdeki benzer çalışmalarla birlikte ele alınarak tartışılmıştır. Son olarak da, bu sonuçlara dayalı önerilere yer verilmiştir.

Sonuç ve Tartışma

Bu çalışmada, Angoff ve Evet/Hayır yöntemleriyle yapılan standart belirleme çalışmalarındaki madde sayısını azaltma noktasında seçkisiz örnekleme yöntemlerinin etkililiği incelenmiş ve farklı seçkisiz örnekleme yöntemleri (basit seçkisiz örnekleme [BSÖ], içerik tabakalı [İ-TSÖ], madde gücü tabakalı [G-TSÖ] ve içerik ve madde gücü tabakalı [İG-TSÖ]) birbirleriyle karşılaştırılmıştır. Çalışmanın sonuçlarının, tüm test geçme puanına benzer kestirimler (tüm test MGP'sinin 1 TSH içerisi) elde edebilmek için test maddelerinin yüzde kaçının yeterli olduğunun belirlenmesi noktasında yönlendirici olması amaçlanmıştır.

Çalışma kapsamında, tüm testten farklı büyüklükte alt testler (%30, %40, %50 ve %70) oluşturmak için dört farklı madde örnekleme yöntemi (BSÖ, İ-TSÖ, G-TSÖ ve İG-TSÖ) kullanılmış, toplam 16 çalışma durumu (4 yöntem x 4 alt test) değerlendirilmiştir. Bu işlemler hem Angoff hem Evet/hayır yöntemi için ayrı ayrı yapılmıştır. Her bir duruma ilişkin kestirilen ortalama Angoff MGP'lerinin tüm testin MGP'sinden (45.12) manidar düzeyde farklılaşıp farklılaşmadığının incelenmesi amacıyla ANOVA analizi yapılmıştır. Bu analizin sonuçlarına göre; İ-TSÖ yöntemiyle oluşturulan %30 ve %40'lık alt testlerin, G-TSÖ ve İG-TSÖ yöntemiyle ise oluşturulan %30'luk alt testlerin ortalama Angoff geçme puanlarının tüm testinkinden manidar düzeyde farklılaştığı görülmüştür. Bunun aksine, diğer 12 durumda ise manidar düzeyde farklılık bulunamamıştır. Benzer şekilde, Evet/Hayır MGP açısından tüm testten (56.25) manidar düzeyde farklılık gösteren alt testler yine aynı alt testler ve buna ek olarak İG-TSÖ yöntemiyle elde edilen %50'lik alt test olmuştur. Bu ANOVA analiz sonuçları Kannan, Sgammato, Tannenbaum ve Katz (2015)'in sonuçlarını nispeten desteklemektedir. Kannan vd. (2015), kestirilen ortalama Angoff MGP'lerinin farklı örnekleme yöntemleri ve farklı alt testler için çok da değişmediğini belirtmiştir. Araştırmasında, sadece İG-TSÖ yöntemiyle oluşturduğu

30 madde (yaklaşık %30) içeren alt testin ortalama Angoff MGP'si tüm testinkinden farklılık göstermiştir.

Yukarıdaki analizler dışında, daha ayrıntılı inceleme yapmak amacıyla her bir çalışma durumu için RMSE değerleri ve belirlenen sınırlar ($\text{ortalama} \pm 1\text{TSH}$; $\text{ortalama} \pm 2\text{TSH}$) içerisinde kalan yüzdelik oranları hesaplanmıştır. Bu işlemler hem Angoff hem de Evet/hayır yöntemi için ayrı ayrı yapılmıştır. Buna göre; RMSE hata değerleri arasındaki fark ölçülebilir olmamasına rağmen, en küçük hata İ-TSÖ yöntemi, en büyük hata ise BSÖ yöntemi kullanıldığında hesaplanmıştır. Dolayısıyla, belirlenen sınırlar içerisinde kalan yüzdelik oranları karşılaştırıldığında; en küçük yüzdelik oranları BSÖ yöntemiyle, en büyük yüzdelik oranları ise İ-TSÖ yöntemiyle elde edilmiştir. Ayrıca, daha da önemlisi tüm testin Angoff MGP'sine çok benzer kestirimler (en az %95 ihtimalle 1 TSH içerisi) elde etmek için İ-TSÖ yöntemiyle maddelerin %50'sinin seçilmesi yeterli olurken, diğer yöntemlerle maddelerin en az %70'inin seçilmesi gerekmiştir. Angoff yönteminin tersine, 16 çalışma durumunun hiçbirisiyle (4 yöntem x 4 alt test) tüm testin Evet/Hayır MGP'sine benzer bir kestirim elde edilememiştir. Ancak, tüm testin MGP'sinin 2 TSH içerisinde (en az %95 olasılıkla) yer alan bir geçme puanı elde etmek için İ-TSÖ ve İG-TSÖ yöntemleriyle maddelerin en az %50'sinin, BSÖ ve G-TSÖ yöntemleriyle ise maddelerin en az %70'inin seçilmesi yeterli olmuştur. Sonuç olarak, Angoff ve Evet/Hayır yöntemleriyle yapılan MGP belirleme çalışmalarının her ikisinde de madde sayısının azaltılması noktasında İ-TSÖ yöntemini kullanmak daha etkili olmuştur.

Bu çalışmadaki (Angoff) madde sayısının azaltılmasına ilişkin bulgular, önceki araştırma sonuçlarıyla (Ferdous & Plake, 2005, 2007; Kannan vd., 2015; Sireci vd., 2000; Smith, 2011) tutarlılık göstermiştir. Sireci vd. (2000), yaptıkları çalışma sonucunda oldukça benzer MGP kestirimleri elde etmek için test maddelerinin 2/3'sinin seçkisiz olarak seçilmesine ihtiyaç duyulacağını kanıtlamıştır. Ferdous ve Plake (2005, 2007) ve Smith (2011) ise yaptıkları çalışmalar neticesinde, tüm testin MGP'sine benzer kestirimler elde etmek için maddelerin yaklaşık %50'sinin yeterli olacağını savunmuşlardır. Kannan vd. (2015) ise, yaklaşık 100 madde içeren tüm teste genellenebilir MGP elde etmek için yaklaşık 45 maddenin yeterli olduğunu göstermiştir.

Madde örnekleme yöntemlerinin etkililiği bakımından çalışmanın sonuçları ele alındığında ise, elde edilen bulguların literatürdeki çalışmalarla uyum içerisinde olduğu görülmüştür. Benzer çalışmalarda da, tabakalı seçkisiz örnekleme yöntemlerinin benzer MGP kestirimleri verme noktasında basit seçkisiz örnekleme yöntemine göre daha etkili olduğu (Ferdous & Plake, 2005, 2007; Kannan vd., 2015, Smith, 2011) ve daha düşük hata değerine (RMSE) sahip olduğu (Kannan vd., 2015) görülmüştür. Ancak, Ferdous ve Plake (2005) ve Kannan vd. (2015), tabakalı yöntem kullanımının etkili olduğunu belirttikleri bu çalışmalarında, içerik-güçlük tabakalı yöntemin (İG-TSÖ) benzer MGP kestirimleri verme noktasında içerik tabakalı yöntem (İ-TSÖ) ve güçlük tabakalı yöntem (G-TSÖ) göre daha etkili olduğunu saptamışlardır. Bu sonuçlar, bu çalışmadan elde edilen içerik tabakalı yöntemin (İ-TSÖ) diğer yöntemlere göre daha etkili olduğu bulgusuyla çelişmektedir.

Yapılan tartışmaların ışığında, bu çalışmada elde edilen bulguların, bazı noktalarda literatürdeki çalışmalarla uyumlu sonuç verdiği, bazı noktalarda ise çelişkili bulgular elde edildiği görülmüştür. Çalışmada kullanılan madde sayısının 40 maddeyle sınırlı olması ve buna bağlı olarak oluşturulan tabakalardaki bazı hücrelerde madde sayısının çok az olması veya hiç olmaması sonuçları olumsuz etkilemiş olabilir. Örneğin, içerik ve güçlük tabakalı madde örnekleme yönteminde toplamda 15 tabaka (3 içerik x 5 güçlük) oluşturulmuş ve 40 maddenin bu 15 tabakaya göre sınıflandırılması yapılmıştır. Ancak madde sayısının az olması dolayısıyla boş kalan tabakalar olmuş, madde sayısı yetersiz kalmıştır. Buna ek olarak testteki maddelerin güçlük parametrelerinin çok düşük ve birbirine yakın olması, güçlük tabakasının etkisini azaltmış olabilir. Ayrıca, güçlük ve içerik sınıflandırmalarının farklı şekillerde yapılması da sonuçlar üzerinde etkili olabilir.

Öneriler

Araştırmanın sonuçlarına yönelik öneriler. Araştırmanın sonuçlarına yönelik öneriler aşağıda verilmiştir.

1. Araştırmada madde örnekleme yöntemleriyle oluşturulan alt testlerden elde edilen MGP'nin, tüm test MGP'ye yakın sonuçlar verdiği, özellikle tabakalı yöntemlerin kullanımıyla oldukça benzer sonuçlara ulaşıldığı görülmüştür. Tüm testin Angoff geçme puanına benzer kestirimler elde etmek için İ-TSÖ yöntemiyle maddelerin %50'sinin seçilmesi yeterli olabilir. Eğitimcilerin, bunu göz

önünde bulundurarak geçme puanı belirlerken madde azaltma yoluna gitmeleri önerilebilir. Böylelikle, hem zamandan hem paradan hem de iş yükünden tasarruf sağlanabilir.

2. Madde sayısını azaltarak standart belirleyecek kişi veya kurumlara, alt test oluştururken tüm testin hem içerik hem de güçlük açısından temsil edilmesi önerilir. Bu sayede, uzmanların tüm test hakkında kapsamlı bilgi edinmeleri sağlanabilir.

İleride yapılacak araştırmalara yönelik öneriler. İleride yapılacak araştırmalara ilişkin öneriler aşağıda sunulmuştur.

1. Madde sayısı artırılarak, çalışma tekrarlanabilir.
2. Yeterlilik testleri gibi daha çok sayıda içerik alanı içeren ve daha geniş madde güçlüğüne sahip testlerle yeni çalışmalar yapılabilir.
3. Bir uzman grubunun bu çalışmadaki tüm test maddeleri hakkında, diğer grubun ise tüm testin %50'si hakkında yargıda bulunduğu yeni bir uygulama yapılabilir. Dolayısıyla, bu standart belirleme çalışmasındaki madde sayısını azaltmanın etkisi görülmüş olur.
4. Çoklu matris örnekleme ve dengeli tamamlanmamış blok tasarımı gibi farklı örnekleme yöntemleri ve farklı sınıflandırma teknikleri (çeşitli sayıda güçlük/içerik tabakaları) kullanılarak tekrarlı örneklem çalışmaları yapılabilir.
5. İçeriğe, madde güçlüğüne ve madde ayırt ediciliğine göre tabakalandırmanın etkisi, halihazır eğitim uygulamalarında incelenebilir.
6. Bu çalışmada Angoff ve Evet/Hayır standart belirleme yöntemleri kullanılmıştır. İşaretleme yöntemi kullanılarak bu sonuçların genellenebilirliği incelenebilir.

Kaynaklar

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. R. L. Thorndike (Ed.), *Educational measurement içinde* (ss. 508-600). Washington, DC: American Council on Education.
- Behuniak, P., Gable, R. K., & Archambault, F. X. (1982). The validity of categorized proficiency test scores. *Educational and Psychological Measurement*, 42, 247-252.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56(1), 137-172.
- Berk, R. A. (1996). Standard setting: The next generation (where few psychometricians have gone before!). *Applied Measurement in Education*, 9, 215-235.
- Buckendahl, C. W., Ferdous, A. A., & Gerrow, J. (2010). Recommending cut scores with a subset of items: An empirical illustration. *Practical Assessment*, 15(6), 1-10.
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Cizek, G. J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement*, 30(2), 93-106.
- Cizek, G. J. (2001). *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J., & Bunch, M. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- Coraggio, J. T. (2008). *A Monte Carlo approach for exploring the generalizability of performance standards*. University of South Florida.

- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley & Sons.
- Cronbach, L. J. (1988). Five perspectives on validity argument. *Test validity*, 3-17.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265-286.
- Çetin, S. (2011). *İşaretleme ve angoff standart belirleme yöntemlerinin karşılaştırılması* (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Downing, S. M. (2006). Selected-Response item formats in test development. T. M. Haladyna , & S. M. Downing (Ed.), *Handbook of test development* içinde (ss. 287-300). Mahwah, New Jersey: Routledge.
- Ferdous, A. A. , & Plake, B. S. (2005). The use of subsets of test questions in an Angoff standard setting method. *Educational and Psychological Measurement*, 65(2), 185-201.
- Ferdous, A. (2007). The use of subsets of test items in a bookmark standard setting method. National Council on Measurement in Education toplantısında sunulan bildiri, Chicago, IL.
- Ferdous, A. A., & Plake, B. S. (2007). Item selection strategy for reducing the number of items rated in an Angoff standard setting study. *Educational and Psychological Measurement*, 67(2), 193-206.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8. bs.). New York: McGraw Hill.
- Freeman, L., & Miller, A. (2001). Norm-referenced, criterion-referenced and DA: What exactly is the point? *Educational Psychology in Practice*, 17(1), 13–16.
- Hambleton, R. M. (1998). Setting performance standards on achievement tests: Meeting the requirements of Title I. L. N. Hansche (Ed.), *Handbook for the development of performance standards* içinde (ss. 87-114). Washington, DC: Council of Chief State School Officers.

- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* içinde (ss. 89-116). Mahwah, NJ:Lawrence Erlbaum Associates.
- Hambleton, R. K., & Pitoniak, M. (2006). Setting performance standards. R. L. Brennan (Ed.), *Educational Measurement* içinde (ss. 433–470). Westport, CT: Praeger.
- Harvey, A. L., & Way, W. D. (1999). A comparison of web-based standard setting and monitored standard setting. National Council on Measurement in Education toplantısında sunulan bildiri, Montreal, Canada.
- Impara, J. C. (1995). *Licensure testing: Purposes, procedures, and practices*. Buros Institute of Mental Measurements, University of Nebraska-Lincoln.
- Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34, 353-366.
- Irwin, P. (2007). *An alternative examinee-centered standard setting strategy* (DoktoraTezi). University of Nebraska, USA.
- Jaeger, R. M. (1989). Certification of student competence. R. L. Linn(Ed.), *Educational measurement* içinde (ss. 485-514). New York: American Council on Education/Macmillan.
- Jaeger, Richard M. (1991). Selection of judges for standard-setting. *Educational Measurement: Issues and Practice*, 10(2), 3–14. doi:10.1111/j.1745-3992.1991.tb00185.x
- Kane, M.T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425-461.
- Kane, M. T. (1998). Choosing between examinee-centered and test-centered standard-setting methods. *Educational Assessment*, 5(3), 129-145.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.

- Kannan, P., Sgammato, A., & Tannenbaum, R. J. (2015). Evaluating the operational feasibility of using subsets of items to recommend minimal competency cut scores. *Applied Measurement in Education*, 28(4), 292-307.
- Kannan, P., Sgammato, A., Tannenbaum, R. J., & Katz, I. R. (2015). Evaluating the consistency of angoff-based cut scores using subsets of items within a generalizability theory framework. *Applied Measurement in Education*, 28(3), 169-186.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). Standard setting: A bookmark approach. D. R. Green (Ed.), *IRT-based standard setting procedures utilizing behavioral anchoring* içinde. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Linn, R. L. (1994). The likely impact of performance standards as a function of uses: From rhetoric to sanctions. Joint Conference on Standard Setting for Large-Scale Assessments toplantısında sunulan bildiri, Washington, DC.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, N.J.: Educational Testing Service.
- McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, 49, 34-42.
- Mehrens, W.A., & Lehmann, I.J. (1991). *Measurement and evaluation in education and psychology* (3. bs.). New York: Holt, Rinehart and Winston.
- Mehrens, W. A., & Popham, W. J. (1992). How to evaluate the legal defensibility of high-stakes tests. *Applied Measurement in Education*, 5(3), 265-283.
- Mehrens, W. A. (1995). Methodological issues in standard setting for educational exams. *Proceedings of Joint Conference on Standard Setting for Large-Scale Assessments* içinde (ss. 221-263).
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741.

- Mills, C. N., Melican, G. J., & Ahluwalia, N. T. (1991). Defining minimal competence. *Educational Measurement: Issues and Practice*, 10(2), 7-10.
- National Assessment Governing Board. (1994). *Setting achievement levels on the 1994 National Assessment of Educational Progress in Geography and in US History and the 1996 National Assessment of Education Progress in Science*. Final Version. Washington, DC: Author
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.
- Norcini J., & Guille R. (2002). Combining tests and setting standarts. G.R. Norman, C. P. M. Van der Vleuten, & D.I. Newble(Ed.), *International Handbook of Research in Medical Education* içinde (ss.811–834). Dordrecht: Kluwer Academic Publishers.
- Norcini, J. J., Lipner, R. S., Langdon, L. O., & Strecker, C. A. (1987). A comparison of three variations on a standard-setting method. *Journal of Educational Measurement*, 24(1), 56-64.
- Norcini, J., & Shea, J. (1992). The reproducibility of standards over groups and occasions. *Applied Measurement in Education*, 5(1), 63-72.
- Norcini, J., Shea, J., & Ping, J. C. (1988). A note on the application of multiple matrix sampling to standard setting. *Journal of Educational Measurement*, 25(2), 159–164.
- Özçelik, D. A. (2013). *Test hazırlama kılavuzu*. Pegem Akademi Yayıncılık.
- Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (2. bs.). Crows Nest, Australia: Allen & Unwin.
- Pant, H. A., Rupp, A. A., Tiffin-Richards, S. P., & Köller, O. (2009). Validity issues in standard-setting studies. *Studies in Educational Evaluation*, 35(2-3), 95-101.
- Plake, B. S. (1998). Setting performance standards for professional licensure and certification. *Applied Measurement in Education*, 11(1), 65-80.
- Plake, B. S., & Impara, J. C. (2001). *The fourteenth mental measurements yearbook*. Lincoln, NB: Buros Institute of Mental Measurements.

- Reckase, M. D. (2000). *The evolution of the NAEP achievement levels setting process: A summary of the research and development efforts conducted by ACT*. National Assessment Governing Board.
- Reckase, M. D. (2001). Innovative methods for helping standard-setting participants to perform their task. The role of feedback regarding consistency, accuracy, and impact. G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* içinde (ss.159-174). Mahwah, NJ: Erlbaum.
- Schultz, E. M., & Mitzel, H. (2005). The Mapmark standard setting method. National Council on Measurement in Education toplantısında sunulan bildiri, Montreal, Canada.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481.
- Sireci, S. G., Patelis, T., Rizavi, S., Dillingham, A. M., & Rodriguez, G. (2000). Setting standards on a computerized-adaptive placement examination. *Laboratory or Psychometric and Evaluative Research Report No. 378*.
- Smith, R. W., & Ferdous, A. (2007). Using a subset of items for a modified Angoff standard setting study—An automated item selection procedure. National Council on Measurement in Education toplantısında sunulan bildiri, Chicago, IL.
- Smith, T. N. (2011). *Using stratified item selection to reduce the number of items rated in standard setting* (Yüksek Lisans Tezi). University of South Florida, USA.
- Turgut, M.F., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme metotları* (4. bs.). Ankara: Pegem Akademi.
- Van der Linden, W. J. (1982). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting. *Journal of Educational Measurement*, 19(4), 295-308.
- Wang, N., Wiser, R., & Newman, L. (2001). An evaluation of standard setting using a Rasch-based IRT model. National Council on Measurement in Education toplantısında sunulan bildiri, Seattle, WA.

Zieky, M.J., Perie, M., & Livingston, S. A. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.



EK-A: Uzman Görüşleri için Gönüllü Katılım Formu

Değerli Meslektaşım,

Çalışmamı gösterdiğiniz ilgi ve ayırdığınız zaman için çok teşekkür ederim. Bu çalışma, bir standart belirleme çalışmasındaki madde sayısını azaltmak için kullanılan farklı seçkisiz örnekleme yöntemlerinin etkililiğinin incelenmesinin amaçlandığı bir yüksek lisans tez çalışmasıdır. Çalışma Doktor Öğretim Üyesi Sevdâ ÇETİN danışmanlığında Yüksek Lisans Öğrencisi Hakan KARA tarafından yürütülmektedir. Araştırma kapsamında, 2012 ve 2013 yılı SBS Matematik testi soruları için değerlendirme yapmanıza ihtiyaç duyulmaktadır. Yapacağınız değerlendirmeler geçme puanı oluşturulması amacıyla kullanılacaktır. Araştırma kapsamında, siz değerli uzmanların sağlayacağı bilgiler hiçbir şekilde başka amaçlar için kullanılmayacaktır. Bu sebeple değerlendirmelerinizi önemle yapmanızı rica ederim. Amacı yukarıda açıklanmış olan bu araştırma için Hacettepe Üniversitesi Etik Komisyonu'ndan gerekli izinler alınmıştır.

- Çalışmaya katılımınız gönüllülük esasına dayanmaktadır.
- Araştırmada katılımcıların isimleri gizli tutulacaktır. Eğer araştırmada isminizin kullanımı gerekirse bunun yerine bir takma ad kullanılacaktır.
- Toplanan tüm veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmanın amacı dışında ya da bir başka araştırmada kullanılmayacaktır.
- Dilediğiniz takdirde sizden toplanan verileri incelemeyi talep edebilirsiniz.
- Katılımınız sırasında herhangi bir rahatsızlık hissettiğiniz takdirde çalışmadan istediğiniz zaman ayrılma hakkına sahipsiniz. Çalışmadan ayrılmanız size hiçbir sorumluluk getirmeyecektir.
- Uygulama esnasında herhangi bir rahatsızlık yaşanması halinde size gerekli destek sağlanacaktır.
- Çalışmadan ayrılmanız durumunda sizden toplanan tüm veriler imha edilecek ve çalışmada kullanılmayacaktır.
- Uygulama esnasında aklınıza gelen tüm sorular ve uygulama sonrası sonuçlarla alakalı her zaman bilgi alabilirsiniz.
- Sizden toplanan veriler güvenli bir şekilde saklanacak ve çalışma bitiminde imha edilecektir.

Gönüllü katılım formundaki tüm açıklamaları okumak üzere ayırdığınız vakit için teşekkür ederim. Uygulama sürecinde veya sonrasında aklınıza takılan sorularınızı ve iletmek istediğiniz yorumlarınızı aşağıdaki iletişim bilgilerinden bana iletebilirsiniz.

Sorumlu Arařtarmacı :

Doktor Öğretim Üyesi Sevda ÇETİN

Adres: HÜ Eğitim Bilimleri Enstitüsü

e-posta: tsevda@hacettepe.edu.tr

Tel: 0312 780 5110

İmza:

Arařtarmacı:

Hakan KARA

Adres :Ressam İbrahim Çallı Ortaokulu

e-posta: hakankaraodtu@gmail.com

Tel: 537 894 3058

İmza:

Bu çalışmaya tamamen kendi rızamla, istediğim takdirde çalışmadan ayrılabilceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.

(Lütfen bu formu doldurup imzaladıktan sonra veri toplayan kişiye veriniz.)

Katılımcı Öğretmen

Ad ve Soyadı:

Adres:

E-posta:

Tel:

İmza:

Tarih:

Minimum Yeterlik Düzeyi

Uluslararası Öğrenci Değerlendirme Programı (PISA) matematik yeterlik düzeylerinde yer alan düzey 2, minimum matematik yeterlik düzeyi olarak düşünülmektedir ve öğrencilerin matematiği etkili olarak kullanmalarına olanak tanıyan becerilerini göstermeye başladıkları düzeydir. Bu düzeydeki öğrenciler, doğrudan çıkarım yapmaktan başka bir beceriye gerek olmayan bir bağlamda ifade edilmiş olan durumları tanıyabilir ve yorumlayabilirler. Bu öğrenciler, tek bir kaynaktan gerekli bilgiyi elde edebilir ve sadece bir gösterim biçimini kullanabilirler. Temel algoritmaları, formülleri, alışlageldik işlem yollarını kullanabilirler. Doğrudan ispat gibi basit akıl yürütmeleri yapabilirler ve sonuçlar üzerinde görülenin ötesine geçmeyen yorumlar yapabilirler. (PISA, 2007).

1. Değerlendirme: “Matematik dersinde minimum yeterlik düzeyinde olan öğrencilerin yüzde kaç soruyu doğru cevaplayabilir?” sorusunu cevaplamanız beklenmektedir.

Örnek: %70 gibi...

2. Değerlendirme: “Matematik dersinde minimum yeterlik düzeyinde olan bir öğrenci soruyu doğru cevaplayabilir mi?” Cevaplarınızı kutucuklara çarpı işareti koyarak belirtiniz.

Örnek: Doğru cevaplayabilir

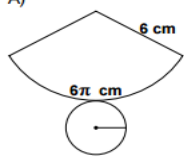
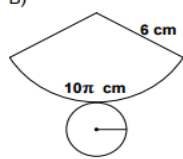
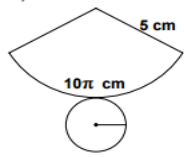
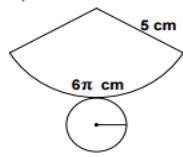
(x)

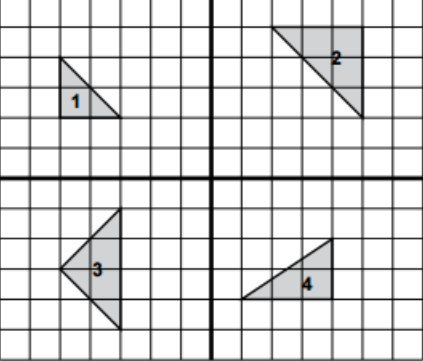
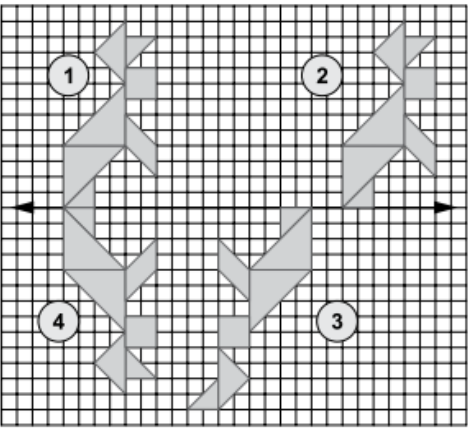
Doğru cevaplayamaz

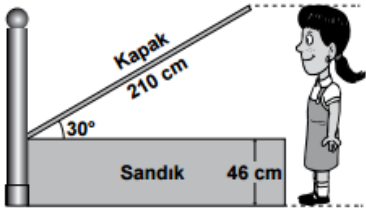
(x)

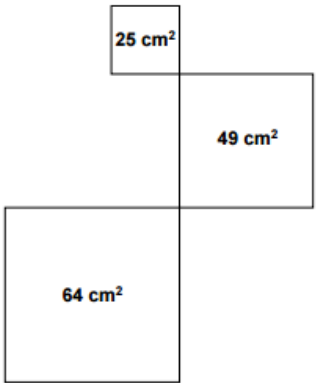
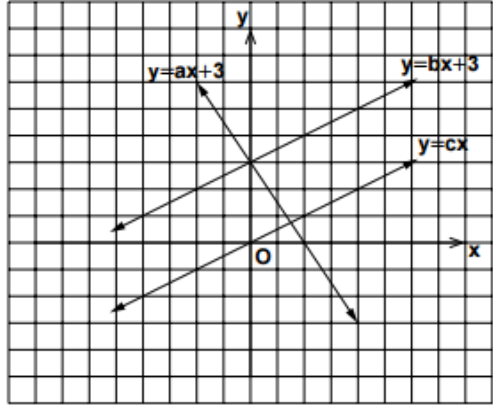
Not 1: Değerlendirmelerinize başlamadan önce lütfen test sorularını dikkatle inceleyiniz.

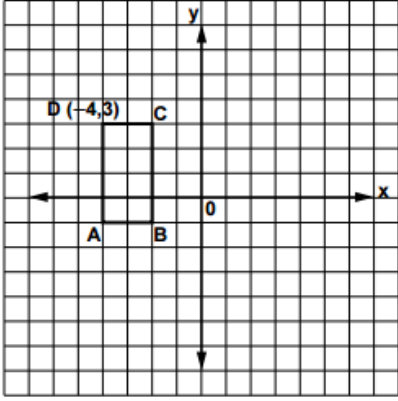
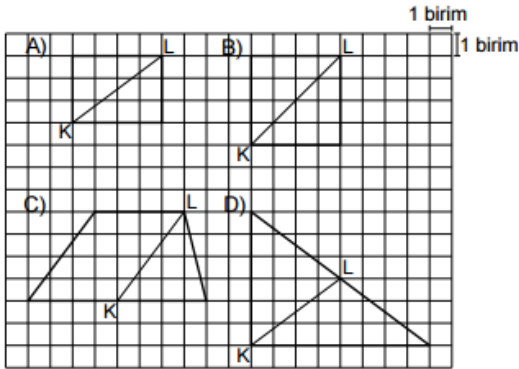
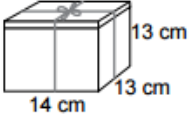
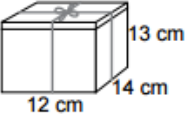
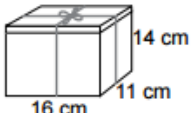
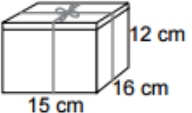
Not 2: 1. değerlendirme bitmeden lütfen 2. değerlendirmeye geçmeyiniz.

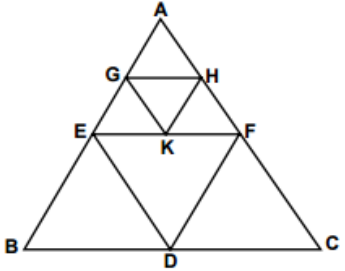
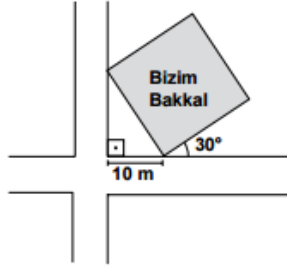
2012 SBS Matematik Testi Soruları	1. Değerlendirme (<u>Minimum yeterlik düzeyinde olan öğrencilerin</u> yüzde kaçını doğru cevaplayabilir?)
1. Türkiye genelinde bir yılda $8,1 \times 10^5$ adet çam ağacının kurtarılması hedeflenmektedir. Yukarıdaki afişe göre, bu hedefe ulaşmak için bir yılda kaç ton kullanılmış kâğıt geri dönüşüme kazandırılmalıdır? A) $4,5 \times 10$ B) $4,5 \times 10^2$ C) $4,5 \times 10^3$ D) $4,5 \times 10^4$	2. $2\sqrt{10}$, $\sqrt{17}$, $3\sqrt{3}$ sayılarının küçükten büyüğe doğru sıralanışı, aşağıdakilerden hangisidir? A) $\sqrt{17} < 3\sqrt{3} < 2\sqrt{10}$ B) $3\sqrt{3} < \sqrt{17} < 2\sqrt{10}$ C) $\sqrt{17} < 2\sqrt{10} < 3\sqrt{3}$ D) $3\sqrt{3} < 2\sqrt{10} < \sqrt{17}$
%	%
3. Şekildeki beş katlı düğün pastasında her kattaki pastanın yarıçap uzunluğu, bir üstündekinin yarıçap uzunluğunun 2 katıdır. En üstteki pastanın yarıçap uzunluğu 2^3 cm olduğuna göre, <u>en alttaki</u> pastanın yarıçap uzunluğu kaç santimetredir? A) 2^{10} B) 2^8 C) 2^7 D) 2^5	4. Şekildeki O noktası, verilen dik dairesel koninin taban merkezidir. Şekil üzerindeki verilere göre bu koninin açını aşağıdakilerden hangisi <u>olabilir</u>? A)  B)  C)  D) 
%	%

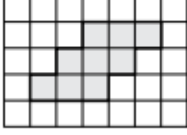
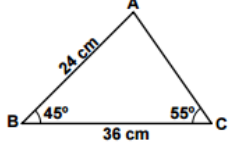
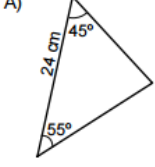
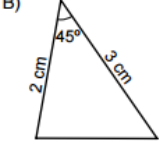
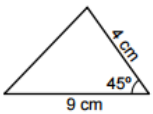
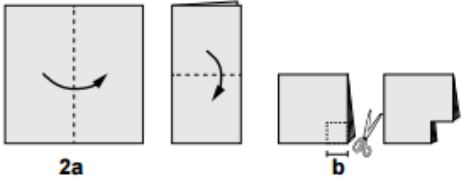
5. Bir mimar bir kenarı $3a$ metre olan kare şeklindeki arsada, her birinin kenar uzunluğu $2b$ metre olan belli sayıda karesel bölge belirlemiştir. Arsada geri kalan bölgenin alanı $(3a - 4b)(3a + 4b)$ metrekare olduğuna göre, belirlenen karesel bölgelerin sayısı kaçtır? A) 8 B) 4 C) 3 D) 2	6. Ahmet, kenarlarından ikisinin uzunluğu 17 m ve 10 m olan üçgen biçimindeki bahçesinin çevresini çit ile çevirmiştir. Ahmet'in kullandığı çitin uzunluğu, metre olarak aşağıdakilerden hangisi olabilir? A) 30 B) 34 C) 40 D) 54
%	%
7. Aşağıdaki şekilde dört farklı bahçede yer alan havuzların kuşbakışı çizimleri verilmiştir.  Bu havuzlardan hangisi diğer üç havuzla benzer değildir? A) 1 B) 2 C) 3 D) 4	8. Seda'nın matematik dersi dönem sonu puanı, ilk sınavda aldığı puanın iki katından 4 eksiktir. Dönem sonu puanı 50 ve üzeri olan öğrenci, o dersten başarılı olmaktadır. Matematikten dönem sonunda başarılı olan Seda'nın ilk sınavdaki puanı <u>en az</u> kaçtır? A) 28 B) 27 C) 26 D) 25
%	%
9. Aşağıdaki numaralandırılmış tangram şekillerinden hangi ikisi, birbirinin ötelemeli yansımasıdır?  A) 2 ve 4 B) 1 ve 4 C) 1 ve 2 D) 2 ve 3	10. Aşağıdakilerden hangisi, $0,4\overline{5}$ ile $0,4\overline{52}$ devirli ondalık kesirleri arasında yer alır? A) 0,451 B) 0,453 C) 0,455 D) 0,457
%	%

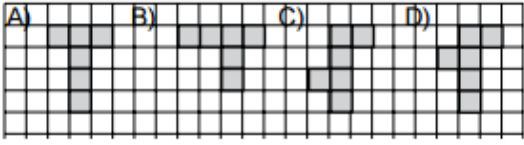
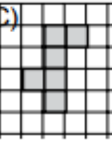
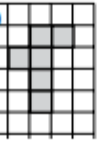
11. Ceren, açılıp kapanabilir bir sandıklı yatağın 210 cm uzunluğundaki kapağını aşağıdaki gibi açtığına, kapağın yerden yüksekliği Ceren'in boyuna eşit olmaktadır.  Sandığın yüksekliği 46 cm olduğuna göre, Ceren'in boyu kaç santimetredir? ($\sin 30^\circ = 0,5$) A) 156 B) 151 C) 128 D) 105	12. Aşağıdakilerden hangisi yapıldığında, taban ayrıntının uzunluğu 8 cm, yüksekliği 12 cm olan kare dik prizma şeklindeki kutunun hacmi yarıya iner? A) Bütün ayrıntı uzunlukları ikişer santimetre azaltıldığında B) Taban ayrıntı uzunlukları dörder santimetre azaltıldığında C) Bütün ayrıntı uzunlukları yarıya indirildiğinde D) Yüksekliği 6 cm azaltıldığında																						
13. Eser, kare dik prizma biçiminde ve ayrıtlarından birinin uzunluğu 30 cm olan yandaki teneke kutunun yan yüzlerini boyadıktan sonra bir yüzüne "çöp" yazısını yazıyor.  Boyanan yüzeyin alanı 1680 cm^2 olduğuna göre, bu çöp kutusunun diğer farklı ayrıntının uzunluğu kaç santimetredir? A) 14 B) 21 C) 25 D) 28	14. Biri silindir, diğeri küp şeklinde olan iki kap, tamamen su ile doludur. Silindir şeklindeki kabın taban çapının ve yüksekliğinin uzunlukları, küpün bir ayrıntı uzunluğuna eşittir. Buna göre, silindir şeklindeki kapta, küpte bulunan suyun kaçta kaç kadar su bulunmaktadır? (π yerine 3 alınız.) A) $\frac{3}{4}$ B) $\frac{2}{3}$ C) $\frac{1}{2}$ D) $\frac{1}{6}$																						
15. Bir torbadaki özdeş topların 11'i kırmızı, 8'i beyaz, 9'u mavi ve 12'si siyahtır. En az kaç top çıkarılırsa, torbada kalan topların renklerine göre çekilme olasılıkları eşit olur? A) 7 B) 8 C) 9 D) 10	16. Bir sınıftaki öğrencilerin günlük kitap okuma süreleri aşağıdaki tabloda verilmiştir. Tablo: Öğrencilerin Kitap Okuma Süreleri <table border="1" data-bbox="954 1370 1404 1765"> <thead> <tr> <th>Kitap okuma süresi (dakika)</th> <th>Öğrenci sayısı</th> </tr> </thead> <tbody> <tr><td>20 - 24</td><td>2</td></tr> <tr><td>25 - 29</td><td>6</td></tr> <tr><td>30 - 34</td><td>5</td></tr> <tr><td>35 - 39</td><td>3</td></tr> <tr><td>40 - 44</td><td>8</td></tr> <tr><td>45 - 49</td><td>1</td></tr> <tr><td>50 - 54</td><td>3</td></tr> <tr><td>55 - 59</td><td>12</td></tr> <tr><td>60 - 64</td><td>6</td></tr> <tr><td>65 - 69</td><td>2</td></tr> </tbody> </table> Tabloya göre kaç kişi günde 1 saatten daha az süre kitap okumaktadır? A) 8 B) 12 C) 40 D) 46	Kitap okuma süresi (dakika)	Öğrenci sayısı	20 - 24	2	25 - 29	6	30 - 34	5	35 - 39	3	40 - 44	8	45 - 49	1	50 - 54	3	55 - 59	12	60 - 64	6	65 - 69	2
Kitap okuma süresi (dakika)	Öğrenci sayısı																						
20 - 24	2																						
25 - 29	6																						
30 - 34	5																						
35 - 39	3																						
40 - 44	8																						
45 - 49	1																						
50 - 54	3																						
55 - 59	12																						
60 - 64	6																						
65 - 69	2																						
%	%																						

17. Üç yol ağzında bulunan bir tavşan, bu yollardan rastgele birine yöneliyor. Bu üç yolun her biri, devamında iki dar yola ayrılıyor. Tavşan yol ayrımına geldiğinde iki dar yoldan birini rastgele seçerek yoluna devam ediyor. Tavşanın, dar yollardan birinde bekleyen kaplumbağa ile karşılaşma olasılığı nedir? A) $\frac{1}{2}$ B) $\frac{1}{3}$ C) $\frac{1}{6}$ D) $\frac{1}{9}$	18. Bir tel, şekildeki gibi kıvrılarak üç tane kare oluşturuluyor.  Karelerin sınırladıkları bölgelerin alanları 25 cm^2, 49 cm^2 ve 64 cm^2 olduğuna göre, bu tel ile oluşturulabilecek <u>en büyük</u> karenin sınırladığı bölgenin alanı kaç santimetrekare olur? A) 100 B) 138 C) 225 D) 400
%	%
19. Bir yarışma programında, verilen her doğru cevaba + 3 puan, her yanlış cevaba – 2 puan verilmektedir. Bu yarışmaya katılan Aysun, sorulan 5 sorunun tümünü cevaplamıştır. Yarışma sonunda 10 puan aldığına göre, Aysun kaç soruyu doğru cevaplamıştır? A) 2 B) 3 C) 4 D) 5	20.  Yukarıda denklemleri ile verilen doğruların grafiklerine göre, aşağıdakilerden hangisi doğrudur? A) $a < b$ ve $b = c$ B) $b = c$ ve $c < a$ C) $c < b < a$ D) $a < b < c$
%	%
2013 SBS Matematik Testi Soruları	1. Değerlendirme (<u>Minimum yeterlik düzeyinde olan öğrencilerin</u> yüzde kaçını doğru cevaplayabilir?)

1. Aşağıdakilerden hangisi çevre uzunluğu 12 cm olan bir üçgenin kenar uzunluklarından biri <u>olamaz</u>? A) 1 cm B) 3 cm C) 5 cm D) 6 cm	2. $\frac{8^4 \cdot 4^2}{2^8}$ işleminin sonucu aşağıdakilerden hangisine eşittir? A) 2^{12} B) 2^8 C) 2^6 D) 2^4
3.  D noktasının koordinatları $(-4, 3)$ olan şekildeki ABCD dikdörtgeni, verilen düzlemde öteleniyor. x ve y eksenleri, elde edilen dikdörtgenin simetri doğruları olduğuna göre, aşağıdakilerden hangisi ötelenmiş dikdörtgenin B köşesinin koordinatları olur? A) $(-1, -2)$ B) $(1, -2)$ C) $(2, -2)$ D) $(-2, -2)$	4. 20 kişiden oluşan bir kafilenin yaş ortalaması 36'dır. Kafileye 5 kişi daha katıldığında, kafilenin yaş ortalaması 37 olmaktadır. Yeni katılanlardan 4'ü kırk yaşında olduğuna göre, 5. kişi kaç yaşındadır? A) 37 B) 40 C) 42 D) 45
5. Aşağıdaki şekillerin hangisinde KL doğru parçasının uzunluğu, bir irrasyonel sayı ile ifade edilir? 	6. Selim, bir yüzünün alanı 150 cm^2 olan küp şeklindeki bir oyuncak hediye kutusuna koymuştur. Bu oyuncak, aşağıda ayrıt uzunlukları verilen dik prizma şeklindeki hediye kutularından hangisinin içinde <u>olabilir</u>? A)  13 cm 14 cm 13 cm B)  13 cm 12 cm 14 cm C)  14 cm 16 cm 11 cm D)  12 cm 15 cm 16 cm
%	%

7. Bir öğretmen deney yapmak üzere fasulye tanelerini, öğrencileri sıralayarak ilk öğrenciye 3, sonraki her öğrenciye bir öncekinden 2 fazla vererek dağıtıyor. n. sıradaki öğrencinin kaç fasulye aldığı, aşağıdaki hangi ifade ile bulunur? A) $2n + 1$ B) $2n + 3$ C) $3n + 2$ D) $n + 2$	8. Şekildeki ABC eşkenar üçgeninin kenarlarının orta noktaları E, D, F ve AEF üçgeninin kenarlarının orta noktaları G, K, H'dir.  Buna göre, ABC üçgeni ile KHG üçgeni arasındaki benzerlik oranı nedir? A) $\frac{1}{16}$ B) $\frac{1}{12}$ C) $\frac{1}{4}$ D) $\frac{1}{2}$
9. Denklemi $y = 10$ olan doğrunun y eksenini kestiği noktadan, eğimi 2 olan bir başka doğru geçmektedir. Bu doğrunun x eksenini kestiği noktanın apsisi kaçtır? A) -5 B) $-\frac{1}{5}$ C) $\frac{2}{5}$ D) 20	10.  Şekildeki planda, zemini kare biçiminde olan bakkal dükkânı kaç metrekarelik alana yapılmıştır? ($\sin 30^\circ = \frac{1}{2}$) A) 25 B) 75 C) 300 D) 400
11. t aşağıdakilerden hangisi olursa $\frac{a^2 - a - 2}{a^2 - t}$ ifadesi sadeleşebilir? A) -4 B) -2 C) 2 D) 4	12. Tabanının bir ayrıtının uzunluğu 6 m ve yüksekliği 10 m olan eşkenar üçgen dik prizma şeklindeki saat kulesinin her bir yan yüzüne, çapı 4 m olan daire şeklinde birer saat yerleştirilmiştir. Bu kulenin yanal yüzeyinin, saatler dışında kalan bölgesinin alanı kaç metrekaredir? (π yerine 3 alınız.) A) 32 B) 48 C) 144 D) 168

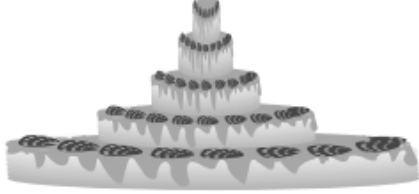
13.  Yukarıdaki boyalı şeklin alanı 243 cm^2 olduğuna göre, çevresinin uzunluğu kaç santimetredir? A) $36\sqrt{3}$ B) $48\sqrt{3}$ C) $64\sqrt{3}$ D) $108\sqrt{3}$	14. Bir torbada aynı özelliklere sahip 10 kalem-den 4'ü sarı, 2'si mavi ve diğerleri beyaz renktedir. Torbadan rastgele 1 kalem çekildikten sonra, torbaya atılmadan ikinci bir kalem daha çekilmektedir. Her iki kalemin de beyaz olma olasılığı nedir? A) $\frac{8}{45}$ B) $\frac{4}{25}$ C) $\frac{2}{15}$ D) $\frac{3}{25}$
%	%
15. Aşağıda, AEN çeşitkenar üçgeni üzerinde yapılan bir çizim anlatılmaktadır: 1. Pergel, EN doğru parçasının uzunluğu kadar açılır. 2. E ve N merkezli çemberler çizilir. 3. Çemberlerin kesiştiği noktalardan geçen bir doğru parçası çizilir. Yukarıdaki işlemler sonucunda, AEN üçgeninde hangi çizim yapılmış olur? A) Kenar orta dikme B) Kenarortay C) Açıortay D) Yükseklik	16.  Aşağıdakilerden hangisi, verilen ABC üçgenine benzer <u>değildir</u>? A)  B)  C)  D) 
%	%
17. Bir kenarının uzunluğu $2a$ santimetre olan kare şeklindeki bir kâğıt, aşağıda görüldüğü gibi üst üste iki kez katlanarak yeni bir kare elde ediliyor.  Elde edilen kareden, bir kenarının uzunluğu b santimetre olan kare kesilerek atılıyor. Kalan kâğıt tamamen açıldığında alanı kaç santimetrekare olur? A) $a^2 - b^2$ B) $2a^2 - b^2$ C) $(2a - b)^2$ D) $4a^2 - 4b^2$	18. Yüksekliği ve tabanının çap uzunluğu a metre olan bir silindirin hacminin, çap uzunluğu a metre olan bir kürenin hacmine oranı nedir? A) $\frac{3}{2}$ B) $\frac{\pi}{2}$ C) $\frac{3a}{2}$ D) $\frac{\pi a}{2}$
%	%

19. Aşağıdakilerden hangisi küpün bir açınımi <u>değildir</u>? (A)  (B)  (C)  (D) 	20. Bir baloncu, tanesini 10 liraya aldığı balonları 12 liraya satmaktadır. Balonların 3 tanesi satılmadan önce patlıyor. Baloncu 400 liradan fazla kâr elde ettiğine göre <u>en az</u> kaç balon satmıştır? A) 198 B) 201 C) 216 D) 231
%	%

➤ 2. değerlendirmeye geçebilirsiniz. (Cevaplarınızı “X” işareti koyarak belirtebilirsiniz.)

2012 SBS Matematik Testi Soruları	2. Değerlendirme(minimum yeterli düzeyinde olan bir öğrenci soruyu doğru cevaplayabilir mi?)
1. Türkiye genelinde bir yılda $8,1 \times 10^5$ adet çam ağacının kurtarılması hedeflenmektedir.  Yukarıdaki afişe göre, bu hedefe ulaşmak için bir yılda kaç ton kullanılmış kâğıt geri dönüşüme kazandırılmalıdır? A) $4,5 \times 10$ B) $4,5 \times 10^2$ C) $4,5 \times 10^3$ D) $4,5 \times 10^4$	2. $2\sqrt{10}$, $\sqrt{17}$, $3\sqrt{3}$ sayılarının küçükten büyüğe doğru sıralanışı, aşağıdakilerden hangisidir? A) $\sqrt{17} < 3\sqrt{3} < 2\sqrt{10}$ B) $3\sqrt{3} < \sqrt{17} < 2\sqrt{10}$ C) $\sqrt{17} < 2\sqrt{10} < 3\sqrt{3}$ D) $3\sqrt{3} < 2\sqrt{10} < \sqrt{17}$
Doğru Cevaplayabilir Doğru Cevaplayamaz () ()	Doğru Cevaplayabilir Doğru Cevaplayamaz () ()

3. Şekildeki beş katlı düğün pastasında her kattaki pastanın yarıçap uzunluğu, bir üstündekinin yarıçap uzunluğunun 2 katıdır.



En üstteki pastanın yarıçap uzunluğu 2^3 cm olduğuna göre, en alttaki pastanın yarıçap uzunluğu kaç santimetredir?

- A) 2^{10} B) 2^8 C) 2^7 D) 2^5

Doğru Cevaplayabilir

()

Doğru Cevaplayamaz

()

5. Bir mimar bir kenarı $3a$ metre olan kare şeklindeki arsada, her birinin kenar uzunluğu $2b$ metre olan belli sayıda karesel bölge belirlemiştir. Arsada geri kalan bölgenin alanı $(3a - 4b)(3a + 4b)$ metrekare olduğuna göre, belirlenen karesel bölgelerin sayısı kaçtır?

- A) 8 B) 4 C) 3 D) 2

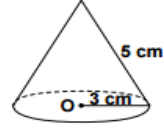
Doğru Cevaplayabilir

()

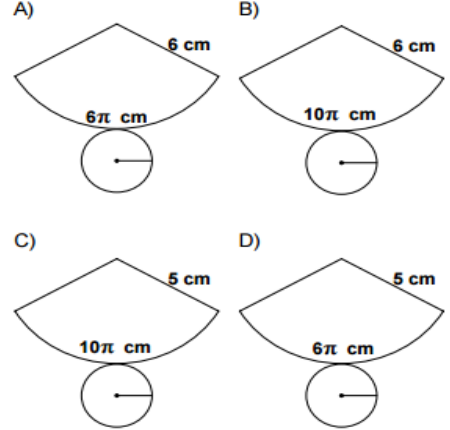
Doğru Cevaplayamaz

()

4. Şekildeki O noktası, verilen dik dairesel koninin taban merkezidir.



Şekil üzerindeki verilere göre bu koninin açınımı aşağıdakilerden hangisi olabilir?



Doğru Cevaplayabilir

()

Doğru Cevaplayamaz

()

6. Ahmet, kenarlarından ikisinin uzunluğu 17 m ve 10 m olan üçgen biçimindeki bahçesinin çevresini çit ile çevirmiştir.

Ahmet'in kullandığı çitin uzunluğu, metre olarak aşağıdakilerden hangisi olabilir?

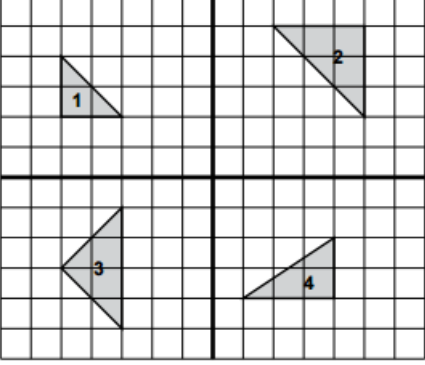
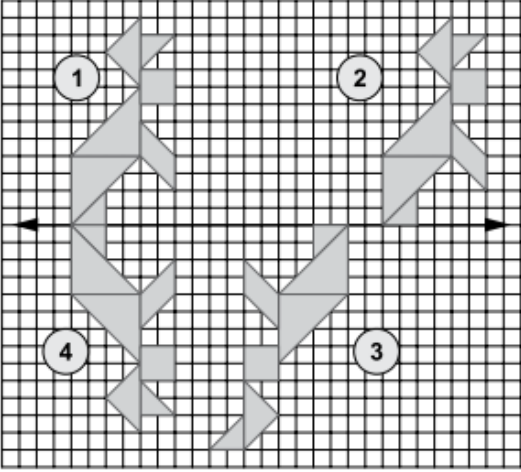
- A) 30 B) 34 C) 40 D) 54

Doğru Cevaplayabilir

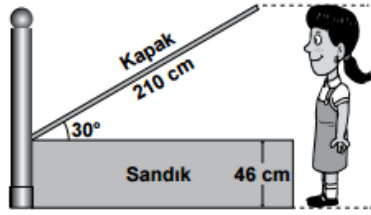
()

Doğru Cevaplayamaz

()

7. Aşağıdaki şekilde dört farklı bahçede yer alan havuzların kuşbakışı çizimleri verilmiştir.  Bu havuzlardan hangisi diğer üç havuzla benzer <u>değildir</u>? A) 1 B) 2 C) 3 D) 4	8. Seda'nın matematik dersi dönem sonu puanı, ilk sınavda aldığı puanın iki katından 4 eksiktir. Dönem sonu puanı 50 ve üzeri olan öğrenci, o dersten başarılı olmaktadır. Matematikten dönem sonunda başarılı olan Seda'nın ilk sınavdaki puanı <u>en az</u> kaçtır? A) 28 B) 27 C) 26 D) 25
Doğru Cevaplayabilir Doğru Cevaplayamaz () ()	Doğru Cevaplayabilir Doğru Cevaplayamaz () ()
9. Aşağıdaki numaralandırılmış tangram şekillerinden hangi ikisi, birbirinin ötelemeli yansımasıdır?  A) 2 ve 4 B) 1 ve 4 C) 1 ve 2 D) 2 ve 3	10. Aşağıdakilerden hangisi, $0,\overline{45}$ ile $0,\overline{452}$ devirli ondalık kesirleri arasında yer alır? A) 0,451 B) 0,453 C) 0,455 D) 0,457
Doğru Cevaplayabilir Doğru Cevaplayamaz () ()	Doğru Cevaplayabilir Doğru Cevaplayamaz () ()

11. Ceren, açılıp kapanabilir bir sandıklı yatağın 210 cm uzunluğundaki kapağını aşağıdaki gibi açtığında, kapağın yerden yüksekliği Ceren'in boyuna eşit olmaktadır.



Sandığın yüksekliği 46 cm olduğuna göre, Ceren'in boyu kaç santimetredir? ($\sin 30 = 0,5$)

- A) 156 B) 151 C) 128 D) 105

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

13. Eser, kare dik prizma biçiminde ve ayrıtlarından birinin uzunluğu 30 cm olan yandaki teneke kutunun yan yüzlerini boyadıktan sonra bir yüzüne "çöp" yazısını yazıyor.



Boyanan yüzeyin alanı 1680 cm^2 olduğuna göre, bu çöp kutusunun diğer farklı ayrıtlarının uzunluğu kaç santimetredir?

- A) 14 B) 21 C) 25 D) 28

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

15. Bir torbadaki özdeş topların 11'i kırmızı, 8'i beyaz, 9'u mavi ve 12'si siyahtır.

En az kaç top çıkarılırsa, torbada kalan topların renklerine göre çekilme olasılıkları eşit olur?

- A) 7 B) 8 C) 9 D) 10

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

12. Aşağıdakilerden hangisi yapıldığında, taban ayrıtının uzunluğu 8 cm, yüksekliği 12 cm olan kare dik prizma şeklindeki kutunun hacmi yarıya iner?

- A) Bütün ayrıt uzunlukları ikiye santimetre azaltıldığında
B) Taban ayrıt uzunlukları dörder santimetre azaltıldığında
C) Bütün ayrıt uzunlukları yarıya indirildiğinde
D) Yüksekliği 6 cm azaltıldığında

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

14. Biri silindir, diğeri küp şeklinde olan iki kap, tamamen su ile doludur. Silindir şeklindeki kabın taban çapının ve yüksekliğinin uzunlukları, küpün bir ayrıt uzunluğuna eşittir.

Buna göre, silindir şeklindeki kaptaki kütte bulunan suyun kaçta kaç kadar su bulunmaktadır? (π yerine 3 alınız.)

- A) $\frac{3}{4}$ B) $\frac{2}{3}$ C) $\frac{1}{2}$ D) $\frac{1}{6}$

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

16. Bir sınıftaki öğrencilerin günlük kitap okuma süreleri aşağıdaki tabloda verilmiştir.

Tablo: Öğrencilerin Kitap Okuma Süreleri

Kitap okuma süresi (dakika)	Öğrenci sayısı
20 - 24	2
25 - 29	6
30 - 34	5
35 - 39	3
40 - 44	8
45 - 49	1
50 - 54	3
55 - 59	12
60 - 64	6
65 - 69	2

Tabloya göre kaç kişi günde 1 saatten daha az süre kitap okumaktadır?

- A) 8 B) 12 C) 40 D) 46

Doğru Cevaplayabilir

Doğru Cevaplayamaz

()

()

17. Üç yol ağzında bulunan bir tavşan, bu yollardan rastgele birine yöneliyor. Bu üç yolun her biri, devamında iki dar yola ayrılıyor. Tavşan yol ayrımına geldiğinde iki dar yoldan birini rastgele seçerek yoluna devam ediyor.

Tavşanın, dar yollardan birinde bekleyen kaplumbağa ile karşılaşma olasılığı nedir?

- A) $\frac{1}{2}$ B) $\frac{1}{3}$ C) $\frac{1}{6}$ D) $\frac{1}{9}$

Doğru Cevaplayabilir

()

Doğru Cevaplayamaz

()

19. Bir yarışma programında, verilen her doğru cevaba + 3 puan, her yanlış cevaba – 2 puan verilmektedir.

Bu yarışmaya katılan Aysun, sorulan 5 sorunun tümünü cevaplamıştır. Yarışma sonunda 10 puan aldığına göre, Aysun kaç soruyu doğru cevaplamıştır?

- A) 2 B) 3 C) 4 D) 5

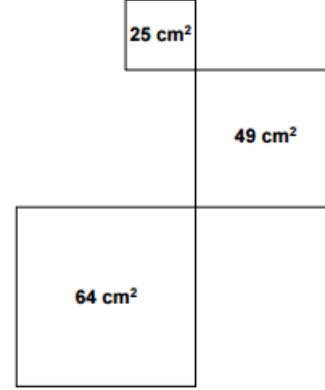
Doğru Cevaplayabilir

()

Doğru Cevaplayamaz

()

18. Bir tel, şekildeki gibi kıvrılarak üç tane kare oluşturuluyor.



Karelerin sınırladıkları bölgelerin alanları 25 cm^2 , 49 cm^2 ve 64 cm^2 olduğuna göre, bu tel ile oluşturulabilecek en büyük karenin sınırladığı bölgenin alanı kaç santimetrekare olur?

- A) 100 B) 138 C) 225 D) 400

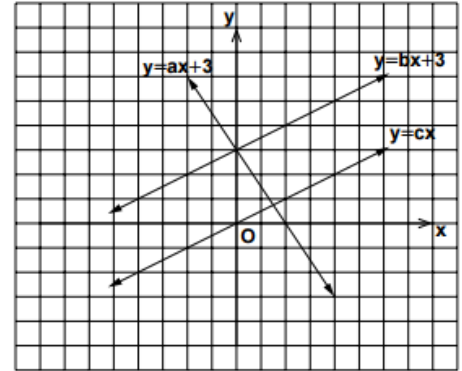
Doğru Cevaplayabilir

()

Doğru Cevaplayamaz

()

20.



Yukarıda denklemleri ile verilen doğruların grafiklerine göre, aşağıdakilerden hangisi doğrudur?

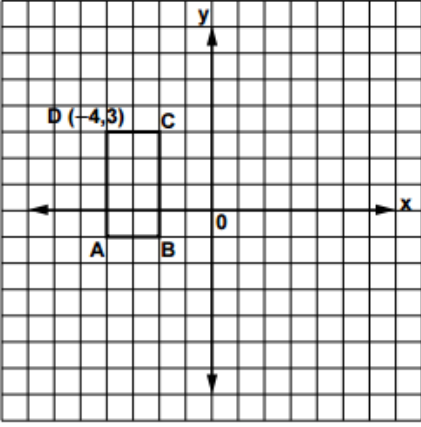
- A) $a < b$ ve $b = c$
B) $b = c$ ve $c < a$
C) $c < b < a$
D) $a < b < c$

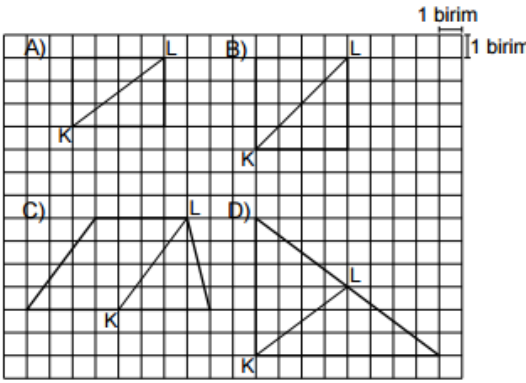
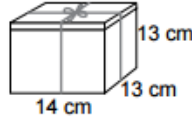
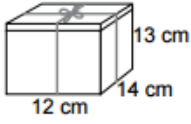
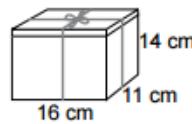
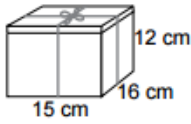
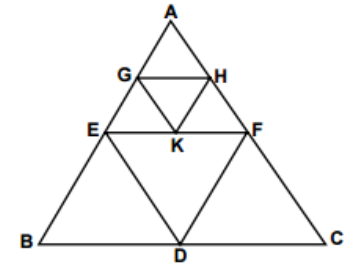
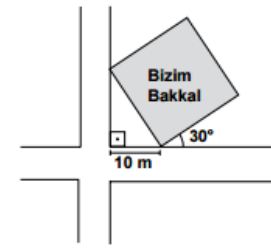
Doğru Cevaplayabilir

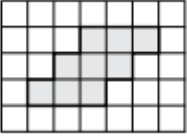
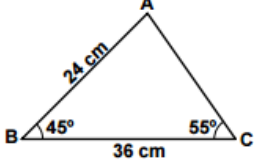
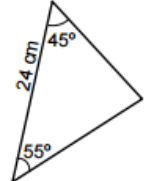
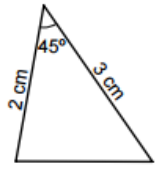
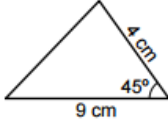
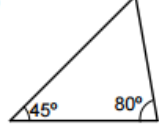
()

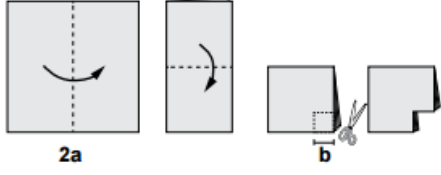
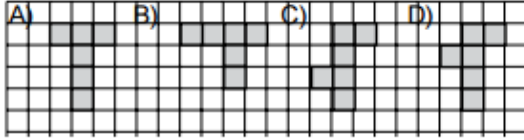
Doğru Cevaplayamaz

()

2013 SBS Matematik Testi Soruları	2. Değerlendirme(minimum yeterli düzeyinde olan bir öğrenci soruyu doğru cevaplayabilir mi?)
1. Aşağıdakilerden hangisi çevre uzunluğu 12 cm olan bir üçgenin kenar uzunluklarından biri <u>olamaz</u>? A) 1 cm B) 3 cm C) 5 cm D) 6 cm	2. $\frac{8^4 \cdot 4^2}{2^8}$ işleminin sonucu aşağıdakilerden hangisine eşittir? A) 2^{12} B) 2^8 C) 2^6 D) 2^4
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
3.  D noktasının koordinatları $(-4, 3)$ olan şekildeki ABCD dikdörtgeni, verilen düzlemde öteleniyor. x ve y eksenleri, elde edilen dikdörtgenin simetri doğruları olduğuna göre, aşağıdakilerden hangisi ötelenmiş dikdörtgenin B köşesinin koordinatları olur? A) $(-1, -2)$ B) $(1, -2)$ C) $(2, -2)$ D) $(-2, -2)$	4. 20 kişiden oluşan bir kafilenin yaş ortalaması 36'dır. Kafileye 5 kişi daha katıldığında, kafilenin yaş ortalaması 37 olmaktadır. Yeni katılanlardan 4'ü kırkar yaşında olduğuna göre, 5. kişi kaç yaşındadır? A) 37 B) 40 C) 42 D) 45
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()

5. Aşağıdaki şekillerin hangisinde KL doğru parçasının uzunluğu, bir irrasyonel sayı ile ifade edilir? 	6. Selim, bir yüzünün alanı 150 cm^2 olan küp şeklindeki bir oyuncak hediye kutusuna koymuştur. Bu oyuncak, aşağıda ayrıt uzunlukları verilen dik prizma şeklindeki hediye kutularından hangisinin içinde <u>olabilir</u>? A)  B)  C)  D) 
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
7. Bir öğretmen deney yapmak üzere fasulye tanelerini, öğrencileri sıralayarak ilk öğrenciye 3, sonraki her öğrenciye bir öncekinden 2 fazla vererek dağıtıyor. n. sıradaki öğrencinin kaç fasulye aldığı, aşağıdaki hangi ifade ile bulunur? A) $2n + 1$ B) $2n + 3$ C) $3n + 2$ D) $n + 2$	8. Şekildeki ABC eşkenar üçgeninin kenarlarının orta noktaları E, D, F ve AEF üçgeninin kenarlarının orta noktaları G, K, H'dir.  Buna göre, ABC üçgeni ile KHG üçgeni arasındaki benzerlik oranı nedir? A) $\frac{1}{16}$ B) $\frac{1}{12}$ C) $\frac{1}{4}$ D) $\frac{1}{2}$
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
9. Denklemi $y = 10$ olan doğrunun y eksenini kestiği noktadan, eğimi 2 olan bir başka doğru geçmektedir. Bu doğrunun x eksenini kestiği noktanın apsisi kaçtır? A) -5 B) $-\frac{1}{5}$ C) $\frac{2}{5}$ D) 20	10.  Şekildeki planda, zemini kare biçiminde olan bakkal dükkânı kaç metrekarelik alana yapılmıştır? ($\sin 30^\circ = \frac{1}{2}$) A) 25 B) 75 C) 300 D) 400
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()

11. t aşağıdakilerden hangisi olursa $\frac{a^2 - a - 2}{a^2 - t}$ ifadesi sadeleşebilir? A) -4 B) -2 C) 2 D) 4	12. Tabanının bir ayrıtının uzunluğu 6 m ve yüksekliği 10 m olan eşkenar üçgen dik prizma şeklindeki saat kulesinin her bir yan yüzüne, çapı 4 m olan daire şeklinde birer saat yerleştirilmiştir. Bu kulenin yanal yüzeyinin, saatler dışında kalan bölgesinin alanı kaç metrekaredir? (π yerine 3 alınız.) A) 32 B) 48 C) 144 D) 168
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
13.  Yukarıdaki boyalı şeklin alanı 243 cm^2 olduğuna göre, çevresinin uzunluğu kaç santimetredir? A) $36\sqrt{3}$ B) $48\sqrt{3}$ C) $64\sqrt{3}$ D) $108\sqrt{3}$	14. Bir torbada aynı özelliklere sahip 10 kalem-den 4'ü sarı, 2'si mavi ve diğerleri beyaz renktedir. Torbadan rastgele 1 kalem çekildikten sonra, torbaya atılmadan ikinci bir kalem daha çekilmektedir. Her iki kalemin de beyaz olma olasılığı nedir? A) $\frac{8}{45}$ B) $\frac{4}{25}$ C) $\frac{2}{15}$ D) $\frac{3}{25}$
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
15. Aşağıda, AEN çeşitkenar üçgeni üzerinde yapılan bir çizim anlatılmaktadır: 1. Pergel, EN doğru parçasının uzunluğu kadar açılır. 2. E ve N merkezli çemberler çizilir. 3. Çemberlerin kesiştiği noktalardan geçen bir doğru parçası çizilir. Yukarıdaki işlemler sonucunda, AEN üçgeninde hangi çizim yapılmış olur? A) Kenar orta dikme B) Kenarortay C) Açıortay D) Yükseklik	16.  Aşağıdakilerden hangisi, verilen ABC üçgenine benzer <u>değildir</u>? A)  B)  C)  D) 
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()

17. Bir kenarının uzunluğu 2a santimetre olan kare şeklindeki bir kâğıt, aşağıda görüldüğü gibi üst üste iki kez katlanarak yeni bir kare elde ediliyor.  Elde edilen kareden, bir kenarının uzunluğu b santimetre olan kare kesilerek atılıyor. Kalan kâğıt tamamen açıldığında alanı kaç santimetrekare olur? A) $a^2 - b^2$ B) $2a^2 - b^2$ C) $(2a - b)^2$ D) $4a^2 - 4b^2$	18. Yüksekliği ve tabanının çap uzunluğu a metre olan bir silindirin hacminin, çap uzunluğu a metre olan bir kürenin hacmine oranı nedir? A) $\frac{3}{2}$ B) $\frac{\pi}{2}$ C) $\frac{3a}{2}$ D) $\frac{\pi a}{2}$
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()
19. Aşağıdakilerden hangisi küpün bir açınımi değildir? 	20. Bir baloncu, tanesini 10 liraya aldığı balonları 12 liraya satmaktadır. Balonların 3 tanesi satılmadan önce patlıyor. Baloncu 400 liradan fazla kâr elde ettiğine göre en az kaç balon satmıştır? A) 198 B) 201 C) 216 D) 231
Doğru Cevaplayabilir () Doğru Cevaplayamaz ()	Doğru Cevaplayabilir () Doğru Cevaplayamaz ()

Katkılarınız için tekrar teşekkür ederim.

EK-B: Angoff Yöntemine göre Uzmanların Güçlük Tahminleri

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28
1	.50	.20	.30	.18	.30	.50	.30	.65	.50	.10	.30	.55	.80	.40	.75	.60	.40	.25	.15	.60	.25	.20	.70	.08	.40	.60	.70	.40
2	.40	.50	.50	.67	.40	.90	.40	.55	.40	.40	.70	.45	.90	.70	.60	.60	.70	.30	.70	.70	.75	.60	.80	.12	.50	.30	.80	.50
3	.40	.10	.60	.59	.10	.60	.30	.75	.70	.10	.40	.40	.80	.60	.80	.40	.60	.15	.50	.80	.60	.30	.70	.14	.50	.50	.70	.60
4	.30	.60	.30	.53	.10	.70	.30	.55	.40	.30	.90	.45	.90	.70	.70	.80	.60	.12	.50	.60	.55	.60	.70	.20	.60	.50	.30	.50
5	.30	.10	.05	.07	.10	.20	.10	.55	.40	.20	.10	.35	.50	.40	.40	.05	.30	.10	.10	.40	.20	.10	.50	.06	.20	.40	.20	.40
6	.50	.30	.15	.33	.30	.70	.40	.60	.70	.40	.80	.45	.95	.50	.60	.60	.50	.20	.20	.30	.55	.40	.50	.14	.30	.50	.50	.50
7	.70	.40	.05	.88	.30	.90	.50	.75	.60	.20	.70	.50	.90	.80	.80	.90	.50	.50	.70	.95	.70	.70	.80	.18	.70	.35	.90	.60
8	.20	.20	.40	.75	.10	.80	.60	.60	.75	.30	.25	.55	.50	.40	.50	.80	.40	.15	.15	.50	.50	.40	.70	.21	.30	.50	.40	.40
9	.30	.40	.40	.94	.20	.50	.50	.70	.50	.40	.50	.60	.80	.70	.70	.95	.70	.15	.70	.75	.55	.30	.80	.22	.70	.40	.90	.50
10	.50	.40	.50	.78	.30	.70	.25	.55	.30	.50	.75	.55	.95	.70	.80	.95	.70	.20	.40	.80	.70	.70	.90	.14	.60	.60	.70	.60
11	.20	.30	.10	.23	.40	.50	.20	.60	.40	.20	.15	.55	.80	.60	.80	.30	.70	.25	.10	.45	.60	.45	.80	.22	.20	.30	.30	.50
12	.10	.30	.10	.07	.10	.60	.15	.55	.60	.20	.70	.40	.60	.50	.60	.30	.40	.12	.30	.20	.50	.60	.30	.04	.60	.40	.70	.60
13	.20	.50	.15	.23	.40	.30	.10	.65	.60	.10	.40	.50	.50	.60	.40	.20	.40	.17	.40	.60	.60	.35	.60	.12	.30	.50	.50	.60
14	.20	.30	.15	.07	.10	.20	.10	.45	.50	.10	.30	.40	.40	.40	.40	.25	.30	.15	.10	.35	.40	.25	.40	.04	.20	.20	.20	.40
15	.40	.40	.30	.07	.10	.80	.25	.60	.60	.30	.90	.35	.40	.70	.50	.90	.60	.20	.30	.40	.70	.40	.70	.20	.30	.40	.90	.30
16	.70	.80	.60	.96	.50	.90	.80	.65	.70	.40	1.0	.60	.90	.90	.80	.95	.80	.80	.80	.80	.60	.70	.80	.30	.70	.70	.90	.60
17	.60	.40	.50	.07	.10	.70	.65	.60	.40	.60	.90	.30	.80	.60	.80	.40	.50	.18	.40	.30	.40	.50	.60	.32	.40	.35	.10	.40
18	.10	.50	.20	.07	.10	.50	.30	.45	.40	.50	.80	.55	.90	.80	.70	.70	.50	.20	.05	.40	.60	.50	.60	.18	.20	.50	.30	.50
19	.30	.50	.35	.73	.50	.80	.80	.60	.50	.40	.60	.35	.50	.60	.50	.50	.40	.30	.40	.25	.60	.40	.80	.10	.50	.45	.10	.60
20	.10	.30	.30	.03	.10	.30	.05	.60	.35	.50	.70	.35	.30	.30	.40	.40	.50	.40	.40	.40	.40	.30	.50	.02	.20	.30	.10	.40
21	.20	.20	.50	.07	.40	.60	.20	.65	.60	.60	.50	.65	.40	.50	.50	.60	.50	.45	.10	.40	.65	.40	.60	.26	.40	.40	.30	.50
22	.60	.50	.40	.39	.40	.80	.20	.60	.50	.60	.70	.70	.80	.70	.80	.60	.70	.50	.80	.85	.80	.60	.90	.20	.80	.60	.80	.60
23	.40	.40	.40	.67	.45	.60	.30	.65	.40	.40	.45	.40	.30	.40	.40	.95	.40	.30	.30	.60	.50	.50	.70	.12	.40	.35	.90	.40
24	.40	.60	.20	.23	.20	.70	.40	.60	.40	.40	.60	.35	.80	.50	.70	.30	.50	.25	.40	.50	.70	.60	.70	.12	.30	.50	.30	.60
25	.10	.50	.15	.07	.10	.50	.15	.55	.30	.50	.30	.40	.80	.80	.85	.40	.50	.20	.25	.75	.70	.20	.70	.16	.50	.50	.20	.60
26	.50	.50	.30	.68	.20	.70	.10	.50	.30	.60	.35	.55	.50	.40	.50	.20	.40	.15	.10	.45	.40	.30	.70	.20	.40	.40	.80	.60
27	.20	.20	.20	.23	.05	.90	.40	.65	.40	.60	.50	.40	.90	.80	.70	.70	.50	.25	.75	.30	.50	.30	.70	.36	.30	.50	.60	.40

28	.40	.40	.15	.07	.05	.60	.35	.45	.30	.40	.25	.50	.80	.70	.80	.30	.50	.25	.30	.60	.60	.40	.60	.12	.50	.45	.30	.50
29	.10	.20	.10	.07	.05	.50	.10	.60	.30	.40	.60	.40	.80	.40	.75	.10	.30	.20	.50	.20	.75	.45	.60	.14	.40	.35	.10	.40
30	.10	.50	.20	.07	.05	.40	.45	.60	.40	.40	.40	.60	.90	.60	.80	.05	.50	.15	.40	.55	.70	.50	.70	.26	.20	.30	.20	.50
31	.30	.40	.25	.17	.20	.70	.15	.45	.20	.60	.40	.55	.90	.60	.80	.05	.50	.15	.20	.55	.60	.30	.70	.08	.40	.40	.10	.50
32	.05	.30	.05	.03	.10	.50	.10	.40	.30	.20	.60	.55	.70	.70	.60	.20	.30	.15	.30	.30	.30	.30	.60	.08	.30	.30	.00	.40
33	.60	.60	.25	.07	.10	.70	.40	.65	.40	.70	.15	.65	.80	.80	.80	.20	.40	.20	.60	.70	.70	.50	.60	.06	.60	.50	.90	.50
34	.60	.60	.60	.23	.05	.90	.60	.50	.30	.70	.40	.30	.80	.60	.80	.60	.50	.25	.75	.40	.65	.60	.70	.19	.40	.30	.00	.40
35	.10	.70	.10	.07	.10	.50	.15	.60	.20	.40	.20	.70	.40	.50	.40	.50	.60	.25	.10	.25	.30	.50	.30	.03	.40	.20	.00	.50
36	.50	.80	.25	.23	.40	.80	.55	.70	.40	.70	.60	.70	.60	.40	.60	.70	.40	.25	.80	.40	.70	.60	.60	.15	.70	.35	.40	.60
37	.20	.70	.05	.07	.05	.30	.10	.55	.30	.70	.30	.45	.50	.70	.50	.10	.60	.17	.10	.20	.40	.50	.60	.04	.40	.25	.00	.40
38	.50	.70	.40	.07	.10	.50	.50	.55	.20	.70	.50	.35	.50	.30	.50	.50	.40	.25	.60	.50	.60	.70	.50	.06	.70	.30	.20	.60
39	.40	.80	.30	.94	.10	.80	.65	.65	.40	.80	.40	.40	.80	.80	.80	.90	.40	.20	.90	.70	.30	.50	.60	.23	.80	.30	.90	.60
40	.40	.40	.10	.07	.10	.70	.60	.55	.20	.70	.30	.60	.70	.40	.70	.80	.60	.30	.20	.40	.60	.50	.50	.34	.20	.50	.20	.40

EK-C: Evet/Hayır Yöntemine göre Uzmanların Kararlarına İlişkin Puanlar

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28
1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	1	0	0	1	1	0
2	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1
3	1	0	0	1	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1
4	0	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1
5	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0
6	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1
7	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0	0	0	1	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0
9	0	1	0	1	0	0	1	1	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1
10	1	1	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1
11	0	1	0	0	1	0	0	1	0	0	0	1	1	1	0	0	1	0	0	1	1	0	1	1	0	0	0	1
12	0	1	0	0	0	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
13	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	1	1	1
14	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
15	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	0
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0
18	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	1	0	1
19	1	0	1	1	0	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	0	1
20	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
21	0	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1
22	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
23	0	1	0	1	1	0	0	1	0	1	0	1	0	0	0	1	0	1	1	1	0	1	1	1	1	0	1	0
24	1	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1
25	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1	1
26	1	1	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1

27	0	0	1	0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0		
28	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	1	1	1	0	0	1	
29	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	
30	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1
31	0	0	0	0	0	1	1	0	0	1	0	1	1	1	1	0	1	0	0	1	1	0	1	0	1	1	0	1	
32	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	
33	1	1	0	0	0	1	1	1	0	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	1	1	1	1	
34	0	1	1	0	0	1	1	1	0	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	
35	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	
36	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	
37	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	
38	0	1	1	0	0	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1	1	0	0	1	0	0	1	
39	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1	1	1	
40	1	0	0	0	0	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	

EK-Ç: Angoff ve Evet/Hayır Yöntemine göre Uzmanların Bireysel Geçme Puanları

	Angoff	Evet/Hayır		Angoff	Evet/Hayır
U1	34.13	45	U15	64.63	61.11
U2	43.75	65	U16	50.75	77.50
U3	27.38	35	U17	50	65
U4	31.88	30	U18	24.40	40
U5	19.63	22.50	U19	39.50	52.50
U6	62	67.50	U20	50.38	57.50
U7	33.75	62.50	U21	55.50	67.50
U8	58.63	87.50	U22	45	70
U9	42.75	27.50	U23	64.50	80
U10	43.25	77.50	U24	15.58	50
U11	50.88	55	U25	43.75	60
U12	48.50	47.50	U26	41.25	50
U13	69.50	47.50	U27	43.50	45
U14	58.75	70	U28	49.75	65
Nihai Geçme Puanı	45.12	56.25			

EK-D: Alt Testlerin Oluşturulmasında Kullanılan Bir R Kodu Örneği

(%50- İTSÖ Yaklaşımı – Angoff Yöntemi)

```
angof=read.table("C:/Users/hakan/Desktop/angof.txt", header=FALSE)
sample50=combine(sample(c(1,2,3,10,18,22,25,26,33),5, replace=FALSE),
  +sample (c(4,6,7,9,21,23,28,35,36,39),5, replace=FALSE),
  +sample (c(11,12,13,14,30,32,38),3, replace=FALSE),
  +sample (c(15,16,17,24,34),2, replace=FALSE),
  +sample (c(5,8,19,20,27,29,31,37,40),5, replace=FALSE))
subset50= angof[sample50,]
puan50=colMeans(subset50)*100
kes50=mean(puan50)
for (kes50 in c(1:1000)) {
  kes50=mean(colMeans(angof[combine(sample(c(1,2,3,10,18,22,25,26,33),5,
replace=FALSE),
  +sample (c(4,6,7,9,21,23,28,35,36,39),5, replace=FALSE),
  +sample (c(11,12,13,14,30,32,38),3, replace=FALSE),
  +sample (c(15,16,17,24,34),2, replace=FALSE),
  +sample (c(5,8,19,20,27,29,31,37,40),5, replace=FALSE)),)]*100)

  print(kes50)
}
```

EK-E: ANOVA Analizi Varsayım Kontrolleri (Normallik)

	Test	Çarpıklık	Basıklık	p değeri (Shapiro-Wilk)
Alt Problem 1	%30	0.19	-0.15	0.05
	%40	-0.04	-0.07	0.94
	%50	0.07	-0.19	0.38
	%70	-0.03	-0.09	0.88
Alt Problem 2	%30	0.11	-0.19	0.08
	%40	-0.16	-0.15	0.09
	%50	-0.02	-0.35	0.05
	%70	-0.06	-0.14	0.50
Alt Problem 3	%30	0.03	-0.07	0.20
	%40	0.05	-0.06	0.96
	%50	0.05	-0.16	0.41
	%70	0.05	0.02	0.11
Alt Problem 4	%30	0.11	-0.09	0.16
	%40	0.08	-0.18	0.39
	%50	0.21	-0.11	0.05
	%70	-0.10	-0.17	0.29
Alt Problem 6	%30	-0.07	-0.13	0.21
	%40	0.01	0.01	0.27
	%50	0.04	-0.11	0.47
	%70	-0.03	-0.05	0.63
Alt Problem 7	%30	0.01	-0.15	0.31
	%40	0.10	-0.09	0.30
	%50	0.16	-0.28	0.06
	%70	-0.05	-0.09	0.36
Alt Problem 8	%30	0.02	-0.23	0.30
	%40	-0.08	-0.18	0.14
	%50	-0.05	-0.05	0.46
	%70	-0.04	-0.22	0.24
Alt Problem 9	%30	-0.06	-0.40	0.09
	%40	0.01	-0.29	0.06
	%50	-0.09	-0.21	0.19
	%70	-0.21	-0.16	0.10

EK-F: ANOVA Analizi Varsayım Kontrolleri (Varyansların Homojenliği)

	Levene İstatistiği	p değeri
Alt Problem 1	620.992	0.00
Alt Problem 2	609.474	0.00
Alt Problem 3	572.366	0.00
Alt Problem 4	573.153	0.00
Alt Problem 6	598.121	0.00
Alt Problem 7	585.546	0.00
Alt Problem 8	596.132	0.00
Alt Problem 9	647.236	0.00

EK-G: İçerik Kategorilerine Göre Alt Testlere Seçilecek Madde Sayıları

Alt Testler	İçerik Kategorileri					Toplam
	Sayılar	Geometri	Ölçme	Olasılık ve İstatistik	Cebir	
%30	3	3	2	1	3	12
%40	4	4	3	2	3	16
%50	5	5	3	2	5	20
%70	6	7	5	4	6	28

EK-Ğ: Güçlük Gruplarına Göre Alt Testlere Seçilecek Madde Sayıları

Alt Testler	Güçlük Grupları			Toplam
	Grup 1 (.00<p≤.20)	Grup 2 (.20<p≤.40)	Grup 3 (.40<p≤.63)	
%30	4	6	2	12
%40	5	8	3	16
%50	6	10	4	20
%70	9	13	6	28

EK-H: İçerik-Güçlük Kategorilerine Göre Alt Testlere Seçilecek Madde Sayıları

Alt Testler	İçerik-Güçlük Kategorileri															Toplam
	Sayılar			Geometri			Ölçme			Olasılık ve İstatistik			Cebir			
	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	Grup	
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
%30	1	1	1	1	2	0	1	1	0	0	0	1	2	1	0	12
%40	2	1	1	1	2	1	1	2	0	0	1	1	2	1	0	16
%50	2	2	1	1	3	1	1	2	0	0	1	1	2	2	1	20
%70	3	2	1	1	4	2	2	3	0	0	2	2	3	2	1	28

EK-I: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Sayı : 35853172/ 433 - 1667

13 Nisan 2018

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 02.04.2018 tarih ve 859 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı tezli yüksek lisans programı öğrencisi **Hakan KARA**'nın **Dr. Öğr. Üyesi Sevda ÇETİN** danışmanlığında yürüttüğü "**Farklı Madde Örneklemeye Yaklaşımlarına Göre Angoff ve Evet – Hayır Yöntemleri ile Belirlenen Kesme Puanlarının Karşılaştırılması**" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun 03 Nisan 2018 tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

EK-İ: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

10/07/2018


Hakan KARA

EK-J: Yüksek Lisans Tez Çalışması Orijinallik Raporu

10/07/2018

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Farklı Seçkisiz Madde Örneklemeye Yaklaşımlarına Göre Angoff ve Evet/Hayır Yöntemleri İle Belirlenen Geçme Puanlarının Karşılaştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
10/07/2018	66	108549	18/06/2018	%3	981483215

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden daha az örtüşme içeren metin kısımları hariç

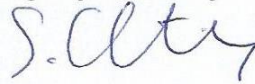
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Hakan KARA
Öğrenci No.: N12228706
Ana Bilim Dalı: Eğitim Bilimleri
Programı: Eğitimde Ölçme ve Değerlendirme
Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.


İmza

DANIŞMAN ONAYI

UYGUNDUR.
(Dr. Öğr. Üyesi, Sevda ÇETİN)



EK-K: Thesis Originality Report

10/07/2018

HACETTEPE UNIVERSITY
Graduate School Of Educational Sciences
To The Department Of Educational Sciences

Thesis Title : Comparison of Passing Scores Set By Angoff and Yes/No Methods With Respect to Different Random Item Sampling Approaches

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using Turnitin plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
10/07/2018	66	108549	18/06/2018	3%	981483215

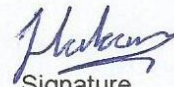
Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

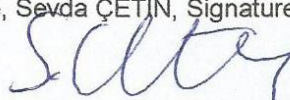
I respectfully submit this for approval.

Name Lastname: Hakan KARA
Student No.: N12228706
Department: Educational Sciences
Program: Educational Measurement and Evaluation
Status: ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.


Signature

ADVISOR APPROVAL

APPROVED
(Title, Sevda ÇETİN, Signature)



EK-L: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversite'ye verilen kullanım hakları dışındaki bütün fikrî mülkiyet haklarım bende kalacak, tezimin tamamının veya bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversite'ye teslim etmeyi taahhüt ederim.

☐ **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirirse bile, teziniz arama motorlarının ön belleklerinde kalmaya devam edebilecektir)

☒ **Tezimin/Raporumun 10/07/2019 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

☐ **Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

☐ **Serbest Seçenek/Yazarın Seçimi:**

.....
.....
.....
...

10/07/2018

Hakan KARA

